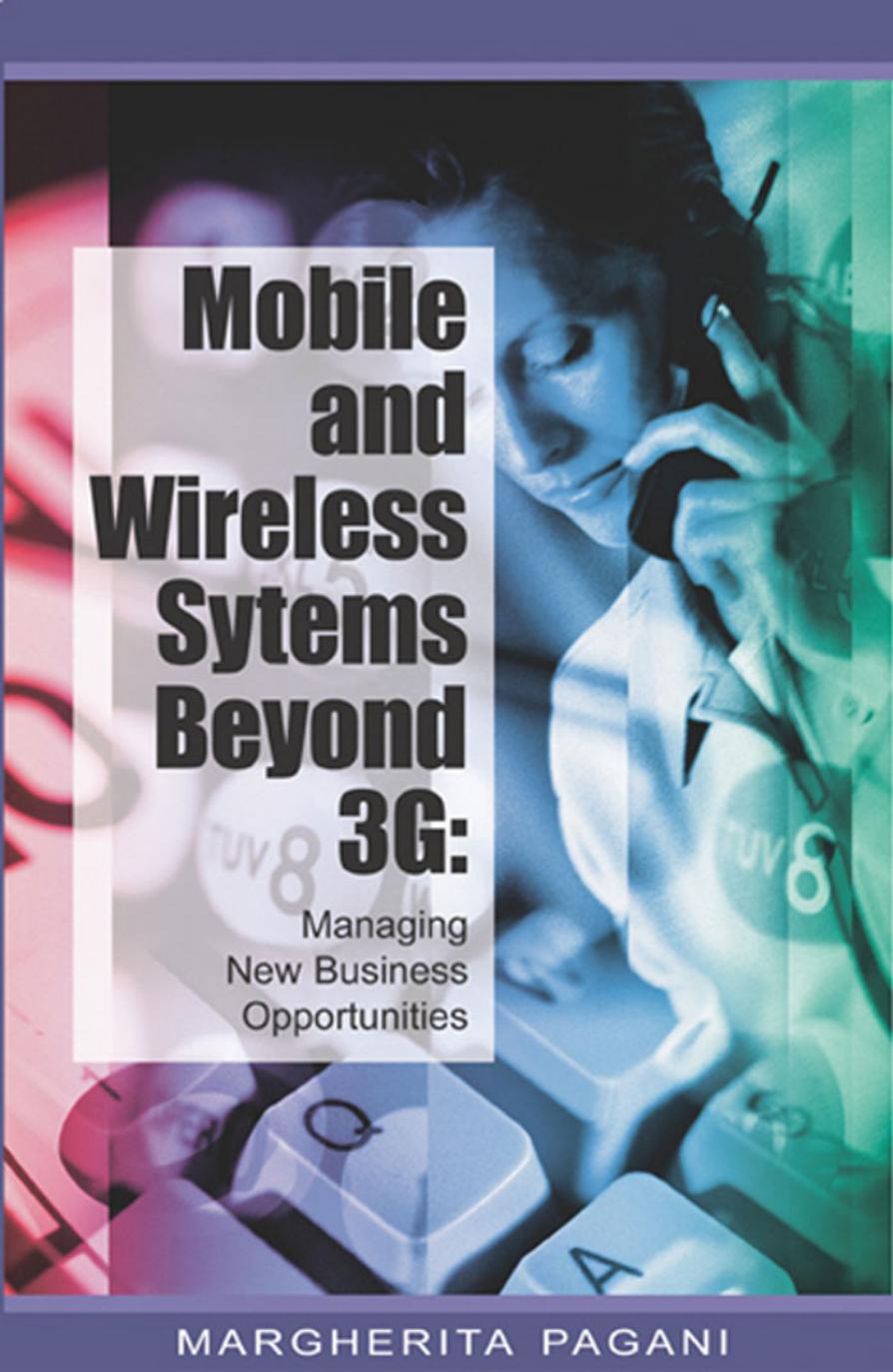

Mobile and Wireless Sytems Beyond 3G:
Managing New Business Opportunities
MARGHERITA PAGANI

Mobile and Wireless Systems Beyond 3G:

Managing New Business Opportunities

Margherita Pagani
I-LAB Centre for Research on the Digital Economy,
Bocconi University, Italy

IRM Press

Publisher of innovative scholarly and professional information technology titles in the cyberage

Hershey • London • Melbourne • Singapore

Acquisitions Editor: Renée Davies
Development Editor: Kristin Roth
Senior Managing Editor: Amanda Appicello
Managing Editor: Jennifer Neidig
Copy Editor: Jennifer Young
Typesetter: Marko Primorac
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
IRM Press (an imprint of Idea Group Inc.)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033-1240
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@idea-group.com
Web site: http://www.irm-press.com

and in the United Kingdom by
IRM Press (an imprint of Idea Group Inc.)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 3313
Web site: http://www.eurospan.co.uk

Library of Congress Cataloging-in-Publication Data

Mobile and wireless systems beyond 3G : managing new business opportunities / Margherita Pagani, editor.
p. cm.
Includes bibliographical references and index.
ISBN 1-59140-570-X (hc) -- ISBN 1-59140-544-0 (sc) -- ISBN 1-59140-545-9 (ebook)
1. Cellular telephone services industry. 2. Wireless communication systems. I. Pagani, Margherita, 1971-
HE9713.M62 2005
384.5'35'0684--dc22
2004023609

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Mobile and Wireless Systems Beyond 3G: Managing New Business Opportunities

Table of Contents

Preface

With the rapid growth of the wireless mobile applications, wireless voice has begun to challenge wireline voice, whereas the desire to access e-mail, surf the Web or download music (e.g., MP3) wirelessly is increasing for wireless data. While second generation (2G) cellular wireless systems, such as cdmaOne1, GSM2 and TDMA3, introduced digital technology to wireless cellular systems to deal with the increasing demand for wireless applications, there is still the need for more spectrally efficient technologies for two reasons. First, wireless voice capacity is expected to continue to grow. Second, the introduction of high-speed wireless data will require more bandwidth.

While the current 2G technologies can support wireless data using cdmaOne circuit switched data or general packet radio system (GPRS), there is clearly the need for more spectrally efficient wireless technology given the limited spectrum available in the wireless bands.

The ability to provide more spectrally efficient voice capacity and spectrally efficient high-speed wireless data has been the focus of third-generation (3G) technologies.

Three important changes have taken place over the last year that will force us to change the way mobile networks develop services to their users:

- Changes in the expectations of users — the boundaries between “core” network services and “value added services” in mobile communications networks are increasingly blurred. Most importantly, the services and content that users expect to receive are no longer the massively produced homogenous things of the past; they are tailored services that are probably only appealing to thinner segments of the population;

- An imbalance between network operators and independent application developers in the "value network" for the provision of network-dependant services; and
- The long-awaited launch of next generation networks and handsets.

The answer to the above challenges lies in leveraging the deployment of next-generation networks to bring in the myriad application developers into an environment that harnesses their nimbleness. Network operators have today the tools to deploy environments that would allow the challengers to work better by working within, and in partnership with, the network operator. The opportunity to the operators runs through a change in their engineering focus that will enable a dramatic change in their business model.

The business model calls for gathering as many of the small "challengers" in as possible; making it worth their while to work with, rather than against, the operator, and insuring that operators get a larger cut of any transaction than simply being a bit pipe.

This change in the business model would imply that today's operators would have to change their current focus and processes greatly. They would need to recognize that their new "partners" are an integral part of the business strategy and treat them accordingly, by providing easier access to operators' business processes.

A sea change in engineering focus will be necessary to allow this business model to succeed. It will be necessary to change our understanding of what "core network" and "value added services" (VAS) are. VAS infrastructure will have to move to the core, at least intellectually. Moreover the nature of the VAS infrastructure will have to change. Operators will steel need to deploy robust "telco grade" systems, but these systems will be tooled to serve as the launching pad for dozens, or thousands, of applications brought forward by the new partners.

This book explores these challenges and their implications on the development of future services.

Purpose of the Book

This book is a pioneering initiative to develop an interdisciplinary view of wireless systems, drawing upon the best work of diverse streams of economic and technological researches.

Researchers have conducted extensive studies and developed theories focused on specific parts of the challenges generated by mobile and wireless systems.

This book draws together these varied perspectives and places them in the hands of managers and students.

These insights have never been more needed. As the competitive environment becomes increasingly dynamic, managers need fresh perspectives and a sharply tuned understanding of business opportunities with mobile services and applications. This is the goal of this book.

Since the perspectives developed from different streams of research and theory, there is not a perfect fit. Nor is the goal of this work to produce one formulaic answer to the complex challenge of mobile and wireless systems. Instead, the following chapters offer diverse perspectives on analyzing strategy and diverse tools for formulating strategy.

Organization of the Book

The book is organized into six main parts and 12 chapters.

Section I (Chapter I) is intended to describe changes in competitive advantages deriving from the development of Third Generation services. **Chapter I** provides the theoretical framework of competitive analysis and it focuses on value chain strategy framework giving an analysis of wireless market attractiveness and changes in competitive advantages. Five scenarios are outlined and validated in order to analyze the behavior of systems not only in management but also in environment change, politics, economic behavior.

Section II (Chapters II-III) considers determinants of mobile technology adoption from the user's context. In this part, **Chapter II** focuses on identifying the technology and non-technology factors that corporations consider important in their decision to deploy devices designed for mobile telephony and mobile data services. It also considers the approval steps in decision-making, the extent and importance of 3G and beyond as it relates to web-enabled cell phones, and the functional areas of use of cell phones.

Chapter III discusses requirements for uptake to occur in specific sectors where a value proposition for mobile data services has been identified and yet adoption rates have varied. Adoption of mobile data services refers to organizational-related solutions as well as service innovations related to the product or service delivered to end-users.

Section III (Chapter IV-VI) surveys the most business opportunities with mobile services and applications. In this part, **Chapter IV** presents a framework derived from the literature to aid the development of viable business models

expected from the amalgamation of mobile telecommunication and space infrastructure. It also identifies the various actors involved in the delivery of these services.

Chapter V introduces the basic ideas and characteristics underlying the concept of ubiquity commerce. It discusses market drivers and applications of u-commerce as well as the underlying technology and the benefits and challenges of u-commerce.

Chapter VI explores the opportunities offered by 3G services/business applications to SMEs, making inferences from the long term research project and providing a broad critical outlook on future opportunities for SMEs to benefit from 3G services.

Section IV (Chapter VII-VIII) explores main technical challenges. In **Chapter VII** a multi-layer ATM architecture is proposed for the interconnection of current and future mobile communications nodes. Moreover, facing the huge expansion of transmission interconnection network that will support current and future generation mobile communications, a modification of the standard ATM cell structure is introduced in order to efficiently support user mobility functional procedures. The proposed ATM architecture is integrated over a suitable, with respect to region and capacity, physical interface, consisting of SDH or SONET for wide area topologies, wireless links for outdoor areas and LED - POF combination for indoor areas. Being an interesting alternative over copper or traditional fiber, POF characteristics, and performance issues are analyzed.

New business opportunities for mobile, wireless and fixed networks are going to require managed packet-based services; this requires SLAs that relate to the level of QoS purchased, and the measurement (monitoring) of information loss and delay at the packet level. **Chapter VIII** investigates the two available measurement techniques: passive and active monitoring and it proposes some ideas which may enhance accuracy.

Section V (Chapter IX-X) deepens security and privacy issues related to mobile and wireless systems development. In this part **Chapter IX** focuses on smart card in mobile communications as a service platform and as a marketing instrument for the network operator. The (Universal) Subscriber Identity Module—(U)SIM—is the network operator's "business card" that is handed out to the end-user. The design of the artwork printed on the smart card, the packaging as well as the functionality directly influence the positioning of the operators' brand in the market. The smart card as used in mobile communications enjoys a high reputation and is very important for the network operators. It does not only provide security and trust thus securing the revenues of the network operator but is also a platform for value added services. **Chapter X** focuses on wireless local area network security evolution and WLAN security threats. Special attention is given to user authentication schemes and to protec-

tion of data communication. WPA is also compared with the present WLAN security protocol WEP. Other covered issues are available WPA supported technology and open source WLAN security software. A WLAN designed according to the new security standards is a real alternative to a secure enterprise LAN and also a reliable platform for e-commerce.

Chapter XI discusses the need for privacy and security in mobile systems and presents technological trends which highlight that this issue is of growing concern.

Finally, in **Section VI**, **Chapter XII** aims to investigate some among the current technical, business, financial, and regulatory visions associated with the effective evolution of third generation (3G) networks and services, in particular to fulfil the great market realities, the expectations and 3G's significant potential in building the EU Information Society. The chapter depicts data related to the current state of play for 3G communications in Europe, with specific emphasis given to the underlying technologies and probable standardization options (both for network and terminal equipment).

Wireless mobile technologies are a major driver to realize the way forward to a knowledge-based economy, in terms of consumer demand, productivity, competitiveness and job creation. Under suitable terms, this may create enormous potential and significant investment incentives, for the full recovery of the wider ICT sector. 3G is likely to play a key role in providing widespread access to the Internet and to interactive services, thus maximising consumer choices and providing flexibility for the market itself.

Concluding Remarks

I should note that all of the chapters were reviewed by either the editor or by external reviewers via a blind review process. Both chapters submitted by academic researchers and by the professionals working from firms in industry, were submitted to external reviewers who did not know the authors' names and affiliations. In this way, papers were given a through scrutiny by experts in the fields of mobile and electronic commerce. In total, we were quite selective regarding actually including a submitted chapter in the book.

I'm delighted to present this book to you and I am proud of the many outstanding chapters that are included herein. I'm confident that you will find it to be a useful resource to help your business, your students, or your business colleagues to better understand the topic of 3G wireless.

Endnotes

1 TIA/EIA/IS-95B. Mobile station–base station compatibility standard for wideband spread spectrum cellular systems. April 1999.

2 Halonen T, Romero J, Melero J. GSM, GPRS and EDGE Performance. John Wiley & Sons, Ltd, 2002.

3 TIA/EIA/IS-136. TDMA cellular PCS. April 1999.

Acknowledgments

I first want to recognize the expertise, enthusiasm, and cooperative spirit of the authors of this volume. Without their commitment to this multidisciplinary exercise, I would not have succeeded.

Mobile and Wireless Systems Beyond 3G: Managing New Business Opportunities, like all fieldbooks, would not exist without a great deal of time and attention from a great many people. The efforts that we wish to acknowledge took place over the course of the last two years, as first the premises, then the project, then the challenges, and finally the book itself took shape.

I owe a great debt to colleagues who have worked with me directly (and indirectly) on the research represented here. I am particularly indebted to all the authors involved in this book who provided the opportunity to interact and work with the best and the brightest from around the world. I would like to thank all of them: James B. Pick and Keith Roberts (University of Redlands, California, USA), Elizabeth Fife and Francis Pereira (University of Southern California, USA), Phillip Olla (Brunel University, UK), Fiona Fui-Hoon Nah and Holtjona X Galanxhi (University of Nebraska-Lincoln, USA), Bardo Fraunholz and Chandana Unnithan (Deakin University, Australia), Jürgen Jun (Uni Duisberg-Essen, Germany), Iossifides Athanassios and Spiros Louvros (Cosmote Mobile Cellular Telecommunications, Greece), John Schormans and Chi Ming Leung (Queen Mary University of London, UK), Claus Dietze (The European Telecommunications Standards Institute, France), Göran Pulkkis, Kaj J. Grahn, Jonny Karlsson, and Mikko Martikainen (Arcada Polytechnic, Finland), Daniel Escartin Daniel (Universitaria Politecnica de Teruel, Spain), Marco Cremonini, Ernesto Damiani, Sabrina De Capitani Di Vimercati, Pierangela Samarati (University of Milan, Italy), Angelo Corallo, Gianluca Elia (University of Lecce, Italy), Ioannis P. Chochliouros and Anastasia Spiliopoulou-Chochliourou (Ellenic Telecommunications Organization, Greece).

Crafting a wealth of research and ideas into a coherent book is a process whose length and complexity I underestimated severely. I owe a great debt to Michele Rossi, the development editor of this book. She organized and carried out the complex tasks of editorial management, deadline coordination, and page production—tasks normally kept separate, but which, in this book, were integrated together so we could write and produce this book.

Mehdi Khosrow-Pour, my editor, and his colleagues at Idea Group Inc. have been extremely helpful and supportive every step of the way. Mehdi took on this project with enthusiasm and grace, and I benefited greatly both from his working relationship with me and his editorial insights. His enthusiasm motivated me to initially accept his invitation for taking on this project.

A further special note of thanks goes also to Jan Travers and Jennifer Sundstrom at Idea Group Inc., whose contributions throughout the whole process from inception of the initial idea to final publication have been invaluable.

I would like to acknowledge the help of all involved in the collation and review process of the book, without whose support the project could not have been satisfactorily completed. Most of the authors of chapters included in this book also served as referees for articles written by other authors. Thanks go to all those who provided constructive and comprehensive reviews. A special thank to Jeimy José Cano Martinez of Universidad de Los Andes, Colombia Steven John Simon of Stetson School of Business and Economics, Mercer University, Atlanta and Danilo Schipani, Valdani Vicari & Associati, Milan.

This book benefited from the support and encouragement of Professor Enrico Valdani Director of I-LAB Centre for Research on the Digital Economy of Bocconi University where the project was developed and Foundation Tronchetti Provera which supported my research project of mobile industry during the last three years.

In closing, I wish to thank all of the authors for their insights and excellent contributions to this book. Working with them in this project was an extraordinary experience.

I dedicate this book to Bocconi University for providing such a stimulating environment where a project as broad as this book became possible to envision and develop.

Margherita Pagani
Milan, Italy
June 2004

Section I

Market View

Chapter I

3G Wireless Market Attractiveness: Dynamic Challenges for Competitive Advantages

Margherita Pagani, I-LAB Centre for Research on the Digital Economy, Bocconi University, Italy

Abstract

Nearly every incumbent operator in the wireless market has experienced business problems in recent years. The reason for this is the opening of the market for competitive operators and the following drop in prices as well as attractive services in the mobile telephony market. This chapter describes changes in competitive advantages deriving from the development of Third Generation services. The remainder of this chapter is organized into the following three sections. The first section provides the theoretical framework

of competitive analysis. Next the chapter focuses on value chain strategy framework giving an analysis of wireless market attractiveness and changes in competitive advantages. Finally the chapter analyzes competitive dynamics and describes five competitive strategies that differ in their aggressiveness in launching new services and deploying new technology.

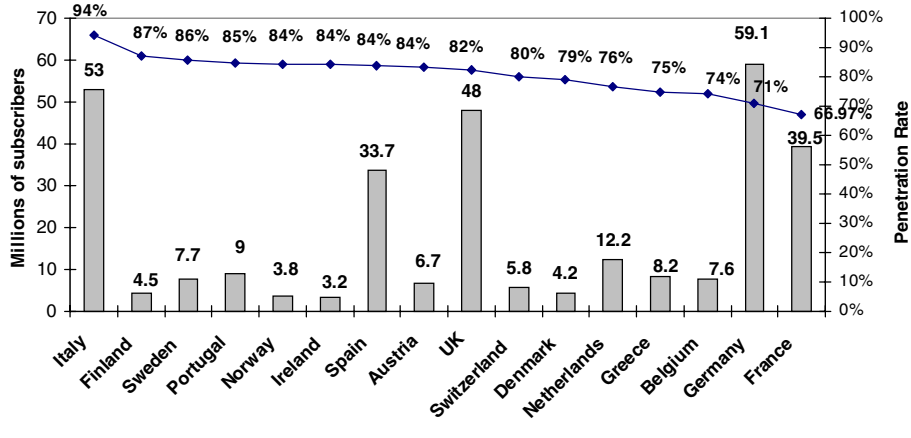
Introduction

The rate of evolution (clock speed[1]) of an industry depends on its products, processes, and customer requirements (Fine, 1999). The wireless industry is one of the most dynamic and demanding industries in the world economy today. Competition is intense. Rapid growth, increasing complexity of technology, globalization, and other changes pose enormous challenges for core business processes such as the supply chain and product/service development.

In many European countries mobile penetration rates are now reaching saturation point (Figure 1) but there are still plenty of opportunities for subscriber growth in South-east Asia and South America. In addition, month-on-month minutes of use continue to grow dramatically worldwide.

As mature markets reach saturation, over 95 percent of subscriber usage remains stuck on voice-only communication. In increasingly aggressive, competitive markets, high-volume usage can only mean falling prices. The conse-

Figure 1. Mobile users and service penetration in Europe (Source: Own elaboration on data ITU 2003)



quence of this is that the premium price for mobility is disappearing and this is clearly reflected in the decline of revenue per minute per mobile subscriber (ARPU — Average Revenue per User) (Figures 2-3).

A saturated market and a slow down in mobility voice-price and in premium SMS means operators must hold onto customers. The market conditions motivate the networks to shift their overall strategy from acquiring customers to retaining them.

There are two major causes for the loss of revenue by the incumbent telephony operators in the market:

Figure 2. Average Revenue Per User (ARPU) in Europe (Source: New Media&Tv-lab — I-LAB Bocconi University 2003)

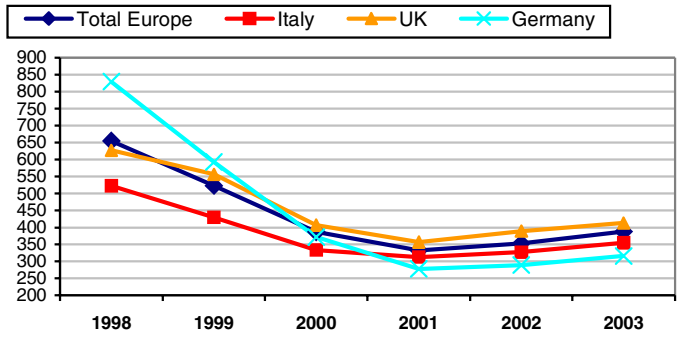
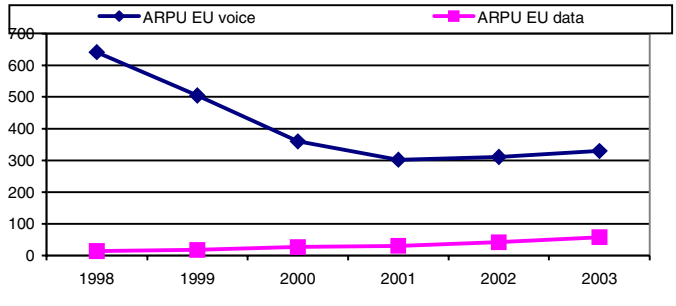


Figure 3. Average Revenue Per User (voice and data) in Europe ($) (Source: New Media&Tv-lab — I-LAB Bocconi University 2003)



1. The liberalization of the market reduces barriers to market entry and
2. High churn rate of customers towards mobile operators because of the technological progress in the telecommunications sector and the high level of competition.

The faster the industry clockspeed the shorter the half-life of any given competitive advantage (Fine 1999).

The purpose of this chapter is to describe changes in competitive advantages deriving from the development of 3G services. The remainder of this chapter is organized into the following three sections. The first section provides the theoretical framework of competitive analysis. Next, the chapter focuses on a value chain strategy framework giving an analysis of wireless market attractiveness and changes in competitive advantages. Finally the chapter analyzes competitive dynamics and describes five competitive strategies that differ in their aggressiveness in launching new services and deploying new technology.

Competitive Analysis: The Theoretical Framework

The Competition Based View Paradigm and the studies focused on dynamic competitive strategies (D'Aveni, 1994; Day & Reibstein, 1997; Valdani, 2003) assume that in high technology sectors the best players are those who are able to anticipate and drive the challenges generated by unexpected new innovative technologies and are able to anticipate and proactively manage the cycle of competitive dynamic.

The different phases of competitive dynamic described by the Competition Based View Paradigm originate three competitive games: movement, imitation, and position (Valdani, 2003).

The mobile sector is characterized by a high competitive dynamic justified and originated by a succession of *quantum leap,*[2] originated by technology innovations (such as the development of a Third Generation standard, the development of i-Mode and voIP, etc.).

The first step in being competitive is to understand one's own competitors. Competitive analysis is an ongoing function, especially in tactical adjustments to the daily sales and promotion activities of the competition, but also is an important component of the annual and strategic plans and must address strategic differences and the current status of the company versus the competition in all areas (Steuernagel, 1999).

What are the feature advantages of their systems? How well are they controlling call quality and system capacity? How is their distribution organized in size, quality, and location? How good are their advertising agencies? What are their marketing and advertising budgets? What happens when you call customer service? What are the revenues and market shares of the other carriers and all players in the market? What segments are they targeting? The answers to questions such as these must be laid out side by side with one's operations to determine the area in which each competitor is the strongest.

This analysis is the prime determinant of the positioning strategy based on strengths and strategies (Day & Reibstein, 1997). These strategies need not directly attack competitors. You can attack the competition head-on, down them with your own unique strategy, and move the competitive battle to a different focus, or sidestep them by concentrating on a different segment, channel, or marketing element (Steuernagel, 1999; Valdani, 2000, 2003).

The New Players

The competitive landscape for network providers is changing fundamentally. While the competition for traditional carriers was in the past mainly caused by competitive local exchange carriers (CLECs), new types of competitors are emerging:

- New broadband voice carriers, such as Vonage and Delatthree/iConnect in the United States, BroadSIP/DigiSIP and Bredbandsbolaget (B2) in Sweden, and KDDI and Yahoo Broadband in Japan, which offer flatrate-based voice services over an existing broadband infrastructure at very low rates.
- CATV providers are offering triple-play services and providing a value add with a unique service bundle of CATV, broadband Internet and voice services (i.e., Fastweb in Italy).
- Software companies such as Microsoft and media giants such as Sony are trying to break into the market as well.

There is also a shift from fixed to mobile, and this shift is going to increase as mobile providers are decreasing prices and are going to offer attractive services such as multimedia messaging (MMS), which are not immediately available in the fixed networks. Furthermore, Internet service providers (ISPs) are turning

the IP pipe into a commodity by making it available at very economic rates. A shift in the revenue chain is happening where providers go from providing simply connection to also providing services and content.

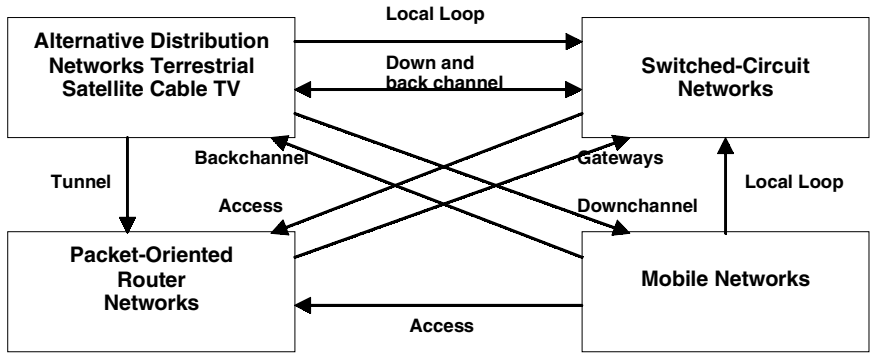
Today we can observe that multiple separated network technologies are heading towards convergence, and it may be expected that in the future there will be a single network to carry all heterogeneous kinds of services.

Figure 4 shows the convergence of different networks:

- Satellite, cable, and terrestrial networks can carry tunnels for IP-based routed networks;
- Access to routed networks from ISPs is possible via switched circuit networks (SCNs) or mobile networks (=dial in);
- Access between IP telephony and the SCN is possible via gateways;
- Local loop access is possible via SCN, cable or mobile networks;
- Down channel for broadband services; and
- Back word channel via SCN or mobile networks, which enable satellite services to be interactive.

It can be seen that the networks are growing together, and offer the possibility for new "convergent" services.

Figure 4. Network convergence. Source: Adapted from Eutelis Consult: "Trends und Konvergenzen im Telekommunikations — und Rundfunk-Markt", 1996

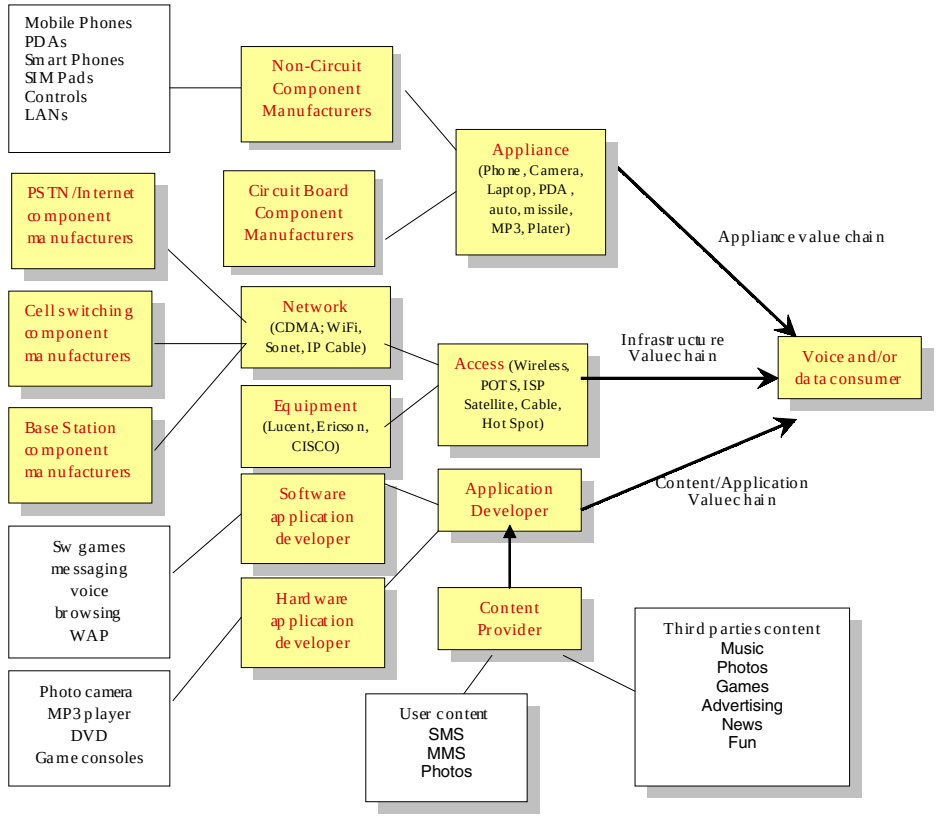


The Value Chain Strategy Framework

To approach competitive analysis the traditional two dimensions of concurrent engineering (Fine, 1998; Fleischer and Liker, 1997; Nevins and Whitney, 1989; Ulrich and Eppinger, 1994;)[3] such as products and processes are insufficient to ensure competitive advantage. For this reason, the research question is what must be added to bring the theoretical model in line with current and future realities? The answer to this question lies in the design and development of the supply chain. As it is evident in the wireless market, the supply chain forms the third axis of concurrent engineering. Taken with process and product design, it invites us to look at concurrent engineering in three dimensions rather than the traditional two, and it thereby offers even very successful companies a significant opportunity to establish and enhance their competitive advantage.

To approach competitive analysis in the 3G wireless industry, we first identify the elements of the value chain. For this reason, in order to bring the theoretical

Figure 5. Wireless supply chain structure



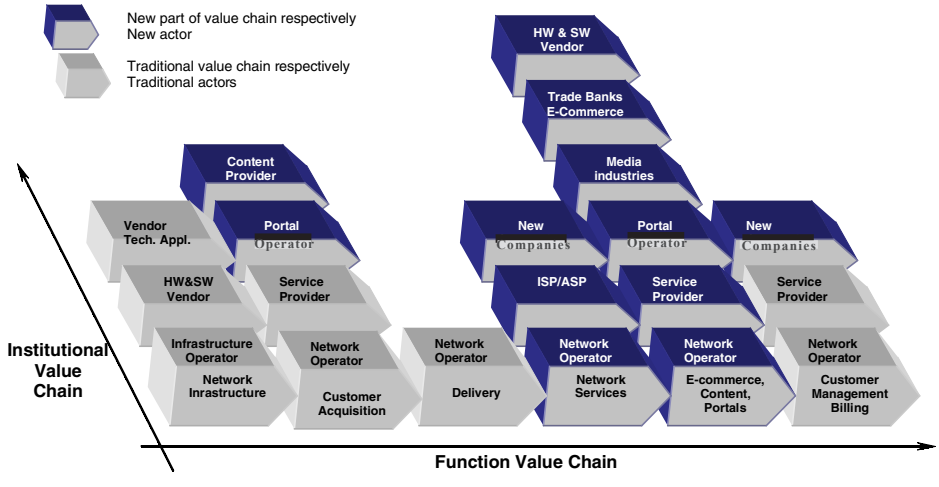
model in line with current and future realities we first need to consider the design and development of the supply chain. The three chain maps illustrated in Figure 5 indicate three levels of supply chain mapping that can be used to identify various pitfalls and opportunities in the wireless chain.

a. Appliance Value Chain
b. Infrastructure Value Chain
c. Content/Application Value Chain

The supply chain forms the third axis of concurrent engineering. Analyzing service and process design problems at the architecture level provides a strategic, high level perspective on how supply chain design can be integrated into concurrent engineering (Fine, 1998).

The value chain in the telecommunications industry has seen a huge change over the past few years. This tendency will intensify over the next period of time. Figure 6 shows the horizontal integrations (telephony companies buy or part-own

Figure 6. Changing value chains. Source: Adapted from Wissenschaftliches Institut für Kommunikationsdienste GmbH, Bad Honnef, Entwicklungstrends im Telekommunikationssektor bis 2010', April 2001



other telephony companies) and vertical integrations (companies start operating in business markets which have not been in their traditional scope).

The emerging industry structure is disintegrated. Three bilateral links are emphasized:

- The service/product needs to reflect what the customer wants, his needs and priorities;
- Processes are influenced by technology evolution;
- Corporate Strategy is influenced by the supply chain analysis.

Market Attractiveness Analysis

The transition to an IP network for voice and data has significant implications for the mobile industry. Not only does an all-IP network introduce the possibility of interoperability with fixed IP networks, it also facilitates market entry by non-traditional competitors. Cost structure, product portfolios, partnership arrangements, and industry structure can be impacted. Some of the potential changes may provide incumbent mobile network operators with renewed growth and profitability. Other changes will increase the intensity of market competition. Expected changes in market attractiveness indicators based on the migration to IP-based services are:

- Barriers to market exit will be lowered since services are less tightly integrated to network elements and new services can be tested and launched more quickly and at lower cost. Unsuccessful services also can be withdrawn more quickly. This is a benefit to the incumbent mobile network operator and increases market attractiveness.
- Customer power will increase. Customers will have more service provider choices and more options of how to access mobile and fixed services. This decreases the market attractiveness for the incumbent mobile network operator.
- The nature of competition will change. Voice will become even more price competitive, but this will be offset by increased differentiation possibilities of value-added mobile data services. This is positive for the incumbent mobile network operator and increases market attractiveness.
- Substitution threats will increase. There are an increasing number of substitutes for mobile IP services such as public WLAN. There also will be

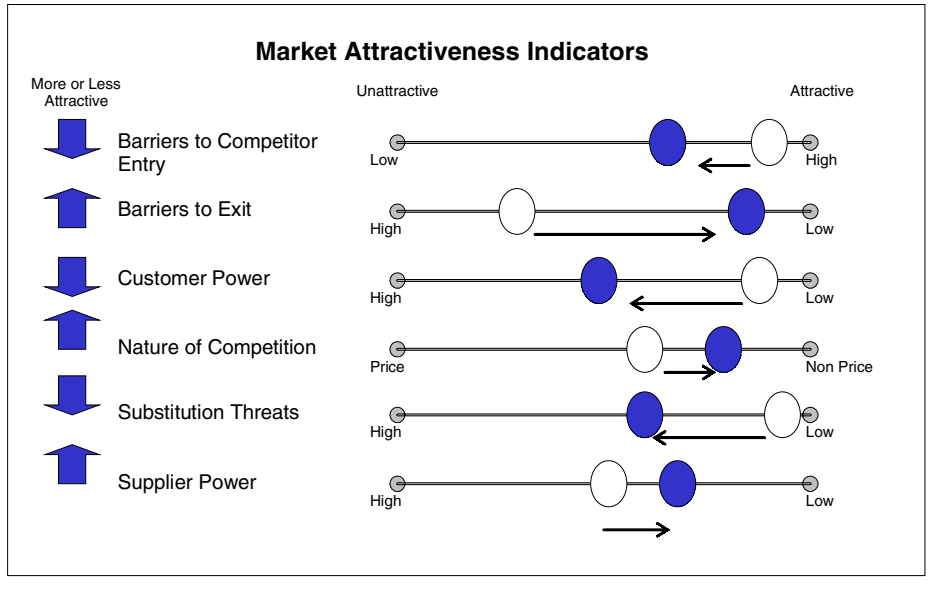
more types of devices from which access the mobile network. The technology and price distinctions between mobile and fixed networks and services are diminishing. Users are more able to substitute one service for the other. This decreases the market attractiveness for the incumbent mobile network operator.

- Supplier power will decrease. Network operators will be less dependent upon specific network infrastructure providers. The wider supplier choice increases the market attractiveness for the incumbent mobile network operators.

Figure 7 illustrates these changes showing the expected movement in each market attractiveness indicator starting from the open circle (present) and moving towards the solid circle (future). The on the left indicate whether the movement is positive (up arrow) or negative (down arrow) from perspective of the incumbent network operator or service provider.

Despite the increased competition and substitution threats, the market will most likely become more attractive with the migration to IP networks and the addition of value added services. In the future, incumbent operators will likely see increased market growth as well as new substitution threats and lower barriers for competitor entry.

Figure 7. Trends in market attractiveness indicators — most likely case



The Competitive Casual Loop

Successful liberalization has significant implications for the mobile industry; in particular it encourages entry of new service providers. Combined with the effects generated by spectrum availability and global standards it influences also costs of access network. Expected changes in market attractiveness indicators based on the migration to IP-based services indicate that barriers to market entry will be lowered. More importantly, the decoupled service creation functions will allow application and other service providers to capture a greater portion of the mobile value chain.

Barriers to market exit will be lowered since services are less tightly integrated to network elements and new services can be tested and launched more quickly and at lower cost.

Following the entrance of new service providers, made easier by low entry barriers, the nature of competition changes, and this will be offset by increased differentiation possibilities of value-added mobile services (service attractive-

Figure 8. The competitive casual loop

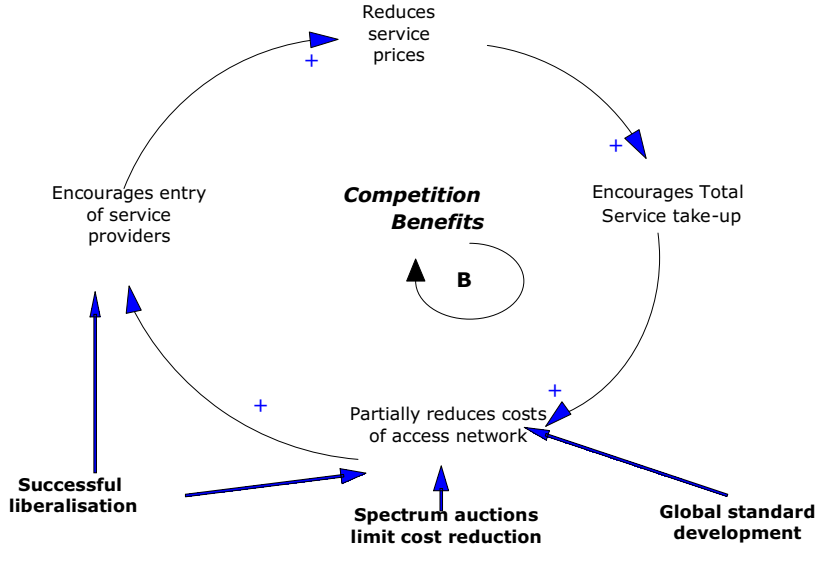
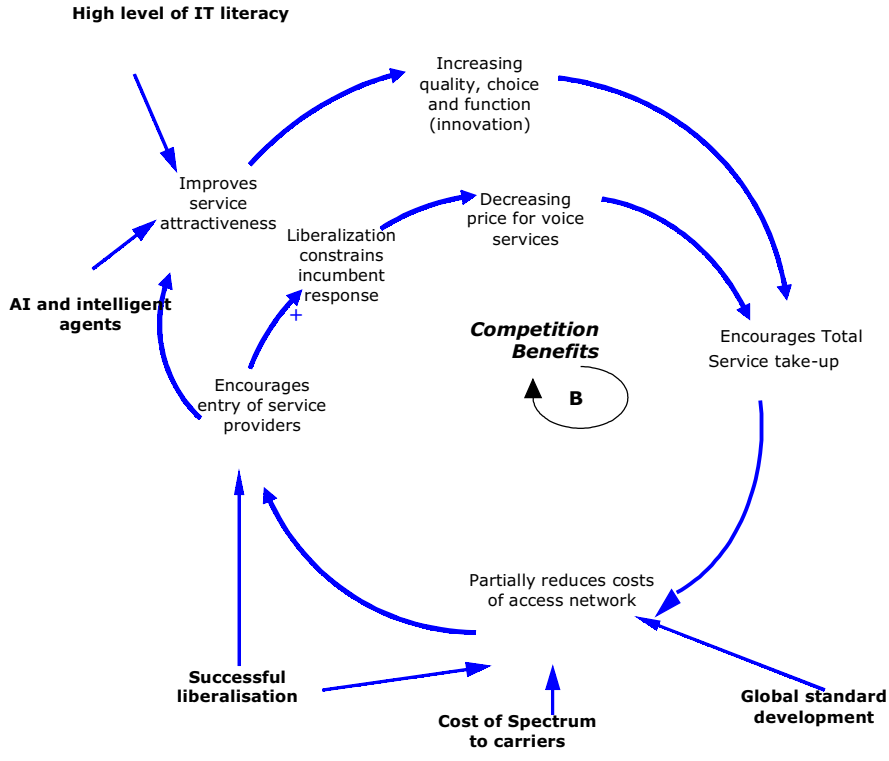


Figure 9. The new competitive casual loop



ness). The following critical questions emerge at the competitor level and application level:

1. *Competitors:* Do we want to differentiate ourselves from the competitors? Can we do so by choosing a different technology? Can we achieve a competitive advantage?
2. *Applications:* What is potential for success of the currently available applications in the served market? How expensive are they?

In order to create a competitive advantage, incumbent operators, and service provider increase service attractiveness.

This is positive for the incumbent mobile network operator and increases market attractiveness. We model this effect adding a new casual loop (Figure 9).

Market Competitive Strategies

Increasing volatility/uncertainty in the current telecommunications environment and speed of technology innovation prospect different future business scenarios for mobile operators. Without any doubt, licensed and unlicensed wireless broadband technologies represent one of the key dynamic variables to drive and characterize heavily the potential value of each scenario. Technology innovation leads towards a fragmented scenario, a complex competitive framework, driven by regulatory decision and incumbent player capability to run the scene.

In order to maintain the leading role for mobile operators, it is necessary to polarize around their capabilities and assets, technological solutions and reference scenario, to avoid value erosion and market slowdown.

The introduction and maturity of disruptive/evolutionary wireless broadband technologies push the transition from a typical telecommunication business model (with high cost, reliability and quality, authentication and security, customer lock in, and vertically integrated structure) to a new disruptive scenario: new players enabled by unlicensed spectrum technology realities, agile business model, and good quality/price mix, new service architecture and technology paradigm, which could heavily challenge the mobile operator. Table 1 shows some probable future scenarios, characterized by different threats and opportunities for mobile operators.

Table 1. Macro scenarios

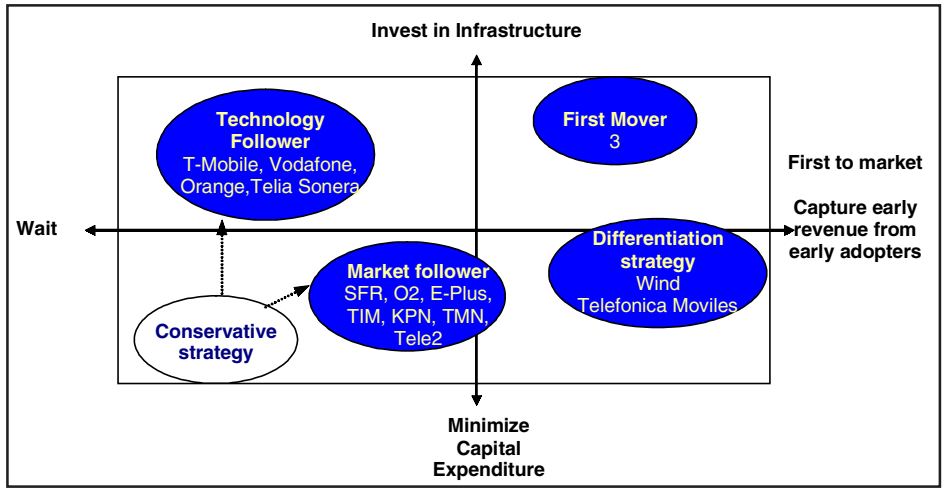
Mild scenario Big players' success in a consolidated telco scenario	**Heavy Scenario** Raise of disruptive phenomenon/technologies in a rapidly growing wireless scenario
- Reduced competition - Big players predominance - Close control of governments/authorities - Big players become moguls in almost all of the reference markets - Operators and service providers become successful as global companies - Focus on user convenience, safety and security - No wireless explosion but satisfaction of the markets confronted	- Rapid growth of industry/use of services - Rapid technological development/evolution - Old telcos lose ground in favour of datacoms - Submission of the wireless industry to disruptive phenomena, therefore a. Remarkable increase in competition b. New spectrum releases, especially for a non-licensed use c. Ad hoc deployed networks d. Do-it-yourself access networks

In analyzing competitive dynamics we can describe five competitive strategies that differ in their aggressiveness in launching new services and deploying new technology:

1. Conservative strategy
2. Technology Follower
3. Differentiation strategy
4. Market Follower
5. First Mover

The "Technology Follower" and "Conservative strategy" approaches illustrate the two Internet Mobile Services network deployment extremes, respectively, of fully deploying Internet Mobile services in accordance with the 3GPP standards (when the standards are commercially developed) and not deploying Internet Mobile Services at all. Both these approaches suggest a conservative market entry strategy that prefers to wait rather than be first to market with new services. The "Technology Follower" approach was adopted in Europe by T-Mobile (in Austria, UK and Germany), Vodafone (in the Netherlands, Portugal, Italy, Germany, Spain, Ireland and UK), Orange (in France and UK) and Telia Sonera (in Finland). The traditional incumbent operators adopted a conservative strategy until 2004.

Figure 10. Competitive strategies comparison in Europe



In both the "First Mover" and "Differentiation Strategy" approaches, the mobile operator's primary objective is to be first to market and to capture as much early revenue as possible. However, in the first case (first mover), the mobile operator intends to comply with IMS industry standards, while the second case (differentiation strategy) mobile operator takes a wildcard approach and deliberately chooses a closed network. An example of first mover is represented by the operator "3" in Italy, UK, Sweden, Austria, Denmark and Ireland.

"Market Follower" represents the most conservative Internet Mobile Services investment strategy and minimizes a mobile operator's infrastructure investments by waiting to build its 3G network until the 3GPP standards are commercially available and the market demand more developed. This approach was followed by national incumbents such as SFR in France, O2 and E-Plus in Germany, TIM in Italy, KPN Mobile in the Netherlands, TMN in Portugal, and Tele2 in Sweden. All of these operators launched 3G services in 2004 (Figure 10).

We assume that for each competitive strategy the operator is making a rational deployment decision based upon the existing individual market conditions and strategies. The profile of a hypothetical operator and market strategy is summarized in Table 2. Each strategy represents a distinct deployment choice and a distinct set of market conditions and strategies.

Table 2. Operator profile summary — five competitive strategies

	Market Strategy	Value Proposition	Examples
First Mover	- Differentiation - Segmentation-based Product Portfolio - Pursue new services and niches - Aggressively beat competition	- Competitive prices - "Get it here first" - Global ubiquity and interoperability	**3**: Italy (3/2003) UK (3/2003) Australia (3/2003) Sweden (5/2003) Austria (5/2003) Denmark (10/2003) Ireland (10/2003) Hong Kong (1/2004) Israel (10/2004) **Cingular AT&T**: USA (7/2004)
Conservative strategy	- Price Parity - Narrow Product Portfolio - Grow Usage/Traffic	- High-speed Mobile Access to Fixed Internet - Good Quality Voice - Network Transmission Quality and Speed	All national incumbent operator
Technology Follower	- Competitive Price - Broad Product Portfolio - Sustain and build from existing customer base - Meet competition as necessary	- Competitive prices - Single Source Provider	**T- Mobile**: Austria (12/2003) UK (2/2004) Germany (4/2004) **Vodafone**: Netherlands (2/2004) Portugal (2/2004) Italy (2/2004) Germany (2/2004), UK (2/2004), Spain (2/2004) Ireland (7/2004) **Orange**: France (2/2004), UK (7/2004) **Telia Sonera** : Finland (12/2003)
Differentiation Strategy	- Differentiation - Unique Product Portfolio - Redefine market	- Competitive value - Unique services	**NTTDoCoMo**: Japan (10/2001) **Wind**: Italy (2004)
Market Follower	- Price Leadership - Voice-focused Product Portfolio - Sustain and build from existing customer base - Meet competition as necessary	- Value pricing - "Good enough" new services - Global interoperability	**SFR**: France (5/2004) **O2**: Germany (4/2004) **E-Plus**: Germany (6/2004) **TIM**: Italy (5/2004) **KPN Mobile**: Netherlands (7/2004) **TMN**: Portugal (4/2004) **Tele2**: Sweden (6/2004)

Service Capabilities for the Five Competitive Strategies

In this section we look closer at five entry strategies in order to discuss in more details network deployment strategies, cost structures and service capabilities. These differing service capabilities support the overall market strategy, value proposition and revenue potential described for each competitive strategy.

Conservative Strategy

In this first option, the operator makes a clear decision not to deploy Internet Mobile Services. It is assumed that a full 3G network is required and that it is implemented in a way that enables full interworking with other fixed and mobile networks.

The 3G network provides conventional circuit-switched voice plus broadband packet data services. Furthermore, the incumbent operator in this scenario has a 2G network that also can deliver voice and data (2.5G). Internet Mobile Services such as Rich Voice are not available.

The long-term direction of fixed, mobile and corporate networks is to move away from ISDN based technology to all-IP implementation. This will ensure the integration of voice communications with data and information services, so enabling advanced services to be provided together with improved efficiency obtained by having only one integrated network to manage rather than two networks employing different technology. This option can be considered as the baseline against which the other competitive strategies are compared with regard to costs and revenues.

Technology Follower

In this option, the operator deploys both 3G and Internet Mobile Services. Both are implemented in a way that enables full interworking with other fixed, mobile and IP networks. Because this operator finds it most valuable to use standards-based technology, Internet Mobile Services are deployed as soon as commercially available. The assumption has been made that the deployment is fully according to 3GPP specifications, including IP based Radio Access Network (RAN) and that the capacity and quality of the IMS network is that it can carry a significant proportion of the voice traffic of the mobile network.

Differentiation Strategy

In this third option, the operator deploys both 3G and Internet Mobile Services. However, both are implemented in a way that discourages or even prevents full interworking with other fixed, mobile and IP networks. The operator believes that by deploying a network ahead of available standards or with features that are not included in any standard, he will be able to deliver unique services ahead of the competition that more exactly meet his target market's needs. A particular example might be an operator who would deploy a Microsoft NetMeeting server together with some of the IMS capability and other proprietary technology in order to provide a popular service to his users. Thus, the Closed IMS scenario could be regarded as a competitive threat to less adventurous providers. There is no deliberate intention to prevent interworking, but the likelihood is that technological incompatibilities due to the early nonstandard deployment would act to make interworking difficult.

Therefore, 3G network is deployed in a way that restricts interworking for data services as well as for IMS Services (Rich Voice). However, it is assumed that voice phone calls would interwork with other networks (particularly fixed networks) and that if a 2G network is available, then it would not be restricted in any way.

Market Follower

In this option, the operator deploys both 3G and IMS, but only when IMS is commercially available.

Because IMS then provides for voice as well as the more advanced data services, there is no need to deploy the Circuit Switched domain of the 3G network. However, the network is implemented in a way that enables full interworking with other fixed, mobile and IP networks. The IMS network in this scenario will be deployed exactly to 3GPP specifications (including IP RAN), which have not yet been completed, with both 3G and IMS services starting later.

First Mover

In this option, the operator fully deploys both 3G and IMS with full interworking with other fixed, mobile and IP networks. However, prior to the availability of the fully specified IMS system, the operator also deploys a small subset of the SIP/IMS network without the full features of 3GPP IMS, but with sufficient

capability to allow the start of a few early innovative but limited IMS services such as the dispatch service (Push-to-Talk). This network strategy is consistent with the operator's "First to Market" and intention to create competitive advantage.

So for example, at the start of the early deployment, the SIP/IMS Push-to-Talk service would likely be more suitable for consumer users. However, at the time that a fully specified IMS is available, the SIP/IMS Dispatch service should be suitable for professional users such as the emergency services, and the IMS network would then also have the quality and capacity to carry a substantial part of the conventional voice telephony services also.

Competitive Strategies in the European Market

In Europe, Hutchison Whampoa's W-CDMA brand "3" was the first mover entering Austria, Denmark, Ireland, Italy, Sweden and UK.

Since launching into the European market in Q1 2003, Hutchison's 3G service has had to resolve a number of device-related issues. Handsets associated with the service were initially criticized for their short battery life, a problem Hutchison has acknowledged and addressed by providing users with complementary extra batteries. The launch also has been troubled by a shortage of compatible handsets. In Italy in Q4 2003, for example, it is alleged that 3 had a waiting list of around 100,000 subscribers all demanding new handsets.

Hutchison believed that focusing on 3G could offer significant advantages in the design and deployment of next generation networks. At times, however, health and environment issues have stood in the way of coverage development. In the UK, for example, the operator has been forced to deal with resistance from local councils, MPs and residents who fear the environmental impact of additional radio masts and perceived but scientifically groundless health risks. As an interim measure 3 has been using a GPRS roaming service provided by O2 to boost its coverage in the UK. On a more positive note, as a new entrant 3 clearly benefits in some respects from the absence of an existing 2G subscriber base—for example, the operator enjoys more flexibility than incumbents with regard to 3G recruitment strategies. In addition, as a first mover Hutchison has been able to capture early-adopter market share and gain valuable field experience of 3G technology.

For the European market as a whole, the arrival of a new operator in well established 2G and 2.5G markets has created some turbulence. To drive take up and build market share, Hutchison has offered increasingly attractive consumer deals. In addition to slashing voice and cross network tariffs, Hutchison has offered its customers free samples of advanced data service offerings and specially designed video clip packages for particular market segments. Incumbents will need to match, better or at least take into account such offerings when they launch rival services.

Despite such marketing initiatives, Europe's new operator has still found it tough to hit its own commercial targets. Now Hutchison was revising its target of one million subscribers in Italy and UK. With more established European operators like Vodafone, Orange, T-Mobile, and TeliaSonera launching 3G products, the big question is whether or not Hutchison be able to endure in the European market.

There is a fine line between the gains of first mover advantage and the dangers of disappointing an embryonic market with underdeveloped product. Well-established operators like Vodafone, T-mobile, and Orange have weighted the risk of an early entry into 3G and decided to put their plans on hold. 3 has undoubtedly gained first mover advantage with video calls, but will need to be extremely competitive and tactically nimble once the novelty wears off and rivals begin to launch competitive offerings.

So what are the key factors that will determine take up of 3G services in Europe? In an enhanced data environment, content will be king. Superior content will be one of the main reasons that subscribers will switch to 3G networks. In time it will be one of the reasons that subscribers switch between 3G operators. Once 3G markets mature, the retention of subscribers in 3G networks will be determined by the quality and range of content and services provided by the operator. Service providers that secure attractive mobile content services at an early stage will have an advantage in attracting and retaining subscribers.

Conclusions: Changes in Competitive Advantages

From this study, we tried to draw dynamic forces that influence competitive dynamics in the 3G wireless industry. It emerges at a first level that the value chain in 3G and 4G wireless systems is highly horizontal, reflecting the multiplication of the required investments and competencies. The industry structure is disintegrated.

Three bilateral links emerge:

- The service/product needs to reflect what the customer wants, his needs and priorities;
- Processes are influenced by technology evolution
- Corporate Strategy is influenced by the supply chain analysis

Within the foreseeable future, the industry structure will most likely remain basically the same — large players remaining large with smaller players gaining some share. There are no significant shifts in market demand or industry structure. Internet Mobile Services gain moderate market acceptance by end users and therefore add revenue benefits to those that have implemented IMS. This is a relatively stable situation, fairly conservative, with no major disruptive forces.

Operators developing a differentiation strategy will gain some market share, but still be a small, niche player. Those operators not offering IMS or offering it later, will lose some competitive differentiation advantage, but may gain some cost advantage due to lower infrastructure investment and possibly lower cost of debt.

References

Barnes, S. J. (2002). The mobile commerce value chain: Analysis and future developments. *International Journal of Information Management, 22*(2), 91-108.

Bouwman, H. & Ham, E. (2003). Designing metrics for business models describing Mobile services delivered by networked organizations. Paper presented at the *16th Bled Electronic Commerce Conference eTransformation Workshop on Concepts, Metrics & Visualization*, Bled, Slovenia.

Camponovo, G. & Pigneur, Y. (2002). Analyzing the actor game in m-business. Paper presented at the *Proc. First International Conference on Mobile Business*, Athens.

D'Aveni, R. (1994). *Hypercompetition*. New York: The Free Press.

Day, G.S. & Rubenstein, D.J. (1997). *Wharton on dynamic competitive strategy*. New York: John Wiley and Sons.

Fine, C.H. (1999). *Clockspeed*. Cambridge, MA: Perseus Books.

Fine, C.H., Vardan, R., Pethick, R., & El Hout, J. (2002, Winter). Rapid-response capability in value chain design. *MIT Sloan Management Review, 43*(2), 69-75.

Fleischer, M. & Liker, J. (1997). *Concurrent Engineering Effectiveness*. Gardner Publications, Cincinnati, OH.

Kleijnen, M.; Ruyter, K. & Wetzels, M.G.M. (2003). Factors influencing the adoption of mobile gaming services. In B.E. Mennecke & T.J. Strader (Eds.), *Mobile commerce – technology, theory and applications* (pp.213-214). Hershey, PA: Idea Group Publishing.

Li, F. & Whalley, J. (2002). Deconstruction of the telecommunications industry: From value chains to value networks. *Telecommunications Policy, 26*(9-10), 451-472.

Maitland, C.F., Bauer, J.M., & Westerveld, R. (2002). The European market for mobile data. *Telecommunications Policy, 26*(9-10), 485-504.

Nevins, J. & Whitney, D. (1989). *Concurrent design of products and processes: A strategy for the next generation in manufacturing.* New York: McGraw-Hill.

Olla, P. & Patel, N.V. (2002). A value chain model for mobile data service providers. *Telecommunications Policy, 26*(9-10), 551-571.

Owen, B. (1999) *The Internet challenge to television* (p. 347). Cambridge, MA: Harvard University Press.

Pagani, M. (2004). Factors that affect adoption of third generation mobile multimedia services. *Journal of Interactive Marketing, 18*(2).

Pigneur, Y. (2000). *An ontology for m-business models*. University of Lausanne, Ecole des HEC, CH-1015 Lausanne.

Sabat, H. K. (2002). The evolving mobile wireless value chain and market structure. *Telecommunications Policy, 26*(9-10), 505-535.

Sterman, J.D. (2000). *Systems thinking and modeling for a complex world.* McGraw-Hill/Irwin.

Steuernagel, R.A. (1999). *Wireless marketing*. New York: John Wiley & Sons, Inc.

Talluri, S., Baker, R.C. & Sarkis, J. (1999). A framework for designing efficient value chain networks. *International Journal of Production Economics, 62*(1-2), 133-144.

Ulrich, K. & Eppinger, S. (1994). *Product design and development.* New York: McGraw-Hill.

Valdani, E. (2003). *Competition-based view — I giochi competitivi di movimento, imitazione e posizione*, ETAS, Milano.

Valdani, E. (2000). *L'impresa Pro-Attiva*. Milan: McGraw-Hill.

Wirtz, B.W. (2001). Reconfiguration of value chains in converging media and communications markets. *Long Range Planning, 34*(4), 489-506.

Endnotes

1 See C.H. Fine, "Clockspeed" (Cambridge, Massachusetts, Perseus Books, 1999)

2 According to Valdani (2003), a quantum leap originates when an extraordinary and unforeseeable event, caused by a new technology or application, happens. This event generates a change that can be transformed into a killer application.

3 See, for example, James Nevins and Daniel Whitney, *Concurrent Design of Products and Processes: A Strategy for the Next Generation in Manufacturing* (New York: McGraw-Hill. 1989); K.Ulrich and S. Eppinger, *Product Design and Development* (New York: McGraw-Hill, 1994); Mitchell Fleischer and Jeffrey Liker, *Concurrent Engineering Effectiveness* (Cincinnati: Hanser Gardner Publications, 1997).

Section II

Determinants of Mobile Technology Adoption

Chapter II

Corporate Adoption of Mobile Cell Phones: Business Opportunities for 3G and Beyond

G. Keith Roberts, University of Redlands, USA

James B. Pick, University of Redlands, USA

Abstract

This chapter identifies the technology and non-technology factors that companies consider important in deciding to adopt and deploy wireless devices designed for mobile telephony and information services, the extent of current use of cell phones, the extent of existing utilization and/or planning for Web-enabled cell phone use, the constraining factors in their deployment decisions, how such decisions are made, and the practical technology implications for decision-making, including beyond 3G. This

chapter seeks to help decision makers by shedding light on the adoption process. The conceptual model combines the TAM and innovation adoption/ diffusion models, adding the factors of security, cost, reliability, digital standards/regulatory environment, technology product suitability, and future Web connectivity. Case study methodology is utilized for five manufacturing and technology firms. A key finding is that the most important technology decision factors are security, reliability, and Web connectivity. Although the current uses are dominated by voice, Web-enabled capability dominates future decision-making.

Introduction

Mobile devices have been among the fastest adopted consumer products of all time (Clarke III, 2001). Subscribers for mobile telephony services in the United States through December 31, 2002, stood at 141.8 million, which equates to a nationwide average population penetration rate of 49 percent (Federal Communications Commission, 2003). While such a penetration rate is significant, there are other areas of the world that are much higher (e.g., 80 percent in Western Europe (Federal Communications Commission, 2003) and over 90 percent in some countries, such as Sweden) for the same time period. It has been estimated that this year (2003) there will be 1.4 billion mobile phones worldwide, with half of them capable of being Internet-enabled (Clarke III, 2001). An estimated 11.9 million in the U.S. subscribe to mobile Internet service and an estimated 21 percent of all Web-enabled mobile phone users in the U.S. (7.5 percent of all mobile phone subscribers) actually use the phones to browse the Internet (Federal Communications Commission, 2003). Jeff Bezos, the CEO and founder of Amazon.com, believes that in five to 10 years almost all e-commerce will be done with wireless devices (Clarke III, 2001). The benefits to users include removal of space and time constraints, better access to decision makers, better reception of information about an organization and its environment, and improved social networking (Davis, 2002; Palen, 2002; Mennecke & Strader, 2003), while disadvantages may include greater security and privacy intrusions, interruption of business work and personal life, and social improprieties (Davis, 2002; Palen, 2002). Regardless of which time estimate is correct, the point is fast approaching when more people will be likely to access the Internet through a mobile device than through a personal computer. Just as the general population is increasingly dependent upon wireless communication devices for both entertainment and commerce, corporations are increasingly considering cell phones as a critical success factor to conducting business. This chapter focuses on identifying the technology and non-technology factors that corporations consider important in

their decision to deploy devices designed for mobile telephony and mobile data services. We also consider the approval steps in decision-making, the extent and importance of 3G and beyond as it relates to Web-enabled cell phones, and the functional areas of use of cell phones.

Background and Literature Review

There has been little research regarding corporate adoption of wireless (mobile) devices, but there is a solid foundation of theories and previous studies on technology adoption (Kleijnen & de Ruyter, 2003; Van Akkeren & Harker, 2003). The decision by a company to utilize cell phones in its business, is in essence a technology adoption issue. A number of theories have been developed to help explain the concept of technology adoption (Mennecke & Strader, 2003; Kleijnen & de Ruyter, 2003). One widely accepted model is the Technology Acceptance Model (TAM) (Davis, 1989, 1993). Davis (1989), in an innovation adoption and diffusion model, emphasized the theoretical constructs of perceived usefulness and perceived ease of use as a means of predicting user acceptance of information technology. Adams, Nelson and Todd (1992) replicated Davis' research for fixed voice and e-mail. They refined the measurement scales and utilized structured equation modeling to explain interactions. In later research using the TAM model, Davis' results indicated that while ease of use is clearly significant, usefulness is even more important in determining user acceptance (Davis, 1993). Lederer, Maupin, Sena, and Zhuang (2000) investigated TAM for work-related tasks involving the Web. Their findings provided support for TAM and also corroborated that usefulness has a stronger effect than ease of use.

Rogers (1995) identifies five attributes of an innovation that help to explain the rate of technology adoption: (1) relative advantage (degree to which innovation is perceived as being better than the idea it supersedes), (2) compatibility (degree to which innovation is perceived as consistent with existing values, past experiences, and needs of potential adopters), (3) complexity (degree to which innovation is perceived as relatively difficult to understand and use), (4) trialability (degree to which innovation may be experimented with on a limited basis), and (5) observability (degree to which results of innovation are visible to others). In his discussion of the attributes of innovation, Rogers states: "Cellular phones have an almost ideal set of perceived attributes, and this is undoubtedly one reason for the innovation's very rapid rate of adoption in the U.S." (1995, p. 245). Rogers then describes how cell phones meet all of his attributes.

The Davis and Rogers models are both widely supported and followed, and also are complementary. Davis's two main constructs can fit quite nicely within the

Rogers model. Specifically, usefulness is similar to Roger's factor of relative advantage and ease of use is similar to Roger's factor of complexity (Agarwal & Prasad, 1997).

The Rogers factors were enlarged to include perceived risk (Eastlick & Lotz, 1999). We include this since cell phones are vulnerable to security and privacy violations. Another specific factor for cellular devices is payment and cost (Kleijnen & de Ruyter, 2003) and we likewise include it. Since studies of mobile adoption (Kleijnen & de Ruyter, 2003; Van Akkeren & Harker, 2003) point to present applications dominated by voice communications and simple Internet, but a future of complex Web, Internet, and e-commerce enhanced uses, we have added Web connectivity as a factor. Our pilot study emphasized concern in businesses for reliability of mobile devices, the importance of technology product suitability, digital standards/regulatory environment, and Web-connectivity, and hence we include them.

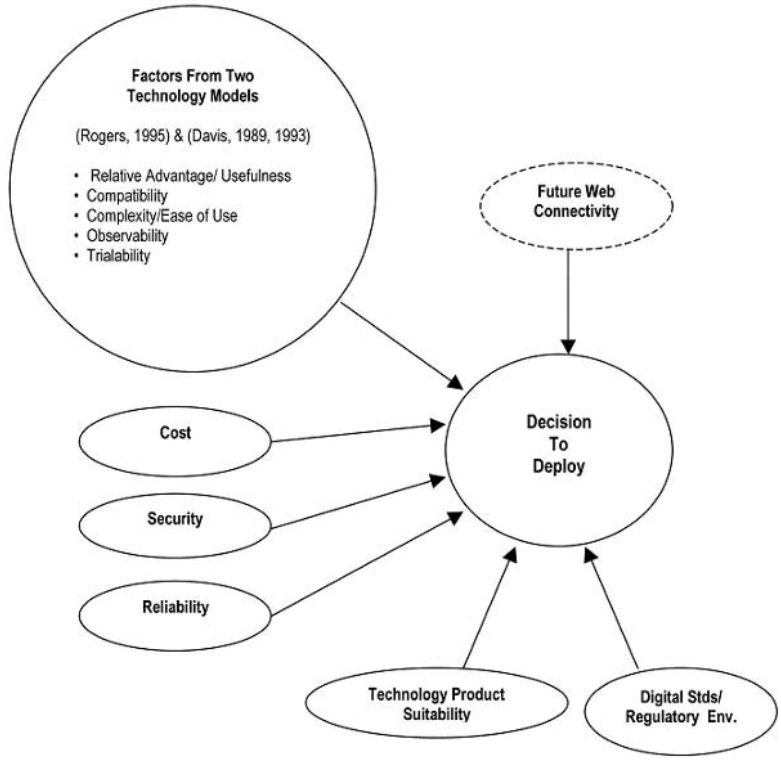
The regulatory environment of the wireless industry in the U.S. is distinctly different versus the wireline industry. A review of the telecommunications regulatory history reveals the telephone industry has transitioned from a period of competition in the late 1800s to de facto monopoly in the early 1900s, for example, AT&T, to regulated monopoly in 1934 (Communications Act of 1934) to a breakup of AT&T into "natural" monopoly pieces in 1984. Eventually the Telecommunications Act of 1996 was passed with the purpose of reversing the concept of natural monopoly and to encourage competition (Black, 2002). Congress realized that evolution from monopoly to competition could take years if it depended upon wired systems. A much quicker approach was to acquire local infrastructure by encouraging competition through wireless (Black, 2002). As a result, Commercial Mobile Radio Services (CMRS), which includes cell phones, was exempted from certain provisions of the 1996 Act. Section 332 (c) of the Act excludes CMRS from the definition of "local exchange carrier" (47 U.S.C. § 153[26]). Additionally, a state or local government is normally excluded from regulating the entry of or the rates charged by CMRS (47 U.S.C. § 332[c][3][A]).

Price competition in the U.S. wireless industry is described as "intense", "fierce", and "ultra-competitive" (Federal Communications Commission, 2003). An important data point to help determine what effect the intense wireless competition is having on the ultimate user of the services in the U.S. is to look at the Cellular CPI, which is the cellular telephone services component of the Consumer Price Index (CPI). From 2001-2002, the annual Cellular CPI decreased by 1.0 percent, while the overall CPI increased by 1.6 percent. The Cellular CPI has declined almost 33 percent since 1997 (Federal Communications Commission, 2003). Given the fact the regulatory environment has encouraged rapid competition in the wireless industry and that wireless service,

according to a number of analysts (Federal Communications Commission, 2003), is now cheaper than wireline service, one of our primary research questions is to determine if the regulatory environment is an important factor in corporate decision-making.

In sum, the Davis and Rogers models and recent studies seek to explain user adoption and acceptance of technology. This paper builds upon that body of research by seeking to identify the technology and non-technology factors that corporations consider important in their decision to deploy mobile devices. The theoretical framework combines the Rogers and Davis models, and the present study adds the factors of cost, security, reliability, digital standards/regulatory environment, technology product suitability, and future Web-connectivity (see Figure 1).

Figure 1. Research model



Research Questions

Corporate decision-making can be quite complicated when adopting new technology. Wireless technology has capabilities, features and challenges that can make adoption decision-making even more difficult. Wireless devices in the corporate environment include cell phones, personal digital assistants with wireless modems, wireless laptops, two-way pagers/short message systems, and wireless networks. The overall goal of this chapter is to give corporate decision-makers better insight and knowledge into making often difficult and complex wireless cell phone adoption decisions. The focus is on cell phones for three reasons. First, cell phones typically provide the company's first wireless experience, since voice-to-voice communication is the primary service adopted by mobile devices (Roberts & Pick, 2003). Second, most of the factors that apply to cell phones are equally applicable to other wireless devices. Finally, cell phones encompass digital standards and regulatory issues that are unique to the cellular industry. The specific research questions are:

1. What are the most important technology factors in making the decision to adopt cell phones?
2. What effect has regulatory policies and the lack of digital standards had on the corporate decision to deploy cell phones.
3. To what extent are companies using or planning to use Web-enabled cell phone devices?
4. What are the constraining factors in cell phone use?
5. What is the decision-making process for cell phone adoption? Who in the organization makes the final decision about cell phone deployment?
6. What are the business functional areas of cell phone use and how is the technology used in those areas?

Research Methodology

The methodology for this research is case study (Stake, 1995; Yin, 1993, 1994). The case study strategy consists of defining the study focus, framework construction, interviews, data collection, and case analysis. Case studies are frequently utilized to gain a greater depth of insight into organizations and their decision-making processes than is available with large sample surveys (Yin, 1993, 1994). Case studies often have very small sample sizes (Yin, 1993, 1994).

The present case study sample frame was determined by narrowing the industry focus to five manufacturing, distribution, entertainment, and technology companies having different size categories, ownership characteristics, and corporate structures. Our frame encompasses this extent of differences in order to encourage a greater range of decision-making factors. The specific criteria for company selection was the following: a mixture of high tech versus manufacturing; public versus private ownership; a majority of cases to have a global presence; and at least one company whose future is closely tied to broadband communication, such as the global entertainment company. As seen in Table 1, the first company employs 450 and is America's largest distributor of an industrial product. The second company is a global leader in information systems with more than 2,600 full-time staff and distributors located around the world. The third firm is a medical systems company that employs 2,600 worldwide. The fourth company is a global technology leader. The fifth one is a global leader in entertainment.

For each firm, the study is designed to interview the chief information officer or equivalent executive, and one or two managers in charge of telecommunications that includes cell phones. Two telecommunications managers were included, if the CIO designated that two people had overlapping responsibility.

To answer the research questions, the case study method is utilized to evaluate the factors in the research model. The study evaluates the relative importance of these factors in the companies' decisions (see Table 2 [Technology Factors] and Table 3 [Non-technology Factors]).

Table 1. Sample of five companies

Case No:	Case 1	Case 2	Case 3	Case 4	Case 5
Industry:	Distributor of Industrial Products	Software Vendor and Services	Medical Products Manufacturing	Networking and Telecom Hardware	Entertainment
Products:	Industrial Plastics	GIS	Pumps/Disposables for Critical Care	Routers, etc.	Entertainment And Movies
Employees:	450	2,600	2,600	35,000	120,000
Offices Outside U.S.	No (First Overseas Office Opens in 2003)	Yes	Yes	Yes	Yes
Ownership	Subsidiary of Larger Int'l Public Firm	Private	Public	Public	Public

Interviews were carried out at both the IT decision-making and IT operational levels of each company. Interviews were conducted in person by the authors based on a standard set of interview questions. (See Appendix for the protocol of interview questions for the telecommunications manager in charge of cell phones.) Each interview lasted one to two hours. The findings were taped as well as hand recorded by the interviewers. The results were transcribed, and the interview transcripts were sent to each interviewee, who validated the information. At the time of the interview, additional supporting documents were gathered or requested. Examples include organization charts, annual reports, product reports, planning reports, and Web sites containing company and product information.

Each case met the validity criteria for case studies, in particular, construct validity, internal validity, external validity, and reliability (Yin, 1994). The construct validity came from multiple evidence sources, review of the case study transcripts by interviewees, and multiple sources of evidence (interviews and documents). Internal validity came from the construction of a detailed research framework, indicating the steps in analysis, ahead of time (Yin, 1994). External validity is limited, since this is an exploratory study, and not replicating other studies. The presence of five cases from different types of firms constrains the external validity at the most to the manufacturing, distribution, entertainment, and technology sectors. Reliability is based on a detailed case study protocol that documents the scheduling, interview procedures, recording, follow-ups, questions, and summary database (Yin, 1994).

The research framework consists of factors under the groupings of organization, cell phone decision-making, and cell phone utilization. Under organization, the factors were industry, primary product(s), firm size, firm organizational structure, and current cell phone dependency. The cell phone decision-making factors consisted of cost, success of units already deployed, bandwidth, e-connectivity, security, reliability, scalability/expandability, digital standards, technology suitability, project promoter, and level of decision-making. The cell phone utilization factors were number of cell phones deployed, extent of anticipated future deployment, uses of cell phones, and anticipated future uses. The research framework has proven to be robust, based on the interviews. Each question elicited values in the ranges expected by the research protocol, and the analyses were realizable.

Findings

This section on findings first considers the prevalence of cell phones and anticipation of Web-enabled devices. Next it examines the findings as they relate to each of the research questions.

The current dependence of the five firms on cell phones is high (Roberts & Pick, 2003). The total prevalence of corporate cell phones (equipment provided or reimbursed) varies from 21 to 40 percent of the workforce. The dominant form of cell phone use in all the firms today is voice. No firm indicated that cell phone geo-referencing was activated. The software firm (Case 2) pointed out that some of its clients do have geo-referencing activated and in use. The proportion of Web-enabled units in use is small, with a range of .025 to 12 percent of units. Many respondents emphasized that they feel that the Web-enabled technology has not arrived yet on the market. At the same time, all firms indicated high or medium/high future dependence on Web-enabled technology.

Table 2 contains the technology factors that were studied and Table 3 contains the non-technology factors. The Rogers and Davis factors were not separately listed in the tables as Rogers found all of his factors to be met in the case of cellular phones (Rogers, 1995, p. 245). The five factors were, however, distributed throughout Tables 2 and 3. For example, relative advantage is included within Productivity (Table 2), compatibility is included within Digital Standards (Table 2), complexity and customer service are included within Convenience/Ease of use (Table 3), observability is included within Outside Perception (Table 3), and trialability is included within Cost (Table 2). As seen in Table 2 (Technology Factors) and Table 3 (Non-technology Factors), firms rated the cell phone deployment factors fairly consistently.

RQ1. Currently, the most important technology factors are security, reliability, and Web-based connectivity (Table 2). Security's prominence is consistent with other studies of mobile technology. It corresponds to perceived risk being most significant for mobile gaming (Kleijnen & de Ruyter, 2003). Cell phones are known to be a less secure form of communication (Dodd, 2002), in particular messaging from cell phones can contain vital corporate information on sales, marketing, strategies, and business areas that may be of crucial importance to competitors. It also may include proprietary information or intellectual property. Hence security would surely be a high concern for any competitive firm. Security was not present in the original TAM model (Davis, 1989, 1993), most likely because security exposure was less for the forms of technology examined in the mid to late 1980s and early 1990s.

RQ2. The least important decision-making factors are federal/state regulation and digital standards. Regulation was consistently rated as of no or low importance, except for the CIOs of the software firm and entertainment

Table 2. Importance of technology fators

Case No: / Technology Factor	Case 1 CIO	Case 1 Tel. Mgr.	Case 2 CIO	Case 2 Network/ Tel. Mgr.	Case 2 Tel Indus. Solutions Mgr.	Case 3 VP-IT	Case 3 Dir. Tech. & Shared Services	Case 3 Tel. Lead	Case 4 IT Super-visor	Case 4 Tel. Mgr.	Case 5 CIO	Case 5 Tel. Mgr.
Cost:	M	M/H	M	M	H	M	H	M	H	M	M/H	M
Reliability:	M+	H	M	H	M/H	H	M/H	H	M	H	H	L
Bandwidth:	L/M	(NA)	L today	M	None	M	None now H future	None	L/M	H	L	L
Security:	H	H -	H	H	(NA)	H	M	(NA)	H	H	H	H
Expandability/ Scalability:	L/M	MI	M	M	(NA)	M	M	L	H	H	M	M
Connectivity to Web:	H	H	M	H	None now M future	L/M	None now H future	None	H	H	L Now H Future	L Now H Future
Digital Standards:**	L Today	L		M	L	L	L-M	L-M		H	L	
Technology/ Suitability:**	M-H			M-H			M	H		M-H	H	
Other - Supportability:			H	L		H	H	H			H	M
Other - Productivity:			H		H			M	M		H	L
Other - Coverage:								H				
Most Important:	Connect-ivity to Web/ Security	Ability to Provide Service to Customer	Produc-tivity	(NA)	Produc-tivity	Relia-bility	Relia-bility, Support-ability	Coverage/ Service Quality	Security	Security	Security	Security

(H) High
(M) Moderate
(L) Low
(NA) Does not know answer
(MI) Missing from interview
Other: refers to factors identified by respondents and not in the theoretical model.
** Conclusion from entire interview as specific question was not asked of interviewees.

Table 3. Importance of non-technology factors

Case No: / Non-Technology Factors	Case 1 CIO	Case 1 Tel. Mgr.	Case 2 CIO	Case 2 Netwrk Tel. Mgr.	Case 2 Tel Indus. Solut. Mgr.	Case 3 VP-IT	Case 3 Dir. Tech. & Sha'd S'vics	Case 3 Tel. Lead	Case 4 IT Super-visor	Case 4 Tel. Mgr.	Case 5 CIO	Case 5 Tel. Mgr.
Convenience to Employees; Employees' Ease of Use	M	L+	M	L/M	L	M	H	M	M	M	M	L
Ability to Provide Service to Customer	M	H	H	H	(NA)	(NA)	M	H	(NA)	(NA)	H	H
Outside Perception	M	H	M	M	(NA)	M	None	H	H	H	M	M
State/ Federal Regulation	L	None	H	L	None	L	None		L/M	M/H	M	L
Other: Opportunity Cost										M		

(H) High (M) Moderate
(L) Low (NA) = does not know answer
Other: refers to factors identified by respondents and not in the theoretical model.

company. For instance, one CIO stated in regards to FCC regulation: "It doesn't create satisfaction, I'm either going to be in a neutral state or dissatisfied." The other respondents pointed to several reasons including that cell phone regulation is the concern of the cell phone equipment makers and of service providers, not of the customer firms using the equipment and services. In other words, "let AT&T worry about it." Another reason for the low rating is that U.S. federal and state regulations of use and content are very limited; hence, why be concerned about it. One network manager rated this factor low in the U.S., but rated it as high for other countries. The reason is that most other nations have cell phone cost structures that are prohibitive, and often the costs are affected by their federal/national regulation.

Another area of impact relates to equipment standards, which may be restrictive overseas. The software firm CIO who rated U.S. federal and state regulation as high was concerned with the realms of privacy and security, which are influenced by regulation. He felt strongly that these realms could not be ignored, but rather that corporate citizens must consider them. The entertainment company CIO rated regulation as medium because of concerns for future outlets in which to push the company's content services through broadband. Further, he felt the firm was quite affected by the FCC, especially regarding regulation and legalities of intellectual property in the Web-enabled cell phone environment. The only significant interest shown by respondents in digital standards (or the lack thereof) related to the GSM standard. While GSM is one of three 2G (second generation) standards available in the U.S., it becomes critically important for the international business traveler, since GSM is the de facto standard in Europe. GSM offers the pluses of enlarged coverage and good acceptance.

RQ3. A forward-looking factor is connectivity to the Web. This was mostly rated as of high importance, although respondents in three different firms specified that it is of low/no importance now, but medium to high importance for the future. Since Web enabled devices have a low level of prevalence today (in the range of 12 percent or less of equipment deployed for this U.S.-based sample), this response is inherently forward-looking. IT management recognizes that these devices, although imperfect today, are likely to improve in their functionality and user-friendliness to become reliable Web devices in the future. Respondents had mentioned the need in sales and marketing for very rapid business communication response times in the field, often 30 minutes or less. The wireless e-mail and Web capability might ease the ability to send business communications rapidly, although, the delays inherent in typing would not change. It also might extend the use of these devices to field applications in more data-intensive sides of the business, such as inventory control, supply chain management, and operations. Web-connectivity was not a factor in the traditional models, which preceded widespread Web use in businesses.

RQ4. Only one non-technology factor, customer service, is at the high level of the three technology factors discussed above. The service to the customers is consistent with the TAM model (Davis, 1989, 1993) and subsequent TAM studies (Adams et al., 1992; Lederer et al., 2000; Gefen & Straub, 1997). For two of our sample firms (Case 1 and Case 3), the importance of service to the customer is rated highest by respondents having the lowest management reporting level. That may be because those interviewees are closer to the customers.

The software provider company put a high emphasis on productivity, a new factor not in our theoretical model. It appears to be aligned with that firm's internal goal to emphasize productivity. It is consistent with the importance in TAM of usefulness (Davis, 1989, 1993). Another firm, the medical manufacturer, added and stressed another new factor of support. This seems tailored to that firm and sector. Medical devices are becoming more sophisticated all the time, which requires increased supportability. From its experience with medical devices, the firm is sensitive to the costs associated with supportability. It is a forward-looking factor, since support of a cell phone for simple voice is not burdensome, but will multiply with Web-enabled cell phones, which that firm is actively piloting. The factors of reliability and cost were rated at medium to high. It is significant that cost is not the primary driver for these cases. We heard from several respondents that companies may not choose the cheapest alternative if the key factors of security and convenience are not met. Reliability also is a leading factor and relates to the strategic importance of cell phones, verified by all respondents. Reliability and convenience also are consistent with the TAM factor of usefulness (Davis, 1989, 1993).

RQ5. The findings also demonstrate that decision-making for cell phone adoption is quite varied and is linked to the corporate culture and organizational structure. The decision process originated with promoters and progressed through approval stages, with different routes to the final approval. The decision-makers ranged from a middle level board to the CIO and, on occasion, the CEO. This finding is practically important for vendors of mobile devices, who need to be cognizant of and adaptable to the variety of ways such adoption decisions are made. For a full discussion of the decision-making process, see Roberts and Pick (2003). Future research, with larger samples, could analyze the link of the adoption decision-making approaches to organizational and cultural types.

RQ6. The findings identify the leading functional business areas of corporate cell phone use. In all cases, sales had a high level of cell phone use. This makes sense since verbal communications are all important in competing for sales in the field. Overall, marketing was moderate to high in use. Executives were also intensive users, while middle managers had moderate to high use. Clearly, anytime-anyplace communications enhances executive strengths in voice communica-

Table 4. Findings summary

Dependence on Corporate Cell Phones	21-40% of Workforce
Dominant Cell Phone Use	Voice
Web-enabled Units	.025-12% of Units
Most Important Technology Factors	Security * Reliability Web-based Connectivity
Most Important Non-technology Factor	Customer Service
Least Important Factors	Federal/State Regulation Digital Standards
Decision-making Process	Varied—Linked to Corporate Culture and Organizational Structure
Leading Functional Areas of Cell Phone Use	Sales Marketing Executives IT

* Most Important

tions (Davis, 2002). All five firms revealed high cell phone use in IT. However, the use emphasis in IT departments was on researching, piloting, and evaluating wireless technology, for example, Roger's trialability. A full discussion of the business activity uses for this research appears in Roberts and Pick (2003).

Discussion

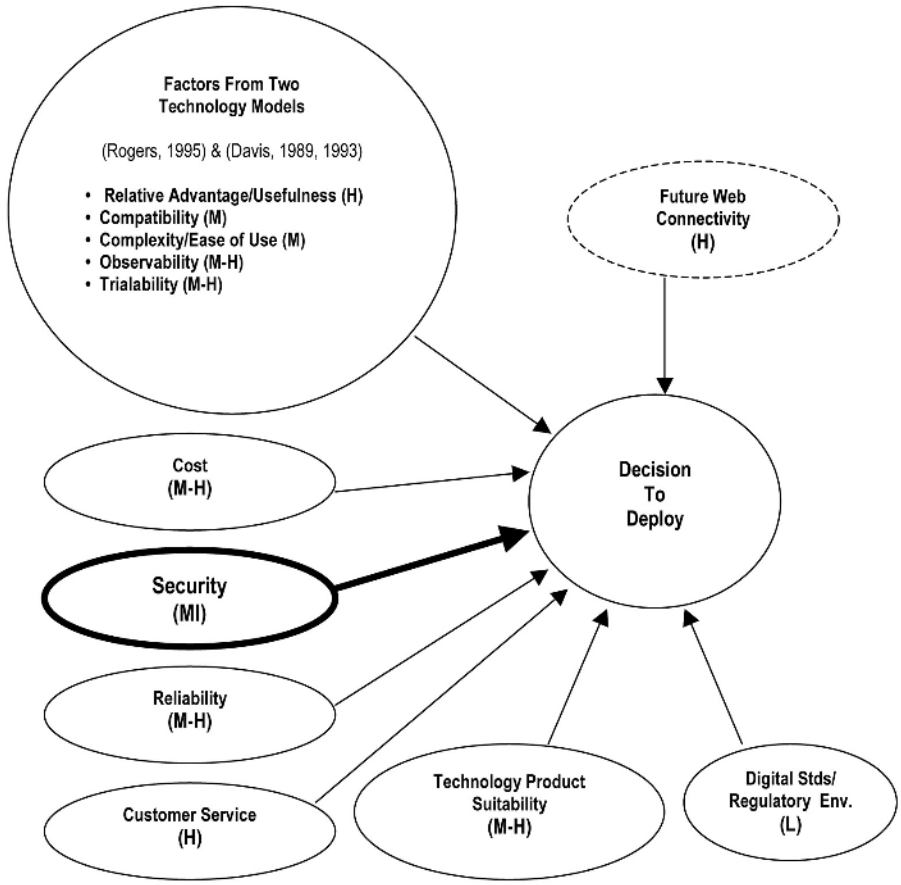
The above findings resulted in the revised Research Model (Figure 2).

The most important technology factors for adoption are security, reliability, and Web connectivity. The security and reliability factors were not in the traditional adoption models, but may have become more significant in the decade or more since those models were introduced. They correspond to the finding for mobile gaming adoption that the most important individual factor was perceived risk (Kleijnen & de Ruyter, 2003). The consensus result of security as the number one factor was summed up the best when the CIO of Case 5 responded by stating: "It's pretty easy. You know that right now for some applications wherever the application might contain proprietary or personal or data, security rules." Reliability is a highly rated factor because it is so intertwined with the concept of coverage and the ability to provide the service required.

The most prominent non-technology factor was customer service. It points to practical steps that management can take to assure good adoption success with

cell phones, such as to build up internally or arrange for external customer service for the cell phone users in the firm. This factor is likely to become more important in the future as Web and Internet applications become more prevalent and also more complex, requiring greater user support. Its importance is related to the ease of use factor, stressed in the TAM models (Davis, 1989, 1993). It is different in that it is corporate actions that can improve customer service.

Figure 2. Revised research model



The technology product suitability ratings were compiled by the authors' post-evaluation of all of the interviews for each company. Case 1 received a medium to high rating. The CIO of Case 1 spoke extensively about security, connectivity to the Web, reliability, user interface, upgradeability, and the search for a combination PIM/cell phone capability. These are factors that when combined result in a medium to high rating for suitability. Case 2 also received a medium to high rating because that company not only provides mobile technology to its own employees, but also extensively advises its customers on the latest mobile technology and the compatibility/suitability of that technology with its own GIS software. For Case 3, it also warranted a medium to high rating. The continuing theme in the Case 3 interviews was reliability and coverage, which is the suitability of the product to provide the required service when and where it is needed. For Case 4, a company whose stock in trade is cutting-edge technology, product suitability, both internally and externally to its customers, is quite important. Case 5's suitability rating was high. One reason is the fact the company has created several "centers of excellence" to leverage the maximum benefit from different technologies, for example, 802.11b. All of the companies mentioned such suitability issues as user interface, display sizes, and geographic coverage.

The findings showed cost to be a medium to high adoption factor. A representative explanation for this came from the CIO of Case 5 who stated that in the case of commodity items, cost could be a huge driver, although in the case of point solutions driven by a unique application, cost is not a major issue.

The mostly low ratings for regulation are due largely to firms' perceptions that cell phone vendors are responsible to consider this, rather than corporate users. The one CIO who gave a high rating for regulation is from the software firm which has a stated high social consciousness and sensitivity. That may have encouraged understanding and valuing regulation beyond just authorization to operate the devices, but delving into privacy, intellectual property, and the international sides of regulation. Cases 4 and 5 rated the regulation factor as moderate. Case 4 is sensitive to regulation because its products are susceptible to regulation worldwide. Case 5, the entertainment company, has very large intellectual property holdings and faces the problem of potential exposure of intellectual property laws and regulations, an exposure that may worsen with Web-enabled uses. Generally, as the uses of this technology become more complex and Web-driven in the future, regulation may become more prominent.

Web connectivity is a forward-looking technology factor, since voice communications dominates in today's cell phone uses for the sample firms. Since the technology is moving so rapidly (Dodd, 2002), it is essential that corporate decision makers look ahead in their adoptions; all of the case companies are doing exactly that. One example was the CIO of Case 1 who stated: "We're moving all of our applications to a Web-based interface." Likewise, the CIO of Case 5

regards the Web "…as an emerging new way of delivering our content." The decision makers in the mid 1990s needed to think forward in assessing cell phone adoption to a much wider prevalence of their use among employees and their customers, and not base decisions on the then modest adoption rates. It is no different for the decision-makers of today. Since the Web-based cell phone uses today are higher than for our sample in certain world regions, in particular Western Europe and Asia Pacific, future studies should try to include those regions. Just as all of the companies in the case study are looking to Web-based connectivity for the future, they also realize they will not reach the full benefits of such connectivity until the high bandwidth capabilities of 3G (third generation) are available, particularly for streaming video.

3G Standards

The primary digital standards for cell phones in the United States are TDMA, CDMA, and GSM. These three standards are known as 2G (second generation) technology. In our results, the protocol standards received a current overall rating of low to medium because the main standards issue today for the present U.S. sample is business travelers need a GSM cell phone for Europe. But all of the companies realize that the low to medium rating for standards today is going to shift to a high rating in the immediate future because of the increasing importance of bandwidth. 3G devices will offer higher bandwidth, packet-based transmission of text, voice, video, and multimedia to support the data intensive applications (Siau, Lim, & Shen, 2001). A 3G phone, for example, may be used as a phone, computer, television, or credit card. The respondents reported that they were intensively prototyping and testing 3G cell phone devices, since they were skeptical of their current technical functionality and robustness.

Table 5 lists the 2G, 2.5G, and 3G cellular services. W-CDMA, cdma2000, and EDGE are collectively known as IMT-2000, which is an International Telecommunication Union (ITU) International Mobile Telecommunications 3G initiative (Dodd, 2002). The 3G standard calls for 144 Kbps or higher in high mobility (vehicular) traffic, 384 Kbps for pedestrian traffic, and 2 Mbps or higher for indoor traffic (FCC/3G, 2003). The primary 3G protocols are cdma2000, promoted and owned by Qualcomm, and W-CDMA, which also is known as Universal Mobile Telecommunications System (UMTS) and is controlled by an industry consortium (Anonymous, 2002).

The company in the case study with the strongest interest in 3G was Case 5 as it realizes that 3G and the increased bandwidth that it provides will be critical to its ability to push content in new and exciting ways. The company's CIO stated

Table 5. Cellular standards: Progression from 2G to 3G. Source: Dodd, 2002 and Federal Communications Commission, 2003

2G	Comments	
CDMA	Code Division Multiple Access	
TDMA	Time Division Multiple Access	
GSM	Global System for Mobile Communications (European Union standard)	
2.5G	**Comments**	**Maximum Rate**
GPRS	General Packet Radio Services	115 Kbps
EDGE	Enhanced Data Rates for GSM Evolution (can be 2.5G or 3G)	384-473 Kbps
3G	**Comments**	**Maximum Rate**
W-CDMA	Wideband CDMA (also known as UMTS)	2-2.4 Mbps
cdma2000	Includes following cdma2000 platforms: 1xEV-DO 1xEV-DV	 2.4 Mbps 3.9 Mbps

Sources: Dodd, 2002 and Federal Communications Commission, 2003

"When you look at 3G structures and Asia Pacific…we can see the opportunity in that. So as soon as we start seeing this upgrading of the infrastructure [in the U.S.] it's going to be very important for the media … side …"

In fact, the Asia Pacific region is an excellent place for a deeper investigation of 3G. Japan's NTT DoCoMo launched the first commercial W-CDMA (3G) network in October 2001. The 3G service provided by the company is known as Freedom of Multimedia Access (FOMA) and it provides a variety of features including:

- Packet data transmission at up to 64 Kbps
- Data reception rate at up to 384 Kbps
- High-quality voice services supported at up to 12.2 Kbps
- Multi-access
- Take images, attach to e-mail while talking on phone
- Simultaneous access for voice and data
- E-mail and Internet access
- Streaming video
- Videoconferencing
- Voice calls same quality as fixed line (*http://www.nttdocomo.com/foma/3g/technology.html*)

3G enhances access to the Internet (the "killer app" after voice) (Schneiderman, 2003, p. 159). The high-speed, always-on, anytime/anywhere characteristics of 3G provide businesses with great advantages, such as:

- Mobile access to corporate information
- Being more responsive to customer needs
- Adopting new working styles

Deployment of 3G

There is no clear 3G leader in the U.S. at this time. The deployment of 3G systems depends on three factors: regulatory policy decisions, existing infrastructure, and popular culture (Banks, 2001). The U.S mobile market has been slower to develop than some other developed countries. While the cell phone penetration rate in the U.S. at the end of 2002 was 49 percent, it was 80 percent in Western Europe for the same time period (Federal Communications Commission, 2003). The slowness can be attributed, in part, to the FCC's policy of approving licenses on a regional basis and the refusal to require a uniform standard, such as the GSM (2G) and W-CDMA (3G) standards in Europe (Banks, 2001). Another reason for slow adoption may be that many features that digital handsets offer customers in Europe are already available through other means in the U.S. GSM handsets in Europe feature strong security measures and are used to make purchases. Consumers in the U.S. do the same through credit cards and cash where ATM machines are more common (Banks, 2001). A recent article, however, highlighted the fact that Japan's mobile network provider, NTT DoCoMo, is testing a system whereby an infrared beam is sent from the phone to a special infrared reader attached to the cash register, and MasterCard and Nokia have teamed up to test a mobile-phone credit card in Dallas, Texas (Kahn and Prystay, 2003). An additional reason the adoption of 3G is expected to be slower in the U.S. is the fact that the U.S., with unmetered phone access, already has one of the highest Internet usage rates in the world. Thus, 3G may initially be seen as less important in the U.S.

If one were to design a phone to have the most worldwide compatibility, the conventional wisdom points to a GSM/cdma2000 cellular phone. For 2G, there are five times as many worldwide GSM subscribers as CDMA, where, for 3G, there are more than 100 times as many worldwide cdma2000 subscribers as W-CDMA (Brodsky, 2003). Regardless of the speed with which 3G is fully implemented in the U.S., it is on its way.

Beyond 3G

The capabilities of a fourth generation (4G) wireless device is only a gleam in someone's eye since most still do not have access to 3G devices. It is believed that 4G will embody the capabilities of 3G, but at even higher transmission speeds of 20-100 Mbps (Schneiderman, 2003). How will business be conducted and or changed by a 3G/4G environment? Mobile services provided now are generally categorized as follows: communication, information, entertainment, and transaction (Mennecke & Strader, 2003). While information and transaction services are important from the corporate perspective, it is voice-to-voice communication that is the primary service provided today by mobile devices (Roberts & Pick, 2003). As a result, sales and marketing personnel and executives were the most intensive mobile-device users in the case study because they are heavily dependent upon voice communications.

But as 3G becomes a commercial reality and the resulting broadband and speed capabilities are significantly increased, one would expect other corporate functional areas will be emphasized. For example, logistics and supply chain management functions that depend upon data intensive applications may benefit from 3G/4G mobile devices. A scenario currently used in Case 1 where the company's traveling sales representatives are paired with sales associates back at the corporate headquarters so the sales representatives in the field can obtain inventory, parts, and delivery information real-time by phone may no longer be needed. That could free additional human resources to meet with customers in the field. Also, companies that want to push streaming content to their customers, such as Case 5, will see their distribution channels dramatically simplified by sending the content directly to the customer over a mobile device. The advertising departments will also have a new capability with 3G/4G mobile devices. The high-speed, always-on, anytime/anywhere characteristics of 3G/4G provide businesses with the opportunity to target customers through Global Positioning System (GPS) technology and to send specific advertisements to a customer depending upon the customer's specific location at that time. As mobile devices, such as a cell phone, are normally considered to be personal devices carried by a single individual, businesses will be encouraged to personalize the advertisement to the owner of the device.

3G/4G technology will encourage businesses to move additional applications and content to the Internet and to become even more dependent upon mobile (wireless) devices. As Web and Internet applications become more prevalent and, arguably, more complex, users of the technology may require even greater support. As a result, the most prominent non-technology factor that came out of the case study—customer service—will become even more important in the future. While there will be obvious flexibility, productivity, and customer/client

satisfaction gains to be reaped from wireless devices, particularly 3G/4G devices, wireless devices are inherently more exposed than wireline devices. Corporations are already very concerned about the security/privacy and intellectual property issues inherent within wireless devices (Roberts & Pick, 2003). There will probably not be a rush by businesses to increase the use of such devices until the privacy/security concerns have been resolved.

The wireless industry is currently on track to overtake the regular (wireline) industry in about two years (Rosenbush, Crockett, Palmeri & Burrows, 2003), and that is in an environment where 3G is not currently the standard. Imagine how that may be accelerated once the speed and other capabilities of 3G become routinely available to businesses and consumers. The U.S., as discussed earlier, has loosened its regulatory grip on wireless in order to encourage competition. The phenomenal growth of wireless means the federal and state governments will not firmly control an increasing percentage (and soon to become a majority) of the U.S. telecom industry. That has obvious benefits for increasing competition, as has already been proven, but may give governments pause to reconsider their largely "hands-off" approach. If for no other reason, the increasing risks of exposure to security/privacy threats and intellectual property theft that are inherent within wireless devices may lead to intensified regulation of the wireless industry.

Conclusion

This study has analyzed technology factors in corporate cell phone adoption and uses. The research questions are answered as follows:

- *RQ1.* The most important technology factors are security, reliability, and Web connectivity. Technology factors are more important than non-technology ones.
- *RQ2.* The regulatory policies had low impact for three respondent firms. The software company had moderate influence, related to its concern about the regulation and ethics of content. On the high end, the large entertainment firm was very impacted by the FCC, in particular regarding regulation and legalities of intellectual property, especially for Web-enabled cell phones. Overall, the lack of digital standards was rated as low to medium in importance.
- *RQ3.* Only one large entertainment firm has significant Web-enabled cell phone uses today. The other four companies plan to add significant Web-enabled uses in the future, emphasizing Internet communications initially.

- *RQ4.* The constraining factors on cell phone use were cost (moderate to high), security, regulation, and business activities, in that certain functional areas were presently not emphasized. Technology is an important influence on security.
- *RQ5.* The decision-making process for cell phone adoption was unique to each company and depended on that firm's organizational structure and corporate culture. The final decision-maker varied considerably by firm and by size of project; decision-makers included a middle level board, technical director, CIOs and CEO.
- *RQ6.* The business functional areas of highest cell phone use were sales, marketing, executives, and IT. The first three were primarily for voice communications to customers and employees, while IT used cell phones for testing and prototyping.

The theoretical framework of the paper for cell phone adoption and deployment is validated as appropriate. Both new factors and traditional ones are shown to be important. The new factors emphasize having a robust and secure use of the devices, with necessary user support. More could be gained from the present research by future follow-up, involving re-interviews at each firm to learn about and respond to the present study findings.

The study is limited by only examining cases for five firms in four industries. Future research needs to encompass larger samples of firms, which would be more robust and enable more sophisticated methodology, such as multivariate statistics. A weakness of the present study is not including the measurement or analysis of the extent of success or failure of corporate cell phone implementation. Including success measures for a large sample would provide more extensive and robust advice for corporate decision-makers. A large sample also would allow robust industry sector comparisons. The present multi-layered interview methodology could be supplemented with large-sample surveys of firms. Future studies also should try to sample advanced-use regions, such as Japan.

The increasing use of 3G wireless devices today and 4G devices in the future will cause the momentum to shift from voice-centric uses to data-centric applications and to streaming content. It also will encourage businesses to become increasingly dependent upon the Web, Internet applications, and mobile devices. When looking at information technology, particularly wireless/mobile devices, from either a legal or ethical perspective (Johnson, 2001), privacy/security is the most important issue. The vulnerability of the large and rapidly growing wireless industry to this issue may cause the governments to increase regulation of the industry.

References

Adams, D.A., Nelson, R.R. & Todd, P.A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly 16*(2), 227-247.

Agarwal, R., & Prasad, J. (1997). The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. *Decision Sciences, 28*(3), 557-582.

Anonymous (2002). Business: Time for plan B; Mobile telecoms [Electronic version]. *The Economist*, *364*(8292), 78.

Banks, C.J. (2001). The third generation of wireless communications: The intersection of policy, technology, and popular culture [Electronic version]. *Law and Policy in International Business*, *32*(3), 585-642.

Black, S.K. (2002). *Telecommunications law in the Internet age*. San Francisco: Morgan Kaufmann.

Brodsky, I. (2003). How to salvage Europe's 3G industry [Electronic version]. *America's Network*, *107*(1), 22.

Clarke III, I. (2001). Emerging value propositions for m-commerce. *Journal of Business Strategies, 18*(2), 133-148.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Davis, F.D. (1993). User acceptance of information technology: System characteristics, user perceptions, and behavioral impacts. *International Journal of Man-Machine Studies, 38*, 475-487.

Davis, G.B. (2002). Anytime/anyplace computing and the future of knowledge work. *Communications of the ACM*, *45*(12), 67-73.

Dodd, A.Z. (2002). *The essential guide to tele-communications* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Eastlick, M.A. & Lotz, S. (1999). Profiling potential adopters and non-adopters of an interactive electronic shopping medium. *International Journal of Retail and Distribution Management, 27*(6), 209-223.

FCC/3G (2003). *Third Generation Wireless*. Retrieved May 6, 2003, from *http://www.fcc.gov/3G*

Federal Communications Commission (2003). 8th Annual Report and Analysis of Competitive Market Conditions With Respect to Commercial Mobile Services. Retrieved July 28, 2003, from *http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-03-150A1.pdf*

Gefen, D. & Straub, D.W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quarterly, December*, 389-400.

http://www.nttdocomo.com/foma/3g/technology.html. Retrieved August 10, 2003.

Johnson, D.G. (2001). *Computer Ethics* (3rd ed.). Upper Saddle River: Prentice-Hall.

Kahn, G. & Prystay, C. (2003). "Charge it!" Your cellphone tells your bank. *The Wall Street Journal*, August 13, B1

Kleijnen, M. & de Ruyter, K. (2003). Factors influencing the adoption of mobile gaming services. In Mennecke, B.E., & Strader, T.J. (Eds.), *Mobile commerce: Technology, theory, and applications* (pp.202-217). Hershey, PA: Idea Group Publishing.

Lederer, A.L., Maupin, D.J., Sena, M.P. & Zhuang, Y. (2000). The technology acceptance model and the world wide web. *Decision Support Systems, 29*, 269-282.

Mennecke, B.E. & Strader, T.J. (2003). *Mobile commerce: Technology, theory and applications*. Hershey, PA: Idea Group Publishing.

Palen, L. (2002). Mobile telephony in a connected life. *Communications of the ACM, 45*(3), 78-82.

Roberts, G.K. & Pick, J.B. (2003). Case study analysis of corporate decision-making for cell phone deployment. *AMCIS 2003 Proceedings*. Atlanta: Association for Information Systems.

Rogers, E.M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

Rosenbush, S., Crockett, R.O., Palmeri, C. & Burrows, P. (2003). A wireless world: In a few years, mobile phones will dominate U.S. communications. *BusinessWeek*. October 27, 110.

Schneiderman, R. (2003). *Technology Lost: Hype and Reality in the Digital Age.* Upper Saddle River: Prentice Hall.

Siau, K., Lim, E. & Shen, Z. (2001). Mobile commerce: Promises, challenges, and research agenda [Electronic version]. *Journal of Database Management*, *12*(3), 4-13.

Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: SAGE Publications.

Van Akkeren, J. & Harker, D. (2003). Mobile data technologies and small business adoption and diffusion: An empirical study of barriers and facili-

tators. In Mennecke, B.E., & Strader, T.J. (Eds.), *Mobile commerce: Technology, theory, and applications* (pp. 218-244). Hershey, PA: Idea Group Publishing.

Yin, R.K. (1993). *Applications of case study research*. Thousand Oaks, CA: SAGE Publications.

Yin, R.K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Appendix

Protocol of Interview Questions for the Telecommunications Manager in Charge of Cell Phones:

- What is your name?
- Job title?
- Primary duties?
- How are you involved in the identification, selection, or deployment of telecom wireless devices (must include voice component) in your company?
- What is the organizational structure of your company, particularly as it relates to telecom?
- Who else is a key player at the managerial or operational level regarding the identification, selection, or deployment of telecom wireless devices in your company?
- How does your job interface/coordinate with the IT Director?
- Who promotes a project to acquire and deploy wireless devices?
- Who is the final decision authority?
    - Dollar level for approval?
    - Other factors affect level of approval?
- What types of wireless devices have you deployed?
- When was the deployment decision made (Mo./Yr.)?
- If you have made second deployment decision, when (Mo./Yr.)?

Questions below must focus on most recent deployment or on future deployment within next three months:

- How many wireless devices have been deployed in U.S.? Overseas?
- Can you break the numbers down by category in terms of cell phones, PDAs, etc.?
- What are wireless devices used for (voice, Web/Internet, e-mail/messaging)?
    - Which is the primary use?
- To what extent, if at all, is Web connectivity (Web-enabled) important for your wireless devices?
- To what extent, if at all, are your wireless devices used for M-commerce (buying and selling with wireless devices)?
- To what extent, if at all, are wireless devices activated for geo-referencing?
- What business purposes/company functions are served with wireless devices? Core/Non-Core?
    - Sales?
    - Marketing?
    - Logistics?
    - Management?
    - Execs?
    - Middle Management?
    - Other?
- What is the bandwidth of your deployed devices?
- Where, geographically, are wireless devices used?
- Is wireless deployed for the optional convenience of your customers or employees or is wireless considered as a critical business "necessity"? Explain.
- Who primarily benefits--customers or employees?
- What factors are important in making the decision to acquire and deploy wireless devices? If not mentioned by interviewee, what about
    - Cost?
    - Bandwidth?

    - Security?
    - Reliability?
        - Communications service
        - Hardware
    - Convenience?
    - Scalability?
    - Productivity?
    - Manageability/supportability?
    - Customer perception that company is up-to-date?
    - Connectivity to Web?
    - FCC/state regulation?
    - Other factors?
- Which of the above factors would be high, medium, low, or of no importance?
- How successful has the deployment of wireless cell phone devices been for your company?
- Do you have plans to deploy additional wireless devices in the near future?
- If so, what types of devices, how many, and purposes/functions served?
- What part does FCC regulation of the wireless industry play in your decision to use wireless?
- What part does state regulation of the wireless industry play in your decision to use wireless?
- What part does regulation in foreign countries play, if applicable?
- How does cell phone deployment strategically benefit the company?
    - How do you measure the benefit?
- How does cell phone deployment strategically cost the company?
    - How do you measure the cost?
- Does the company have a policy regarding the use of personal cell phones for business?
    - If so, what is it?

- Does the company have a policy regarding the use of business cell phones for personal use?
  - If so, what is it?

Chapter III

Adoption of Mobile Data Services: Towards a Framework for Sector Analysis

Elizabeth Fife, University of Southern California, USA

Francis Pereira, University of Southern California, USA

Abstract

The evaporation of dramatic growth forecasts for mobile data services highlights the need for greater understanding of user's behavior, needs and attitudes to technology, as well as their environment and other contextual factors. By examining sectors where a value proposition for mobile data services has been identified and yet adoption rates have varied, we discuss requirements for uptake to occur in specific sectors. Adoption of mobile data services refers to organizational-related solutions

as well as service innovations related to the product or service delivered to end-users who in these cases include customers, patients, and students. Using frameworks for innovation diffusion, we examine promising mobile services in the areas of health, construction, and education. The underlying behavioral, cultural, and economic factors affecting demand for mobile technology in these markets is investigated. This exploratory research contributes to theory-building for understanding technology adoption from the user's context.

Introduction

The adoption of mobile communications and data technologies offers the potential of fundamental life and work changes for business and personal users. Interactions between people, networks, companies, and organizations can quicken, deepen and expand vastly through always-on devices that are flexible and easy to use. Although the envisioned transformations have not yet taken place, they are still anticipated to occur with the implementation of high-speed mobile networks. The varied rates of mobile service diffusion on a global basis are being closely watched, especially in the consumer market. For example, dramatic variance is noticeable in the take-up of short-messaging service (SMS), which gained rapid popularity in Europe and parts of Asia, but has lagged thus far in the U.S. The need to consider context, demographics and other social factors has been accentuated by these regional discrepancies in the adoption of mobile data services. Overly optimistic forecasts for mobile data services in both consumer and enterprise markets in the United States in particular, has revealed a need for analytic frameworks to understand consumer adoption of technology in general and mobile technology in particular.

Technology diffusion frameworks have been adapted specifically for mobile services. The Input-Process-Output model (IPO), for instance, focuses on consumer adoption and usage of mobile devices (Sarker & Wells, 2003). This model includes user characteristics, features of mobile technologies, contexts for use, the process of trying out the technology, assessment of the experience, and the outcomes of usage. Gilbert & Kendall (2003), on the other hand, have developed a research model for studying market segmentation of mobile data services that include supply, demand and contextual forces. As Venkatesh, Morris et.al (2003, p.426) observe, "researchers are confronted with a choice among a multitude of models, and find that they must pick and choose constructs across the models, or choose a factored model and largely ignore the contributions from alternative models." The growth and development of new models to

analyze mobile technology adoption attests to the level of uncertainty faced by industry decision-makers trying to make business and investment strategies.

Our preliminary framework builds on Rogers and Venkatesh's Unified Theory of Acceptance and Use of Technology (UTAUT), and is applied to three areas where the value of mobile data services has previously been identified. We focus on mobile technology adoption in three promising sectors: medicine, construction and education.

The process of adoption is examined in these sectors by utilizing models of diffusion previously applied to the consumer side of the market. This investigation suggests that greater emphasis on organizational norms that characterize specific industries would be helpful for understanding the adoption of mobile data services. An initial framework is presented which seeks to account for circumstances of technology adoption or non-adoption that are not easily handled by other current models. Below in Table 1 the most common frameworks and the modified framework proposed here are compared in terms of their explanatory power.

Most models of technology diffusion tend to focus on individual users or adoption at the firm level, whereas this effort attempts to generalize users and user needs and requirements within specific domains. By examining mobile services among discrete sectors with users who have a degree of similarity in terms of needs, context, and social structure and who operate in a more definable environment than individual users, we suggest that technology diffusion can be usefully investigated within specific fields such as medicine or education.

Review of Technology Diffusion Models

Table 1 compares the dimensions of commonly used models for technology diffusion.

First, the Diffusion of Innovation framework is prominent in the innovation literature and has long served as a useful explanatory device. The framework covers the major attributes of an innovation, such as its benefits, compatibility with existing ways of doing things, its ease of use, as well as ease of experimentation. Influence from the social system is also a factor that will determine the ultimate fate of an innovation (Rogers, 1995).

The individual user is the unit of analysis in the Diffusion of Innovation, Technology Acceptance Model (TAM), and UTAUT models. The TAM is based upon intentions to use technology and uses two key constructs "perceived usefulness" and "perceived ease of use" (Davis, 1989). The former refers to the

Table 1. Models of technology diffusion

	MODELS OF DIFFUSION
Diffusion of Innovation	
Relative Advantage	The social and economic advantage that can be derived from adopting the new product
Compatibility	The degree to which an innovation is perceived as consistent with existing values and past experiences of the adopter
Complexity	The extent to which the innovation is perceived as difficult to understand and use
Trialability	The degree to which the innovation can be experimented with on a limited basis
Observability	The degree to which the results of an innovation are visible to others
Social System	The set of interrelated units engaged in joint problem solving, its structure (formal and informal) and its norms
Type of Decision	Innovations can be adopted by individual members of the social system or by the entire social system.
Communication Channels	Effects of change agents on social system

Technology Acceptance Model	
Perceived Usefulness	The extent to which person believes using a particular technology which increase job performance
Perceived Ease of Use	The extent that using a particular technology would be easy to use
Subjective Norm	The individual's perception that people who she think important should adopt the technology
Perceived Usefulness	The extent to which person believes using a particular technology which increase job performance
Perceived Ease of Use	The extent that using a particular technology would be easy to use
Subjective Norm	The individual's perception that people who she think important should adopt the technology

Unified Theory of Acceptance and Use of Technology (UTAUT)	
Performance Expectancy	Degree to which individual believes using the system will help her attain gains in job performance
Effort Expectancy	Degree of ease associated with use of system
Social Influence	Degree to which individual perceives that important others believe he should use the technology
Facilitating Conditions	Degree to which an individual believes that organizational and technical infrastructure exists to support use of the system

degree to which the individual believes that using the technology will enhance their job performance. However, the link between the individual's behavior at the enterprise/organizational level to the overall sector has not been explicitly developed. TAM has been used to understand skill training (Venkatesh, 1999), consumer behavior in an online environment (Koufaris, 2002), and in telemedicine (Hu & Chau, 2001). Hu and Chau find that the TAM model provides a fairly accurate picture of doctor's willingness to use telemedicine technologies. They

find that perceived usefulness was the most significant factor while perceived ease of use was not influential (Hu & Chau, 2001).

Venkatesh, Morris, et.al. (2003) provide a recent review of the user acceptance literature and systematically compare eight well-known models and the various predictive factors that each specify. Their Unified Theory of Acceptance and Use of Technology (UTAUT) is a unified model that synthesizes and adds to previous models. Most important among the determinants of user acceptance of technology are "performance expectancy" which is the degree to which a user believes that using a technology will provide gains in job performance.

Next, "effort expediency" is the degree of ease in using a system. "Social influence" is the degree to which individuals perceive that it is important that others believe that they should use a new system. Finally, "facilitating conditions" are the degree to which individuals believe that there is organizational and technical support for using the system. Additionally, it is argued that these direct determinants of user acceptance are moderated by gender, age, voluntariness and experience (Venkatesh, Morris, et. al., 2003).

The UTAUT model provides a useful starting point for analyzing technology adoption by an individual and by people within a firm. Our cases suggest however, that moving beyond the individual firm level to adoption at the sector or industry level requires a greater emphasis on socialization, organizational culture and the cultural factors related to the sector, including government and regulatory barriers all of which can pose significant obstacles to technology adoption.

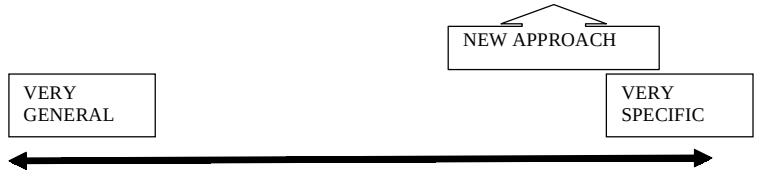
The models listed do not adequately account for the following five situations:

1) Different adoption rates of innovations across national markets or within societies;
2) Different adoption rates of innovations across different sectors in the same national market;
3) Different adoption rates of innovations across different firms/organizations in the same sector;
4) Different adoption rates of innovations in the same ethnic groups across different national markets;
5) Different adoption rates of innovations within the same age groups across different national markets.

Our proposed model, the Global Adoption of Technology (GAT) incorporates and weights cultural norms — social and organizational to a greater degree than

Table 2. Various models

	Diffusion of Innovation	**UTAUT**	**Global Adoption of Technology (GAT) Model**	**TAM**
Different adoption rates across national markets	Not specifically but can be assumed under *type of innovation decision*	Not specifically but can be assumed under *facilitating conditions*	yes	no
Different adoption rates across sectors in the same national market	Not specifically but can be assumed under *comm. channels* or *social system*	Not specifically but can be assumed under *performance expectancy*	yes	no
Different adoption rates across firms /organizations in the same sector	Not specifically but can be assumed under *comm. channels*	Not specifically but can be assumed under *facilitating conditions*	yes	yes
Different adoption rates of innovations in the same ethnic groups across national markets	Not specifically but can be assumed under *comm. channels* or *type of innovation* or *social system*	no	yes	no
Different adoption rates within the same age groups across different national markets	Not specifically but can be assumed under *social system* or *comm. channels*	Not specifically but can be assumed under *social influence*	yes	no



the other models discussed here. The Diffusion of Innovation model is the most general and is often used because it is widely applicable and can explain many types of outcomes. The importance of culture for instance, is noted, but encompasses a wide range of factors that in the end dilutes the models

explanatory power. On the other hand, the TAM model is the most parsimonious model listed below, but as a result can also be viewed as somewhat restrictive.

Different Adoption Rates of Innovations Across National Markets

The Diffusion of Innovation model classifies users and their interactions with new products (Rogers, 1995). This model encompasses a broad range of factors, and doesn't weight any particular factor as more salient than another. Social factors are acknowledged, but are not delineated, particularly at the organizational level. Initially, Rogers felt that a general framework should suffice to explain diffusion and that other factors like culture, profession and other social influences were not important as diffusion would be the same in different contexts (McGrath & Zell, 2001). However, further research suggests that social and environmental factors may indeed have great significance. Consumer behavior studies have found for example, that when countries converge in respect to national income, cultural variables explain differences in country-level consumer behavior (de Mooij & Hofstede, 2002).

Some studies attribute varying adoption rates of mobile services to differences in economic performance. For example, the most comprehensive study of the diffusion for mobile services has been carried out by Gruber and Verboven (2001) who employ a model based upon data for 140 countries. Two variables are used, including a measure of how technologically "advanced" a country is measured by its adoption rates together with the growth rate of diffusion. These variables are modeled as functions of country-specific factors, including per capita GDP, per capita fixed main lines, the level of competition, standards, etc.

They find that countries with higher a GDP and larger fixed networks appear to have higher adoption rates for mobile services. Also, use of mobile services tends to be complementary to fixed line services (Banerjee and Ros, 2002). Their study of the diffusion of mobile communications in the European Union finds that technological developments, namely the transition from analogue to digital technology in the early 1990s along with the increase in spectrum capacity has had a dramatic impact on mobile telecommunications diffusion (Gruber & Verboven, 2001). Overall, technological factors are considered the significant force driving the diffusion of mobile services in the consumer market.

In addition to the factors that are prominent in most theoretical analyses, such as the overall GDP of a country, network size, and the market structure of the telecom industry, other factors such as the level of privatization, regulatory schemes, and pricing arrangements have also been examined (Banerjee and Ros, 2002). These analyses focus primarily on consumer mobile telephony markets,

and highlight the critical factors for mobile technology growth. They find that a technological base must provide performance, ubiquity and reasonable pricing. Mobile technologies have an "epidemic" or "viral" quality as adoption by many, fuels further adoption. The importance of technological readiness is clear, and network capacity and the availability of spectrum are general prerequisites for the adoption of mobile services.

It has been often noted that specific characteristics of national markets are important in determining mobile subscription rates, (Ahn and Lee, 1999), but these have been generally defined in terms of regulatory policies, technological innovation, and competition. Although these factors have been recognized as catalysts accelerating the adoption of new technologies, they nonetheless provide a less than complete explanation for the rapid growth of mobile services in certain select markets throughout the world, such as Finland, South Korea and Japan. To explain mobile service adoption in these markets, social factors and the perception of the relative value seem to provide a more complete explanation. Cultural factors also help explain the lower than expected uptake for broadband services in Singapore, where despite strong government backing, the perceived value of online services and applications has remained low.

Models have also been proposed to specifically explain adoption rates of mobile telephony to individual consumers. Overall, the main drivers of growth are identified as national income, competition, and the ubiquity and quality of the fixed network. Economic drivers including income, price and network externalities have been investigated in terms of determining critical mass and the overall growth of mobile telephony markets (Madden, Coble-Neal and Dalzell, 2002).

From a consumer behavior perspective, Levitt (1983) contends that new technology and the media homogenize consumer's wants and needs. Based on the assumption that user behavior is rational, consumers will show a preference for standardized products. However, this argument is challenged by researchers who find that differences in consumption have to be viewed in light of cultural context. Consumers sometimes do not make adoption choices based upon maximizing utility, but often for emotional reasons and through the veil of cultural context. For instance, exploratory research into the preferences of mobile consumers indicates that status and appreciation for brand names are motivating factors for mobile consumers in the Korean market, but U.S. consumers are driven by the perception of utility and convenience and not the desire to support their social status. (Kim, Fife, et. al, 2004)

Consumer surveys are beginning to investigate the influence of culture on mobile data service adoption on a global basis (Kim, Hong, & Tam, et. al., 2003). Kim, et. al. find that usage and adoption patterns for the mobile Internet vary from country to country due to cultural, economic, as well as other related factors. Although it is suggested that culture is important as a factor for technology

adoption, there are few empirical studies to substantiate this idea. Hall characterizes different cultural orientations as "high context" such as Japan and Korea, and others like the United States as "low context" (1987). Additionally, consumer behavior and cultural factors are identified by Hofstede as significant elements that determine the adoption of innovations (Hofstede, 1993). These theories have not yet been tested in the realm of mobile data applications, however where it appears that cultural context as well as government policy and regulation may play strong roles in the diffusion process.

Different Adoption Rates of Innovations Across Different Sectors in the Same Society

Although the TAM, UTAUT, and Diffusion of Innovation models all can account for innovation adoption at the sectoral level, these models do not give sufficient weight to organizational and social norms that can explain collective adoption decisions. This may be in part due to the still limited degree of testing of these models at the sectoral level. Much testing of these frameworks has relied upon a relatively small *n* and has often used students as test subjects. Investigations have occurred in sectors such as medicine (Hu & Chau et. al., 1999), and by Denis, Hebert et. al. (2002) who investigate the adoption of innovations in medicine, as well as by Mitropoulos and Tatum, (2000) who examine general cases of technology adoption in the field of construction and building. However, there currently is a lack of large scale studies that span sectors.

Different Adoption Rates of Innovations Across Different Firms in the Same Sector

There is still little research examining the link between user adoption and adoption at the organizational or firm level. Organizational knowledge and innovations are often localized and are not visible to other similar organizations. Roger's includes the "change agent" or champion of an innovation as the important entity for communicating the relative advantage of an innovation to others (Rogers, 1995). How this process works within a sector requires further investigation. It is speculated that the diffusion process is facilitated within a sector to the degree that organizations or firms are networked or integrated. (Robinson, Savage, et al., 2003).

Different Adoption Rates of Innovations by the Same Ethnic Groups in Different National Markets

Cultural and societal influences are increasingly recognized as important variables for understanding technology adoption. Research is still in the exploratory stages; studies have looked at a small number of individual users across several cultures to provide comparative analysis. For example, Jarvenpaa, et. al. (2003), conducted a focus group study in Finland, Hong Kong, Japan, and the U.S. to look at motivations and perceived benefits of using mobile devices and services. However, we have not found studies that examine the behavior of mobile users from a specific culture such as Korea, when they are re-located to a different market, such as the United States. Investigating the behavior of transplanted technology users would shed light on the importance of contextual factors through examining the extent to which attitudes and use of mobile devices has been altered. Overall, culture as a variable in diffusion models tends to be too general to provide explanatory power.

Different Adoption Rates of Innovations within the Same Age Groups Across Different National Markets

The age of a user is considered an important variable in some technology diffusion models, such as the UTAUT (Venkatesh, Morris et al., 2003) which finds differences in technology adoption between older workers and younger workers, suggesting that this is a key moderating influence. They suggest that facilitating conditions such as the provision of instruction and support are more necessary for older than younger workers.

Small scale empirical studies have examined mobile users specifically to help understand the influence of age. The ITU (Selian, 2004) has examined the youth market for mobile data services in the U.S. However, there is little research that examines differences in uptake among similar age groups across various cultures, or compares the behavior of users of different generations. For instance, a comparative study of professionals between the ages of 25-65 across societies and disciplines has not been carried out. Venkatesh and Morris also note that their model, UTAUT could be refined through study of different user groups, “such as individuals in different functional areas or organizational contexts” (Venkatesh, Morris et al., 2003, p.470)

Application of User Adoption Models to Business Sectors

The diffusion of mobile services and applications seems to be influenced by numerous factors, including culture, economics, geography, organizational structures, government policies and regulation. In addition, the timing of a technologies' introduction is yet another variable (Rogers, 1995). The rate of technology diffusion between different countries is thought to be affected by time, as changes in design, pricing, and communication systems have influenced the rate of adoption in countries where a service was introduced later (Eliashberg & Helsen, 1996).

It is thought that communication of ideas takes place more frequently between people who are alike, "homophilous," than in situations where individuals have differing beliefs, cultural and socioeconomic situations, "heterophilous," groups (Takada & Jain, 1991). Adoption rates for products may be higher when there is more communication among similar people, and the word-of-mouth effect can be supported. The importance of "word of mouth" as a means of diffusion is noted by several researchers as significant (Bass, 1969; Moore, 1995). This factor has obvious relevance to mobile technology, which depends on critical mass to create value. Intra-industry communication about technology is a factor included in our modified framework as well.

Framework for Analysis: Global Adoption of Technology (GAT) Model

The model used here, the Global Adoption of Technology, or GAT model builds on the Diffusion of Innovation model, but assigns greater weight to cultural socialization including organizational and social norms as factors driving technology adoption in different sectors. Differences between individual's preferences and choices and technology adoption on a sector level are also considered. For example, while it is generally accepted that increased productivity and profitability are the primary drivers motivating firms and industry to adopt technologies, individuals may adopt a technology for a different, albeit wide variety of needs and desires, ranging from curiosity, social and emotional needs, or utilitarian values like increased convenience.

In the case of both construction and medicine, generally the individuals who are adopting innovations are not experienced with technology, and are in fact, often

resistant. In higher education, on the other hand, more openness and experimentation with new technologies can be observed. Despite a difference in user profiles, however, mobile education at this point, appears to be moving slowly, as execution has been problematic. The medical field on the other hand, faces the greatest institutional and regulatory barriers which has clearly hindered efforts in the U.S. In the field of construction, information technologies have historically been adopted slowly, yet there is evidence of growing use of mobile applications.

In construction, medicine and education, individual entrepreneurial efforts have been able to succeed, demonstrating that it is possible to create a user base if an innovation is perceived as having high value, is compatible with existing practices, is user-friendly and is consistent with social and cultural norms of the group. Although individual efforts can be sustained by early adopters, mass take-up in a sector requires visibility and communication between organizations, (*observability* in Roger's model).

Key Features of Analyzed Industry

To assess our model, we examine the adoption of mobile services in sectors where the value in adopting this technology has been identified. Benefits include lowered costs and greater efficiency in accessing and transmitting information. The ability to transcend time and space constraints in accessing information, as well as processing and communicating it are key advantages of mobile services in these sectors. Overall, these three areas have been identified as well-suited to using information technologies, based on several characteristics including:

1) information intensity
2) a high degree of interactivity with clients
3) movement in time and space of workers, clients and the work itself

Although the suitability of mobile technology has been identified for these sectors, adoption has varied, and furthermore when adoption has occurred, it has met with varying results. Despite enthusiasm in the education sector for new models of delivery, mobile education at this point is encountering roadblocks. Institutional opposition to change exists to varying degrees in all three sectors, although medicine faces the greatest institutional and regulatory barriers. This has surely hindered initiatives in the U.S. and Canada.

The importance of social networks in supporting and shaping the diffusion process of mobile technologies and applications is demonstrated in these three sectors. Word of mouth and interactions via the Internet have demonstrated feasibility, and have had a significant effect in educating potential users and providing motivation to investigate and try out mobile services.

In our three cases we see that the medical, education, and construction sectors face the same pressures to achieve greater efficiency, satisfy customer's needs, and engage in competition with other entities within the sector, yet the degree to which new technology can diffuse will be explained by the following determinants:

Percieved Relative Value

Perceived relative value is defined as the social and economic advantage that can be derived from adopting a new technology. The overall benefits of a technology to an industry are in the end determined by increased efficiency. A technology can provide an economic advantage by improving efficiencies, reducing labor costs, speeding up delivery of a service or product and improving the "customer's" experience. These reasons provide a basis for rational decision-makers to adopt a technology, if benefits and the technology are accurately understood. To help explain why an industry or firm is slow to adopt a technology, despite evidence of strong potential for relative value, other factors should be considered.

Usability - Compatibility Drivers

It is easier for an innovation to be incorporated into business practices if it is consistent with the existing and past experiences and mind-set of the adopter(s). This assumption is valid for individuals within an institution and for the overall social system of the institution itself.

Cultural Socialization

Social Norms

Social norms are generally defined as the tacit or explicit social rules that govern interaction between individuals in a society. Also, they define an individual's behavior as a member of a social group or as a citizen. The formal and informal

education system of a society transmits and reinforces these norms. Social interaction between individuals within different parts of the social structure of an industry or discipline vary greatly. The cases included in this study focus on the U.S., and for the sake of simplicity, it is assumed that similar social norms apply in all cases examined here. Commonly agreed on social norms in the U.S. might include wariness of government, valuation of "privacy" over "general good, appreciation of confidentially for information, (which is significant in the field of medicine). Also important is the belief in equality of opportunity, and the generally litigious nature of dispute settlement. All the cases investigated here operate amidst these and other social norms that affect decision-making and attitudes towards technology adoption.

Organizational Norms

Organizational norms are the tacit or explicit rules that define an individual's behavior within a company or group. Organizational norms govern the individual's behavior in the workplace and can vary between companies or other kinds of groups within the same sector. In education, student-professor interactions have specific rules and structures. In addition, varied interactions occur between professors, and within and between administrators and other parts of an academic institution. It is not possible to create absolute categories for organizational norms within an industry as there will be variance.

However, organizational culture and values, like acceptance of risk and uncertainty will influence decisions to try out and utilize new technologies and will influence the provision of training and the collection of skills within a discipline or industry. These circumstances will shape views towards "technology-willingness" and thus, should be considered, acknowledging the difficulties in creating generalizations for organizational norms in an industry or sector.

Technology Adoption Catalyst

Progress in adopting a technology can occur through individual members of a social system (industry or discipline), or by the entire social system. When an individual firm or person adopts an innovation, this is a voluntary process. When an entire social system adopts an innovation however, this happens either by consensus or through an authority. In general, it is thought that authority-based decisions tend to promote faster adoption.

The adoption of new technologies can be influenced to varying degrees by a central governing body or authority that sets the rules and organizational structure of the industry. In medicine in the U.S. for instance, the American Medical Association, (AMA) is the body that sets the rules governing the medical field. Their support of telemedicine is a vital element to diffusion in health systems in the U.S. As is shown, the most popular applications in mobile medicine are personal digital assistants, (PDA's) that are being used on an individual basis by doctors. In the construction industry decisions to use mobile applications also appear to be on an individual basis and voluntary. In the case of telemedicine and e-learning, it appears that the mandated decisions to implement systems have not had universal success, nor have they spurred mass adoption at this point.

The likelihood of success for new technology introduction will be gauged by viewing sectors through this framework.

Case Studies: Mobile-Building, Mobile Medicine, and Mobile Education

Mobile-Building and Construction

The construction industry is an area where adoption of mobile computing is growing. Mobile services have been found to bring cost-savings and improved efficiency. (Jain, 2003). Building and construction can be considered an industry sector which requires a high degree of mobility.

The construction industry overall, has fewer social or regulatory obstacles to adopting information technology, relative to impediments present in telemedicine. Mobile applications are compatible with existing values, practices and systems. As a result, where relative value can be identified, mobile wireless applications have faced low barriers to adoption, despite the fact that this industry is not usually considered technologically-savvy. The attributes of mobile technology have been found to be well-suited to the environment of a construction site and generally can be considered relatively low cost, trialable and easy to learn.

Perceived Relative Value

Many tasks in the construction industry are amenable to automation. Site inspection and checking the status of projects for instance, is considered a high value mobile application because site visits can potentially be reduced. None-

theless, until recently, automation has come slowly. Change is occurring at a more rapid rate now however, with realization in the industry that cost-savings and efficiency gains can be substantial (Mobileinfo.com, 2003).

The cost of handheld devices along with the availability of new applications has helped encourage this industry to automate and use mobile field services. The ability to monitor schedules, transfer equipment, get project status, and dispatch crews to sites using mobile tools has provided a relative value in reducing costs and time.

Usability: Compatibility Drivers

A lack of user-friendliness has been cited as an obstacle in the past to technology adoption. User-friendliness is improving as new devices are introduced along with improved technologies to integrate systems. Handheld devices and notebooks for the construction industry environment have been customized to be highly "ruggedized" and able to withstand extreme temperatures and shock. The availability of low cost handheld devices has helped propel application development by start-up companies.

Many tasks in construction are extremely amenable to automation, including the following (Mobileinfo.com, 2003):

- Project monitoring and reporting
- Dispatching crews to job sites
- On-site estimates
- Scheduling, ordering and transferring of construction equipment
- Tracking labor time and materials used through wireless devices

In the home-building sector, a major issue is efficiency in the supply chain, and scheduling problems due to miscommunications (Cotter, 2003). Mobile applications can allow contractors to coordinate the schedules with subcontractors and communication with suppliers so that materials arrive at the correct time. Also, subcontractors can record their progress frequently to allow scheduling of the next subcontractor after the first one has finished. The detection of scheduling changes, determination of the effects of the change, and then communication with the other contractors can enhance efficiency greatly, leading to less down time and ultimately lower costs. Integration of these kinds of services with back-end applications can contribute to an integrated workflow management system (Jain, 2003).

A firm in the U.S., that has successfully provided field service applications for mobile workers in a variety of industries is BlueVolt (Blank, 2003). They have developed a wireless work order application that allows contractors to send work instructions to PDAs that their field workers carry. PDAs are connected to the Internet via a wireless cellular connection. In the U.S. this service works with most wireless carriers.

In the field, workers are able to track their time and the materials that they used and sync back into the contractor's accounting system. The company took pains to make sure that applications would be consistent with pre-existing ways of doing business and systems. For example, their offerings are tightly integrated with Intuit's QuickBooks, a program that is widely used in the U.S. construction industry.

The main value proposition of BlueVolt is to save customers at least one hour per day on unproductive and many times unbillable activities such as paperwork, driving to and from the office to get work orders, and time-card processing. User adoption was anticipated to be an obstacle as field workers are typically not used to carrying computers. BlueVolt however, found that this has not been as much as a problem as originally believed. By emphasizing the user interface design, they have seen that construction crews have found the devices easy to learn.

Cultural Socialization

Organizational Norms

The construction industry has traditionally been slow to adopt new technologies (Mitropoulos & Tatum, 2000). Overall, studies of the organizational norms in the construction industry have suggested that managing information and knowledge is often a problem, as much construction knowledge is in the minds of people working in the sector. Decision-making remains hidden often as decisions may not be recorded or documented. Also, data had tended to not be managed, and is archived only at the end of the construction project. Often the key people involved move from one company to another. Without reporting of knowledge about projects, it is suggested that decision-making is difficult, and understanding of specific needs might be missed (Vakola & Rezgui, 2000). These factors would seem to mitigate against technology adoption and investment.

Technology Adoption Catalyst

Adoption of mobile construction applications appears to have occurred on an individual level and has spread through word of mouth. A central authority has

not mandated or recommended use, although competition within the industry has motivated companies to find new methods and technologies to increase efficiency and to cut costs.

The construction industry faces fewer restrictions surrounding decisions to utilize mobile applications than the medical field, yet there are still hurdles to cross. Civil engineering personnel who are traditionally conservative are often in charge of technology decisions. Observable benefits in terms of productivity of on-site personnel, however have provided encouraging evidence that has generated visibility for mobile solutions in this sector.

Mobile Medicine

Handheld devices have seen rapid adoption in the U.S. healthcare system. A 2003 report by Spyglass Consulting Group finds that over 90% of doctors under the age of 35 use handheld devices and software such as the Physicians Desk Reference, manuals and medical calculators on a regular basis. The utility of mobile computing has driven clinicians to purchase handheld devices at their own expense (Rosenberg, 2004) Despite the popularity of handhelds, hospitals have not been avidly adopting mobile technology or attempted to link handhelds to existing systems. This is due to several obstacles including funding, compliance with patient privacy rules (HIPAA), and integration with legacy-based systems (Rosenberg, 2004). Ironically, surveys of hospital administrators cited physician unwillingness to adopt technology as another significant obstacle (Rosenberg, 2004).

Under the present individual state licensure system the practice of telemedicine is hampered by an inability to cross state borders. Physicians are often required to have medical licenses in each state in which they practice. (Betty, 2000). Next, there are legal issues associated with telemedicine malpractice liability when services span state borders. Finally, there are concerns regarding the security of personal medical information stored in telemedicine systems.

Perceived Relative Value

Wireless mobile applications hold the promise of greater convenience, efficiency and cost-savings to the medical establishment. Overall, in the United States, the relative value for telemedicine applications that could increase patient access and lower costs seems attractive. Even though telemedicine technology has existed since the 1920s, however, usage has remained limited. A number of organizational and social constraints are often cited. The most prominent

obstacles include, low compatibility with existing medical practices, complexity of telemedicine equipment and interfaces, the absence of reimbursement by third party agencies, and incompatibility of state laws regarding telemedicine and licensure issues.

A study prepared by Arthur D. Little Consulting estimated that annual cost reductions could reach $36 billion (Moore, 1995). Savings could be generated from the following:

1) Reduced costs for serving patients, through reductions in time and travel for doctors and patients, fewer unnecessary referrals, and reduction in the numbers of medically trained personnel
2) Cost reductions from early diagnosis and treatment (Moore, 1995)

Even if perceived value can be calculated, organizational and institutional barriers within the sector of medicine pose formidable barriers to adoption. For instance, a web-based wireless software application designed to manage home-care patients in their own environment called Pixalere, has shown great potential for saving money in the health care system. The Fraser Health Authority in British Columbia is phasing in the Pixalere wireless device, based on estimates that a 10% reduction in home-care visits would save this health authority $450,000 annually. Patients and nurses are advocates of this enabling technology. Doctors however, are less enthused because the Canadian reimbursement system makes it impossible for reimbursement to occur for advice and diagnoses that are provided online (Tausz, 2003).

Usability: Compatibility Driver

Applications that build on the way that medical professionals traditionally do their job, but make the process faster and easier have seen rapid growth, however. Currently, it is estimated that almost 20% of physicians in the U.S. use PDAs (E-Healthcare Connections, www.e-healthcare-connections.com, 2003). Growth is predicted however, as power, storage, and better network connectivity can be expected to create more usable and versatile devices.

Use of handheld mobile devices in the medical establishment continues to grow dramatically. Medical residents are writing their own programs, and it is estimated that 80,000 medical applications have been developed for the Palm.

ePocrates, the prescription drug reference software has seen rapid penetration as 90,000 doctors have downloaded the free version of this software. The

software permits doctors and trainee doctors to look up protocols quickly on a PDA. Useful activities include, looking up drug names, side effects, and formulations. Since the database is available free of charge for basic information, it is possible to test it out without making an investment. Additionally, it is highly automated, so it is easy to use, overcoming any lack of computer skill and experience. Finally, the information provided by ePocrates is considered trustworthy, a critical attribute in the medical domain. A significant feature of this application is the fact that the information it provides is considered accurate and current.

Cultural Socialization

Organizational Norms

Overall, the social system surrounding the adoption of telemedicine is very structured and complex. Notably, a lack of clear support from key institutions, such as the American Medical Association and most medical colleges and medical schools, save the American College of Radiology is a serious impediment for many comprehensive mobile solutions or those involving patient records.

Strong barriers exist at the organizational level of the medical sector, but it does not seem to be a serious obstacle at the individual level. Dr. Thomas Fogarty, inventor of the balloon catheter and often considered the medical profession's "Thomas Edison," has commented on the medical profession's in-grained aversion to adopting new technologies. "There's an incentive not to be innovative," In addition, he maintains that there is an innate conservatism in the medical community which is threatened by innovation (Hiltzik, 2003). This assessment appears to have some merit at the organizational level at least where there is great sensitivity to hidden risk.

Technology Adoption Catalyst

When successful telemedicine projects are examined, often it has been found that the driving force behind them is an early adopter — a highly placed, motivated leader within the medical community. Interviews with telemedicine directors have revealed that project leaders tended to be charismatic entrepreneurs, enthusiastic and impatient for change and true believers in their cause (Pereira, Fife et al., 1997). An energetic leader is often able to overcome the strong structural and organizational barriers that are constraining the growth of telemedicine.

Leadership from a prominent individual or department is necessary often due to the cost of implementing a telemedicine infrastructure. Currently, a large majority of telemedicine initiatives are sponsored by organizations where reimbursement is not crucial, like research centers, the Armed Forces or state-owned hospitals, and these initiatives are frequently financed by demonstration grants. Only a small number of for-profit medical centers are involved in telemedicine and many of these, like the Mayo Clinic, are employing closed telemedicine systems (Tangalos, 1994).

The success of handheld reference software suggests that when conditions exist that make it possible for an application to be used — its compatible, user-friendly, and doesn't require broad consensus, adoption is possible.

Access to information through a handheld has been found to be extremely useful, but extensions of its use could further increase the value of this mobile data application. If it were possible to for a medical professional to write up notes and send them or to write prescriptions, speed and efficiency could be greatly improved. However, writing prescriptions presents institutional and structural impediments. Many doctors have suggested that it would be helpful if images could be transmitted by doctor wirelessly and attached to a record (Sawcer, 2003). However, this cannot be accomplished wirelessly, due to federal regulations that govern what is acceptable for handling patient data.

It is clear that significant non-technological barriers exist to the widespread adoption of telemedicine and mobile medical applications in the United States which principally are related to the cultural and social norms of the medical profession. The growth of handheld mobile applications in hospital systems reflects this obstacle. Use of the handheld reference software has come from the ground-up, as hospitals and health systems have not integrated mobile technologies with their core processes on a sustained basis (Rosenberg, 2004).

Mobile Learning

Mobile learning can be considered as an extension of electronic learning, and like e-learning has several contexts:

1) as an extension of traditional learning blending face-to-face interaction and remote access for the student, i.e., "blended learning,"

2) as a way to provide greater convenience to students on a campus in terms of communicating and accessing information, such as a wireless campus, and
3) as a form of distance education using a wireless device.

Initiatives for m-learning include a broad range of activities including downloading courseware onto a laptop or PDA to online teaching from an instructor. Despite technology advancements and curriculum development however, the potential of electronic and mobile education is still a matter of debate. Although applications are currently available for mobile education through a handheld device, we include laptops accessed through a wireless link such as wi-fi in this discussion.

Within the traditional university setting, e-learning initiatives are a supplement to standard means of learning, or provide specialized training and adult learning.

Perceived Relative Value

The benefits of m-learning and to some extent online education in general, are the "anytime" and "anywhere" attributes. Berg's (2003) study found in fact, that increased access is the principal reason that higher education institutions initiate e-learning projects. As with medicine, education can be considered a field with a medium level of mobility — users and providers of educational materials move around, but usually return to a fixed office or place for working.

Emerging m-learning services that have found an audience include guides and reference tools that support classroom instruction. Quick access to data is an advantage cited by students. Nonetheless, the dream of the "virtual classroom" has still not materialized on a mass scale, as students have been disappointed by offerings, and educational institutions have experienced high costs and other difficulties in providing quality distance education. The perceived value of m-learning, like online education is an extension of traditional education to lifelong learning that allows inclusion of a wider range of students.

The flexibility and customization of an electronic or mobile learning experience is suitable for corporate training and may be sufficient for some kinds of students, such as those who would not have access to traditional educational opportunities. Thus, although the electronic learning experience may not provide some of the richness of a university experience, the flexibility and customization that is possible, may make it good enough for some (Huynh, Umesh & Valacich, 2003).

One of the lessons from distance education efforts is the necessity of creating a new way of engaging students, rather than providing a traditional experience

through electronic means. Effective means of engaging students in an electronic education experience are still being discussed and explored. Given that pc-based electronic learning is in its formative stages, the challenges for mobile learning through a handheld device are all the more daunting since e-learning content through a pc is not necessarily transferable to a mobile device with its limited functionality and screen size.

As is the case with telemedicine, there are still questions about the cost-effectiveness and utility of distance learning or mobile learning activities. Benefits have not been clearly and visibly communicated to key decision-makers although there are individual examples that seem to be successfully deployed.

Usability: Compatibility Drivers

M-learning is usually viewed as complementary to e-learning, and analogous to mobile games and information retrieval through mobile devices. The relative advantage of m-learning appears greatest for adult students seeking specialized training or extension education. Although growth in m-learning can be anticipated, it is most likely that it will be an accompaniment to traditional educational methods, rather than a substitution.

In fact, electronic learning is still an emerging field, as providers grapple with feasibility issues. Although e-learning has much potential, overcoming challenges related to cost and curriculum development will have an impact on the development of m-learning models. Working adults with full time employment are the largest student group seeking online education since the primary attractions of e-learning are the convenience and flexibility that are possible (Huynh, Umesh & Valacich, 2003).

Cultural Socialization Driver

Drivers of electronic learning have initially come from outside the traditional university environment and include corporate universities such as Disney and Motorola, or virtual universities, for-profit education for adult students, such as the University of Phoenix. Finally, start-up ventures and strategic alliances between traditional universities and educationally-oriented technology companies have all sought the adult working professional. The on-campus student is most apt to use m-learning as a supplement to a traditional learning experience although target students for m-learning and e-learning will likely remain professional and corporate workers or working adults.

Organizational Norms

Implementation of e-learning initiatives typically resides in continuing education or distance learning departments, yet assessment and approvals tend to go through the traditional channels of universities (Berg, 2001). Overall, distance learning tends to be administered outside of the central university.

Issues associated with faculty motivation to provide online or mobile education are an oft-cited factor to the slow adoption of e-learning. It is thought that faculty compensation is a problem since additional time and resources are required to develop course material. Additionally, the use of e-learning is a drastic change from the traditional classroom model for education, and faculty see the vision of a small number of faculty delivering education to thousands of students as antithetical to an academic culture (Morrissey, 2002)

Technology Adoption Catalyst

Enrollment in distance learning programs has grown in recent years. In 2003, it was estimated that over 500,000 students were earning degrees in e-learning programs (Symonds, 2003). It has been seen as a great untapped market by many, and it is estimated that $6 billion in venture capital has flowed into the e-learning business during the 1990's to corporations and universities (Huynh, Umesh & Valacich, 2003).

Besides entrepreneurial activities, like telemedicine, many e-learning programs are subsidized by institutions or government. A champion is required to motivate participation. Despite many products and services, it is fair to say that the m-learning and e-learning environments have not achieved centrality in the educational system in the U.S. As of yet, the perceived relative value for students and for educational systems is somewhat ill-defined and organizational norms have also hindered enthusiastic development of online offerings. The high degree of learning required on the part of faculty and support staff has also been neglected in some analyses of e-learning.

Ultimately, however, the cost issue is of paramount importance for e-learning and mobile learning — and the means for delivering mobile and e-learning in a commercially sustainable manner has for the most part not been established. Berg (2002) notes that surprisingly few programs have conducted cost-benefit analyses and that a large percentage of e-learning programs are subsidized.

Berg's (2002) study which found that the primary motivation for institutions to provide distance learning is to increase access also shows that the idea of distance education providing a distinct learning advantage was a distant second

in terms of motivation. In fact there is heated debate over the general effectiveness of online education compared to traditional face-to-face interaction. Some suggest that the motivation to learn should be further investigated and that the value of taking classes on a campus should not be underestimated. For example, mentoring, conversations with professors and interacting with students outside of class are extremely important, in addition to the social atmosphere of a university which makes students reluctant to drop out for fear of losing face (Hamilton, 2001).

Overall, the perceived relative value for the most part has not been demonstrated. Both pedagogy and cost issues are still relevant concerns that have not been visibly reconciled and thus, m-learning and e-learning for the most part are still supplementary, rather than principal means of learning in higher education settings.

Conclusion

This exploratory study suggests that several commonly used models for understanding the diffusion of technology, although helpful, do not provide complete understanding of technology diffusion on a sectoral basis. We have examined three sectors in the U.S. where mobile technologies are diffusing to different degrees in order to examine current frameworks and suggest modifications that may provide more depth of understanding. Our cases are intended to help substantiate the modified framework we propose, and also to shed light on the useful elements of current models and show where new measures can be included.

A significant determinant of mobile technology adoption in these three cases is social interaction; as word spread about the benefits of using a mobile service or application, growth increased. Circumstances that make it easier for word of mouth to spread are networking among organizations in a sector and a medium to high level of mobility of both clients and suppliers of the service or product. Social factors and communication have helped the spread of mobile services and devices that are compatible with the activities of the organization, and are thus easily incorporated into current practices. Successful mobile services are user-friendly, are cost-effective and are visible improvements to conventional means of carrying out particular activities. In the case of mobile medicine and mobile construction applications, it appears that word of mouth, or socio-contagion is an extremely important factor in explaining the growing use of mobile applications, i.e. PDAs in medicine, and mobile networks and devices in construction.

Investigation of medicine, education and construction, three diverse fields that all have technology requirements, reveals that there must be acute awareness of social norms, especially when a new technology is somewhat ill-defined in nature and in terms of costs and benefits. Also, even in fields that have a hierarchical organization, often technology decisions can come from the bottom ranks. In our cases of mobile medicine and mobile construction for instance, technology adoption did not seem to come through the prompting of a "change agent" or champion.

References

Abernathy, D. (2001). Get ready for m-learning. *Training and Development*, *55*(2), 20-22.

Ahn, H. & Lee, M. (1999). An econometric analysis of the demand for access to mobile telephone networks. *Information Economics and Policy*, *11*, 297-305.

Banerjee, A. & Ros, A. (2002). Drivers of demand grown for mobile telecommunications services: Evidence from international panel data. *ITS 14th Biennial Conference*, Seoul, Korea.

Bass, F. (1969). A new product growth model for consumer durables. *Management Science*, *15*, 215-227.

Berg, G. (2002). *Why distance learning? Higher education administrative practices*. Westport, CT: Oryx Press.

Betty, C. (2000). Telemedicine can lower costs and improve access. *Healthcare Financial Management*, *54*(14), 66.

Blank, J. (2003). BlueVolt, E-mail interview.

Cotter, M. (2003). A mobile construct. *Field Force Automation*, 28-34.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340.

de Mooij, M. & Hofstede, G. (2002). Convergence and divergence in consumer behavior: Implications for international retailing. *Journal of Retailing*, *78*(1), 61-69.

Denis, J. & Hebert, Y. (2002). Explaining diffusion patterns for complex health care innovations. *Health Care Management Review, 27*(3), 60-73.

E-Healthcare Connections, The Rise of Palmtop Technology in Medicine, Part 11, *E-Healthcare Connections*.

Eliashberg, J. & Helsen, K. (1996). Modeling lead/lag phenomena in global marketing: The case of VCRs. Working paper, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Gilbert, L.A. & Kendall, J. (2003). A marketing model for mobile data services. *Proceedings of HIICS-36*, Honolulu, USA, 89-98.

Gold, B. (1981). Technological diffusion in industry: Research needs and shortcomings. *Journal of Industrial Economics*, *29*(3), 247-269.

Gruber, H. & Verboven, F. (2001). The diffusion of mobile telecommunications services in the European Union. *European Economic Review*, *45*, 577-588.

Hall, E. (1987). *Hidden differences*. New York: Doubleday.

Hamblen, M. (2002). Home builder relies on wireless for construction schedules. *ComputerWorld*. *http://www.computerworld.com/mobiletopics/mobile/story/0,10801,72444,00.html*

Hamilton, D. (2001). E-Commerce report: The classroom-no substitute: The Internet does not change everything. *Wall St. Journal*, R32.

Hiltzik, M. (2003). Medicine's own Thomas Edison. *Los Angeles Times,* C1.

Hofstede, G. (1993). Cultural constraints in management theories. *Academy of Management Review, 7*(1), 81-94.

Hu, P., Chau, P. et al. (1999, Fall). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of Management Information Systems*, *16*(2), 91-112.

Huynh, M., Umesh, U. & Valacich (2003). E-Learning as an emerging entrepreneurial enterprise in universities and firms. *Communications of the Association for Information Systems*, *12*, 48-68.

Jain, R. (2003). Enterprise mobile services: Framework and industry-specific analysis. *Ninth American Conference on Information Systems*.

Jarvenpaa, S., Lang, K.R., Takeda, Y. & Tuunainen, V. (2003). Mobile commerce at crossroads. *Communications of the ACM*, *46*(12), 41-44.

Kay, D. (2003). E-learning: Drivers, developments, and decisions. *Multimedia Information and Technology*, *29*(1), 26-29.

Kim, J., Lee, I. et al. (2003). Exploring e-business implications of the mobile Internet: A cross national survey in Hong Kong, Japan and Korea. *International Journal of Mobile Communications, 2*(1), 1-21.

Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research*, *13*(2), 205-223.

Kumar, V. & Krishnan T. (2002). Multinational diffusion models: An alternative framework. *Marketing Science, 21*(3), 318-332.

Lee, Y. & Kim, J. (2004). What is the mobile Internet for? A cross-national study on the value structure of the mobile Internet. (unpublished – in review, *Communications of the ACM*)

Levitt, T. (1983). The globalization of markets. *Harvard Business Review*, *61,* 2-11.

Luna, D. & Gupta, S. (2001). An integrative framework for cross-cultural consumer behavior. *International Marketing Review, 18*(1), 45-69.

Lundblad, J. F. (2003). A review and critique of Rogers' diffusion of innovation theory as it applies to organizations. *Organization Development Journal*, *21*(4), 50-64.

Madden, G., Coble-Neal, G. & Dalzell, B. (2002). Economic determinants of global mobile telephony growth. *14th Biennial Conference of the International Telecommunications Society*, Conference paper, Seoul, Korea.

McGrath, C. & Zell, D. (2001). The future of innovation diffusion research and its implications for management. *Journal of Management Inquiry, 10*(4), 386-391.

Mitropoulos, P. & Tatum, C.B. (2000). Forces driving adoption of new information technologies. *Journal of Construction Engineering and Management*, *126*(5), 340-348.

Mobileinfo.com (2003). Vertical applications, *Mobileinfo.com*

Moore, G. (1995). *Crossing the chasm*. New York: Harper Collins.

More, M. (1995). Telehealth cost justification. *http://naftalab.bus.utexas.edu/nafta-7/costjust.html*

Morrissey, C. (2002). Rethinking the virtual university. *Communications of the AIS*, 9.

Muirhead, G. et al. (2000, March 30). An update on telemedicine. *Patient Care*, 16, 96-109.

Pereira, F., Fife, E. & Schuh, A. (1997). Telemedicine: An inquiry in the economic and social dynamics of communications technologies in the medical field, Conference Paper, Webnet, Toronto, Canada.

Robinson D, Savage G., & Campbell, K. (2003). Organizational learning, diffusion of innovation and international collaboration in telemedicine. *Healthcare Management Review, 28*(1), 68-93.

Rogers, E.M. (1995). *Diffusion of innovations*. New York: The Free Press.

Rogers, E.M. (2002). The nature of technology transfer. *Science Communication, 23*(3), 323-341.

Rosenberg, R. (2004). Early A-doctors: Improving patient care and saving lives-one handheld at a time. *Mobile Enterprise*, 34-38.

Saaksjarvi, M. (2003). Consumer adoption of technological innovations. *European Journal of Innovation Management, 6*(2), 90-100.

Sarker, S. & Wells, J. (2003). Understanding mobile wireless device use and adoption. *Communications of the ACM*, *46*(12).

Sawcer, D. (2003). Personal interview, (M.D., Keck School of Medicine, University of Southern California).

Selian, A. (2004). Mobile phones and youth: A look at the U.S. student market, *ITU/MIC Workshop on Shaping the Future Mobile Information Society*, Switzerland: International Telecommunications Union.

Symonds, W. (2003). Cash-cow universities for-profits are growing fast and making money. Do students get what they pay for? *Business Week*, *3858*, 70-74.

Takada, H. & Jain, D. (1991). Cross-national analysis of diffusion of consumer durable goods in Pacific Rim countries. *Journal of Marketing*, *55*, 48-54.

Tangalos, E. (1994). Telemedicine: An information highway to save lives. Written testimony to the Telemedicine hearing before the Subcommittee on Investigations and Oversight, Committee on Science, Space and Technology, U.S. House of Representatives, 103th Congress. U.S. Government Printing Office.

Tausz, A. (2003). Software permits 'virtual' house calls. *Toronto Star*.

Vakola, M. & Rezgui, Y. (2000). Organizational learning and innovation in the construction industry. *The Learning Organization*, *7*(4), 174-183.

Van den Bulte, C. & Lilien, G. (2001). Medical Innovation Revisited: social contagion versus marketing effort. *The American Journal of Sociology*, *106*(5), 1409-1435.

Venkatesh, V., Morris, M., et al. (2003, September). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478.

Venkatesh, V. & Speier, C. (1999). Computer technology training in the workplace: A longitudinal investigation of the effect of mood. *Organizational Behavior and Human Decision Processes*, *79*(1), 1-28.

Vinson, D. & Scott, J. et al. (1977, April). The role of personal values in marketing and consumer behavior. *Journal of Marketing*, *41*(2), 44-50.

Whitten, P., Mair, F.S., et al. (2002, June 15). Systematic review of cost effectiveness studies of telemedicine interventions. *British Medical Journal*, *324*(7351), 1434-1437.

Worldwide Mobile Internet Survey (2004). Survey findings presented at conference, Yonsei University.

Wootton, R. (2001). Telemedicine. *British Medical Journal, 323*(7312), 557-560.

Section III

Business Opportunities with Mobile Services and Applications

Chapter IV

Incorporating Commercial Space Technology into Mobile Services: Developing Innovative Business Models

Phillip Olla, Brunel University, UK

Abstract

This chapter will describe how space technologies can be incorporated into terrestrial 3G /4G mobile telecommunication infrastructure to provide convergent innovative applications and services. The utilization of space applications for non-military use has the potential to generate significant economic, social and environmental benefits on a global scale. The satellite infrastructure will become a key enabling factor in a growing range of

mobile products such as: voice services, broadband Internet services, navigation, and observation systems. The chapter presents a framework derived from the literature to aid the development of viable business models expected from the amalgamation of mobile telecommunication and space infrastructure. The chapter also identifies the various actors involved in the delivery of these services which include: technology actors, service providers, network operators, consumers, and regulators.

Introduction

There are significant benefits that can be realized from incorporating space technology into the terrestrial communication technologies. Currently these benefits are not being realized due to a lack of technical and economical integration of the various network technologies. The current business models exhibited by the various telecommunication providers are focused on competition, ignoring the huge potential that can be achieved by convergence and cooperation. This problem is inherent in the business models that are created independently by various types of network providers. There is no consideration for convergence opportunities. Most of the satellite and mobile network providers that provide communication capabilities via Low earth Orbit (LEO) satellite providers and GSM technologies are often competing in the same space rather than concentrating on their core capabilities and cooperating to generate sustainable business models in the current harsh economic environment.

There are views from various organizations such as, (ESA-Homepage, 2003), (OECD, 2003; UN-Program, 2002; UNESCAP-Report, 2002) that space technology infrastructure will become a key enabling factor for a convergent global mobile telecommunication infrastructure. A growing range of mobile products and services currently in use today or under development will incorporate space technology such as: voice services, radio, broadband Internet services, navigation, and observation systems and gravitational research. The current trend of developing business models for applications and services does not go far enough to investigate generic business models for mobile applications and services that are network independent and which incorporate space technology.

The literature offers various explanations for deriving business models on mobile networks in an ineffective manner due to the evolution of the mobile value chain and market structure outpacing the research (Sabat, 2002). This chapter aims to address this confusion by providing an integrated view of the evolving mobile and satellite markets, and uses the business model framework, to identify market actors to encourage the business world to deliver on the full potential of space technologies in the global mobile arena.

The commercial use of space technologies is a promising and emerging industry characterized by large numbers of technological and strategic uncertainties. This chapter will aim to address the strategic uncertainties caused by inadequate business models. The new breed of business models will need to cater for the increase in the number of actors trying to accurately position themselves in an advantageous position in the value chain. Similar to the mobile commerce business models (Camponovo & Pigneur, 2002; Maitland, Bauer & Westerveld, 2002; Pigneur, 2000; Sabat, 2002; Tsalgatidou & Pitoura, 2001). Successful space technology business models are likely to be the ones that address the economic peculiarities such as mobility, precision positioning, network effects, broadcasting, and communication in a flexible manner.

All actors in the mobile and satellite arena need to explore new revenue generating opportunities to increase their market share and sustain their competitive advantage. Although this chapter describes the implementation of futuristic new business models, the key enabling factor for the success is not the advancement of the technology but the convergence of existing technologies and ideas. This chapter proposes the use of a business model framework for the creation of innovative business model by analyzing the existing actors technical capabilities, portfolios, strengths and competencies and adapting their current business models to harness the full possibilities for new revenues and market share. The chapter is structured as follows. The first section provides an overview of the business model literature followed by a discussion on the business model framework presented in this chapter. The next section identifies the advancements of the mobile communication industry over the last three decades and sets the scene for the discussion on what future technological infrastructure is likely to be. This is followed by the presentation of the business model framework along with examples. A brief conclusion provides a summation of the concepts presented in this chapter.

Background: Using Business Models to Create Innovative Propositions

This section investigates the business model literature to understand what a business model is, along with various components that make up business models, and to understand the various uses of business models. There are various definitions in the business literature of what constitutes a generic business model but some fail to pay explicit attention to technology (Weill & Vitale, 2001). While others fall short in the area of defining the multiplicity of actors. An appropriate definition proposed is as follows:

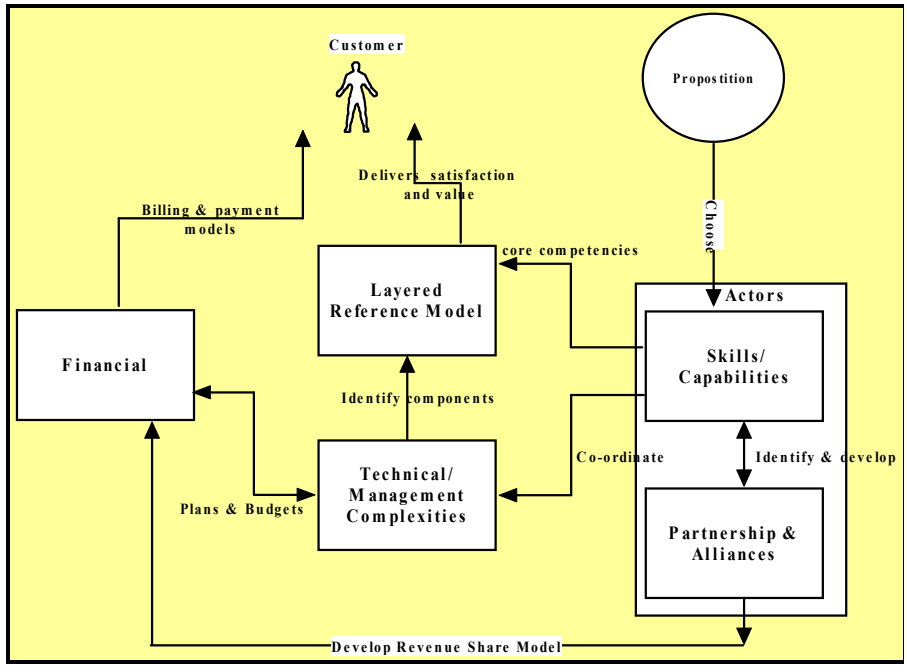
> *"A business model is the architecture of an organization and its network of partners for creating, marketing and delivering value and relationship capital to one or several segments of customers in order to generate profitable and sustainable revenue streams" (Pigneur, p. 2).*

Some researchers from the e-business and Internet arena focus on the revenue aspects of a business model, this use of a business model identifies how an organization generates revenue by its positioning in the value chain (Rappa, 2000). Other approaches look at the business model from the business actor perspective (Afuah & Tucci, 2001; Amit & Zott, 2001) describing the business model as an architecture for products, services and information flow, which provides a description of the various business actors and of their roles, the potential benefits of these actors and the sources of revenue (Bouwman & Ham, 2003). Afuah and Tucci (2001) introduce an approach to business models which emphasis the value perspective and considers the creation of value through several actors (Afuah & Tucci, 2001). While Amit and Zott (2001) describe an e-business model as the architectural configuration of the components of transactions designed to exploit business opportunities.

Literature for generating business models for mobile or wireless propositions is just as complex and perplexing as the generic e-business models. The work of Camponovo and Pigneur (2002) focused on the mobile market actors landscape, and proposed a conceptual tool for identifying key actors, their business models, interactions and their dependencies (Camponovo & Pigneur, 2002). The mobile business model ontology provided by Camponovo and Pigneur (2000) is the conceptualization and formalization into elements, relationships, vocabulary, and semantics. The ontology is structured into several levels of decomposition with increasing depth and complexity. The first level of decomposition of their ontology contains the four main pillars of a business model, which are the products and services offered by the organization , the relationship the organization maintains with customers, the infrastructure necessary in order to provide this and finally, the financials, which are the expression of business success or failure. An important contribution of approach by Bouwman & Ham (2003), Camponovo & Pigneur (2002) is the emphasis placed on the increasing significance that the organizations in the mobile business market attach to building partnerships. Participants of 3rd generation mobile network operators TMobile and MMO2, who are congenital competitors, have been given permission to share their 3G network infrastructures to save network development costs (OFTEL, 2003). This approach to a discreet mutual interest is key to speeding up investments and roll-out of new technological infrastructure (Maitland et al., 2002).

Members of value webs cooperate in the development of enabling technologies, the integration of corporate information systems and the development of middleware solutions, open platforms and standards (Camponovo & Pigneur, 2002). In addition to the technical cooperation at the network development and generation of standards, the 3G mobile business market have other resources that can be used to create a competitive advantage in the market. It is believed by some researchers (Bouwman & Ham, 2003) that the cooperation of network providers and content providers from fixed communication, Internet and mobile services of 2G will generate the highest quality of (Maitland et al., 2002). The former tendency of network operators to develop mobile content in-house has diminished due to a shortage of skills and expertise which is increasing the role played by potential partners. With new potential markets opening up, due to the emergence of new technologies, organizations are looking to partners to accomplish the complex mission of service delivery. Identifying partners with access to key functions such as billing and information sharing, appears to be of great importance in the competition and creation of viable business models for the organizations (Bouwman & Ham, 2003).

Figure 1. Business model framework



In order to correlate the mobile business value chain models (Barnes, 2002; Li & Whalley, 2002; Maitland et al., 2002; Olla & Patel, 2002) with the emerging technological innovation in the satellite and space technological domain, this chapter proposes a business model framework which takes four elements into consideration as illustrated in Figure 1. These elements include: actors, financial arrangement, mobile systems reference model, and technical/managerial complexity. These elements were arrived at by understanding the fundamental business drivers of mobile value chains (Bouwman & Ham, 2003; Camponovo & Pigneur, 2002; Olla & Nandish, 2002; Talluri, Baker & Sarkis, 1999; Wirtz, 2001), along with an understanding of the value chain elements of space technologies and their integration capabilities. To correlate the current mobile business value chain (Bouwman & Ham, 2003; Camponovo & Pigneur, 2002) with the emerging technological innovation from the convergence with space technologies, the next section proposes a business model framework which operates along four dimensions deemed important from the business model literature presented above.

The model developed takes into consideration information from various sources. The first input was from "mobile casual model proposed" by Bouwman (Bouwman & Ham, 2003). This model links the organizational, technical and financial arrangements to a clearly defined business model. Customer value will be decisive with regard to the viability of a business model of a specific mobile service. The next component was the wireless reference model proposed by Olla & Patel (2003) which defined the various layers of a viable mobile system proposition. The reference model is used to understand the various system constituents of wireless applications defining the central elements and proposes a common vocabulary of terms for discussing a mobile proposition. The financial contribution was the derived from literature on the Ontology of M-business and e-business models (Camponovo & Pigneur, 2000, 2002; Pigneur, 2000; Osterwalder, Lagha & Pigneur, 2003). Pigneur (2002) proposes that the best products and services and the finest customer relationships are only valuable to a firm if it guarantees long-term financial success. The financial aspects element is composed of the company's *revenue model* and its *cost structur*e, which determines the *profitability* of a company (Pigneur, 2000).

An important factor in a successful business model is the alliances formed by the various actors involved in developing a proposition. A strategic technology alliance or partnership is a long term, continuous, and mutually beneficial vertical non-equity relationship (Keil & Vilkamo, 2003). Confidential information on plans and visions is shared openly and proactively in order to help all organizations involved, to focus their resources for a particular cause. Since the companies commit to each other and thus become interdependent, they typically also strive to align strategies and support each other's development in order to maximize the outcome of the relationship.

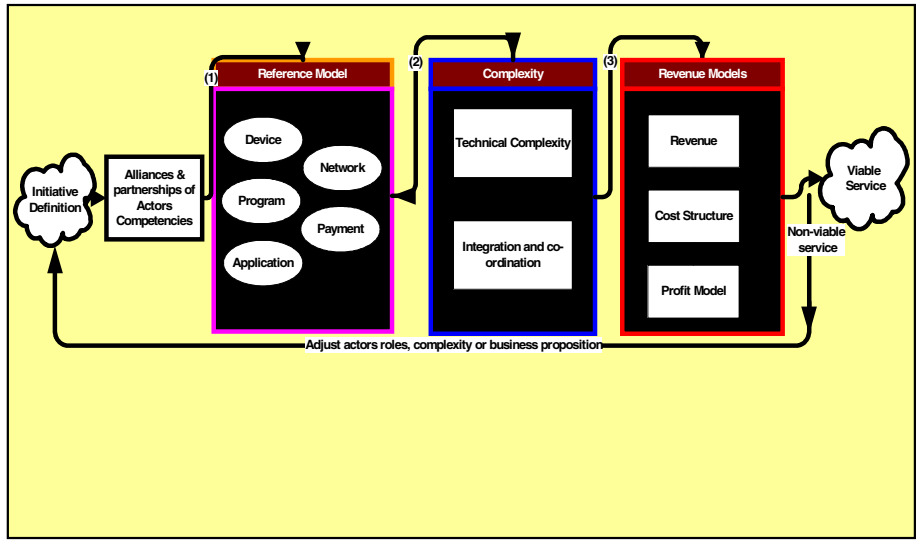
The space technology sector which includes technologies such as location monitoring, television broadcast and communication, and similar to the mobile technology sector is extremely fragmented with a large number of potential market actors. The primary actors are device manufacturers, content providers and payment aggregators, satellite network operators and mobile network operators and space agencies. Other actors that have an indirect impact are the regulation authorities, standardization groups, consumer groups, transit localities authorities (such as airports and train station authorities) and other retail venues such as coffee shops and conference centers. As with the mobile value chain, (2002), due to the complex nature of the market no single actor can provide a service to the customers with an end-to-end solution on its own (Barnes, 2002; Pigneur, 2000). There is a need to sustain viable alliances, however creating a value chain with the right partners positioned in the right part of the value chain is a challenging feat. Partnership management capabilities will have to be a core competence that new mobile-satellite business actors must posses. It is not enough to examine the actor's individual role in the chain, but relationships and interactions among the other actors in the chain have to be assessed concurrently (Pigneur, 2000). Pigneur recommends the use of business models for a brief and clear description of the roles of the different key actors.

Before a coherent discussion can occur on the use of business models to create innovative and viable propositions, it is important to provide an overview of the progression of mobile telecommunication infrastructure. The next section identifies the advancements of the mobile communication industry over the last three decades and sets the scene for the discussion on what future technological infrastructure is likely to be.

Developing a Business Model Framework to Generate Viable Mobile Satellite Propositions

The literature offers various, explanations for business models on mobile networks in a disconcerted manner. The evolution of the mobile wireless value chain and the market structure has outpaced research (Sabat, 2002). The aims of the framework presented in this section is to encourage the business world and policy makers to deliver on the full potential of space technologies in mobile environments, by providing new ways of thinking about creating innovative business propositions. All actors operating in the mobile and satellite arena need to explore new revenue generating opportunities to increase their market share and sustain their competitive advantage. This section provides an integrated

Figure 2. Business model framework

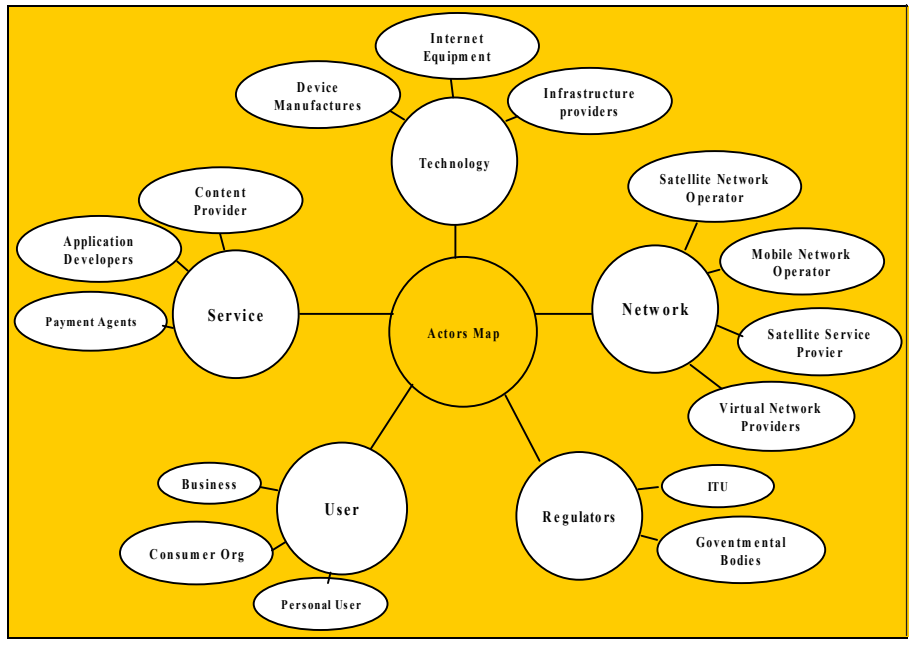


view of the evolving mobile and satellite value chain, and uses the business model framework to identify market actors that have developed alliances to offer mobile applications and services.

Using Alliances to Achieve Competency

This section of the framework requires the identification of all the potential actors along with the role they will be expected to perform in order to contribute to the creation of a viable business proposition. Some organizations may incorporate one or more business roles, such as network operator, content aggregator and content provider (Barnes, 2002), but generally the typical set of actors involved are illustrated in Figure 3. Organizations have to find their position in the value chain, service providers and content providers have to agree on their respective roles for each of the business models (UMTS-Forum-Report21, 2002). Allowing each organization to concentrate on what its does best to the development of a viable proposition. An important undertaking that needs to be carried out during the business modelling stage is the identification and differentiation of the actors' competencies, capabilities skills along with the alliances the are created.

Figure 3. Actors map for mobile-satellite propositions



According to Hamel and Prahalad, competence is "a bundle of skills and technologies rather than a single discrete skill or technology" (Prahalad & Hamel, 1990). The content of competence is typically divided into technological and managerial components, and both need to be explicitly defined during the modeling stage. Sanchez, Heene & Thomas (1996) view capabilities as "repeatable patterns of action in the use of assets to create, produce and/or offer products to a market." A *skill* is "a special form of capability, with the connotation of a rather specific capability useful in a specialized situation or related to the use of a specialized asset." Competence, at a general level, refers to knowledge and skills needed to choose what task to perform as well as why and how to perform the chosen task (Seppanen, 1998). Within the context of the framework the competence required by the actors in the lead organization relates to possessing the appropriate knowledge and skills to select a the appropriate organizations to partner with to deliver value. Each organization in the value chain should only focus on its core competencies and activities and should rely on partner for other non-core competencies and activities, this is an important potential for cost savings in the value creation process (Osterwalder et al., 2003).

When considering the formation of alliances the evidence from the literature suggests that organizations behavior varies across environments (Burns & Stalker, 1961). For high-velocity environments, such as mobile telecommunication and satellite environments authors have argued that organizations exhibit behavior that differs from behavior in more static environments (Thomas, 1996). In high velocity environments, strategies are often more concerned with change (Eisenhardt & Brown, 1998), speed (Eisenhardt, 1989) and flexibility (D'Aveni, 1994) rather than trying to build up sustainable strategic positions. Another typical characteristic is that the risks of cooperation shared for example by sharing the development costs or other investments. Partners share their resources, knowledge and capabilities with the objective of enhancing the competitive position of each partner (Spekman, Forbes, Isabella & MacAvoy, 1998.).

In high-velocity environments, changes are not only fast but they are also discontinuous requiring a management that stresses flexibility. Keil and Vilkamo (2003) identified several important elements of managing strategic technology alliances in high velocity environments. Elements include the management of multiple time scales, balancing exploration and exploitation, integrating technology collaborating into technology strategy, and managing the balance of continuity and change (Keil & Vilkamo, 2003).

With the new alliances and convergence of satellite and mobile services there is the risk that current privacy and security issues could be further amplified. Regulators such as International Telecommunications Union (ITU) will need to monitor the situation to implement new privacy legislations relating the location information being widely available and the extent of information sharing amongst operators. Another issue that will need to be addressed in the future is that different legal regimes apply in the various convergent environments such as: Space Law (1967 Outer Space Treaty), Radio Communications Regulations (ITU), International Private Satellite Organizations (INTELSAT) Broadcasting Law/Regulation (TV without Frontiers Directive).

Reference Model

An important element in research of mobile computing is the production of a reference model (Kleinrock, 1997). Using a reference model in the definition of the business model allows for a consistent discussion of the potential initiatives attributes and features. It structures the discussion in a way that characterizes the view of the system as seen by the user and the view of the user as seen by the system. The dimensions of this reference model depicted in Figure 2 (arrow 1) include the following: Application layer, Program layer, Network layer and

Device layer. The purpose of the reference model was to provide the ability to describe with consistency, each type of project.

(a) **Device Layer:** The device layer deals with issues such as the user interface design, navigation and device software; this layer allows the user to interact with the system. Some example devices are:

- Voice Centric (VC) device with music (e.g. MP3) these display of images display and JAVA capability.
- Smartphones these have voice capability with camera, PIM software and larger display.
- Personal Digital assistants (PDA), they are similar to smartphones with easier input for data, more storage, and larger screens.
- Data card is used to connect a laptop to a mobile network at varying speeds dependant on network availability.

(b) **Network Layer:** The second layer of the reference model is the transmission backbone involved in communications, including transportation, transmission, and switching for voice and data. The 2.5G,3G and satellite 25 networks discussed in the previous section have the bandwidth to support wireless data applications and provide mobile Internet access. This is fuelling the demand for innovative mobile Internet data applications and services.

(c) **Program Layer:** This layer deals with the issues on security, business logic, systems logic, data management issues and integration of the devices from the applications.

(d) **Payment Layer:** This layer describes the payment model to be applied for the service or use of the application. The method for collecting the payment from the subscriber should be explicitly stated when defining the proposition, by all parties to allow the revenue share model to be agreed (UMTS-Forum-Report21, 2002). This layer will feed into the Financial part of the framework. Example of payment models terminating short message service, subscription, premium short code, pre payment model and event billing.

(e) **Application Layer:** In today's environment of wireless applications systems most of a system's components are acquired ready to be installed via systems configuration. The applications layer represents the explanation of what services will be available to the user.

Complexity: Coordination and Integration Management

The portion of the framework concentrates on describing the complexity of the innovation required to fulfill the initiative. This task is very subjective and only provides an indication of what needs to be done on a technical level. The next element in the framework asks the business owners to consider how the work should be coordinated and managed. Using the technological infrastructure to provide business value, sustain competitive advantage, and enable novel and adaptive organizational forms is well recognized by practitioners and academics (Orlikowski & Robey, 1991). The management of end-to-end processes for acquiring suitable products and partners and identification of the skills and competencies that are required is the role of the lead firm in the value chain. The coordination of the various actors means the broker in the chain must have a full view of all the activities performed by the independent actors. For example, a location sensing application can potentially incorporate the following actors mobile network infrastructure, content providers, content developers, content aggregators and hosting providers, software and application platforms, customers segments, customer data, payment and billing, customer support, and management. The role of a broker in the value chain is a colossal task which is normally carried on by the actor responsible for managing the customer relationship.

Financial and Billing Considerations

The commercial element of the framework is formulated by defining the value proposition of the business initiative. Creating three separate models, Revenue Model, Cost Structure Mode and Profit Model fulfills this activity, determining this portion of the framework determines the propositions profit model and the ability to survive and compete (Osterwalder et al., 2003). The revenue model is an element that measures the ability of a firm to translate the value it offers its customers into money and therefore generate incoming revenue streams. The organizations revenue model can be composed of different revenue streams with all having different pricing models

With the mobile convergent model described in more detail in the next section, there are three possible sources of revenue, interconnect traffic from roaming activities, standards call charges and value added services offered. The cost structure model measures all the costs the firm incurs in order to create, market and deliver value to its customers. It sets a price tag on all the resources, assets, activities and partner network relationships and exchanges that cost the organization money (Osterwalder et al., 2003). This Profit model is outcome of the

difference between revenue model and cost structure. Currently the voice telecommunication centric model divides propositions at a high level into pre-paid and post-paid services. In the future models this approach will need to change, allowing service transactions to be managed in real-time or near real-time, in order to control expenditure and eliminate credit risk especially due to potential of ubiquitous roaming. The billing and collections functions will ultimately become a single role within the business environment. There are considerable challenges to be defined around the settlements and interconnection charging if the true cross networks (Mobile –> Satellite, Satellite –> Mobile) roaming becomes a reality. Currently there is the capability to roam between mobile and satellite networks but this is restricted to one way, that is some satellite mobile communication users can roam onto the GSM network to make voice calls but GSM users don't have the capabilities to roam onto a Satellite network for voice or data calls. Considering the number of satellite users to GSM users the beneficiary from opening up the networks would be the satellite operators and the customers, this would be a serious competitive advantage to any mobile operator or device manufacturer to develop such an initiative.

Using the business model framework the next section provides examples of two mobile initiatives that incorporate space technologies. The framework helps to provide an understanding of how the nature and complexity of these developments can be explored, to aid the decision maker in the business world to appreciate the underlying technical and integration issues.

Trends in Mobile Communication Technologies

The wireless high speed packet data has received enormous attention in both the academic domain and the mobile industry under the context of 3G standardization (third generation) This trend is driven by the mobile Internet; the mobile Internet evolution is achieved largely through the Global System for Mobile Communications (GSM) technology platform (GSM-Information, 2004). To fulfil the needs of the wireless Internet higher bandwidth will be required, as in the current situation with the wireline Internet. There is a strong belief by some (Qiu, & Zhang, 2002) that the mobile Internet will make wireline and wireless converge and fuel the development of new applications demanding an even higher bandwidth. Some applications will require data rates of up to 10-20Mbps for applications like real-time streaming video and the mobile office concept. Mobile communications can be divided into three distinct eras identified by the increase in bandwidth as illustrated in Figure 4. These eras relate to the implementation

of technological advancements in the field. The industry is currently on the verge of implementing the 3rd technological era and at the beginning of defining what the next step for the 4th era should be.

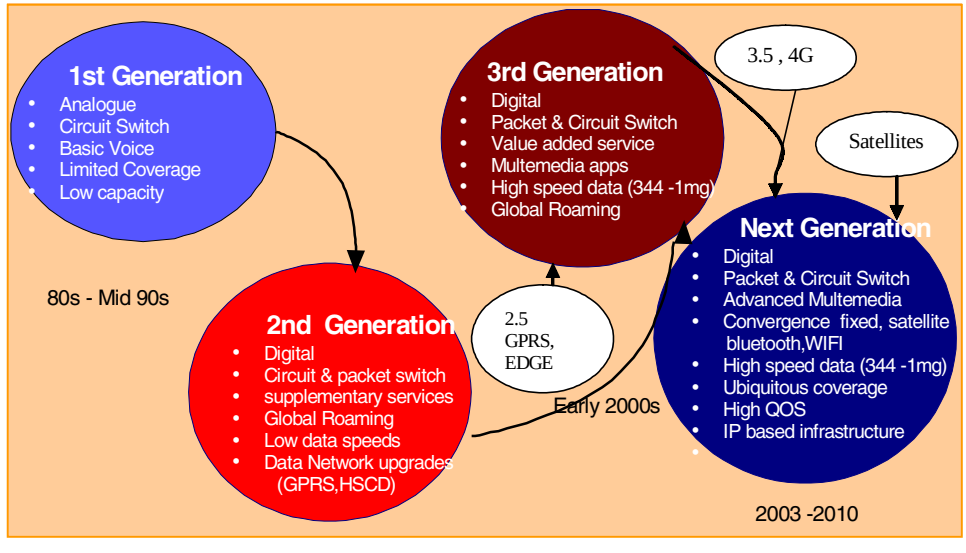
First Generation (1G) Mobile Networks

The first-generation cellular systems (1G) were the simplest communication networks deployed in the 1980s. The first-generation networks were based on analog-frequency-modulation transmission technology. Challenges faced by the operators included, inconsistency, frequent loss of signal and low bandwidth. The 1G networks were also expensive to run due to a limited customer base.

Second Generation (2G) Networks

The second-generation cellular systems (2G) were the first to apply digital transmission technologies such as Time Division Multiple Access (TDMA) for voice and data communication. The data transfer rate was on the order of tens of kbits/s. Other examples of technologies in 2G systems include Frequency

Figure 4. Mobile telecommunication eras



Division Multiple Access (FDMA) and Code Division Multiple Access (CDMA). The second-generation networks deliver high quality and secure mobile voice and basic data services such as fax and text messaging, along with full roaming capabilities across the world. Second Generation technology is in use by more than 10% of the world's population and it is estimated that at the end of 2002 there were 787 million GSM subscribers across the 190 countries of the world, global GSM subscribers are expected to reach one billion in the near future (GSM-Information; GSM-Information, 2004).

Enhancement of Second Generation Networks (2.5G)

The later advanced technological applications are called 2.5G technologies and include networks such as General Packet Radio Service and EDGE. GPRS enabled networks provides functionality such as: "always-on", higher capacity, Internet-based content and packet-based data services enabling services such as color Internet browsing, e-mail on the move, visual communications, multimedia messages and location-based services. Another complimentary 2.5G service is Enhanced Data rates for GSM Evolution EDGE which offers similar capabilities to the GPRS network. Another 2.5G network enhancement of data services is High Speed Circuit Switched Data (HSCSD). It allows you to access non-voice services at three times faster, which means subscribers are able to send and receive data from their portable computers at a speed of up to 28.8 kbps; this is currently being upgraded in many networks to rates of and up to 43.2 kbps. The HSCSD solution enables higher rates by using multiple channels, allowing subscribers to enjoy faster rates for their Internet, e-mail, calendar and file transfer services. The GSM High Speed Data service is now available to more than 100 million customers across 27 countries around the world in Europe, Asia Pacific, South Africa, and Israel (GSM-Press-Release, 2002)

Third Generation Networks (3G)

The most promising period is the advent of 3G networks which are also referred to as the Universal Mobile Telecommunications Systems (UMTS). The global standardization effort undertaken by the ITU is called IMT-2000. The aim of the group was to evolve today's circuit switched core network to support new spectrum allocations and higher bandwidth capability. Over 85% of the world's network operators have chosen 3G as the underlying technology platform to deliver their third generation services (GSM-Information, 2004). Efforts are underway to integrate the many diverse mobile environments in addition to blurring the distinction between the fixed and mobile networks. The implemen-

tation of the third generation of mobile systems has experienced delays in the launch of the service. There are various reasons for the delayed launch, ranging from device limitations, application and network related technical problems to lack of demand. A significant factor in the delayed launch, that is frequently discussed in the telecoms (Maitland et al., 2002; Melody, 2000; Klemperer, 2002) is the extortionate fees paid for the 3G spectrum license in Europe during the auction process, the technical hitches with the devices and applications have been overcome but the financial challenges caused by the high start-up cost and the lack of a subscriber base due to the market saturation in many of the countries launching 3G. In 2002, industry experts revealed lower than expected 3G forecast. The continued economic downturn prompted renewed concerns about the near-term commercial viability of mobile data services, including 3G. The UMTS forum re-examined the worldwide market demand for 3G services due to the effect of September 11 and the global telecommunication slump, and produced an updated report. (UMTS-Forum-Report18, 2003). The re-examination highlighted the fact that due to the current negative market conditions, the short-term revenue generated by 3G services will be reduced 17% through 2004—a total reduction of $10 billion. Over the long term, however, services enabled by 3G technology still represent a substantial market opportunity of $320 billion in 2010, $233 billion of which will be generated by new 3G services (Qiu & Zhang, 2002).

The eventual full implementation of 3G worldwide will be a stepping-stone towards global mobile convergence. Standards and network technologies have already been developed to meet the challenges posed by 3G. Inter-operability issues for equipment from different manufacturers have also been addressed, leading to significant cost reductions in handsets which allow operators to offer a reasonable discounted rate (Olla & Patel, 2002). Currently efforts are underway to create and deliver viable applications and services to mobile user over a packet-switched IP network. The ultimate goal is to eliminate circuit switching and the cellular technology used in the current 2G networks. This means that the future vision and trends in mobile network evolution are directed towards an all-IP core network infrastructure. This all-IP end-to-end solution is referred to as the fourth-generation (4G) systems.

Potential Enhancement to 3.5G Networks

There is a view in the industry (Qiu & Zhang, 2002) that there is a need for a phase prior to the beginning of the 4th era in the evolution and beyond the 3G phase of today. This is being called HSDPA (High Speed Downlink Packet Access) or 3.5G (Inforcom-Reserach, 2002; Qiu & Zhang, 2002)). HSDPA promises a data rate of up to 10Mbps and higher spectrum efficiency (Qiu &

Zhang, 2002). The 4G systems are expected around 2010–2015. They will be capable of combining mobility with multimedia-rich content, high bit rate, and IP transport. In general the 4th generation technology supports broadly similar goals to the 3rd generation effort, but starts with the assumption that future networks will be entirely packet-switched using protocols evolved from those in use in today's Internet. Today's Internet telephony systems are the forerunners' of the applications that will be used in the future to deliver telephony services.

Fourth Generation (4G) Mobile Networks

The benefits of the 4th generation approach are described by Inforcom-Research (2002), Qiu & Zhang (2002) as: voice-data integration, support for mobile and fixed networking, enhanced services through the use of simple networks with intelligent terminal devices and a flexible method of payment for network connectivity that will support a large number of network operators in a highly competitive environment. Over the last decade, the Internet has been dominated by non real-time, person-to machine communications. According to UMTS report (UMTS-Forum-Report14, 2002) the current developments in progress will incorporate real-time person-to-person communications, including high quality voice and video telecommunications along with extensive use of machine-to-machine interactions to simplify and enhance the user experience. Currently the Internet is used solely to interconnect computer networks, IP-compatibility is being added to many types of devices such as set-top boxes to automotive and home electronics. The large-scale deployment of IP based networks will reduce the acquisition costs of the associated devices. The future vision is to integrate mobile voice communications and Internet technologies, bringing the control and multiplicity of Internet applications services to mobile users. The creation and deployment of IP-based multimedia services (IMS)allows person-to-person real-time services, such as voice over the 3G packet-switched domain. Described in (UMTS-Forum-Report20, 2002), IMS enables IP inter-operability for real-time services between fixed and mobile networks solving current problems of seamless converged voice/data services. Service transparency and integration are key features for accelerating end-user adoption. Two important features of IMS are: IP-based transport for both real-time and non-real-time services, and a multimedia call model based on the Session Initiation Protocol (SIP). The deployment of an IP based infrastructure will encourage the development of Voice-over-IP (VoIP) services.

The current version of the Internet Protocol (IPV4) is being upgraded due the constraints of providing new functionality for modern devices. The pool of Internet addresses are also being depleted. The new version, called IP Version 6 (IPv6), resolves IPv4 design issues and is primed to take the Internet to the next

generation. Internet Protocol Version 6 is now included as part of IP support in many products including the major computer operating systems. IPv6 has also been called "Ipng" (IP Next Generation). The most evident enhancement in IPv6 over the IPv4 is that IP addresses are being lengthened from 32 bits to 128 bits. This extension anticipates considerable future growth of the Internet from both fixed and mobile devices and provides relief for what was perceived as an impending shortage of network addresses. IPv6 also describes rules for three types of addressing: unicast (one host to one other host), anycast (one host to the nearest of multiple hosts), and multicast (one host to multiple hosts) (Microsoft, 2003).

Mobile Satellite Networks

The next potential advancement in the mobile telecommunication arena is the convergence of the next generation mobile technologies with space technologies. Incorporating space technology into mobile communications offers two main advantages. The first is ubiquitous access to voice and data services anywhere in the world. The second is accurate positioning information which is used to provide location sensitive information used for navigation and location based services.

The fast commercialization process has also brought in large-scale private investment in space technology, not only in the application market but also in the manufacturing of satellites and launch vehicles, introducing a paradigm shift in the traditional roles of government and industry, and calling for a new look into the regulatory framework, which had largely been set up with governmental actors in mind (UNESCAP-Report, 2002). Satellite operators such as Inmarsat are developing services to address the increased high-speed data needs (Imrasat-Homepage, 2002). The introduction of regional Broadband Global Area Network (BGAN) provides 144kbit/s IP connection and is seen as the first evolution towards a global broadband infrastructure which is expected to deliver voice and data services at speeds up to 432kbit/s, and currently due for launch in 2004. The technology operates via a lightweight, A4 sized portable satellite IP modem, and uses standard interfaces for ease of use. Its key features include: continuous global coverage, "always on" access to IP-based networks, including the Internet and corporate data networks, Bluetooth, USB, and Ethernet ports.

Space technology have contributed to the technological leaps the human race has made over the last three decades. Space technology is present in applications areas such as remote sensing, communications, and navigation. Space technology has touched every facet of human life helping modern society to cope with its problems of sustainable development, preservation of the environment, global connectivity, entertainment, education, tele-health services, disaster manage-

ment, and information management (UNESCAP-Report, 2002) and this is extending to mobile communication. The enabling factor vital to the success of the convergent vision is the inclusion of space technology elements, which will essentially drive the need for a seamless integration. The Satellite portion of the 3G system utilizes the S-band Mobile Satellite Service (MSS) frequency allocations set aside for satellite by the IMT2000, and provides services compatibility with the terrestrial UMTS systems. For situations where rapid and easy installation is required, satellite-based services offer greater advantages over other communication and Internet connectivity technologies, because they can bypass network congestion and provide high quality, large bandwidth connectivity. The use of hybrid broadband techniques consisting of copper wire, optical fibre and satellites are believed to provide viable alternatives to bride the digital (UNESCAP-Report, 2002).

With a proven record around the world (OECD, 2003; Space-Business-News, 2000; UNESCAP-Report, 2002; UN-Program, 2002) space technology and application activities have also become a multi-billion dollar business with enormous investments made in space systems, ground infrastructure and down-stream applications markets. In the future, there will be the possibility of the use of commercial location information due to the creation of the Galilieo network that aims to complement the American Global Positioning System (GPS). Galileo is a global navigation satellite system being developed by the European Space Agency (ESA), it will provide high accuracy with 99% availability and up to 4m vertical accuracy level (Benedicto, Dinwiddy, Gatti, Lucas & Lugert, 2000). The network will be capable of supporting applications where safety is crucial, such as running trains, guiding cars and landing aircraft and personal navigation. The fully deployed Galileo system, which aims to be operational by 2008, would consist of 30 satellites, positioned in three circular Medium Earth Orbit (MEO) planes in 23616 km altitude above the earth. One of the biggest challenges is to develop ideas and concepts that allow the creation of viable model that support the integration of space and mobile communication technologies (ESA, 2003). For the convergent vision to be truly complete, the role of various enabling technologies must be addressed along with the regulation, business models and policy issues.

Future: Vision of a Convergent Environment

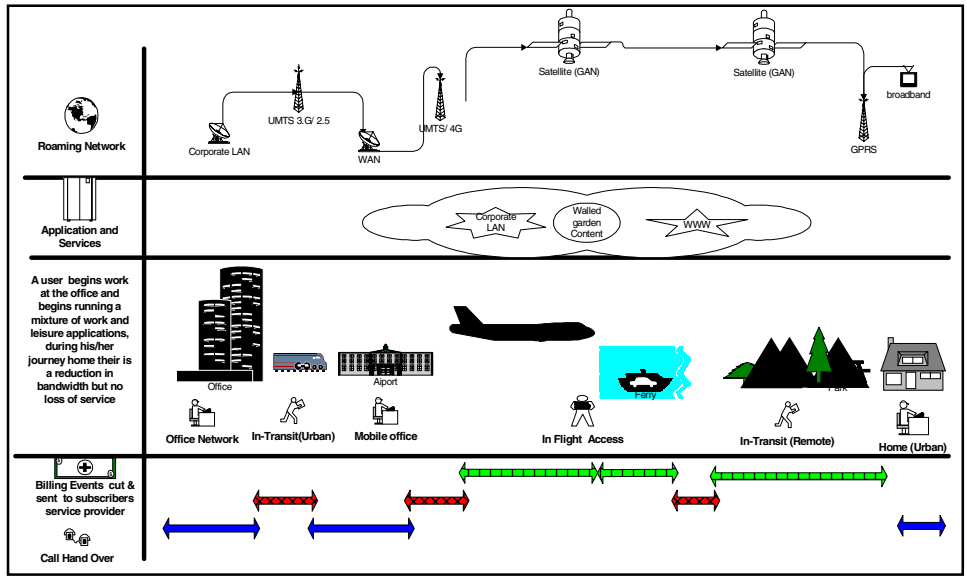
The dreams of the pioneers such as Mark Weiser (1993), Normand Klenrock (1997) that led to the development of concepts such as ubiquitous computing and

nomadic computing was of a global digital convergent environment. The vision describes a convergent environment covering both the developed and developing nations, rural and urban dwellings. To fulfil this vision today would mean developing the business models to incorporate satellite, mobile, Wireless Local Area Networks (Wi-fi or WLAN) as well as wired delivery channels, facilitated by both terrestrial and space-based systems. This vision creates both opportunities and threats to organizations and countries trying to reap the benefits. The expansive range of technological alternatives presents an opportunity to become more globally competitive if the right partners and models are chosen, and for others, it provides a means for accelerating the provision of basic services at a lower cost using appropriate technologies (UNESCAP-Report, 2002). The advent of computers and the advancement of digital technology worldwide were the prime movers of this convergence phenomenon around the world. It has resulted in an entirely new perspective on communication and technology, requiring organizational re-orientation around the world, not only cooperatively through newer alliances and mergers, but also competitively through intrusion into one another's markets (Hukill, Ono & Vallath, 2000). The notion of convergence discussed in this chapter goes beyond the typical notion of convergence from a technological perspective. The important aspect of convergence relates to the convergence of technological services, business processes but another important factor that is normally omitted from this discussion is the business model. Using the business model framework described in the previous section, this section will demonstrate how the use of business model framework can aid the visualization and definition of a viable business proposition.

The global satellite services industry is slowly transforming toward the creation of convergent network infrastructures. A convergent network combines both space and mobile and fixed-line connections to deliver a ubiquitous customer signals efficiently and economically. Broadcasting and content service providers are now using both terrestrial and satellite links to provide serves. With the new convergent networks, the emphasis is on service, not the technology through which the service is carried. In contrast, conventional satellite-only (or terrestrial-only) network operators put the delivery technology first, and then try to fit the service into their technology's specific parameters and limits (Careless, 2004). Satellite service providers such as Intelsat have reported revenue-generating result from convergent networks. The organization integrated 25 satellites with numerous terrestrial Points of Presence (PoPs). In turn, these PoPs link to local high-speed telephone loops, which deliver efficient, cost-effective global access.

Satellite operator Panamsat acquired "Sonic Telecom", an international provider of high-definition multimedia transmission services and business applications is another example of the creation of a convergent network. From this merger Panamsat now provides clients with a satellite/fibre network delivering video

Figure 5. Convergent mobile environment



content throughout the United States, Europe and Asia. Panamsat services now can transfer video traffic, connect to a global network as well as conduct videoconferencing, bridging and video content management (Careless, 2004).

Based on the notion of a convergent architecture illustrated in Figure 5, there are a variety of new business models that emerge from a convergent environment. This section will pick on two models to concisely demonstrate how the various elements of the business model framework presented in the previous section are addressed.

Model 1: Mobile Telecoms Convergence Model

UMTS/GSM Mobile communications are largely metropolitan based, satellite broadband services that extend beyond the reach of urban centres of population offering a celestial extension to terrestrial networks. Therefore, the integration of space technologies with mobile communications technologies encourages the overlapping and ubiquitous use of computer systems, satellite network infrastructure and mobile network infrastructure. The converged technology is used

to deliver meaningful digital content in any of the following formats: communications, broadcasting or information technology.

The diagram in Figure 5 illustrates the day of a working executive called Bob. During the course of the day Bob works in his office in the morning accessing files on the Internet and office Intranet and running an application that is downloading information from an Internet site. At lunch time he begins his long journey home for the weekend, but continues to download the application on the journey. There is currently the technical competence available to seamlessly integrate all forms of wireline and wireless networks and media as illustrated in Figure 5, to allow access to an array of information from anywhere in the world at any time ideally using the (3G/4G) Internet protocol infrastructure described in the previous section.

Once Bob leaves the office, his mobile device will latch on to the GSM (2.5, 3G) network and the bandwidth would depend on his travelling speed and other users on the network. On arrival to the train station the system would hand-over to "In train" or "On Platform" Mobile Internet service. An example of a railway station mobile Internet infrastructure is the system being trailed by the UK Train Operator GNER. The technical solution includes satellite downlinks, multiple cellular uplinks running in parallel, a GPS receiver to track position and an onboard management server. The service scans for the best cellular signal and creates as many GPRS and HSCSD links as necessary to deliver broadband speeds between 100-500Kbps, to end-users.

On arrival at the airport, Bob's device would automatically attach to the wireless area network he subscribes to via his service provider and this service will continue until his flight takes off. When he is in the sky, the in-flight system provides Bob with the access to office systems and the Internet. The technical complexity to deliver in-flight mobile data services is not as far-fetched as it seems. Imrasat, the global mobile satellite communications provider (Imrasat-Swift64, 2002), has already implemented the In-flight Passenger Entertainment and Communications systems (IPEC). Imrasat has announced the commercial availability of Swift64, a service which gives aircraft passengers the ability to access Internet-based applications such as e-mail, video streaming and file transfer whilst in the air at ISDN speeds of 64kbits/s. Currently up to 80% of modern long haul commercial aircraft and over 1,000 corporate jets already have the Inmarsat satellite communications antenna infrastructure needed to carry Swift64 services. The platform uses existing aircraft antennas and satellite communication avionics. At the end of the flight, Bob continues his journey home in his car via remote routes and the GSM (2.5/3G) mobile network is supplemented with coverage via satellite networks. Once Bob arrives at home, access to the Internet will be provided by his fixed line broadband operator. If Bob lived in a remote area then the broadband access could be provided by the a satellite service provider as illustrated in business model presented in the next section.

For UMTS to live up to its name and achieve a true "universal" status, instead of an international presence, which is the best, that can be hoped for on the current route taken by mobile network operators, more consideration and effort is required to instigate a program to integrate UMTS with space technologies such as satellite communication capabilities and global positioning techniques. Universal mobile satellite systems are touted as the ultimate solution to the problem of covering large areas economically, and serving widely scattered or remote rural customers in both developing and developed countries (Muratore, 2001). For this to happen from a technical basis the satellite component (UTRA) must be integrated with the UMTS to create a more integrated and advanced Mobile Broadband System (MBS) capable of 2mbs. UMTS was designed such that it could be easily integrated into existing 2.5G, 3G and other GSM networks. The mobile broadband communications systems must be capable of the different mobility requirements ranging from stationary (for wireless local loops) to quasi-stationary (outdoors, office, and industrial environments). From a commercial perspective there needs to be an introduction of innovative business models, which support revenue share partnerships and joint ventures. The scenario created by Bobs journey home will lead to the development of new business models as illustrated in Table 1. There are at least four new business models that can be identified in the diagram such as:

- *Transit Corporate Business Model:* providing access to mobile office applications, downloading files, and video conferencing functionality while the user is traveling in the air or on the sea.
- *Transit Leisure Business Model:* streaming video, online games, news and, multimedia content for the traveler.
- *Satellite Broadband:* delivering broadband services to remote users using a combination of satellite and GSM technology. This model is described further in Table 2.
- *Data & Voice Convergent Model:* allowing voice and data services to be handed over to appropriate networks such as GSM, GPRS, and satellite depending on the task being performed, network availability, agreements between operators and the associated costs to the user. This model is described in Table 1.

For the realization of the convergent model described in Table 1, there are some fundamental challenges that need to be addressed. There will be a need for new interfaces between organizations such as content providers and service providers to exchange relevant charging information. This is not completely new as operator and service providers currently exchange billing data. However, there

Table 1. Mobile convergent business model

Initiative Definition:
This initiative brings together various types of mobile networks to provide true ubiquitous global roaming capabilities to mobile users.

Actors	Competencies and Capabilities	Complexity
Satellite Network Operator	Handover capability Roaming capabilities with UMTS networks Inter operator billing capabilities	Medium technical complexity High complexity will involve network changes
Mobile Network Operator	Handover capability Roaming capabilities with Satellite networks	New roaming billing model required
Device Manufacturers	Provide data devices that are capable of roaming on mobile and satellite networks	Device capabilities similar to Tri band devices but with capabilities to latch on to satellite networks
Content Providers	Provide user content, this could be multimedia such as live streaming, Internet portals or access to other users.	The complexity will vary depending on the type of content being delivered
Application Developers	Provide innovative applications such as location sensing applications, financial,	The complexity will vary depending on the type of service
Network Infrastructure Vendors	Network operators infrastructure vendors develop interface and gateways to the Internet and other appropriate networks e.g Nortel,	The complexity will vary depending on the type of network and roaming agreements
Service Provider	They also must work closely with content and application providers in order to differentiate their offering.	The complexity will involve developing appropriate interfaces and managing the customer billing and care activities

Layers

Device	Wireless	Program	Payment	Application
Triband Mobile handsets, or data cards	Wi-Fi GSM GPRS UMTS Satellite	Network centric hand-off capabilities Security Disconnection control	Subscription to Service + Usage dependant on which network + Event charge for a particular service	Location based services Mobile Internet facilities Mobile Office Multimedia applications

Financial Element
Revenue will be generated from the providing complimentary services to mobile subscriber user. The revenue will be collected from the customer by the mobile operator / service provider.

Revenue is earned from user subscriptions and traffic agreements with other ISPs and operators. Currently the Telci centric model divides propositions at a high level into pre-paid and post-paid services. This model will need to change, allowing service transactions to be managed in real-time or near real-time, in order to control expenditure and eliminate credit risk especially due to potential of ubiquitous roaming. Billing and collections will ultimately become a single role within the business environment, handling the account balance for end customers regardless of whether the account happens to be prepaid or post paid, there are considerable challenges to be defined around the settlements and interconnection charging.

Customer Satisfaction

- The minimal loss of service when hand-off occurs.
- Being informed of the charging differential when moving from wif-> mobile –> satellite if it likely to vary significantly, with the option to suspend the session
- No loss of data during the network change over.
- Clear and concise charging rules as opposed to the current bytes model.

will be the need for new charging protocol for the availability and exchange of real-time (rather then batch oriented) charging and authorization information. Charging information has to be available from the network elements or from the application servers through to the billing system. There are currently applications that perform this task but there are no mature interface standards. Also the settlements and retail charging facility will face a range of changes. Work is currently in-progress by the UMTS forum produce a dedicated report to highlight the settlements' related issues (UMTS-Forum-Report21, 2002). Due to the potential services likely to emerge, Quality of service (QOS) must meet customers' expectations or services will fail to be taken up. Customer billing should be performed incorporating the QoS measurements achieved against what was guaranteed. The current implementations of 3G do not take into account QOS billing features.

Another key factor to the success of the mobile convergent model is the formation of strong alliances between organizations. An example of an alliance that would benefit from partnership with satellite operators is the "Starmap Mobile Alliance", recently launched in February 2004. The Starmap mobile alliance currently has nine members: Amena (Spain), O2 (Germany, the UK and Ireland), One (Austria), Pannon GSM (Hungary), sunrise (Switzerland), Telenor Mobil (Norway) and Wind (Italy), covering a subscriber base of more than 41 million. The starmap management board comprises representatives from each operator. The Starmap mobile alliance cooperates to provide an environment for innovative and easy-to-use services offering a "home-away-from-home" experience for subscribers, encompassing the convenience and quality of service to which customers are accustomed to at home. The group have both technical and commercial agreements between the operators, customers benefit from GPRS and Mobile Media Messaging (MMS) roaming, as well as access to familiar services such as voice-mail and short-codes whilst travelling in other alliance countries. Alliance members are cooperating on the development of 3G handsets, and a common distribution agreement has been established providing availability of a standard PDA (Xda II Pocket PC) across alliance networks. The aim of the alliance is to provide seamless mobile voice and data services across the alliance footprint. The alliance worldwide footprint could be significantly increased with an agreement with a mobile satellite operators who would join the alliance to provide ubiquitous accesses to services when roaming in rural areas and travelling (both sea and air).

Model 2: Mobile Satellite Broadband

One of the most promising mobile innovative applications that use satellite technology to deliver mobile services is Mobile Broadband Service (MBS)

provided by "Connexion by Boeing". The service provides real-time, high-speed Internet access to air travellers while in flight. The planes are modified and equipped with either an Ethernet Local Area Network (LAN) connection or a wireless 802.11b network. The service provides travellers with high-speed Internet access allowing users to check e-mail, retrieve information securely from corporate Intranets or browse the Internet. The service will be launched commercially in March 2004 with Lufthansa airline. Using the same satellite and ground-based network, provides the same revolutionary capabilities for robust, high-speed connectivity to the maritime industry. Boats of various sizes can be equipped broadband speeds in excess of 1 Mbps. Full coverage to all the world's major shipping lane is expected by 2006.

There are signs that the voice Average Revenue Per User (ARPU) could continues to fall for most mobile operators, but there is evidence that non-voice services can help to reverse the ARPU decline (Olla & Patel, 2002). Some UK operators are already reporting non-voice revenues contributing up to 20% of

Table 2. Mobile satellite broadband

Initiative Definition: The initiative involves using satellite technologies to provide Internet services at varying speeds (1.44,GPRS and broadband to remote locations, which have difficulties gaining access to traditional services.				
Satellite Network Operator	Piggyback broadband capabilities on TV broadcast			
Mobile Network Operator	Use the network for uploading			
Device Manufacturers	Provide devices that can receive the signal at a reasonable cost			
Service Provider	Provide access to the Internet and value added services			
Application Developers	Provide innovative applications			
Layers				
Device	Wireless	Program	Payment	Application
Receiver	Kaband Satellite network	Internet Access Voice over IP Security Protocols	Initial one-off Equipment Cost & Installation. + Monthly Service charge + Tariff Time of Day based unlimited usage. Usage banding based on Data transferred. Unlimited usage	Content to include Value added services: Entertainment, news, weather.
Financial Element Revenue will be generated from providing access to Internet and other additional telecommunication services to users in remote areas. The revenue will be collected by the satellite service provider and distributed to other partners such as the device manufacture's and satellite network operator.				
Customer Satisfaction Providing a service where normal providers may struggle can be considered to be reasonable customer satisfaction. However, the key to the proposition involves increasing the subscriber figures to a reasonable level, to keep the costs comparable to what urban dwellers pay for equivalent services.				

their total mobile service revenue; supported by 2.5G networks and by a new generation of advanced mobile devices (Raja, 2004). In the UK more than 111 million person-to-person SMS messages were sent over UK networks on New Year's Day (January 1), according to figures released by the Mobile Data Association (MDA) nearly twice the 2003 daily average and an 8% increase on the previous year. Meanwhile, Wireless Application protocol (WAP) page impressions reached an all-time monthly high in November, averaging 31 million per day, compared with 12 million for the same period in 2002 (Gallagher, 2004).

Another example of an MBS application is the satellite-based digital multimedia broadcasting (DMB) system. The system uses satellite networks to beam down broadcasts to hand-held devices, and is due for launch during 2004 in South Korea. This has the potential to be a promising new revenue source for telecom firms. With global telecoms firms facing market saturation in core voice services, this approach is likely to receive keen interest worldwide. The South Korea's $16 billion telecom services market is seen as an ideal test-bed for the technology. Seventy percent of its 48 million people use mobile phones for voice as well as data services and it also has the world's highest broadband penetration rate (Reuters-News, 2004).

Conclusion

This chapter presented a vision of the future of mobile communication, which is a convergence of mobile and satellite technologies to create a truly universal seamless network. The environment creates new business opportunities which can be realized by using a appropriate modeling techniques to identify viable propositions.

A concerted effort is required in the research and development area to achieve the convergence of telecommunication equipment integration, incorporating local networks (WIFI,Bluetoothe), satellite networks, GSM 2.5, 3G and 4G networks. It would be highly unrealistic to assume that every type of communication technology can be integrated but as long as each scenario such as urban stationary, urban motion, maritime, air, and rural, is covered then a global communication infrastructure vision becomes achievable.

For UMTS and 4G to live up to its name and achieve a true "universal" status, instead of an international presence, which is the best, that can be hoped for on the current route taken by mobile network operators, more consideration and effort is required to instigate a program to integrate UMTS with space technologies such as satellite communication capabilities and global positioning techniques. Further work is required to define and develop common service stan-

dards with consistent transmission parameters and a radio interface between satellite and terrestrial implementations, along with the billing capabilities. Mobility rules will be required by the participating network operators to decide the rules for call handovers. This should cover in-call handoff between cells of different network types including land-based and satellite networks. The mobility rules should be tightly linked to the business models.

This chapter has described a structured approach to business model development in the mobile-satellite communication sector which may be used to aid policy makers and network operators develop innovative business models.

References

Afuah, A. & Tucci, C. (2001). *Internet business models and strategies*. Boston: McGraw Hill.

Amit, R. & Zott, C. (2001). Value Creation in eBusiness. *Strategic Management Journal, 22*, 493-520.

Barnes, S. J. (2002). The mobile commerce value chain: Analysis and future developments. *International Journal of Information Management, 22*(2), 91-108.

Benedicto, J., Dinwiddy, S.E., Gatti, G., Lucas, R. & Lugert, M. (2000). *GALILEO: Satellite System Design and Technology Developments*: European Space Agency.

Bouwman, H., & Ham, E.V.D. (2003). *Designing metrics for business models describing Mobile services delivered by networked organisations.* Paper presented at the 16th Bled Electronic Commerce Conference eTransformation Workshop on concepts, metrics & visualisation, Bled, Slovenia.

Burns, T. & Stalker, G.M. (1961). *The management of innovation*. London: Tavistock.

Camponovo, G. & Pigneur, Y. (2002a). *Analyzing the Actor Game in m-Business.* Paper presented at the Proc. First International Conference on Mobile Business, Athens (2002).

Camponovo, G. & Pigneur, Y. (2002b). *Business Model Analysis Applied to Mobile Business.* Paper presented at the International Conference on Enterprise Information Systems (ICEIS), Anger 2003.

Careless, J. (2004). Hybrid Networks: A Winning Partnership For Satellite. *Via Satellite*.

D'Aveni, R. A. (1994). *Hypercompetition: Managing the dynamics of strategic maneuvering*. New York: Free Press.

Eisenhardt, K. M. (1989). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal, 32*(3), 543–576.

Eisenhardt, K. M. & Brown, S. L. (1998). Time pacing: competing in markets that won't stand still. *Harvard Business Review, 76*(2), 59–69.

ESA. (2003). *European Space Agency*. Retrieved from the World Wide Web: *http://www.esa.int/export/esaSA/GGGMN850NDC_navigation_0.html.*

ESA-Homepage. (2003). *European Space Agency. http://www.esa.int/export/esaSA/GGGMN850NDC_navigation_0.html*.

Gallagher, R. (2004). Business Intelligence for Mobile Industry Executives. *Mobile Communications,* (369).

GSM-Information. (2004). *http://www.gsmworld.com/index.shtml*.

GSM-Press-Release. (2002). High-speed data communication now available to over 100 million GSM users in 27 countries worldwide.

Hukill, M., Ono, R. & Vallath, C. (2000). *Electronic communication convergence:Policy challenges in asia*. (Eds.) London.

Imrasat-Homepage. (2002). *http://www.inmarsat.com*.

Imrasat-Swift64. (2002). Inmarsat Announces Availability of the 64KBIT/S Mobile Office in the Sky. *http://www.inmarsat.com/swift64/press_1.htm*.

Inforcom-Reserach. (2002). The Dawn of 3.5 and 4G next Generation Systems. *Gateway to N+1 Generation Networks, 1*(4), *http://www.icr.co.jp/nG/src/0104_contents.pdf*.

Keil, T. & Vilkamo, T. (2003). Strategic technology partnering in high-velocity environments lessons from a case study. *Technovation 23, 23,* 193–204.

Kleinrock, L. (1997). Nomadic computing (keynote address). *Telecommunication Systems, 7*(1-3), 5-15.

Klemperer, P. (2002). How (not) to run auctions: The European 3G telecom auctions. *European Economic Review, 46*(4-5), 829-845.

Li, F. & Whalley, J. (2002). Deconstruction of the telecommunications industry: from value chains to value networks. *Telecommunications Policy, 26*(9-10), 451-472.

Maitland, C. F., Bauer, J. M. & Westerveld, R. (2002). The European market for mobile data. *Telecommunications Policy, 26*(9-10), 485-504.

Melody, W. H. (2000). Telecom development. *Telecommunications Policy, 24*(8-9), 635-638.

Microsoft. (2003). *IPV6 Technologies*. Retrieved, from the World Wide Web: *http://www.microsoft.com/windowsserver2003/technologies/ipv6/introipv6.mspx*.

Muratore, F. (2001). *UMTS Mobile Comunication of the Future*. Chicester: Wiley.

OECD. (2003). *Organisation for Economic co-operation and Development (OECD).* Unpublished manuscript, International Futures program.

OFTEL. (2003). *Director General's statement on the Competition Commission's report on mobile termination charges*. London: OFTEL.

Olla, P. & Patel, N. (2003). Framework for Delivering Secure Mobile Location Information 1, No3 Page 289-300, 2003. *International Journal of Mobile Communications, 1*(3), 289-300.

Olla, P. & Patel, N. V. (2002). A Value Chain Model for Mobile Data Service Providers. *Telecommunications Policy, 26*(9-10), 551-571.

Orlikowski, W. J. & Robey, D. (1991). Information Technology and the Structuring of Organizations. *Information Systems Research, 2*, 143-169.

Osterwalder, A., Lagha, S.B. & Pigneur, Y. (2003). An Ontology for Developing e-Business Models. *http://inforge.unil.ch/aosterwa/*.

Pigneur, Y. (2000). *An Ontology for m-Business Models*: University of Lausanne, Ecole des HEC, CH-1015 Lausanne.

Pigneur, Y. (2002) *An ontology for m-business models*, in S. Spaccapietra et al. (Eds.) Conceptual Modeling - ER 2002, Tampere, Lecture Notes in Computer Science, 2503, October 2002. *http://inforge.unil.ch/yp/Pub/02-ER.pdf*.

Prahalad, C. & Hamel, G. (1990). The Core Competence of the Corporation. *Harvard Business Review May-June 1990*.

Prahalad, C. & Hamel, G. (1994). Competing for the Future. Harvard Business School Press, 1994.

Qiu, R. C. W. Z. & Zhang, Y.Q. (2002). Third-Generation and Beyond (3.5G) Wireless Networks and Its Applications,. *IEEE International Symposium on Circuits and Systems (ISCS), Scottsdale, Arizona, May 26-29, 2002*.

Raja, S. (2004). Mobile Communications management reports. *Informa Telecoms Group*.

Rappa, M. (2000). Managing the digital enterprise - Business models on the Web, 2000. *http://ecommerce.ncsu.edu/business_models.html [Accessed on March 22nd, 2002]*.

Reuters-News. (2004). S.Korea satellite project may stir up telecom sector. March 12, 2004.

Sabat, H. K. (2002). The evolving mobile wireless value chain and market structure. *Telecommunications Policy, 26*(9-10), 505-535.

Sanchez, R., Heene, A. & Thomas, H. (1996). Dynamics of Competence-Based Competition: Theory and Practice in the New Strategic Management. *Oxford, England, Elsevier Pergamon, 1-36.*

Seppänen, Veikko, Matti Kurki, & Kimmo Alajoutsijarvi. (1993). Competance-based Evolution of R&D Relationships. WG8.2 &WG8.6 Joint Working Conference on Information Systems: Current Issues and Future Changes. Helsinki, Findland, December 10-13. *http://is.Ise.ac.uk/helsinki/seppanen.pdf*

Space-Business-News. (2000). *http://www.space.com/businesstechnology/business/iridium_chapt11.html.*

Spekman, R. E., Forbes III, T. M., Isabella, L. A., & MacAvoy, T. C. (1998). Alliance management: A view from the past and a look to the future. *Journal of Management Studies, 35*(6), 747–772.

Talluri, S., Baker, R. C. & Sarkis, J. (1999). A framework for designing efficient value chain networks. *International Journal of Production Economics, 62*(1-2), 133-144.

Thomas, L. G. (1996). The two faces of competition: dynamic resourcefulness and the hypercompetitive shift. *Organization Science, 7*(3), 221–242.

Tsalgatidou, A. & Pitoura, E. (2001). Business models and transactions in mobile electronic commerce: requirements and properties. *Computer Networks,37*(2), 221-236.

UMTS-Forum-Report14. (2002). *Support of Third Generation Services using UMTS in a Converging Network Environment. http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/Resources_Reports_index: UMTS.*

UMTS-Forum-Report18. (2003). *The UMTS 3G Market Forecasts - Post September 11, 2001. http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/Resources_Reports_index: UMTS.*

UMTS-Forum-Report20. (2002). *IMS Service Vision for 3G Markets. http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/Resources_Reports_index: UMTS.*

UMTS-Forum-Report21. (2002). *Charging, Billing and Payment Views on 3G Business Models. http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/Resources_Reports_index: UMTS.*

UNESCAP-Report. (2002). *Towards a policy framework for integrating space technology applications for sustainable development on the information superhighway*.Unpublished manuscript.

UN-Program. (2002). *Environment and Natural Resources Management: Space Technology Applications Section*.Unpublished manuscript.

Weill, P. & Vitale, M. R. (2001). *Place to Space. Migrating to e-business Models*. Boston.

Weiser, M. (1993). Ubiquitous Computing. *Computer, 26*(10), 71-72.

Wirtz, B. W. (2001). Reconfiguration of Value Chains in Converging Media and Communications Markets. *Long Range Planning, 34*(4), 489-506.

Chapter V

Ubiquitous Commerce: Beyond Wireless Commerce

Holtjona Galanxhi-Janaqi, University of Nebraska – Lincoln, USA

Fiona Fui-Hoon Nah, University of Nebraska – Lincoln, USA

Abstract

Ubiquitous commerce, also referred to as "u-commerce" or "über-commerce", is the combination of electronic, wireless/mobile, television, voice, and silent commerce. It extends traditional commerce (geographic, electronic, and mobile) to a world of ubiquitous networks and universal devices. This chapter introduces the basic ideas and characteristics underlying the concept of u-commerce. It discusses market drivers and applications of u-commerce as well as the underlying technology of u-commerce. It highlights the benefits and challenges of u-commerce and provides specific research directions for future research.

Introduction

Ubiquitous commerce, also referred to as "u-commerce" or "über-commerce", extends the traditional commerce (geographic, electronic and mobile) to a world of ubiquitous networks and universal devices (Junglas and Watson, 2003b). It is a new paradigm that broadens and extends the Internet era. It has the potential to create a completely new environment in business. In the era of u-commerce, it will be possible to execute interactions and transactions anywhere and at any time without being constrained to stay connected to power and telephone lines.

The purpose of this chapter is to introduce the basic ideas underlying the concept of u-commerce. U-commerce applications offer many benefits, but there are also challenges of business, technological, and social nature. The last part of the chapter will discuss implications of u-commerce and how this new vision can be successfully achieved and managed.

Characteristics of Ubiquitous Commerce

U-commerce emerges as a continuous, seamless stream of communication, content and services exchanged among businesses, suppliers, employees, customers and products (Watson, Pitt, Berthon & Zinkhan, 2002). In this way, through the convergence of the physical and the digital means, higher levels of convenience and value can be created.

To begin with, u-commerce will be *ubiquitous*. Ubiquity builds upon the ideas of accessibility and reachability (Junglas & Watson, 2003b). Therefore, computers will be everywhere and every device will be connected to the Internet. The ubiquity, or omnipresence, of computer chips means that they are not only everywhere, but also in a sense "nowhere", for they become invisible, as we no longer will notice them (Watson et al., 2002).

U-commerce will also add *universality*. Universality aggregates the aspect of network and devices into one logical construct (Junglas & Watson, 2003b). It will eliminate the problems of incompatibility caused by the lack of standardization, like the use of mobile phones in different networks. A universal device will make it possible to stay connected at any place and any time. Current devices are limited in their usefulness because they are not universally useable. Laptops and PDAs will also gain universality and be constantly connected to the Internet via wireless network or satellite, wherever the owner is (Watson et al., 2002). One could say that the Internet has become universal, since one can be almost anywhere and be connected.

U-commerce will add *uniqueness* of information. Uniqueness builds upon the ideas of identification and localization (Junglas & Watson, 2003b). This means that the information provided to users will be easily customized to their current context and particular needs at certain time and place. Mass customization of information is already available, and the next IT generation will add contextual customization (Watson et al., 2002). Therefore, for the same individual, there will be customization depending on such variables as place, time, preference and context.

Finally, *unison* aggregates the aspects of application and data in one construct (Junglas & Watson, 2003b). In a u-commerce environment, it will be possible to integrate various communication systems so there is a single interface or connection point (Watson et al., 2002). This means that there will always be consistency and matching in the data regardless of the device, network, or place from which the data is extracted or updated. This would also mean that when a change is made in one record, this change is reflected instantaneously on all devices containing the record.

Background

Components of Ubiquitous Commerce

The following subsections provide a description of each type of commerce that makes up u-commerce. U-commerce is a new environment that combines wireless, voice, television, and silent commerce with traditional e-commerce (see Figure 1).

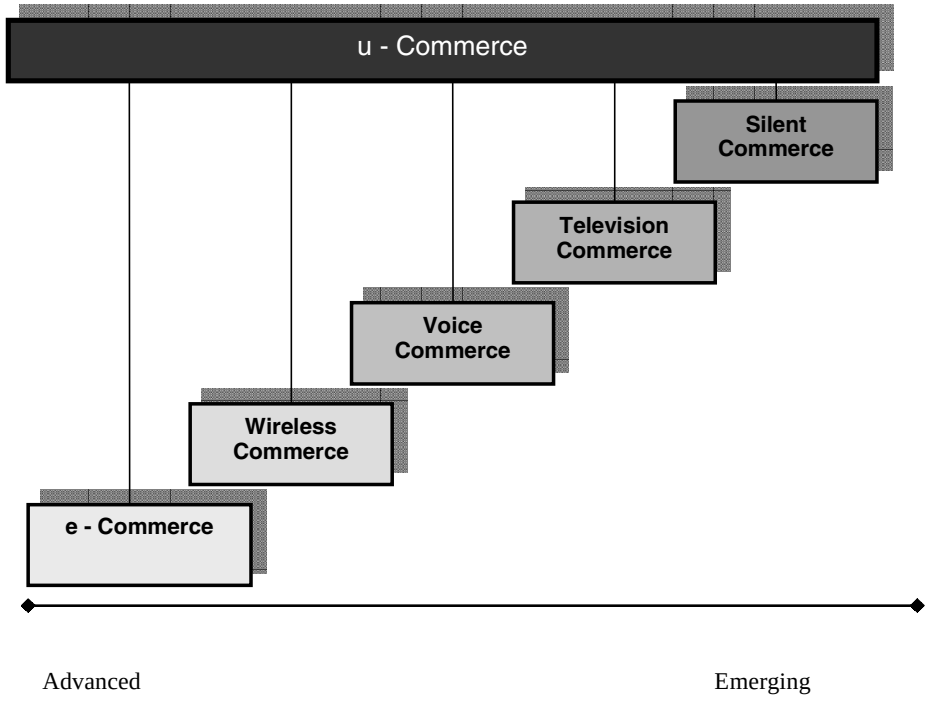
Electronic Commerce

Electronic commerce (e-commerce) is the use of the Internet and the Web to transact business. There are three main types of e-commerce: business-to-consumer, business-to-business, and consumer-to-consumer. In addition, government-to-government, government-to-consumer, and consumer-to-government have emerged.

Laudon and Traver (2002) identify seven characteristics of e-commerce:

- **Ubiquity:** Internet/Web technology is available everywhere—at work, at home, and elsewhere.

Figure 1. U-commerce (adapted from Accenture, 2001)



- **Global reach:** The technology reaches across national boundaries, around the earth.
- **Universal standards:** There is one set of technology standards, namely Internet standards.
- **Richness:** Video, audio, and text messages are possible.
- **Interactivity:** The technology works by interacting with the users.
- **Information density:** The technology reduces information costs and raises quality.
- **Personalization/Customization:** The technology allows personalized messages to be delivered to individuals as well as groups.

E-commerce is the most established type of commerce performed through digital means. Nonetheless, there are problems and the types of problems differ from sector to sector (Duffy & Dale, 2002). Companies involved in e-commerce still face obstacles such as choice of business model, security, and trust issues,

integration with legacy systems, interoperability of systems with other organizations' systems, assessment of the effectiveness of their e-commerce investments, and management of information overload. Furthermore, a comprehensive and unambiguous legal framework regarding online transactions is missing.

Wireless Commerce

The additional characteristics of m-commerce—*reachability* (a person can be in touch and be reached at any time or place), *accessibility* (the user can access the device and the network from any location and at anytime), *localization* (the geographical position of the user can be localized), *identification* (the device is usually personal and can be more easily associated with a specific user),and *portability* (the devices can be carried by the user virtually everywhere)—make wireless commerce distinct from e-commerce. Among these characteristics, portability has a unique standing, because it makes the other four characteristics unique to wireless commerce (Junglas & Watson, 2003a).

Wireless commerce is a key component of u-commerce as it creates the possibility for communications between people, businesses and objects to happen anywhere and at any time. Mobile and wireless devices are enabling organizations to conduct business in more efficient and effective ways (Nah, Siau & Sheng, 2004). Wireless devices can offer many advantages for companies and individuals such as: empowering the sales force, coordinating remote employees, giving workers mobility, improving customer service, and capturing new markets. For example, Wells Fargo, a major financial institution in the United States, uses wireless technology to offer banking services to its customers and to provide access to corporate applications to its employees.

Although most of today's wireless applications in organizations are used to support basic tasks such as e-mail, messaging, calendar, and contact management, there are two main trends toward greater sophistication in their applications:

- integration with outside users such as customers, vendors, and partners; and
- carrying out business functions and transactions with wireless connections, such as supply chain management, sales force automation, work force automation, and totally customized banking applications (IC2 Institute, 2004).

Nevertheless, there are challenges to the full realization of wireless commerce (Siau, Nah & Sheng, 2003a), and consequently ubiquitous commerce, such as:

different standards in different countries that do not allow the global use of devices, limited computational power, slow data transmission rate, difficult navigation, 2G telephony being optimized for voice rather than data, and problems faced by the 3G technology.

Designing user-friendly interfaces is very important. On one hand, portability calls for smaller devices; on the other hand, easy-to-navigate interfaces are an equivalent necessity and hence, larger devices are more appropriate. Wireless service quality issues such as limitations in bandwidth, instability of connections, low predictability and the lack of a standardized protocol (Nah et al., 2004) remain central. Therefore, technology-related factors remain particularly important for wireless commerce. Trust is another major obstacle in its adoption and development (Siau & Shen, 2003a; Siau, Sheng & Nah, 2003b; Siau, Sheng, Nah & Davis, 2004). Trust in wireless commerce shares some characteristics with trust in e-commerce, but at the same time it has some new characteristics of its own that are closely related with technological issues of mobile devices and networks (Siau et al., 2003b).

The enthusiasm on the wireless industry must be tempered by taking into account the challenges (IC2 Institute, 2004). For companies that wish to be involved in wireless commerce, the challenges include the potential need to change their business strategies, investment risk, computer network infrastructure, technological and usability limitations of devices, and security and trust issues (Siau & Shen, 2003b).

Voice Commerce

An increasing number of businesses are using computerized voice technologies: speech recognition, voice identification and text-to-speech. Voice commerce enables businesses to reduce call-center operating costs and at the same time, to improve the customer service. Voice commerce can also be used to generate new sources of revenue, but this will probably take longer to materialize. Companies are mostly pursuing voice commerce as part of a multi-channel strategy.

The challenges to voice commerce include (Accenture, 2001):

1. It is best suited for transactions that are simple, standard and frequent, such as simple account enquires, requests for information, placing orders and account payments.
2. Customer resistance may be high in cases where the services cannot recognize speech accurately enough (basic voice-recognition systems can achieve accuracy rates of up to 97 percent.)

3. The US is the global leader in the use of voice commerce. It is a big country with a large population sharing a common language. But, for other countries, language differences are another obstacle that needs to be overcome.

Television Commerce

The spread of interactive digital television will provide a platform for two-way personalized communication in the center of most homes. This will make television commerce a big opportunity for business and a critical component of u-commerce. Television commerce is mainly used as an end-consumer channel. Since it can reach a big range of the population, governments may also use it to deliver their services. Digital television is also a suitable method to deliver innovative services. The interactive TV (TiVo) integrates software and set-top boxes to facilitate digital television with many capabilities, including 'time-shifting' content and filtering advertisements.

Compared to wireless and voice commerce, successful television commerce requires companies to build up partnerships with broadcasters, which is expensive. The increasing number of businesses involved may decrease the costs in the future. Another concern that businesses have is that television commerce may only divert sales from their existing sales channels. Also, infrastructural factors may influence the adoption rate of television commerce technology.

Silent Commerce

Silent commerce refers to the business opportunities created by making everyday objects intelligent and interactive. Radio frequency identification (RFID) chips allow the tagging, tracking and monitoring of objects along an organization's supply chain. By granting objects the gift of reasoning and communication, silent commerce can increase the efficiency of supply chains, enhance the health and safety aspects of many processes, and offer new opportunities for revenue generation (Accenture, 2001). An important advantage of RFID as compared to technologies like barcodes is its ability to identify and track individual assets while barcodes identify classes of assets.

Micro-electro-mechanical systems (MEMS) chips combine the capabilities of an RFID tag with small, embedded mechanical devices such as sensors. Nowadays, researchers are even talking about *nano*electromechanical systems (NEMS) or structures, which have dimensions below a micron.

Most of today's silent commerce applications are simple solutions that deal with an isolated problem within a business. Closed solutions provide considerably

more advantages since they can capture information at a number of points in a supply chain of a single business, but there is also the need to link the new data and devices with legacy applications. Open solutions offer the option of operating across multiple businesses. Silent commerce applications can improve productivity and service: material management, inventory tracking, supply-chain management, theft prevention, asset management, production management, vehicle management, employee safety, access control, micro payment, customer convenience, and customer service.

Over the years, RFID chips have been used in everything from antitheft devices in clothing stores to Exxon Mobil Corp.'s Speedpass, which allows customers to pay for their gas wirelessly (*http://www.sltrib.com*). In a 2001 test of RFID technology, the San Francisco-based Gap Inc. equipped some of its stores with "smart shelves" containing RFID readers. The system used built-in readers to instantly monitor the inventory on the smart shelves, gathering information on each garment through layers in a stack, a task that would be impossible with a bar-code scanner. Industries that rely on physical assets and extensive supply chains are more likely to adopt silent commerce applications early on. The gigantic Wal-Mart has declared that they expect their top 100 suppliers to incorporate the RFID technology in their products by the end of 2004 and the rest of its suppliers to do so in 2005 (*http://www.emergic.org/archives/indi/005657.php*).

In addition, telematics is another emerging technology. Telematics is the ability to wirelessly provide information to or extract information from vehicles and industrial equipment such as generators, pumps, and heating and cooling systems (Riaz, Osman, Gorra-Stockman & Shoup, 2002). They can be business-to-consumer, business-to-product and business-to-business applications. These applications have received the most attention in the automotive industry. IBM, for example, is collaborating with leading automotive companies and helping the Automotive Multimedia Interactive Collaboration (AMI-C) standards body to offer automakers a comprehensive choice of technology and implementation services to complete the value chain (*http://www-306.ibm.com*). Complete telematics solutions would include customer service, support, billing, technology infrastructure, application integration, and data mining and management. This usually requires revamping back-end infrastructure and establishing alliances with technology and service providers (Riaz et al., 2002).

Silent commerce applications can improve productivity and service: material management, inventory tracking, supply-chain management, theft prevention, asset management, production management, vehicle management, employee safety, access control, micro payment, customer convenience, and customer service. With more advanced silent commerce applications, it will be possible for organizations to identify, track, and monitor every single product along the entire

supply chain and even after the sale, up to the point when the product is recycled. These more complex solutions could completely transform the businesses of tomorrow and create a stream of information and value.

However, some of the requirements for a wider application of silent commerce are: greater collaboration between supply-chain participants and a common strategy to determine how the costs of the technology will be shared; integration of legacy systems with the new technology of silent commerce; and strategies for managing the massive amounts of data that will be generated. Although the move to more complex silent commerce solutions will take time, companies that are already testing or using the technology are likely to gain first-mover advantage.

The New Concept of Ubiquitous Commerce and the Drivers for Its Growth

So far, we have covered the different components of u-commerce and issues related to them. It is also important to emphasize that u-commerce can bring value that is greater than the simple sum of its individual components. Ubiquitous commerce can be defined as: "The use of ubiquitous networks to support personalized and uninterrupted communications and transactions between a firm and its various stakeholders to provide a level of value, above and beyond traditional commerce" (Watson et al., 2002).

Schapp and Cornelius (2001) identify three global phenomena that will accelerate the growth of u-commerce:

Pervasiveness of Technology

The past has clearly shown that, if properly applied, technology drives efficiency, productivity, and value. The explosive growth of nanotechnology and continuing capital investments in technology at the enterprise level expand the platform on which to leverage innovations and new applications by making the technology more pervasive.

Ubiquitous computing will make it possible to integrate data that is directly linked to and sometimes not even distinguishable from the physical world with data in the virtual world. The term "ubiquitous computing" signifies the omnipresence of tiny, wirelessly interconnected computers that are embedded almost invisibly into just about any kind of everyday object (Mattern, 2001).

One of the main barriers for ubiquitous computing to become fully omnipresent is size. Today's devices are much too large to be embedded into small items.

Another barrier is power—for a device to send a signal, it needs power supply. The existing power devices are not ideal for ubiquitous computing devices. Bluetooth and micro-electro-mechanical systems (MEMS) technology can help to overcome these barriers in the future. Bluetooth is a low-power wireless network standard that allows computer, peripherals, and consumer electronic devices to talk to each other at distances of up to 30 feet. MEMS chips combine the capabilities of an RFID tag with small, embedded mechanical devices such as sensors.

In addition, software agents, which are autonomous software components that perform tasks on behalf of a user or other agent, can offer a broad range of applications and functions (Du, Li & Chang, 2003; Lange & Oshima, 1999) such as news-filtering agents, shopping agents, supply chain scheduling agents, intelligent travel assistants, to learning assistants, and many others. Agent-based technologies could play a critical role in this regard with the potential of eventually delivering unprecedented levels of autonomy, customization and general sophistication in the way electronic commerce is conducted (Sierra, Wooldridge, Sadeh, Conte, Klusch & Treur, 2000).

Growth of Wireless Technology

Wireless is one of the fastest-growing distributed bases: wireless networks have expanded around the globe; mobile phone usage and new applications have exploded. Wireless commerce is therefore a critical component of u-commerce. It is of vital significance to solve the issues that relate to it in order to capture the full advantage that u-commerce can offer.

Table 1 provides an overview of the current state of different generations of cellular voice and data services.

Table 1. Current state of different generations of cellular voice and data services (adapted from IC2 Institute 2004)

Generation	Transmission technology	Current location
1G	AMPS (Advanced Mobile Phone Service)	US, but declining usage in metro areas
2G	CDMA (Code Division Multiple Access)	Mostly metro areas
	TDMA (Time Division Multiple Access)	Being phased out
	GSM (Global System for Mobile Communications)	Most of the world except US
2.5G	GPRS (General Packet Radio Service) and CDMA 2000 1x	Current changes in the US and some other areas
2.75G	EDGE (Enhanced Data rated for Global Evolution)	In deployment phase in the US
3G	CDMA2000 (Broadband CDMA)	Current push for use in the US
	W-CDMA (Wideband CDMA)	Standard in Japan and Europe

Increasing Bandwidth and Connectivity

Bandwidth has been doubling every nine months, or roughly at twice the growth rate of computing power. Increasing bandwidth will lead to the creation of what is being called the "evernet," where billions of devices will be connected to the hyper-speed, broadband, multiformat Web. In the future, the Internet will always be "on" (Schapp & Cornelius, 2001).

The high-speed networks of the 3G generation will provide additional capacity and enhanced functionalities. There is a strong need to combine the wireless (LAN) concept and cell or based-station wide-area network design. 4G is seen as the solution that will bridge the gap and therefore provide a much more robust network (IC2 Institute, 2004).

Benefits and Challenges of Ubiquitous Commerce

The new era of ubiquitous commerce will mean that it will be possible to stay always on (e.g. unlimited interconnectivity), always aware (e.g. business and consumer value), and always active (e.g. sustained competitive advantage) (Gershman, 2002). The connectivity created by ubiquitous networks means that the nature of communication between an organization and its suppliers, customers and other stakeholders will change. The nature of competition will change and technology might enable new entrants to take up large shares of existing markets at relatively short notice (Watson et al., 2002). It is important that companies understand the nature and impact of u-commerce. Of course, the effects will be different in different markets depending on the industry. Then, the appropriateness of current and proposed strategies can be evaluated and speculated (Watson et al., 2002). U-commerce will have very broad implications. It may require changes on the structure of the firms and even on the business model itself.

U-commerce will create new levels of convenience and value for buyers and sellers. It is about the integration of more value-added information into each transaction, in ways that benefit both consumers and businesses. Ultimately, it is about minimizing friction in the commerce chain, creating new efficiencies and higher levels of productivity (Schapp & Cornelius, 2001). There are also significant macroeconomic benefits that can be gained from u-commerce. By facilitating the exchange of goods and services, they enable the different components of an economy to interact with one another. By removing friction from the exchange process, u-commerce can help economies to operate in a

more fluid and efficient manner. U-commerce will enable: improved operating efficiency, enhanced customer services, increased service personalization, continuous supply chain connectivity, and continuous interactivity. The companies that will be able to create a higher level of convenience and value through the technology will put themselves in a better competitive position.

There are a number of obstacles that need to be overcome in order to fully realize the u-commerce vision such as the lack of standardization, difficulty of reading from and writing on very small devices, as well as privacy and security issues.

Table 2. Research issues for u-commerce (adapted from Galanxhi-Janaqi & Nah, 2004)

Research Issues	Related Questions
Research issues regarding assessment of the true value of u-commerce and its deployment	• Will companies and individuals (e.g. consumers) benefit from u-commerce applications? • How should companies go about determining what combinations of components of u-commerce would provide the greatest value to the company? • How can contextual customization be realized in such a way that maximum benefits are obtained with minimum disruption of people's privacy? • Are the right devices and applications being introduced into markets to bring value to businesses?
Research issues relating to privacy, trust and security in u-commerce environment	• What kinds of information can be gathered without invading customers' privacy? What and how should/could companies optimize the use of information they have gathered while preserving customers' privacy? • How should companies develop trust (i.e. both initial and long-term trust) with customers? In what ways is trust in a brick-and-mortar setting similar to and different from trust in businesses in the u-commerce era? What additional variables should be taken into account in conducting u-commerce? • How can security be strengthened? Can the security technologies used in online e-commerce applications be adapted for other u-commerce applications? • How can customers' privacy be assured in the u-commerce environment? • Can security be improved without reducing the convenience of operations? Can information about the context be obtained in such a way that it gives accurate information, but at the same time protects people's privacy and anonymity?
Research issues regarding strategy in adopting u-commerce	• What is an appropriate or recommended strategy for companies that wish to take initiatives in u-commerce? Can a "Start Big" strategy be successful or is a step-by-step approach better or more appropriate? What strategy should be adopted for digital firms? What role does u-commerce play for a "still-alive".com?
Research issues relating to the pace of u-commerce adoption and the underlying technology	• Are businesses ready to adopt the new and somewhat unproven u-commerce technology? • What are the priorities and directions in solving the existing technical problems and impediments regarding systems, standards, security, and simplicity? • Which of the components of u-commerce (e.g. e-commerce, m-commerce, silent, and television commerce) is the most critical in realizing its full vision? • In what ways do the different components of u-commerce interact with and support one another? • Who would be accountable for potential breakdowns? How reliable is smart and agent technology? What is the calculated risk?

In addition, culture and lifestyle are two factors that highly influence the adoption rate of u-commerce in different regions of the world. Furthermore, each component of u-commerce applications offers many benefits, but also faces many challenges and raises many questions.

Mobile commerce faces the same problems troubling e-commerce—plus a few of its own (Siau & Shen, 2003b). The same is true for u-commerce. Table 2 summarizes some of the research issues and challenges of ubiquitous commerce. These issues are discussed under four categories (Galanxhi-Janaqi & Nah, 2004):

- issues concerning assessment of the true value of u-commerce and its deployment;
- issues relating to privacy, trust and security in the u-commerce environment;
- issues regarding strategy in adopting u-commerce; and (iv) issues relating to the pace of u-commerce adoption and the underlying technology.

Future Trends

U-commerce, by definition, implies the continued existence of traditional payment forms such as cash and checks. It represents an expansion, not a complete replacement, of traditional commerce (Schapp & Cornelius, 2001). U-commerce is something that is actually happening and it is a natural evolution of e-Commerce, mobile and other forms of digital commerce. Solving the following issues would accelerate u-commerce adoption:

1. Companies must overcome the difficulties they have experienced with their e-Commerce programs; see where they are and plan the u-commerce path.
2. At the same time, many technological impediments have to be solved.

A great deal of efforts is required in several important fields in order to realize the complete vision of u-commerce.

Essential factors that need to be addressed include systems, standards, security, and simplicity (Schapp & Cornelius, 2001). The full realization of the u-commerce vision will require new system interfaces, new customer service systems, etc. In order to take advantage of the innovations, developing countries must be prepared to evolve their basic infrastructure. In fact, not all countries are ready for the u-commerce wave.

Most of the previous discussions may only be valid for US, European countries, Japan and some other developed countries. Differences in the adoption of u-commerce among countries may be due to different business cultures, infrastructure, legal framework, and their experiences with its components (e.g. electronic, mobile, voice, television, and silent commerce).

Conclusion

This chapter provides an outlook on the significance of ubiquitous commerce and its characteristics. It also examines the relations of ubiquitous commerce with other forms of commerce. The chapter also provides a basis for future research in ubiquitous commerce. Since u-commerce is a new phenomenon and trend, it is necessary to study and understand how the different components of u-commerce complement one another to create business value and how such value can be increased through innovative applications.

From a practical standpoint, this chapter can help organizations and individuals better understand implications from the emergence of u-commerce and better manage the change. The last part of the chapter provides an overview of what this change may involve. U-commerce will widely affect many aspects of a business. Firms must understand how ubiquity, universality, unison, and uniqueness will affect their business. The deployment of u-commerce in the real world has implications beyond the technically obvious ones, such as issues relating to social, economic, and legal perspectives. Companies need to understand and know how they are going to manage the change. Privacy issues raised by ubiquitous commerce may also be an increased concern. The main privacy concerns include: the kind of information that can be gathered about a person; persons who have access to the information; how the information will be used; protection of personal information against theft or other unauthorized use; accountability of the entities that gather important and sensitive information.

Finally, it is important to emphasize that u-commerce is not a replacement of other types of commerce, but an extension of them. Since the u-commerce concept is very broad and all-encompassing, it will be a directive, not only an option. In the ideal situation, u-commerce—like an artery—will uninterruptedly connect the parts, and make the world live and function as one. Before such unity can be achieved successfully and swiftly, we need to address the many challenges it faces.

References

Accenture (2001). *The unexpected eEurope*. Retrieved November 20, 2002 from the Web site: *http://www.accenture.com/xdoc/en/ideas/eeurope2001/Full_Survey.pdf*

Du, T.C., Li, E.Y. & Chang, A. (2003). Mobile agents in distributed network management, *Communications of the ACM*, *46*(7), 127-132.

Duffy, G. & Dale, B.G. (2002). E-commerce processes: A study of criticality, *Industrial Management and Data Systems*, *102*(8), 432-441.

Galanxhi-Janaqi, H. & Nah, F. (2004). U-commerce: Emerging Trends and Research Issues, *Industrial Management and Data Systems*, forthcoming.

Gershman, A. (2002). Ubiquitous Commerce - Always on, always aware, always proactive. S*ymposium on Applications and the Internet*, Nara City, January 28 - February 01. Retrieved November 20, 2002 from the Web site: *http://www.accenture.com/xdoc/en/services/technology/publications/UbiCommerce-AINT2002.pdf#search='Gershman%20and%20Ubiquitous%20Commerce%20'*

IC2 Institute (2004). *Austin's Wireless Future*, University of Texas. Retrieved on May 1, 2004 from the Web site: *http://www.wirelessfuture.org/AustinsWirelessFuture.pdf*

Junglas, I.A. & Watson, R.T. (2003a). U-commerce: a conceptual extension of e-commerce and m-commerce," *Proceedings of the International Conference on Information Systems,* Seattle, WA, 667-677.

Junglas, I.A. & Watson, R.T. (2003b). U-commerce: An experimental investigation of ubiquity and uniqueness, *Proceedings of the International Conference on Information Systems,* Seattle, WA, 414-426.

Lange, D.B. & Oshima, M. (1999). Seven good reasons for mobile agents. *Communications of the ACM*, *42*(3), 88-89.

Laudon, K.C. & Traver, C. (2000). *E-Commerce: Business. Technology. Society.* Boston: Addison Wesley.

Mattern F. (2001). The Vision and Technical Foundations of Ubiquitous Computing. *Upgrade*, *2*(5), 2-6.

Nah, F., Siau, K. & Sheng, H. (2004). The value of mobile applications: A study on a public utility company. *Communications of the ACM*, forthcoming.

Riaz, U., Osman, J.A., Gorra–Stockman, M.R. & Shoup, C.A. (2002). Why telematics is moving into the realm of transforming technologies. *Outlook Point of View*, January. Retrieved December 2002 from their Accenture's

Web site: *http://www.accenture.com/xdoc/en/ideas/outlook/pov/USLtr_telematicsPoV.pdf.*

Schapp, S. & Cornelius, R.D. (2001). U-commerce: Leading the world of payments. White Paper, Retrieved December 2002 from Visa International Web site: *http://www.corporate.visa.com/av/ucomm/u_white paper.pdf*

Siau, K., Nah, F. & Sheng, H. (2003a). Values of mobile applications to end-users. *European Research Consortium for Informatics and Mathematics (ERCIM) News*, (54), 50-51.

Siau, K. & Shen, Z. (2003a). Building customer trust in mobile commerce. *Communications of the ACM*, *46*(4), 91-94.

Siau, K. & Shen, Z. (2003b). Mobile communications and mobile services. *International Journal of Mobile Communications, 1*(2), 3-14.

Siau, K., Sheng, H. & Nah, F. (2003b). Development of a framework for trust in mobile commerce. *Proceedings of the Second Annual Workshop on HCI Research in MIS*, Seattle, WA, 85-89.

Siau, K., Sheng, H., Nah, F. & Davis, S. (2004). A qualitative investigation on consumer trust in mobile commerce. *International Journal of Electronic Business*, *2*(3), forthcoming.

Sierra, C., Wooldridge, M., Sadeh, N., Conte, R., Klusch, M. & Treur, J. (2000). Agent research and development in Europe. *Information Services and Use*, *20*(4), 189-203.

Watson, R.T., Pitt, L.F., Berthon, P. & Zinkhan, G.M. (2002). U-commerce: Expanding the universe of marketing. *Journal of the Academy of Marketing Science*, *30*(4), 333-348.

Chapter VI

Tracking and Tracing Applications of 3G for SMEs

Bardo Fraunholz, Deakin University, Australia

Chandana Unnithan, Deakin University, Australia

Jürgen Jung, Uni Duisburg-Essen, Germany

Abstract

With dynamic growth and acceptance of mobile devices, many innovative business applications are beginning to emerge. Tracking and tracing seems to be one of the popular applications which many organisations have initiated, often facilitated by location based services provided by mobile network operators. However, there are many issues associated with the provisioning of this application with current technologies and business models. Small and Medium-size Enterprises (SMEs) that make up a significant segment of businesses worldwide do not yet seem able to benefit widely from these services. In this chapter, we initially review current technologies/ applications and the issues associated with them, drawing from research and the experiences of a long term ongoing action research project with

SMEs in the trade sector. Subsequently, we explore the opportunities offered by 3G services/business applications to SMEs, and provide a broad critical outlook on future opportunities for SMEs to benefit from 3G services.

Introduction

With a plethora of wireless handheld computing devices such as Personal Digital Assistants (PDAs), pocket PCs, Tablet PCs, along with more than one billion cellular phones in the world (Steinfield, 2003), mobile communications and commerce has become a significant market prospect. Mobile network operators continue to look for potential revenue generating business models to increase the demand for services in their area—as there is increased competition reducing prices for voice services. At the same time the cost of transitioning into the new generation infrastructure has risen (D'Roza & Bilchev, 2003). There have been many new developments over the past years and one of them is the arrival of location based services (LBS) for the Global Systems for Mobile communications (GSM) networks.

LBS provide customers with a possibility to get information, based on their location. Such information may be for example, the nearest petrol station, hotel or any similar services that may be stored by the service provider, in relation to any particular locality. These services are somewhat location-aware applications (VanderMeer, 2001) that take the user's location into account, in order to deliver a service. Other location based service applications have been applied for tracking and tracing of vehicles and people, especially by corporations (Paavalainen, 2001) where this activity is an integral part of management. Major freight operators are a typical example for such applications.

In recent years, SMEs with personnel in the field or on demand (for example, plumbers who are called in on demand/ given an assignment on phone based on their locality) are contemplating the introduction of application based on LBS within their organisations. However, the cost of implementing location determination technologies that support LBS is considered expensive by SMEs. The evolution of mobile networks into their third generation or 3G might generate the potential for SMEs to apply LBS — to get an affordable system to track their employees and improve the overall efficiency of the business.

With the current technology, tracking and tracing usually requires an additional GPS antenna to determine the precise location, as the position information from the GSM network is too fuzzy to provide accurate location data. Also communication with mobile units is often established by Short Message Service (SMS),

with a fixed fee associated with every query. There is no permanent flow of data communications (SMS is not an "always on" service) between the field staff and the base—and every query incurs relevant cost. With the evolvement of 3G networks, there is an expectation of change in the infrastructure or rather; there have already been some changes. 3G will allegedly provide "always on" data communication at high speed, which will enable easy and quick queries at any time, even continuous tracking and tracing. It is also assumed that this service will be at a comparatively lower rate.

This chapter initially reviews technology in relation to the accuracy of location data, data communication infrastructure and briefly studies relevant security issues in tracking/tracing applications. Inferences have been drawn from a long term action research project with SMEs in the trade sector as well as current literature. Some business models building on location based technology such as sending job information or changes directly to field staff are examined briefly. Subsequently, we explore opportunities that 3G service offers to SMEs in trade, especially in the tracking/tracing applications area, as compared to existing technologies/applications. Furthermore, a broad critical outlook on 3G provisioning from the SME perspective is provided.

Location Based Services: Context

Location based applications have developed into a substantial business case for mobile network operators during the last few years (Paavalainen, 2001; Steinfield, 2003). ITU estimates worldwide revenues from LBS would exceed $2.6 billion in 2005 and reach $9.9 billion by 2010 (Leite & Pereira, 2001). Market research by Strategy Analytics in 2001 indicated that location-based applications have huge revenue potential for operators, with an expected $6 billion of revenue in Western Europe and $4.6 billion in North America by 2005 (Paavalainen, 2001). An ARC Group study suggests that LBS will account for over 40 percent of mobile data revenues worldwide by 2007 (Greenspan, 2002) and there might be 748 million users worldwide for LBS as early as 2004. According to Smith (2000), more than half of the US mobile customer base was willing to accept some form of advertising on a mobile handset, if they were able to use location services for free. An Ovum study predicts Western European market for LBS to touch $6.6 billion by 2006 and as much as 44 percent of mobile subscribers to be using LBS (Greenspan, 2002).

Mobile subscribers, especially in industrialized societies are unwittingly using a location determination technology, (Steinfield, 2003) due to the fact that regulators in most of these nations have initiated rules requiring network operators to

deliver information about location of a subscriber, to public safety answering points in the event of an emergency. In the U.S., the Federal Communications Commission requires operators to provide the location of all mobile emergency calls and therefore, the market itself was government driven (FCC, 2003). The EU is developing a similar requirement for its emergency services (D'Roza & Bilchev, 2003). Corporations have begun to realize the benefit in deploying location based technologies, as they benefit from cost savings as well as increased efficiency for their existing mobile applications (Schiller, 2003).

Defining Location Based Services

As with the concept of mobile applications and -commerce, there is not one specific definition of LBS. Prasad (2003) purports that location based service or LBS is the ability to find the geographical location of the mobile device and provide services based on this location information. Magon and Shukla (2003) agree that it is the capability to find the geographical location of the mobile device and then provide services based on this location information.

> *"Location Based Services can be described as 'applications, which react according to a geographic trigger.' A geographic trigger might be the input of a town name, zip code or street into a Web page, the position of a mobile phone user or the precise position of your car as you are driving home from the office…" (Whereonearth, 2003).*

Turban (2002) suggests that LBS refers to localisation of products and services or rather applications that are specific to a user location. All of these broad definitions point to one critical component in the LBS—the user location. Technologies that support the determination of user location—commonly termed as positioning technologies are examined in the next section.

Mobile Technologies for LBS

The critical application of LBS is the determination of a user's location—using positioning technologies. Drane and Rizos (1998) emphasize three conceptually different approaches to generic positioning technologies such as signpost, wave-based systems and dead reckoning. Within the mobile communication networks,

Röttger-Gerigk (2002) distinguishes between network-based and specialized positioning services. In the following sections we elaborate on these approaches and further discuss the GPS as a specialized positioning technology and selected network based systems.

Basic Characterization of Positioning Technologies

Sign-post systems represent the simplest sort of positioning and are based on an infrastructure of signposts (i.e., landmark or beacon). Positions are measured by determining the nearest beacon to the mobile object. Therefore, positioning is reduced to the statement that a mobile object is nearby or in certain proximity of a certain beacon. The accuracy of signpost systems is given by the distance between two neighbouring signposts. Currently, signpost systems are used for automatic toll collection on highways. Such an approach is presented as part of the PAMELA-project (Hills & Blythe, 1994). Road-side charging stations communicate with in-vehicle transponder logic circuits for the exchange of a car id with signposts assigned to a certain road stretch. Electronic signpost systems depend on an infrastructure of electronic beacons and a facility for connecting to the next signpost (i.e., a transponder). Advantages of signpost systems are their robustness and low costs for vehicle-mounted devices (Drane & Rizos, 1998). On the other hand, there are a lot of costs for the installation of the signpost infrastructure of automated systems. Those systems are usually installed along (possibly busy) roads and do not cover wide-spread areas

Wave-based positioning systems use propagation properties of (usually electro-magnetic) waves to determine the position of a mobile object. Locations of mobile objects are determined relative to one ore more reference sites. Main criteria for the positioning quality of wave-based systems are accuracy and availability. The accuracy is mainly limited by technical restrictions and additionally by non-technical issues. Technical restrictions will be discussed later. An example for non-technical issue has been the *selective availability* (SA) in the Global Positioning System (GPS). The SA has been a noise, which interfered with the GPS signal and restricted its accuracy for non-military users. The availability of wave-based positioning systems is limited by an undisturbed reception of the radio waves sent by the reference points.

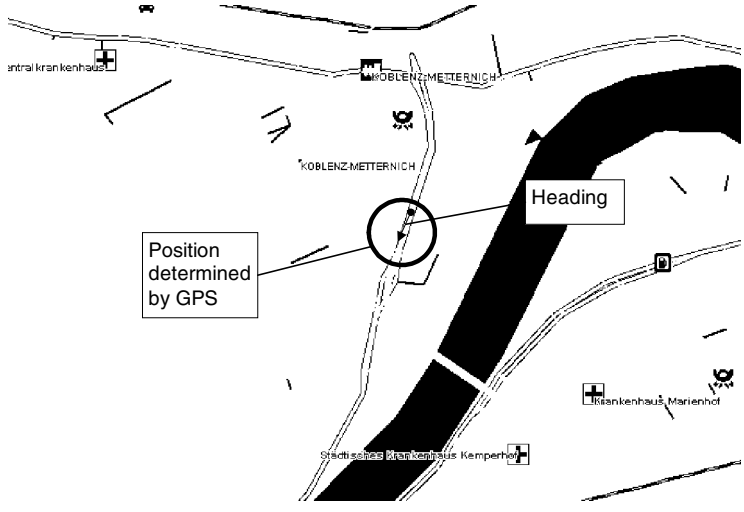
Dead reckoning systems consist of several vehicle-mounted sensors for the detection of a mobile object's movements. These sensors are used for the continuous determination of a vehicle's velocity and heading. Starting from an initial reference point, a mobile object can be located by logging its speed and heading over time. The velocity can usually be determined by evaluating the speed signal of a vehicle. The heading can be measured by a build-in electronic

compass. Acceleration and direction changes may be resolved from evaluating changes in velocity and heading periodically. Other sensor technologies used in dead reckoning systems include accelerometers and gyroscopes (Drane & Rizos, 1998).

Yet another classification of positioning technologies uses the approach as to where the location of a mobile object is determined. Here, positioning systems are characterized as self-positioning or remote positioning (Drane & Rizos, 1998). In *self-positioning* systems the position is determined in the mobile device itself. Hence, the position is primarily known by the mobile object itself. Complementary, the information about the location may be transmitted to external systems or partners over a mobile communication infrastructure. *Remote positioning* systems provide positioning services only for external systems, which can than use this information for customized location-based.

The hitherto presented types of positioning technologies usually result in an absolute specification of a mobile user's location. Signpost systems specify a position basing on a network of landmarks; wave-based systems on basis of properties of the propagation of electro-magnetic waves. Dead reckoning systems record movements, acceleration and the velocity of mobile objects by using special sensors. Nevertheless, mobile users (especially the ones going by car) are moving along roads.

Figure 1. Map matching in positioning services



According to Drane and Rizos (1998) the exact determination of a mobile user's position is supported by its estimated position in relation to given map data. One heuristic might be that a user in a car might only drive on a given road. One example for the combination of established positioning services and map matching is shown in Figure 1. The estimated position of the mobile user is given by the red circle in the diagram. According to the simple rule, that a mobile user in a car can only be located on a road results in the positioning of this user on the given position in the figure.

Practically, several of the given positioning services are combined. The result is a high-value positioning service. Popular navigation systems for example depend on GPS, dead reckoning and map matching: A GPS-antenna is used for the determination of a vehicle's position and this information is adjusted with the information given by dead reckoning and map matching. Hence, different positioning systems can not be discussed in an isolated manner. Current systems basically depend on basic kinds of positioning technologies as well as valuable combinations of those technologies. Special positioning technologies (like GPS) and different kinds of network-based positioning services will be discussed in the next sections. We emphasize on basic conceptualizations and conceptual differences.

Global Positioning System

The Global Positioning System (GPS) is a self-positioning, wave-based positioning system. GPS has been launched by the U.S. Department of Defence in the 1970 (Drane & Rizos 1998). Currently, GPS consists of at least 24 satellites revolving around the earth on six orbits (Lechner & Baumann, 1999). All satellites send a continuous radio signal (every second) including its position and the sending time. A special GPS-receiver uses the signals of at least three satellites for the determination of its global position (Janecke, 1999). The position is computed by propagation delays of the signals sent by the satellites. Similar—but less popular—systems are the Russian GLONASS and the future European satellite-based positioning system GALILEO (Lechner & Baumann, 1999). With respect to accuracy, satellite-based positioning systems are expected to play an important role in the long term.

The accuracy of GPS was formerly divided into two classes: U.S. military profited from maximum accuracy of about 10 metres. The GPS-signal was distorted for civil use by the so called *selective availability* (SA). SA means an interfering signal overlapping the regular GPS-signal, so that devices used for civil purposes may not determine the position in a better accuracy than about 100 metres. Selective availability has been switched off in June 2000. Thus, the

currently available accuracy of GPS-based positioning systems is about 10 metres for civil and military use (Durlacher, 2001).

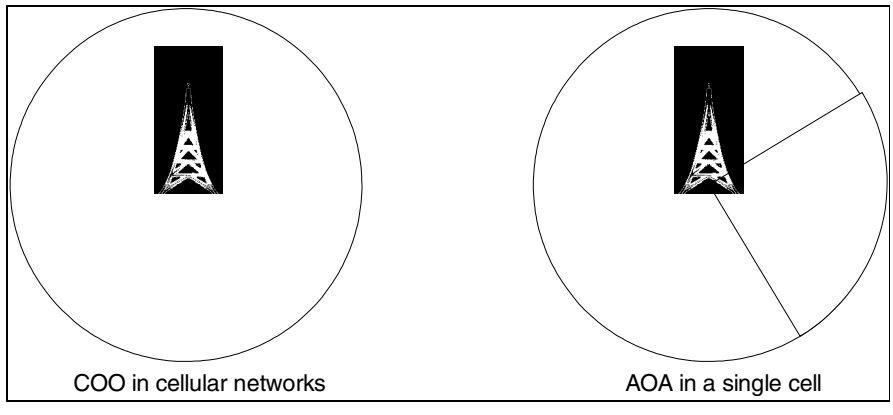
The accuracy of GPS can be improved by *Differential GPS* (D-GPS). D-GPS uses a reference station with a well-known global position for the correction GPS data (Röttger-Gerigk, 2002). This reference station receives the satellite signal and calculates the deviation between its exact position and the GPS signal. This deviation is transferred to the mobile objects and used for the correction of their positioning data.

Network-Based positioning

In contrast to special positioning-systems, network-based positioning is usually part of another given network. Examples for such kinds of networks are cellular communication networks such as GSM (Global System for Mobile telecommunication) and UMTS (Universal Mobile Telecommunication System) as well as WLAN (Wireless Local Area Network). WLAN-positioning as presented in Roth (2002) is not discussed in detail within this chapter as we rather present general characteristics of cellular networks for the determination of a user's location.

Cell of Origin (COO) determines a mobile user's location by the identification of the cell in which the person's mobile device is registered (Röttger-Gerigk,

Figure 2. Positioning by cell ID (left) and arc of a circle



2002; Steinfield, 2003). Hence, the accuracy COO is given by the size of a cell. According to Roth (2002), this positioning method is also known as Cell Global Identity (CGI). Despite of its comparatively low accuracy, this technology is widely used in cellular networks. The reasons are simple: The accuracy is sufficient for some applications and the service is implemented in all GSM-based networks. COO is a remote-positioning service (like most network-based positioning services) but information about a location can also be transferred to the mobile device by cell broadcast.

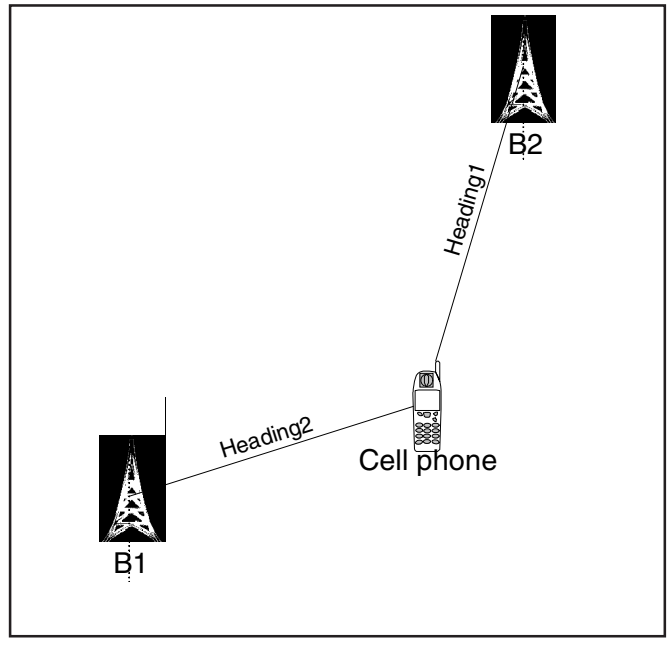
Angle of Arrival (AOA) bases on traditional positioning techniques and uses the bearing of at least two base-stations (Röttger-Gerigk, 2002; Steinfield, 2003). In most cellular networks such as GSM, the antennas of a base-station might be used for the determination of the angle of an incoming signal. The antenna of a base station in GSM only covers a part of the area of a circle (i.e. 120° of a whole circle).

Figure 2 illustrates the difference between COO and AOA: COO covers the whole of a network cell whereas AOA only covers the arc of a circle. AOA is like COO available in most cellular networks and, thus, already implemented. Using a single base station, the positioning accuracy is better than using COO and can be improved by combining the information of at least two base stations. Using the bearing of two base stations is displayed in Figure 3. Each of the base stations B1 and B2 in Figure 3 receives the signal sent by cell phone from a different angle (represented by heading 1 and 2).Figure 3. AOA with two base stations (Röttger-Gerigk, 2002)

Timing Advance (TA) is a very important function in GSM because a time-multiplexing transmission method is used (Röttger-Gerigk, 2002; Steinfield, 2003). Every data package has to fit into a given time slot. Because of the light-speed the radio signal sent by a mobile device needs some time to reach the base station. Such a delay of a data packet has to be taken into account. TA determines the signal's running time and causes the mobile device to send the data some microseconds in advance. The timing advance allows for the determination of the distance between a base station and a mobile device in multiplies of 550m. Positioning using TA is shown in Figure 4: The diagram on the left hand side illustrates TA in a single cell and the one on the right hand side combines TA and AOA. TA is actually a GSM-specific method for the determination of the distance between a base-station and a mobile device. Nevertheless, it demonstrates the basic idea of distance measurement in cellular networks. TA is not only a hypothetical method but practically used in GSM. Similar methods are used in other cellular networks.

According to *Time Difference of Arrival* (TDOA), the time-difference of the arrival of a signal sent by one single mobile device at several (at least three) base-stations is recorded. In other words: a mobile unit sends a specific signal at a

Figure 3. AOA with two base stations (Röttger-Gerigk, 2002)



given time. This signal is received by several base-stations at a later moment. Given the expansion speed (light speed) and the time differences of arrival at the base stations allows the positioning of the mobile unit. Essential for this positioning service are a precise time basis and a central unit (called: Mobile Location Center) for the synchronisation of time data between base-stations. TDOA is a remote-positioning service which needs no upgrade at the mobile unit and minor changes at the net infrastructure.

Time of Arrival (TOA) is a similar method to TDOA. In contrast to TDOA the running time of a radio signal will be measured and not the time-difference (Röttger-Gerigk, 2002; Steinfield, 2003). The mobile unit is sending a signal which will be received by at least three base stations. The position of the mobile device will be calculated on basis of time-differences of the received signal at each base station. A schematically drawing is given in Figure 5. Four base stations are used for the positioning of a mobile user. This user can be located in discrete distances from these four base stations. The distances of the user from the base stations combined with the absolute positions of the base stations allow the absolute localization of the mobile unit.

Figure 4. Positioning based on timing advance

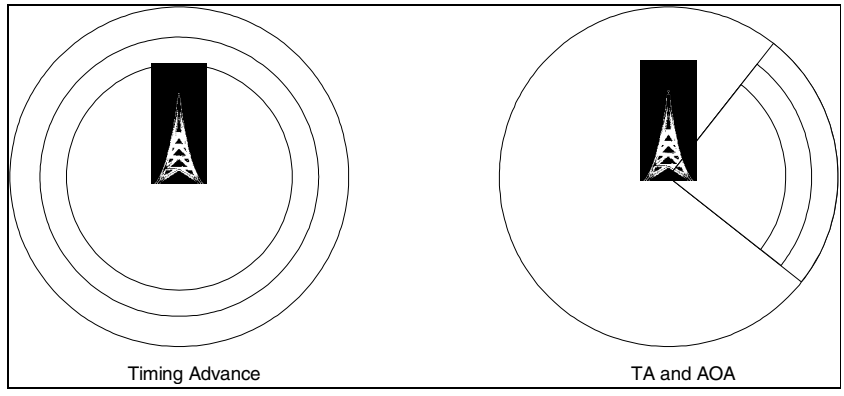
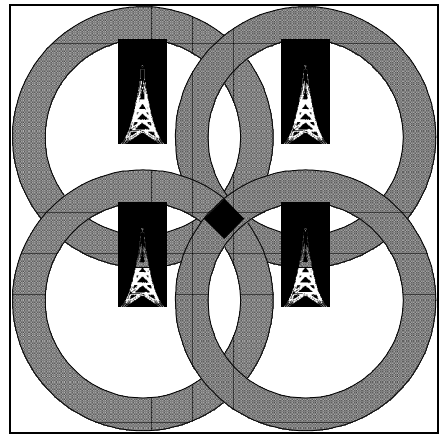


Figure 5. TOA with four base stations



All positioning systems depend on conceptual strengths and weaknesses. GPS profits from high accuracy in conjunction with high user device costs. Nevertheless, the positioning quality of GPS-based systems is hindered by a reduced reception of the satellite signal on certain roads. High buildings and trees may keep away the GPS-reception from the receiver. In current navigation systems, such differences from the GPS-signal are compensated by dead reckoning systems. Combining GPS-position signals with the data from in-vehicle velocity and direction sensors lead to a more precise positioning quality.

Table 1. Comparison of GPS and network based systems

GPS	Network
Network of its own	Part of a popular network
Special end-user devices	Wide spread end user devices
High accuracy	Lower accuracy
Global availability	Availability restricted by network coverage

The most obvious technology behind LBS are the positioning technologies and the widely recognized Global positioning System (GPS). However, there are network based positioning technologies that typically rely on triangulation of a signal from cell sites serving a mobile phone and the serving cell site can be used as a fix for the locating the user (Mobilein, 2003). There is a need to support multiple location determination technologies (LDT) and applications for locating the mobile device. An integrated solution should support many types of available LDT technologies such as Cell ID, AOA, TDOA, GPS and TOA (Infoinsight, 2002).

Geographical data is critical for any LBS and Geographical Information Systems (GIS) provide tools to provision and administer base map data such as built structures (streets and buildings) and terrain (mountain, rivers, etc.) (Mobilein, 2003). GIS is also used to manage point-of-interest data such as location of petrol stations, nightclubs, hotels etc. GIS also includes information about the radio frequency characteristics of the mobile network—which allows the system to determine the serving call site of the user.

Searby (2003) suggests that location is a simply a useful bit of data that can be used to filter access to many types of GIS. It is powerful when combined with user profile information to offer personalized and location sensitive responses to customers.

The location management function processes the positioning and GIS data on behalf of LBS applications, acting as a gateway/ mediator between positioning equipment and LBS infrastructure (Mobilein, 2003). The location manager (LM) or gateway will aggregate the location estimates for the mobile device from the various LDTs, compute the user location and estimate the certainty of that location before being forwarded to the application. Proper location services solutions need to provide interfaces which translate location results to user applications, integrate with content providers and provision query results to MAP servers, clients and databases. A location services solution must provide multiple invocation (or request) options such as SS7 or XML and delivery options such as voice, data and voice/data based on user requirements (Infoinsight, 2002).

The combination of location and content is a big driver for the handset manufacturers to deliver easy to use interfaces at the handset for mapping, video whilst providing response time appropriate for a mobile society. Investment for the operators is larger than the software requirements as the effective use of this technology extends beyond the investment in location services software. There will be increased need to profile customers, customer data base management now including location information, data mining and integration to existing Customer Relationship Management (CRM) software to map content (Infoinsight, 2002).

Crisp (2003) suggests that LBS is a confusing array of changing requirements, emerging standards and rapidly developing technologies. There seem to be confluence of previously independent technologies: mobile communications, positioning systems, mobile computing, the storage and manipulation of spatial information in relational databases and the availability of large spatial datasets. The mobile communications technology has developed as the core of mobile telephone and wireless LAN markets; mobile computing has developed as the core of laptops/notebooks, tablet PCs, PDA markets; spatial RDBMS technology has developed within the GIS market and spatial datasets have been collected largely by research, government, mapping and utility organisations for their own purposes. Confluence has been unpredictable as each technology develops at a different rate, as per the demands of its market, while being constrained by standards specifications (Crisp, 2003).

Steinfield (2003) points out that there are many different players that are involved in LBS (see also Pearce, 2001; Sadeh, 2002; Spinney, 2003) including GIS and other content providers who offer mapping services, geographic content, often accessed via a server; service providers who aggregate GIS and other content to create services; application vendors who package services for mobile operators; location middleware providers who provide tools to facilitate mobile operator's use of various applications from different providers; mobile operators who manage infrastructure, collect position data, offer the service to end subscribers and perform billing/collection services; location infrastructure providers who sell the mobile location centres and other hardware and software to network operators; and handset manufacturers who sell devices capable of interacting with location based services.

Since all are stakeholders who potentially earn revenue from LBS, they require standard formats and interfaces to work efficiently (Spinney, 2003; Steinfield, 2003). Otherwise, the costs of launching a separate set of services would be passed on to end users—and that would be destructive for mobile operators. Therefore, the ability to create, implement LBS, maintain service quality and enable roaming across mobile networks depends on the development of global, industry-wide standards. For example, if there are competing networks using

different air interfaces and network infrastructures with pockets of coverage on a market by market basis rather than national licences—as is the case with US—there is a significant issue in the development of positioning technologies (Steinfield, 2003).

The global third generation partnership project (3GPP) through which various standard bodies are attempting to create a smooth transition to third generation wireless networks deals with LBS (Adams, Ashwell & Baxter, 2003). The Open Mobility Alliance is another attempt for creating global standards not only for positioning technologies, but also for the services and interfaces among content and application providers, for privacy related procedures and for testing systems accuracy (Adams et al., 2003; Hatfield, 2002; Steinfield, 2003).

It appears that the second generation of mobile technology development, with all other related technologies developing at different pace and catering to different standards have hampered the development of LBS applications. In the next section, we examine some LBS business applications particularly dealing with tracking and tracing and relate the future possibilities with third generation mobile technologies.

Security Issues

One of the critical issues in the context of this chapter is the security in the wireless environment. Prasad, Wang and Choo (2003) list three main issues from the user perspective. First, security of the mobile terminal, that is, the device should only be activated by the authorized user, protected against viruses or worms and also theft.

Privacy of data, communication and location is another major issue. These issues have different connotations in economies. Another security challenge is service provisioning—to prevent denial of service attacks and provide secure mobile infrastructure. According to Prasad et al. (2003), network operators, regulatory bodies, manufacturers and other participants need to pull together in creating security standards for future wireless systems.

There are a number of issues in mobile security and the above listed are only a few examples. Initiatives are underway across the world to address them. However, this chapter does not focus on these issues and therefore, examining wider issues/initiatives are beyond its scope and relevance.

Business Applications: Current and Future with 3G

Location services are added value services that depend on a mobile user's geographic position (Infoinsight, 2002). According to Steinfield (2003) there are numerous ways in which location based data can be exploited, especially in combination with user profiles, to offer solutions to customers. Table 2 summarises the classifications offered by different authors (D'Roza and Bilchev 2003; infoinsight 2002; Levijoki 2001; mobilein, 2003; Steinfield 2003; Van de Kar and Bowman 2001).

NEXTBUS in San Francisco offers its customers a location based service. Using an Internet enabled mobile phone or PDA, bus riders can find estimated arrival time at each stop in real time and also location based advertisements will pop up on your mobile. For example, you have the time to get a cup of coffee before the bus arrives and Starbuck's is 200 feet to the right (Turban, King, Lee, Warkentin & Chung, 2002).

Another example is an LBS application that interacts with location technology components to determine the user's location and provide a list of restaurants. Hotelguide.com stores user profiles—specifically business travellers. At a new location, the user is able to search for a suitable hotel using the WAP phone, make a reservation and book a taxi to get them to the hotel. Travellers in unfamiliar cities needing immediate accommodation find this business model very useful. The current technologies however do not have the ability to recognize where the person is and then do the search accordingly (Turban et al., 2002).

Kleiman (2003) refers to two concepts that may be possible in the near future—location awareness and sensitivity. Location awareness refers to applications or services that make use of location information—where location need not be the primary purpose of the application or service. In contrast, location sensitivity refers to location enabled devices such as mobile phones, PDAs, or pagers. In the future, the phone will be able to locate a person, where that person is and search for a suitable hotel, without the need for the person entering the search. This is expected to be even more of a possibility with 3G technologies just around the corner.

Galileo operates one of the largest computerised travel reservation systems and offers a service to enable travellers re-book and monitor the status of flights using WAP phones. They have the provision to notify the customer if the flights are delayed or cancelled. The ability to track people wherever they are and to notify customers of cancelled flights well in advance is yet another future possibility with 3G. Weather forecasts, tourist attractions, landmarks, restaurants, gas stations, repair shops, ATM locations, theatres, public transportation

Table 2. Categorisation of business applications in LBS

Source	Classification
Mobilein (2003)	Location based information, location sensitive billing, emergency services, tracking.
InfoInsight, (2002)	Location based billing, location information services, banking, traffic, entertainment, travel, home shopping, emergency/safety services, tracking services.
Van de Kar and Bowman (2001)	Emergency services, mobile network operator services and value added services which include information provision, entertainment, communication, transaction, mobile office and business process support services.
Levijoki (2001)	Billing, safety, information, tracking and proximity services.
D'Roza and Bilchev (2003)	Pull services - requested by users once their location is determined and Push Services - triggered automatically once a certain condition is met (when a boundary is crossed). OR Five application areas : communication, fleet management, routing, safety, security and entertainment.
Steinfield (2003)	Consumer based – Business or Employees in a firm.

options (including schedules) are some examples of information provision filtered to the user locations (Steinfield, 2003).

Emergency services are one application that triggered the growth of LBS. In the U.S., the Federal Communications Commission issued the E911 mandate requiring every network operator to be able to detect the location of subscribers within 50 metres for 67 percent of emergency calls and 150 meters for 95% of calls (FCC, 2003). Dialling 911 from a mobile phone pinpoints your location and relays it to appropriate authorities and the FCC mandates a degree of accuracy in the pin-pointing for all mobile users in US (Paavalainen, 2001). The European Union has developed similar requirements for their E112 emergency services.

Many governments are moving to require that mobile operators develop the capability to automatically identify subscriber location, so that in the event of emergency the data may be forwarded to the public safety answering point to coordinate dispatch of emergency personnel. Combined with telemedicine techniques that allow psychological data transmission back to health care providers, this is another useful application. With the provision of 3G, it may also be possible to trace the person automatically, without the need for dialling 911—being a context aware, always on technology (Fraunholz, Hoffmann & Jung, 2003).

Proximity services inform users when they are within a certain distance from others, businesses etc. NTT DoCoMo offers a "friends finder" service on iMode

where you can find a predefined friend's location. With 3G context awareness services, it is possible that your mobile device can detect and let you know that you are near a friend or a business, without even trying to find them. However, this provision raises important privacy issues in some economies where the laws do not allow a person to be tracked consistently (for example, Australia). A deeper insight into these issues are beyond the scope of this chapter.

Tracking is rather a large category that contains everything from fleet applications typically entailing tracking of vehicles for the purposes of the owing company knowing the whereabouts of the vehicle and or operator. A successful example is dynamic vehicle routing. Dynamic fleet dispatch is assigned according to the location of a driver—as trucks are equipped with the Global Positioning System (GPS). The GPS points the location of the truck using up to seven satellites; location coordinates are sent to the Internet using a mobile network (mainly GSM) and through to a call centre. The call centre then spots the location of the truck and is able to optimize the route in real time and changed driving directions are sent to a mobile terminal of a driver (Paavalainen, 2001).

Trucking companies are fitting systems that not only track the location of vehicles, but also contents of delivery trucks so that last minute changes can be made based on inventory changes and location (Brewin, 2001). Combined with navigation services, tracking can help optimize deliveries and prevent theft of valuable items as well as locate people. GM's Onstar is using vehicle-based GPS receivers and mapping/route guide services in selected cars. These services can be integrated with real time traffic data, to make routes contingent on traffic conditions (Steinfield, 2003).

Chatterjee (2003) refers to fleet management systems i.e. a vehicle tracking system which is part of the fleet management. The system is fitted into the fleet of vehicles and generates data/ computes the exact distance travelled in a given time span, speed at a given location, analysis of time taken by the vehicle to cover distances etc, This system becomes a powerful tool for the operating agencies to manage their fleets and deploy human resources optimally.

GPS North America (Gpsnorthamerica, 2003) has a Web application called MARCUS which has the ability to locate and find a single vehicle, a fleet of vehicles and the closest unit to a particular location address. This is updated every five minutes and can be seen in real-time as well as historical track or "bread-crumb" trial in the past three months. This application is designed for occurrences to allow remote monitoring of the fleet and crew. Automatic vehicle location in transit is another application that is growing and is expected to benefit in increased overall dispatching and operating efficiency and more reliable service, as the system operates by measuring the real-time position of each vehicle.

Infomove (2003) has a smart telematics software that allows for integration of number of services from a host of service providers. This means that the product would work with different network standards such as GSM, CDMA or 1XRTT. End users of the system are allowed traffic updates, estimated time of arrival (ETA), and so on while driving from a mobile device or voice portal. Highlighting an address on the Palm and using "find me" search triggers the system to navigate the end user. Although the solution is interoperable with different network standards, it still combines the GPS capabilities with the second generation networks, to provide this service. Provision of 3G services in new cars renders automatic tracking and tracing possible, being a context aware technology, and it is not too distant in the future.

Similar solutions are deployed by the utilities industry such as electricity or power plants (Paavalainen, 2001). A remote control unit sends a notice to the central system in case of a failure, which goes through a mobile network and Internet on to an automated system which sends the request dynamically to the closest field worker. The field person then assesses the problem and any spares required are requested through the mobile terminal and the system dynamically orders and dispatches the spares (Elliott & Phillips, 2004). Increased efficiency results from the fact that workers are able to put in their work hours/availability into the mobile terminal--which is stored with the central system as well as availability and order processing capability of the central system (Paavalainen, 2001). With 3G based automation, automatic triggers may be more cost effective and possible without manual intervention.

Yet another useful LBS application is targeted at employees who are in the field and require access to internal information system applications. This application may need to involve a network operator as a partner for implementation. For example, it is possible that the employee is in close proximity to a client and the internal information database suggests critical updates to the client details. Currently, with limited screen spaces on mobiles, even small emails need filtering so that only relevant information is passed on to the field personnel. Alternatively, an SMS may be sent with the critical update, but there is the possibility of it not reaching the recipient in real time. With 3G based LBS, context awareness capability can be combined with always on facility to work around this cumbersome application.

Similarly, scheduling/rescheduling employee tasks in the field—taking into account their current location should become relatively easy and cost effective with 3G. Currently, if an employee has been scheduled for four tasks, it is scheduled in advance. Although, the employee movements can be relatively predicted using this schedule, it is always possible that the employee has finished early or is at another location, perhaps stuck in a traffic jam. There is no way to

track the movement, except if the person calls in on a mobile device or uses SMS to inform the office. There is a strong possibility of mobile employees to default their schedules claiming different reasons, of which the businesses have no control over. It is also possible that a job has finished early and yet another job has occurred in close proximity of the employee's current location. If this condition occurs, again if the employee calls in, the office might be able to reschedule the timings, thus optimizing the effort of the field person at a location. However, this is only possible if the employee movements are traced by the office. With current technologies the mobile calls are expensive and SMS is inconvenient and not always in reliable. Therefore, many organisations are not able to track their employees and optimize their time out in the field. With the provision of 3G it is possible to see the person, regardless of the location, being an "always on" and context aware technology, which enables tracing the person quickly.

Crisp (2003) purports three approaches to LBS applications: consumer services, field operations, and location enabled applications. Field operations, being the emphasis of this chapter, are discussed further in detail. It is an area with a strong business case for investment in LBS. Gathering location information from the crew in the field, without having them report back to the office, thus getting more jobs done each day; would reduce significant overhead costs. Receiving work order forms electronically, filling them out on site and sending them back in real time, immediately with electronic marked maps, ensures all work is done on site—without extra data entry, or staff travel back and forth. Having key locations, tracked vehicles and personnel visible on electronic maps gives the controller a clear view. In reviewing the work flow of field personnel, it may be seen that there is ample opportunity for operations stream line and achieving significant cost saving by using LBS (Crisp, 2003).

The Intelliware white paper (WP1020A, 2003) refers to a mobile resource management solution which allows the management of resources in time and space, where optimising the use of resources is dependant on location. This involves awareness of the location of resources (tracking them); knowing where they have to be (situation assessment) and monitoring incoming jobs/tasks; and matching incoming tasks with available mobile resources (assignment of jobs/tasks). This application namely Mobile Resource Management is applicable to SMEs as well as large organisations. Further, it underpins the aspects of integration into corporate systems, automation of job assignment, and feedback and logistics. It is evident that cost effective solutions are beginning to emerge perhaps also taking into account the capabilities of 3G standards. Having synthesized some business applications and possibilities of 3G, we now focus on the SME scenario as to what is the current and expected future possibilities with 3G.

SME and LBS: Current and Future with 3G

Small and medium enterprises are a significant segment of businesses worldwide and many of them are in the trade sector with typically 10-50 employees. This section examines the current technologies/applications from the SME perspective and compares them to what 3G provisioning promises to deliver.

Take the scenario of an SME in the plumbing trade with employees out in the field. A customer may contact the organisation with an emergency repair request. A technician in the field may be very close to the location of this customer in an emergency. However, with the existing technologies/services, it is not really possible to identify the employee's current location. Instant contact is required with the employee and often employees are equipped with pagers. However, messages do not seem to reach employees in time or are largely ignored as the employee can easily feign to be in a different locale each time. A short message or SMS sent on the mobile phone is another cost efficient alternative with current GSM services. However, there is no guarantee that the SMS will get to the employee instantly on the network. Perhaps, then the best way of contacting a person out in the field is conceivably by a mobile voice call. However, it is still not possible to trace the location of this employee.

Alternatively, many organisations are experimenting with providing PDAs to employees where the provision of sending an e-mail is possible via GPRS services on GSM networks. However, small interfaces, slow download speeds on 2.5 generation data communication network combined with the expensive package costs of GPRS services make it significantly cumbersome and not cost effective. In addition, there is the administration overhead of keeping a person within the office of the SME, to take customer calls and then trace a technician out in the field via mobile voice calls or SMS. Moreover, all of these options tend to be isolated and not integrated with the enterprise resource management system. By the time the organisation manages to trace the employee out in the field, the customer's emergency situation may escalate to the point of getting out of control. In this case, the SME may even lose a customer permanently.

In our preliminary action research on SMEs in plumbing trade, fleet vehicles were fitted with GPS navigation systems for tracking and tracing personnel. In order to facilitate tracking/tracing they were connected via black box to a GSM network so that navigation systems could be incorporated via SMS. This is not ideal but cost effective. Eventually the data gathered by these systems was to be integrated into an enterprise resource management system and thus enable tracking/tracing applications that support SMEs. For example, there is the possibility of tracking /tracing the employee with a Web interface, and redeploy

the person to attend to an emergency situation close by. In addition, there is the possibility of tracking employee data as to distance travelled, for billing purposes.

However, culturally, most owners/managers in SME trade sector belong to a more techno phobic generation where installing satellite based positioning systems, setting up internal information systems software that integrates databases, partnering with a network provider who uses 2.5 generation technologies that offer expensively packaged data services may seem a little excessive. In addition, provisioning an integrated enterprise system may incur significant out of pocket expenses from a sole proprietor, as extra resources may be required adding overheads as well as initial set up costs of provisioning the service.

However, the provision of 3G, with economic scenarios, "always on" and context awareness which claims to be relatively inexpensive, seem to offer significant opportunities for SMEs. 3G service not only enables instant contact with personnel—being "always on"—but is also expected to provide "context aware" applications. The service could have video monitoring provision so that employees cannot claim to be unaware of calls or not respond to calls—as is the case now with second generation mobile networks. With this, an integrated enterprise resource management is close to reality.

However, it remains to be seen if 3G will live up to its promises. This can only happen if the global 3G standards become a reality and the 3GPP forum and individual national legislating is able to address the privacy issues relating to the use of 3G services. Specifically, consistent tracking/tracing of employees are against the privacy laws of some countries in the European Union. It also remains to be seen if 3G will provide the promised "value for money" service as the frequencies had to be purchased at a high premium—possibly making 3G network provisioning more expensive than 2.5G. The issue of speed also becomes significant with an "always on" service if there is high load on the network.

Outlook

It is anticipated that 3G technologies are to be low cost, "always on" and "context aware" with vast potential for the SME context when they become a reality. Steinfield (2003) points out the development of location aware devices and technologies that incorporate "sensing" human factors, physical environment, and so on, which would add significant value to 3G. However, privacy and standardization are major issues that are emerging with the "always on", possibly video monitored 3G applications. Tracking and tracing of employees may be violating privacy laws in some countries, in which case the business models using 3G provisioning may not even be relevant.

This chapter is inspired by a long term research project on developing applications for SMEs in trade and draws inferences from the initial experiences on what 3G can offer to SMEs in the trade sector as compared to the current technologies/applications. If promises are to hold true 3G networks offer viable tracking/tracing applications with integration into the SME's enterprise resource management system. More specifically, it may be the perfect time to integrate and optimize resources without incurring significant additional overheads.

This chapter is clearing the path to enhance the performance/competitiveness of SMEs in trade and chambers involved. Additionally, it is an opportunity for technology providers involved in developing and integrating 3G solutions to develop viable alternatives for SMEs. As of now, business applications in 3G networks are still nascent and service providers need to transition from the offering of "video calls" to say good night to babies, to more viable business models.

References

Adams, P., Ashwell, G. & Baxter, R. (2003). Location based services – An overview of standards. *BT Technology Journal 21*(1) 34-43.

Brewin, R. (2001). Penske outfits fleet with wireless terminals. *GeoInformatics*, September.

Brewin, R. (2001). Trucker McLane Rolls out dual-mode wireless vehicle system, *Computer World*. May 1. Retrieved December 2, 2004 from *http://www.computerworld.com/networkingtopics/networking/story/0,21802,60113,00.html*

Chatterjee, A. (2003). Role of GPS navigation, fleet management and other location based services. Retrieved on December 11, 2003 from *http://www.gisdevelopment.net/technology/gps/techgp0045pf.htm*

Crisp, N. (2003). Open Location based services, an *Intelliware* report. *Integraph Copyright*. Retrieved on December 11, 2003 from *www.intelliware.com*

D'Roza, T. & Bilchev, G. (2003). An overview of location based services, *BT Technology Journal, 21*(1), 20-27.

Drane, C. & Rizos, C. (1998). *Positioning systems in intelligent transportation systems.* Boston: Artech House.

Durlacher (2001). UMTS Report - An Investment Perspective, *Durlacher Research,* London.

Elliott, G. & Phillips, G. (2004). *Mobile commerce and wireless computing systems.* England: Addison Wesley.

FCC (2003). Enhanced 911. Federal Communications Commission. Retrieved on December 11, 2003 from *http://www.fcc.gov/911/enhanced/*

Fraunholz, B., Hoffman, J. & Jung, J. (2003). Evaluation of mobile grameworks - Conceptual and technological aspects. *Proceedings of the10th European Conference on Information Technology Evaluation*, Instituto de Empresa, Madrid, Spain.

Gpsnorthamerica (2003). How GPS North America works for you. GPSNorthAmerica.com. Retrieved on December 11, 2003 from *http://www.gpsnorthamerica.com/how.htm?trackcode=bizcom*

Greenspan, R. (2002). Locating wireless revenue, value. *CyberAtlas Wireless Markets.* Retrieved on December 17, 2003 from *http://cyberatlas.internet.com/markets/wireless/article/0,,10094_1454791,00.html*

Hatfield (2002). A report on technical and operational issues impacting the provision of wireless enhanced 911 services. Prepared for *Federal Communications Commission*. Retrieved on December 11, 2003 from *http://www.fcc.gov/911/enhanced/reports*

Hills, P. & Blythe, P. (1994). Automatic Toll Collection for Pricing the Use of Road space - using microwave communications technology. In Catling, I. (ed.) *Advanced Technology for Road Transport* (119-144). Boston, London: Artech House.

Infoinsight (2002). What are location services? *Info Insight*. Retrieved on December 11, 2003 from *http://www.infoinsight.co.uk/etsi.htm*

Infomove (2003). Smart Telematics. *Infomove website*. Retrieved on December 11, 2003 from *http://www.infomove.com/solutions/SmartTelematics.asp*

ITU (2003). ICT free statistics. *International Telecommunication Union*. Retrieved on December 11, 2003 from *http://www.itu.int/ITU-D/ict/statistics/*

Janecke, J. (1999). Rechnergesteuerte Betriebsleitsysteme RBL im öffentlichen Personen-Nahverker. In H. Evers & G. Kasties (eds.), *Kompendium der Verkehrstelematik - Technologien,* Applikationen, Perspektiven, TÜV-Verlag, Köln ,Germany.

Kleiman, E. (2003). Combining wireless location services with enterprise ebusiness applications. Retrieved on December 11, 2003 from *http://www.gis development.net/technology/lbs/techlbs007pf.htm*

Lechner, W. & Baumann, S. (1999). Grundlagen der Verkehrstelematik. In Evers, H. & Kasties, G. (eds.), *Kompendium der Verkehrstelematik - Technologien, Applikationen, Perspektiven*, TÜV-Verlag, Köln,Germany.

Leite, F. & Pereira, J. (2001). Location based services and emergency communications in IMT-2000. *ITU News 7*. Retrieved on December 11, 2003 from *http://www.itu.int/itunews/issue/2001/07/mobility.html*

Levijoki, S. (2001). *Privacy vs location awareness.* Helinski University of Technology, Unpublished.

Loeb, L. (2001). *What's up with WEP*. Retrieved on October 12, 2001 from *http://www-106.ibm.com/developerworks/library/s-wep/*

Magon, A. & Shukla, R. (2003). LBS, the ingredients and the alternatives. Retrievd on December 11, 2003 from *http://www.gisdevelopment.net/technology/lbs.techlbs006pf.htm*

Millar, W. (2003). Location information from the cellular network – An overview. *BT Technology Journa*l, *21*(1), 98-104.

Mobilein (2003) Location based services. *Mobile in a Minute.* Retrieved on December 11, 2003 from *http://www.mobilein.com/location_based_services.htm*

Paavalainen, J. (2001). *Mobile Business Strategies.* London: Wireless Press, Addison-Wesley.

Pearce, D. (2001). Location enabled context and applications. *Proceedings of Mobile Location Services Workshop*, Rome, June 19-20. Retrieved on December 11, 2003 from *http://www.openmobilealliance.org/lif/presentations.htm*

Prasad, M. (2003). Location based services. Retrieved on December 11, 2003 from *http://www.gisdevelopment.net/technology/lbs/techlbs003pf.htm*

Prasad, A., Wang, H. & Choo, P. (2003). Network operator's security requirements on systems beyond 3G. *Proceedings of WWRF8*, Beijing, China, April.

Roth, J. (2002). *Mobile Computing*, dpunkt, Heidelberg, Germany.

Röttger-Gerigk, S. (2002). Lokalisierungsmethoden. In W. Gora & S. Röttger-Gerigk (eds.). *Handbuch Mobile-Commerce* (419-426). Springer, Berlin (Germany et al.).

Sadeh, N. (2002). *M-commerce: Technologies, services and business models.* New Yokr: Wiley.

Searby, S. (2003). Personalisation – An overview of its use and potential. *BT Technology Journal, 21*(1) 13-19.

Schiller (2004). *Mobile Communications*, 2nd ed. UK: Addison-Wesley.

Smith, B. (2000). France, Japan differ on location strategies. *Wireless Week*, June 26th.

Spinney, J. (2003). A brief history of LBS and how OpenLS fits into the new value chain. *Java Location Services Newsletter.* Retrieved on December 11, 2003 from *http://www.jlocationservices.com*

Steinfield, C. (2003). The development of location based services in mobile commerce. In Preissl, B., Bouwman, H. & Steinfield, C. (eds.), *Elife after the dot.com bust.* Berlin: Springer. Forthcoming. Retrieved on December 11, 2003 from *http://www.msu.edu/~steinfie/elifelbschap.pdf*

Turban, E., King, D., Lee, J., Warkentin, M. & Chung, H. M. (2002). *Electronic commerce – A managerial perspective*. New Jersey: Pearson Education International.

Van de Kar, E. & Bowman, H. (2001). The development of location based mobile services. *Proceedings of the Edispuut Conference*, Amsterdam, October 17.

VanderMeer, J. (2001). What's the difference between m-commerce and l-commerce? *Business Geographics*, March/April. Retrieved on December 2, 2004 from *http://www.geoplace.com/bg/2001/0401wire.asp*

Walke, B. (2000). *Mobilfunknetze und ihre Protokolle.* vol. 1, 2nd ed. Stuttgart, Germany: Teubner.

Whereonearth (2003). What are location based services? *Whereonearth website.* Retrieved on December 11, 2003 from *http://www.whereonearth.com/lbs*

WP1020A (2002). Mobile Resource Management. *An Intelliware Report*, April 11. Retrieved on December 11, 2003 from *http://www.intelliware.com*

Section IV

Technical Challenges

Chapter VII

Next Generation Cellular Network Planning: Transmission Issues and Proposals

Spiros Louvros, COSMOTE S.A., Greece

Athanassios C. Iossifides, COSMOTE S.A., Greece

Abstract

In this chapter, a multi-layer ATM architecture is proposed for the interconnection of current and future mobile communications nodes. Consisting of different ATM node types with respect to switching capability, the proposed architecture is adapted to current 2G and evolving 3G systems as well as future 4G wireless systems, as a common and shared backbone transmission network interconnecting core and access nodes between each other and Internet or PLMN/PSTN. Moreover, facing the

huge expansion of transmission interconnection network that will support current and future generation mobile communications, a modification of the standard ATM cell structure is introduced in order to efficiently support user mobility functional procedures. The proposed ATM architecture is integrated over a suitable, with respect to region and capacity, physical interface, consisting of SDH or SONET for wide area topologies, wireless links for outdoor areas and LED-POF combination for indoor areas. Being an interesting alternative over copper or traditional fiber, POF characteristics and performance issues are analyzed.

Introduction

Chapter Overview

3G and 4G mobile communication systems should provide the subscribers flexibility to multimedia services, including voice, constant or variable data rates and video, in conjunction with increased quality of service, high bandwidth reservation, and increased bit rate transmission. The proposed network plan, in the wireless part of next generation mobile networks, consists of a multi-layer architecture of macro, micro, and pico cells. Cell planning is adapted to the topographical background (hilly, mountainous, flat terrain) and the population-demographic basis (urban, suburban, agricultural areas).

In this chapter, a multi-layer ATM architecture is proposed for the interconnection of radio APs (e.g. Base Stations in 2G, Node B's in 3G glossary, Access Points in WLAN glossary) to the access network controller (e.g. BSC in 2G, RNC in 3G) and the core network. This architecture consists of different types of ATM nodes/switches regarding their capacity and switching capability with respect to the area and traffic load that they will serve. Existing ATM networks are designed to support wire-line users with fixed locations. Consequently, current ATM protocol does not support mobility functionalities like *location registration* and *handover* that are required to support users' mobility; these procedures are fully supported and directed by higher layers. Location registration is required to locate a user prior to information exchange (switching on/off mobile equipment, moving to different locations within the network, etc.). Handover is a function that supports mobility during information exchange, allowing users to move beyond the coverage area of a single cell without disturbing their communication. Since different radio access technologies may be supported by the proposed ATM networking topology and ATM nodes are going to be extended in number (together with the extension of subscribers and

services), it is desirable to engage some portion of the aforementioned procedures to the ATM protocol in order to ensure their proper functionality. Thus, a slight modification of the ATM cell is proposed in order to accommodate new information fields supporting the aforementioned mobility procedures in the standard ATM cell structure. Additionally, modifications of the cell structure are introduced for wireless ATM access from and towards the end user.

The physical layer is mainly over wire or optical fiber and ATM switches may be embedded in the Node B's and RNC (BSC)'s, avoiding the crucial problem of designing new interconnection network and thus minimizing the cost. However, in crowded areas, like city centers or indoor business-shopping centers, the number of cells is extremely increased. In order to achieve the desired level of quality of service and bit rate, the cost for wired or optical fiber infrastructure is prohibitively augmented. A more convenient way is to differentiate the physical layer according to specific geographical data. Thus, while keeping optical fibers for long distance interconnections, a wireless physical layer can be used for outdoor short distance interconnections. On the other hand, regarding indoor interconnections, a new, cheap and reliable optical fiber network is investigated and proposed. Semiconductor lasers have been proved to be the most promising devices in nowadays-optical communication networks. Light emitting diodes (LED) are their counterparts, providing moderate efficiency in bandwidth and lower bit rates compared to semiconductor lasers. Considering though the cost efficiency in integrated multi-optical networks (UMTS, B-ISDN in-building internet links, etc.), LEDs consist a strong candidate for indoor backbone. Moreover, the use of Plastic Optical Fibers (POF) in recent years, mainly for short distance optical networks, increased the interest in LED as a transmission device. An optical network designed for short distance in-building optical links, using LED as transmitter, POF as the transmission medium, is proposed.

Chapter Outline

This chapter proposes new transmission solutions for the interconnections of the separate parts of next generation mobile and wireless cellular networks. It is separated into different sections in order to explain in a more convenient way all the necessary technical and planning issues.

In the first section, the technical background information of mobile cellular networks and their evolution up to day are presented. The ATM philosophy also is introduced and the idea of using optical fibers is mentioned. Additionally, POF technology is introduced.

In the next section, the proposed ATM architecture suitable for mobile applications is examined in detail. Existing ATM protocols do not support mobility of

subscribers, a very important characteristic of the mobile and wireless cellular networks. Necessary corrections are proposed and a modified ATM cell structure is examined to support all the mobility messages exchanged during a connection.

The following section discusses optical fibers and presents a special optical fiber, POF. Details about POF's response, materials and applications are given and a discussion about optical links takes part. Finally, the appropriate characteristics of POF for short distance applications are mentioned.

In the next section, the proposed interconnection architecture of future mobile and wireless cellular networks is presented. Interconnections of separate parts of the wireless network are achieved through ATM switches following a multi-layer architecture adapted to cell planning multi-layer architecture. The multi-layer architecture provides radio coverage that consists of two separate scenarios: indoor and outdoor coverage.

Conclusions are drawn in the final section, emphasizing the future trends of mobile/wireless ATM networks and optical POF fibers. In the Appendix, a detailed technical presentation of a derived channel model of POF is examined. Although quite technical for a manager, it is however important for engineers and readers interested in technical details regarding POF's response.

Technical Background Information

Digital Cellular Networks Overview (GSM-GPRS-UMTS-nGeneration)

In 1991 European Telecommunication and Standardization Institute (ETSI) accepted the standards for a new upcoming mobile, fully digital and cellular communication network, GSM. It was the first Pan-European mobile telephone network standard that replaced all the existing analogue ones. Some of the advantages of GSM are summarized:

- Increased capacity of subscribers
- Improved communication quality due to digitized speech and channel coding
- Improved quality of service
- Roaming possibility

- Compatibility with PSTN networks
- Security policies referring to the subscriber conversation
- New services
- Adaptation to variable traffic load

GSM architecture is developed over the well-known technologies of PSTN switches. It consists of two main branches, the fixed network and the mobile-radio environment part. Fixed network part is the one responsible for the interconnection of the cellular network with the existing PSTN and PLMN ones. Since compatibility is demanded, the switches of GSM network are based on the mostly used technology of PSTN networks, SPC switches. SPC (Stored Program Control) is a fully computerized switch, able to perform traffic management, routing, billing and network Operation and Maintenance (O&M) tasks. Hence, subscribers could take advantage of the interconnectivity of GSM with PSTN, and communicate with other PSTN subscribers all over the world.

The mobile-radio part is based on two innovative ideas, cellular coverage, and frequency reuse. Within cellular coverage the mobility of subscriber is guaranteed and the adaptation to variable traffic load is feasible. With frequency reuse on the other hand, the coverage of unlimited geographical area is possible with the constraint of limited bandwidth in the air interface. Combination of cellular coverage and frequency reuse contributes to the capability of GSM network to be flexible with variable traffic load.

MSC is the heart of the GSM system. It is an SPC switch interconnecting GSM network with all the other PSTN, PLMN and data networks (through MGW in new 3GPP GSM standards), and also is responsible for the traffic management, routing and billing. In newer standards of 3GPP, MSC role is restricted in mobility and call management. Real CS switching takes part in MGW through proper signaling from the MSC Server. BSC is another SPC switch responsible for interconnecting the mobile-radio part with the rest of the network and sustaining all the necessary operations of radio part (handover, location updating, air interface signaling, etc.). BTS (Base Station) is a radio switch responsible for the radio coverage.

During the last two decades of 20th century, PSTN networks have been rapidly developed. ISDN networks have been presented as the more compact solution to voice-data communication. ISDN technology offers the subscriber the possibility to use the telephone network not only for voice transmission but also for data applications as teleconference, calling line identification number presentation or restriction, video transmission, and video telephony. These innovations have seriously affected GSM network. GSM architecture does not support ISDN applications due to poor bandwidth and spectrum restrictions in the air

interface. As a consequence the development of PCN networks seemed inevitable. PCN networks are structured on a hierarchical functional platform. Based on a microcellular and picocellular model, the demands for these enhanced cellular networks were:

- Extension of cellular services
- Access to services, independently of the architecture of the telecommunication platform

DCS1800 is the second-generation cellular mobile network enhancing GSM functionality towards the aforementioned PCN demands. Its main network architecture is based on GSM for compatibility reasons. Bandwidth is increased, new frequencies are allocated in the microwave band, quality of service is increased, more BTS's are needed to cover the same area as GSM and more subscribers (traffic load) are served. Multimedia services and internet access are not fully provided yet on mobile handsets due to limitations in bandwidth of the air interface and absence of corresponding protocols in GSM platform. However, during the last few years the development of special protocols (WAP) has permitted some primitive applications using internet.

B-ISDN networks are the "state of the art" technology in nowadays wired telecommunication links. The main feature of B-ISDN concept is the support of a wide range of voice and non-voice applications in the same network. B-ISDN networks extend the concept of PSTN networks by incorporating additional functions and features of current circuit and packet switching networks for data, to provide both existing and new services integrated. Bit rates of conventional PSTN networks (64 Kb/s per subscriber) have been realized to be insufficient for integration over one network of voice, image, video, and data applications. Hence B-ISDN networks can support up to 622 Mb/s. Though, such rates are still far beyond of the individual end-subscriber needs. In such cases HDSL and ADSL technology has lately dominated for over 2Mb/s bit rates exploitable by the end home user.

Mobile communication networks have to follow the evolution of fixed networks in order to provide moving subscriber with all the services and applications of fixed subscribers. The dream of telecommunications engineers was a mobile/wireless network with capability of services equal to fixed B-ISDN networks. This however is unfeasible due to restrictions and limitations imposed by the hostile radio channel. The ideal mobile network would be able to provide moving subscribers continuous access to every possible voice or data networks, leading to the realization of "mobile office". The result of this effort (although somewhat restrictive in terms of realizable bit rates), was another evolution in mobile

networks, the GPRS and the EDGE network (usually referred as 2.5G) with rates of up to 115Kb/s and 384Kb/s, respectively, when fully exploited.

Universal Mobile Telecommunications System (UMTS) is the realization of a new generation of telecommunications technology for a world in which personal services will be based on a combination of fixed and mobile services to form a seamless end-to-end service for the subscriber. Its realization at least requires:

- Provision of a unified presentation of services to the end user.
- Mobile technology that supports a very broad mix of communication services and applications.
- On demand flexible bandwidth allocation reaching 2Mb/s per subscriber.
- Exploitation of pure (not tunneled) IP interconnection of network elements between each other for data exchange and O&M purposes.
- Provision of flexible end-to-end all-IP connectivity in terms of user information.

Exploiting full capability of UMTS, as the integral mobile access part of B-ISDN, mobile networks will begin to offer pure Packet Switched (PS) services that have been traditionally provided by fixed networks. Generally speaking, UMTS follows the demand, posed by moving subscribers, of upgrading the existing mobile cellular networks (GSM, DCS1800, GPRS) in non-homogeneous environments. The integration of UMTS refers not only to services but also to latform and protocols. 3G radio access controllers (RNC) are already based on ATM switching rather than SPC technology.

3.5G and 4G systems are already under investigation. Aiming to "context-aware personalized ubiquitous multimedia services" (Houssos, Alonistioti, Merakos, Mohyeldin, Dillinger, Fahrmair & Schoenmakers, 2003, p. 52), 3.5G systems promise rates of up to 10Mb/s (3GPP Release 5), while with the use of greater bandwidth these rates may raise even more in 4G (Esmailzadeh, Nakagawa & Jones, 2003). On the other hand, in the last five years, a standardization effort has started for the evolution of WLAN's in order to support higher bit rates in hotspots or business and factory environments with cell radius of the order of 100m. For example, IEEE 802.11 variants face rates of up to 11Mb/s (802.11b), 54Mb/s (802.11a/g) while rates in excess of 100Mb/s have already been referred (Simoens, Pellati, Gosteau, Gosse & Ware, 2003). European HIPERLAN/2 supports somewhat lower rates up to day but with greater cell coverage and enhanced MAC protocols. In any case, 4G and WLAN's technology are going to be based on an IP backbone between APs and access controllers or routers and the internet. Mobile IPv4 and IPv6 are already under investigation (Lach,

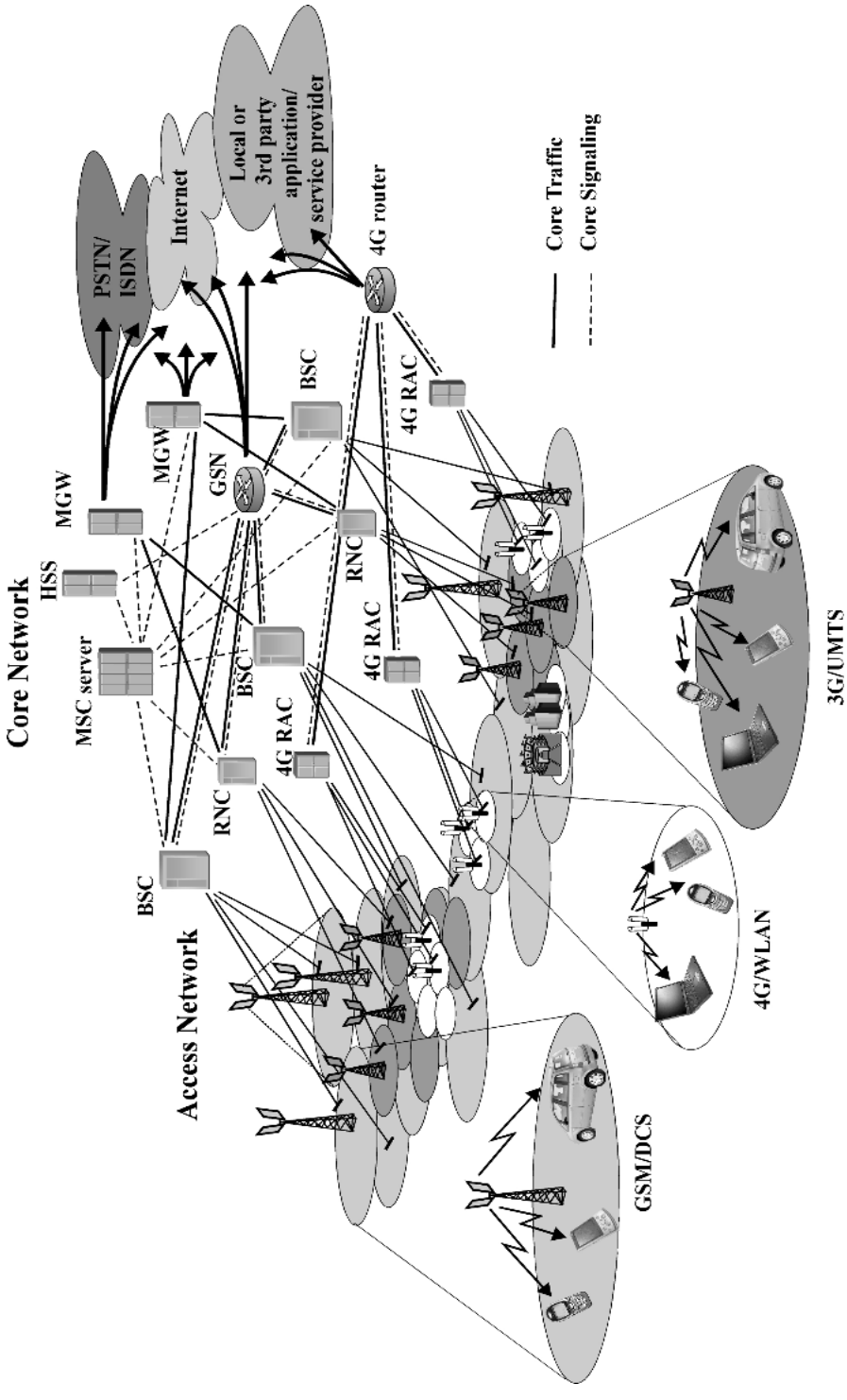


Figure 1. Future mobile network architecture with different technologies (2G/3G/4G)

Janneteau & Petrescu, 2003) to provide user mobility support for context-type services.

An example of future mobile network architecture, engaging almost all forms of aforementioned technologies, is presented in Figure 1. All the technologies will coexist in the next decade and smooth transfer of end user services and information rate among them has to be considered carefully (3GPP is already under a standardization process of interoperability between UMTS and WLAN). In any case, and irrespectively of the radio interface or MAC protocol towards the end user, interconnection between the nodes (elements) of each technology can be based on ATM, following a multi-layer architecture according to the area and capacity needs of the environment to be covered.

ATM Overview

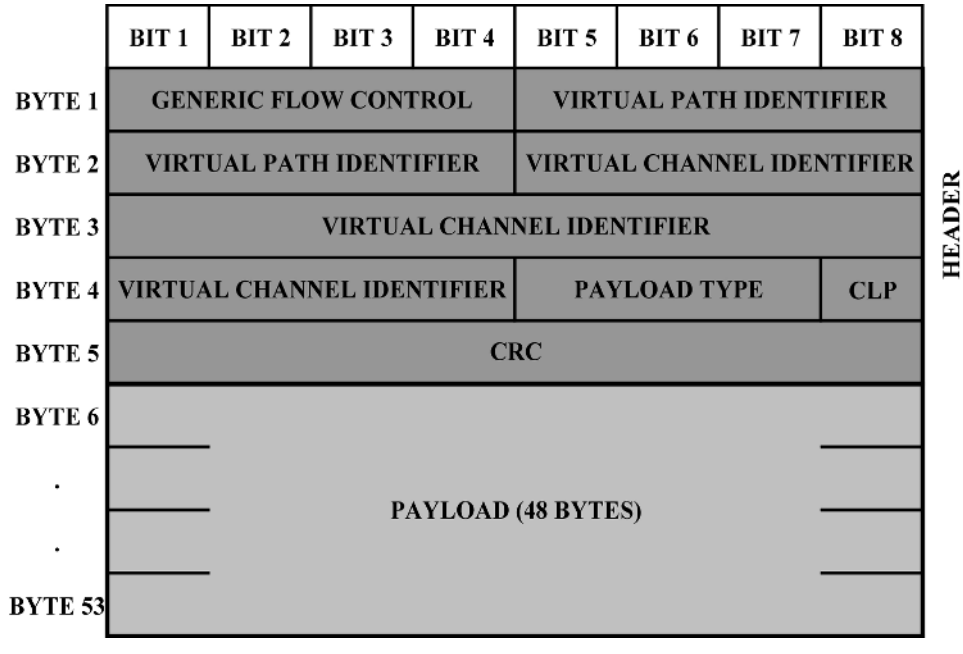
The possible candidate (among applied fixed networks) considered for the interconnection implementation of current and future mobile and wireless networks, is B-ISDN or IP, with the last one being the most preferable due to its huge and mature use in internet applications. The underlying technology that make, for example, IP possible nowadays in UMTS, is ATM. Within the last decade, the world of telecommunications has started to change. Data traffic, where the information is transmitted in form of packets and the flow of information is bursty rather than isochronous (even voice has taken the road to packetization, e.g., VoIP). An additional contributing factor to the evolution of telecommunications is the almost unlimited bandwidth provided by modern fiber optical transmission equipment. ATM technology is proposed by the telecommunications industry to accommodate multiple traffic types in a very high speed wireline network. Due to fixed packet size and very fast packet switching, ATM meets very strict timing and delay requirements. This makes the transmission of time-sensitive traffic, such as voice, through the ATM network possible. Since ATM is based on packet switching, it also accommodates data traffic. ATM networks are designed to support multiple traffic types with different priorities and quality of service requirements. They are first expected to be deployed in the backbone networks and then progress to the edge (air interface) of current telecommunications networks.

Basic idea behind ATM is to transmit all information in small, fixed size packets called *ATM cells* over all transmission channels (wired or wireless). Having fixed-size packets of information for transmission, it can emulate the circuit switching technique of traditional telephony networks and on the same time take advantage of the best utilization of transmission lines bandwidth. Hence, it

operates asynchronously and it can switch continuously information from/to different networks (voice, video, data) with variable bit rates. The responsible nodes for asynchronous operation are called ATM switches. They consist of interfaces in order to communicate with various inhomogeneous networks as LANs, WANs, etc. All these networks transmit information in different bit rates and ATM switches (through the ATM layer of B-ISDN or IP protocol hierarchy) divide this inhomogeneous information (using special ATM Adaptation Layers in terms of OSI layer structure) into fixed size packets of 48 bytes to accommodate them into the ATM cells. The output of the ATM switch is therefore constant bit rate information exploiting fully the physical layer bandwidth that can be shared among different services or users through proper labeling (addressing) of the ATM cells. Labeling is incorporated into the 5-bytes long ATM cell header (VPI – Vitrual Path Identifier and VCI – Virtual Channel Identifier) that also contains fields for signaling purposes (GFC – Generic Flow Control, Payload Type and CLP – Cell Loss Priority) and header error detection (CRC). In Figure 2 the structure of a standard ATM cell is presented.

ATM cells are usually transmitted in the physical layer of a telecommunication network either wireless or through optical lines. For mobile and wireless applications, special care has to be taken in order to protect cell contents from

Figure 2. Standard ATM cell structure



corruptions and to incorporate the handover and location update information of a moving subscriber in the header field of the cell.

Plastic Optical Fiber (POF) overview

Nowadays cabling based on symmetrical copper cables is dominant in LAN applications; glass fibers predominate in long distance networks. Whereas just a few years ago l0Mbit/s Ethernet (l0BaseT) had the main share of interfaces in star or tree structures, today's pure star networks are predominantly set up on the basis of 100 Mbit/s connections. The basis of modern LAN topologies are the standards for structured cabling, for example, IN 11801.

With structured cabling, the LAN is divided into different segments for which there are corresponding recommendations. Within buildings, vertical cabling, such as between cellar and upper floors, and horizontal cabling on the individual floors are separated. Various categories are established and standardized regarding the quality of copper cables adequate to support different bit rates and distances.

Data networks in office buildings are planned and set up very carefully. The use of shielded cables rather than unshielded cables dominates mostly in Europe – in contrast to the U.S. Paying careful attention to a unified ground potential throughout the entire building allows optimal use of the advantages of shielded cables. Consequently, electromagnetic disturbances do not constitute a major problem in data networks, at least when properly installed. Data cables in office buildings are usually laid on grids below the respective floor ceilings.

Plastic optical fiber (POF) is a promising candidate for optical cabling infrastructure due to its low price, large cross section area, easy connectorization/coupling with optical source and simple use. For the proposed optical links a PMMA Graded Index Plastic Optical Fiber (PMMA GIPOF) has been chosen. PMMA material is the best plastic material for this application due to low attenuation, bending flexibility and optical window at 520 nm and 650 nm. This is convenient for the red RC-LED with center wavelength of 650 nm and bandwidth of 5 nm.

The following arguments could be used for the use of polymer optical fibers in LAN applications:

- less space required for the cables
- lower susceptibility to disturbances
- galvanic isolation of the components

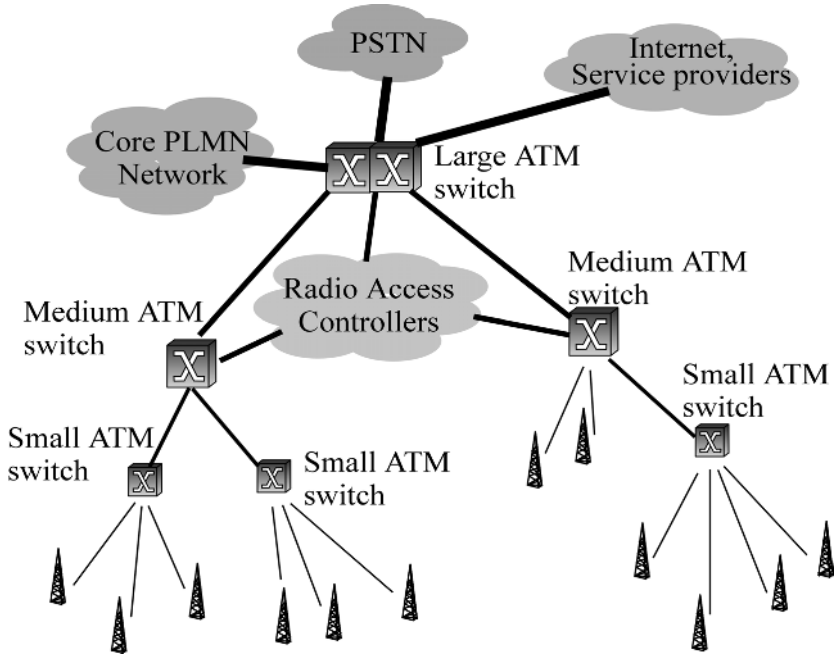
Connecting electronic devices to the electric circuit and through data networks with copper cables, always produces loops which can act as antennas or even create undesired current paths. In commercial use these problems should always be taken into consideration. Above all, the problem of induction, for example, caused by lightning striking, has to be solved by means of appropriate protective grounding. In such a case POF would be an interesting alternative, which could surely be used in special applications. Practical and proven solutions do exist for copper cables, too.

ATM Proposal for Mobile and Wireless Applications

Introduction to Multi-Layer Architecture

Mobile and wireless telecommunications networks have broken the tether in the wireline networks and allow users to be mobile and still maintain connectivity to

Figure 3. Proposed multi-layer interconnection architecture



their offices, homes, and so on. An ATM backbone for mobile and wireless applications consists a natural extension to the development of ATM based wireline networks by providing full support for multiple traffic types including voice, data, and multimedia traffic in a mobile/wireless environment. The proposed interconnection system architecture is multi-layered in structure and consists of three basic ATM node types with decreasing switching capability: Large ATM switch, Medium ATM switch, and Small ATM switch or ATM Multiplexer. Their exact capability can be defined based on the telecommunication traffic load of the access or core network part that they are going to support. Each ATM switch can accept and integrate traffic from different types of radio access technology and forward it to the proper Radio Access Controller (RAC) or to the core network (PLMN or PSTN/ISDN) and the internet through identical physical layer cabling, as shown in Figure 3.

Supporting mobile/wireless users in an ATM network presents two sets of challenges to the existing ATM protocol. The first set includes problems that arise due to the mobility of the wireless users. The second is extending ATM usage to the interface between the radio AP (e.g. base station) and the end user.

Challenges Related to the Mobility of Wireless Users

The ATM standards proposed by the International Telecommunications Union (ITU) are designed to support wireline users at fixed locations. Current ATM standards do not include any provisions for support of location update and registration transactions that are required by mobile users, and also do not support handover functions.

Location information for mobile users of 2G and 3G systems is usually stored in a database structure (HSS) that is distributed across the network. This database is updated by registration transactions that occur as users move within the mobile network. If a mobile user moves while he is communicating with another user or a server in the network, the network may need to transfer the radio link of the user between radio APs in order to provide seamless connectivity to the user (handover process). These procedures are basically maintained by upper layer signaling. Though, the expansion of 2G and 3G systems, the increase of mobile and wireless subscribers and the integration of different radio access technologies (2G, 3G, 4G, WLAN, etc.) to the handsets, lead in extreme growth of the interconnection (backbone) traffic and the necessary ATM nodes. The handover mechanism that is nowadays supported by signaling messages exchange between the subscriber's terminal and the radio access controller is transferred transparently through ATM switches. The growth of the network, the high cost of transmission lines and the different priorities of future supported services will result in longer traveling (through ATM backbone switches) and processing time

of such messages, thus facing greater loss percentages and subscribers displeasure. So, it is important, in order to maintain the higher priority of handover mechanisms over other procedures (e.g. on-going calls or contexts, new call attempts, etc.), to declare their priority in the interconnection layer and the ATM protocol. In this way, ATM cells carrying handover information will travel faster to their destination, a fact rather crucial for circuit switched or time sensitive services.

In this context, a modification of the standard ATM header is proposed, introducing two new identifiers: the Handover Identifier (HOI) to be used for supporting handover mechanisms and prioritize them and the Location Update Identifier (LUI) for supporting location registration/update in a similar manner. It should be, additionally, mentioned that such identifiers, with a proper defined protocol frame structure, can be used in WLAN's in order to provide seamless handover which is not by now fully supported. Since the ATM header is restricted in length, in order to provide space for the new identifiers, VPI and VCI can be shortened. When ATM switches will be distributed to end of the network (reaching radio APs), the need for large amount of identification numbers is decreasing, since identification can take part from layer to layer, that is, a small ATM switch will use an identification range according to the number of ATM multiplexers it supports, an ATM multiplexer will use an identification range according to the number of radio APs it supports, and so on.

The transmission rates that large ATM switches can support are nowadays up to 155 Mb/s. The bandwidth of information is already too large to be transmitted through copper cables. Moreover, ATM protocol is not supporting error control coding techniques or packet retransmission protocols since it is supposed to transmit data reliably and higher layers are assigned this task. Hence, the use of fiber optics is compulsory, and since we are interested in a wide area of data and transmission rates, the use of SDH or SONET networks is recommended, at least to the level of ATM multiplexer.

Wireless ATM

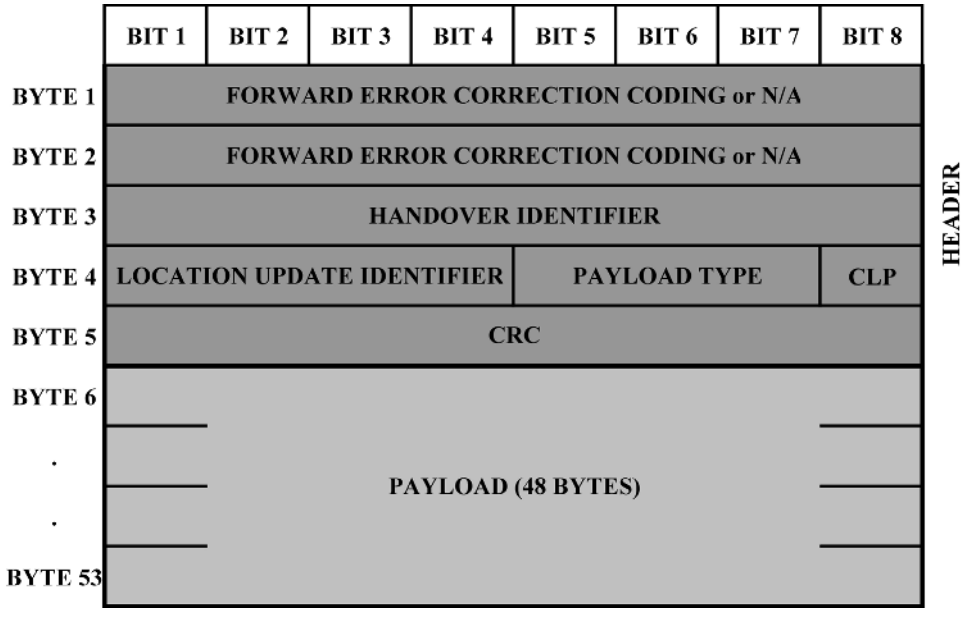
A key benefit of a wireless network is providing tetherless access to the subscribers. The most common method for providing tetherless access to a network is through the use of radio frequencies. A major problem that needs to be addressed when using ATM in the air interface between the radio AP and the subscriber's terminal is the error performance of the radio link. ATM networks are designed to utilize highly reliable fiber optical or very reliable copper-based physical media. These physical links utilize digital transmission techniques where information is encoded into bits. ATM does not include error correction or checking for the user information portion of an ATM packet. Compared to the wireline networks,

wireless networks may achieve an average bit error rates on the order of 10^{-3} to 10^{6}. In order to support ATM traffic in a wireless ATM network, the quality of radio links needs to be improved through the use of equalization, diversity and error correction and detection to a level that is closer to the wireline networks. There are a number of solutions that combine these techniques to improve the error performance of wireless networks.

A wireless ATM network needs to support multiple traffic types with different priorities and quality of service guarantees. In contrast to the fiber optical media used in wireline networks, radio bandwidth is a very precious resource for the wireless ATM network. A Medium Access Control (MAC) protocol that supports multiple users, arranges multiple connections per user and service priorities with quality of service requirements must be developed in order to maintain full compatibility with the existing ATM protocols. This medium access control protocol needs to make maximum use of the shared radio resource and needs to achieve full utilization of the radio frequencies in a variety of environments.

In wireless ATM, cells do not need VCI or VPI information in header since many subscribers must have access to the channel of one transceiver. The idea of VCI or VPI has no sense in the air interface; what has sense is a MAC protocol over a multiple access technique of the physical layer. The proposed radio ATM cell is presented in Figure 4. In order to keep compatibility with the ATM technology and B-ISDN protocol layer hierarchy, the radio ATM-cell has length of 53 bytes

Figure 4. Proposed cell structure for wireless ATM



with empty (not assigned, N/A) the fields of VCI and VPI. Instead of them a Forward Error Correction (FEC) scheme could be used, with at least two (out of 53) redundancy bytes. Although not enough for high performance, such a coding scheme can correct one byte in error providing an extra countermeasure to the hostile radio environment.

POF Proposal for Short Distance Networks

Use of POF in Short Distance Networks

Today's in-building regions are mostly equipped with three different cable-based networks: the telephone network, the connection to the broadband cable network or an antenna system and the 230 V electrical power supply. Each of these networks is adapted to its own specific, albeit very different purpose. Only the electrical power supply effectively connects all regions. The telephone and broadband networks do in fact provide a connection to the access network, but not the possibility of cross-linking different terminal devices within an in-building structure.

The list of possible devices requiring cross-linking could be expanded at will. Surveillance and control systems, for example, for heat, windows and doors, have increasingly gained in importance. The tenant is thus confronted with the problem of establishing data connections between devices with the lowest possible expenditure of time and money. The possibilities for completely overcoming such a situation without installing cables is to use PowerLine technology or to set up a radio system. Both options are technically advanced and thoroughly affordable. However, the possible bit rates and the attainable quality are subject to definite limitations. Cable-based systems are preferable when transmitting high-quality moving pictures in real time or with a broadband connection of computers, for example, when working at home. Different copper cables as well as optic fibers can be considered. Regarding the simplicity of installation, radio systems cannot be surpassed. Among the cable-based systems, POF is distinguished as having the easiest cable setup and the most reasonably priced connection technology.

Besides the question of transmission media, a point of great interest is the interface to the consumer. A system can only gain general acceptance when terminal devices are equipped with appropriate connectors, the services desired can be supported with sufficient quality and the components for setting up the

Table 1. Interfaces supported by POF

Interface	Bit Rates	Advantages/Disadvantages
ATM	25 Mb/s, 155 Mb/s, 622 Mb/s, 2.5 Gb/s	supports high-quality services and is already employed in long-distance networks, up until now too expensive for home use
Ethernet	10 Mb/s, 100 Mb/s, 1 Gb/s	used above all for IP applications, wide-spread and good value, dominant in LAN field difficult with video transmission
USB	12 Mb/s (new 480 Mb/s)	wide-spread standard for PCs very simple operation requires running PC up until now data rates too low
IEEE 1394	100 Mb/s, 200 Mb/s, 400 Mb/s, 800 Mb/s, up to 3.2 Gb/s planned	universal system for all applications (incl. video) multi master network with extremely easy operation

network are available at reasonable prices. In Table 1 there is a list of some of the interesting interfaces.

POF systems have already been created for all four interfaces mentioned. The ATM forum has already specified the use of PMMA POF for 155 Mbit/s. Of particular interest is the inclusion of POF in the IEEE 1394 specification (up until now 100 Mbit/s and 200 Mbit/s over 50 m; 400 Mbit/s over 100 m is in preparation). This interface could gain acceptance not only with computers, but also in diverse multimedia devices such as game consoles, cameras and video cameras, televisions and DVD players and with computer peripherals. IEEE 1394 standard is intentionally not fixed to a medium, but provides the user with the option of selecting his own cable. Therein lies great application potential especially for POF as illustrated in the overview.

Study of the Proposed POF Link

An optical transmission system essentially consists of three components: the transmitter that converts the electrical signal into an optical one to be fitted into the optical transmission channel; the transmission channel, which might contain further active or passive components and guides the optical signal towards the receiver; the receiver side where the optical signal is converted back into an electrical one that is available for further processing with well-known tech-

niques. The goal of an optical transmission system is to transfer effectively the maximum possible information, with the less possible distortion, to the receiver. This is never the case however, since the optical channel inserts attenuation and phase distortion and the transmitter-receiver contains several internal imperfections and mechanisms of errors. Moreover, the available transmitted bit rate is bounded to a certain maximum value due to two main reasons: the specific physical principles that govern the electrical-to-optical transformation of signals at the transmitter (recombination time, imperfections, non-radiative recombinations, interband recombinations) and the specific ways the optical signals are transmitted through the channel (modes, differential mode attenuation, differential mode delay, source bandwidth, chromatic dispersion and time dispersion). As a consequence, it is important to study separately the different components of optical transmission systems in order to have a deep understanding of the imposed limitations.

Nowadays, the possible transmitter elements that are used extensively in optical communications are Semiconductor Lasers (SL) and LED's. The reasons are the very small size of construction (considerably smaller than 1 mm^3), very fast switching times, high efficiency, great number of available transmitted wavelengths, limited bandwidth, small radiation angle. Semiconductor lasers have a series of advantages compared to LEDs. Because of the stimulated emission involved, the external efficiency is considerably higher, the modulation speed is higher due to smaller recombination time of carriers, the radiation angle is smaller and the spectral efficiency is considerably higher (lower emitted bandwidth). These advantages usually make SL the most preferable optical source for transmitter. The most common optical guide for a non-integrated optical network is the optical fiber. Usually optical fibers made of silica are used because of the well-known construction processes, the low attenuation in a wide range of wavelengths, and the ability to support single mode transmission. Within the vast development of local area data networks and ISDN networks, optical fibers made of silica are the most reliable transmission media for extremely high bitrates over long distance transmission networks. This is not the case though for short distance transmission networks due to sensitivity of bending and high cost of installation and purchase. POF's have been recently proposed to replace silica optical fibers in several applications including short distance telecommunication networks. The core and the cladding are constructed from plastic substances as polystyrene (PS), polycarbonate (PC), and lately polymethylmetacrylate (PMMA). The easiest way to construct POF is Step-Index POF but lately construction processes have been improved to enable the production of PMMA Graded-Index POF as well; hence reducing the time dispersion and improving the bandwidth-length product. Optical communication links, based on POF are under investigation nowadays. The behavior of POF differs from usual silicon fibers because the attenuation is extremely high (for PMMA Graded-Index POF

at 660 nm is almost 230 dB/Km), the Differential Mode Attenuation (DMA) differs from silica fibers and the Differential Mode Delay (DMD) contributes to the increase of dispersion of optical pulses. These effects restrict the applications of POF transmission links in very short distance optical links, covering in-house applications to business-on-floor indoor network architectures. Additionally, the core-to-cladding diameter ratio is extremely high compared to silica fibers (in ¼m is 980/1000) and the numerical aperture is high enough (approximately 0.5). The benefit of large core to cladding ratio in POF enables application of LED as the transmitter, reducing the cost per link and allowing easy coupling of optical radiation in POF. A detailed presentation of POF's response exists in the Appendix, where a channel model, according to the above stated principles, is described.

Applications to Next Generation Cellular Systems

Network Multi-Layer Architecture

Future mobile and wireless communication networks should provide to the subscribers flexibility to multimedia services, including voice, constant or variable bit rate data, video, increased quality of service, high bandwidth reservation, increased bit rate transmission, and compatibility with B-ISDN and IP networks. Application of the above requirements in a wireless digital channel is more difficult than fixed broadband networks due to physical restrictions of the wireless channel. Nevertheless, it is important to built such a network and provide qualitative services to subscribers even though a quantitative equation of services in bit rate is impossible. The proposed network architecture in the wireless part of next generation cellular networks consists basically of micro and pico cells. Their interconnection to the main network is based on the multi-layer approach previously presented with the use of ATM switches. Figure 5 presents the proposed multi-layer interconnection architecture applied to a future mobile/wireless network. All the switches are ATM switches. The use of ATM switches for the interconnection of the radio APs to the core network avoids the crucial problem of designing new interconnection transmission links and minimizes the cost for UMTS application for which ATM is already integrated. Adaptation of GSM/DCS to ATM could be based on ATM terminal equipment (such as single plug-in cards of BTS's and BSC's or low price ATM converters). As illustrated in Figures 3 and 5, Large ATM switches are used as gateways to the core network and between the core network and PSTN or other PLMN or

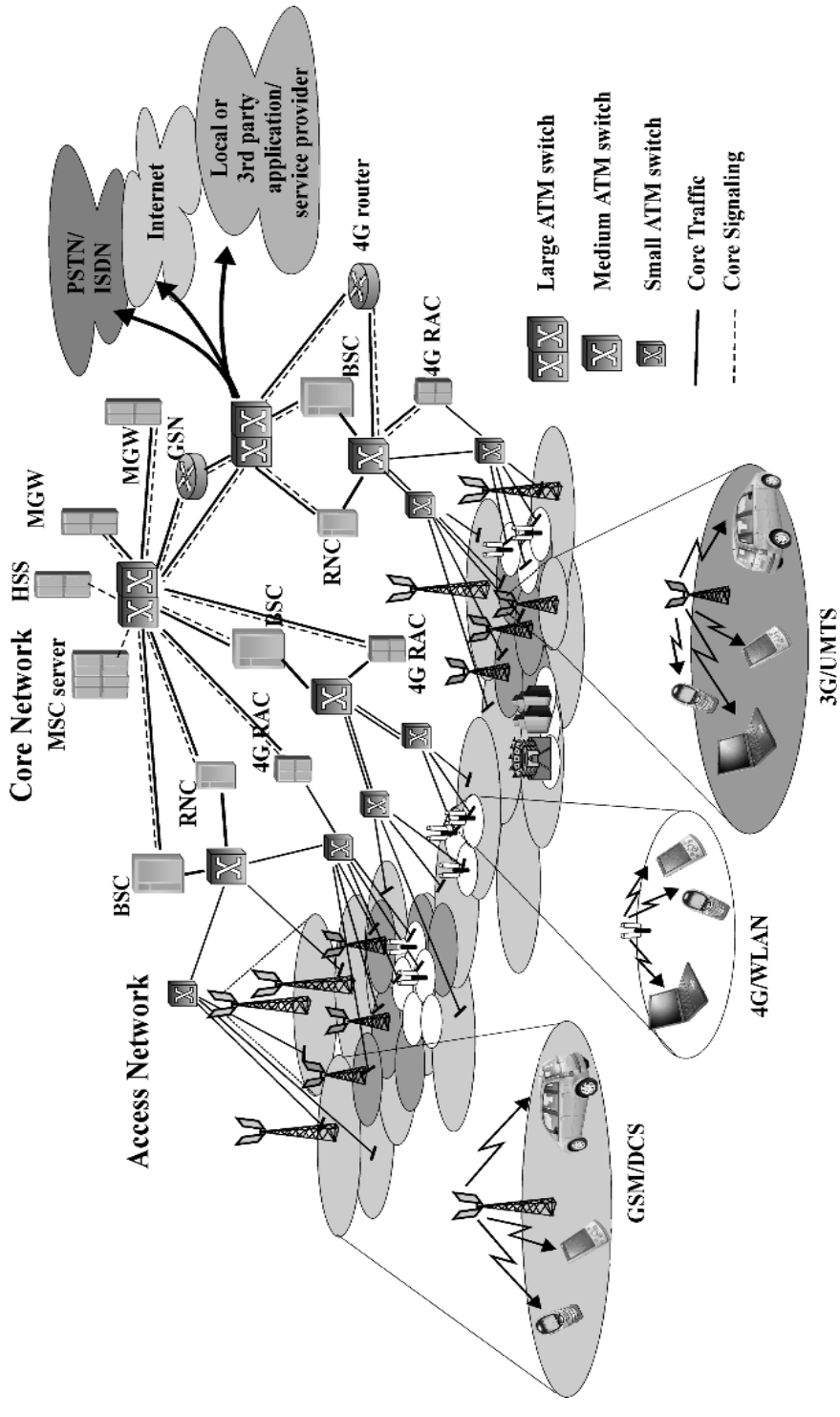


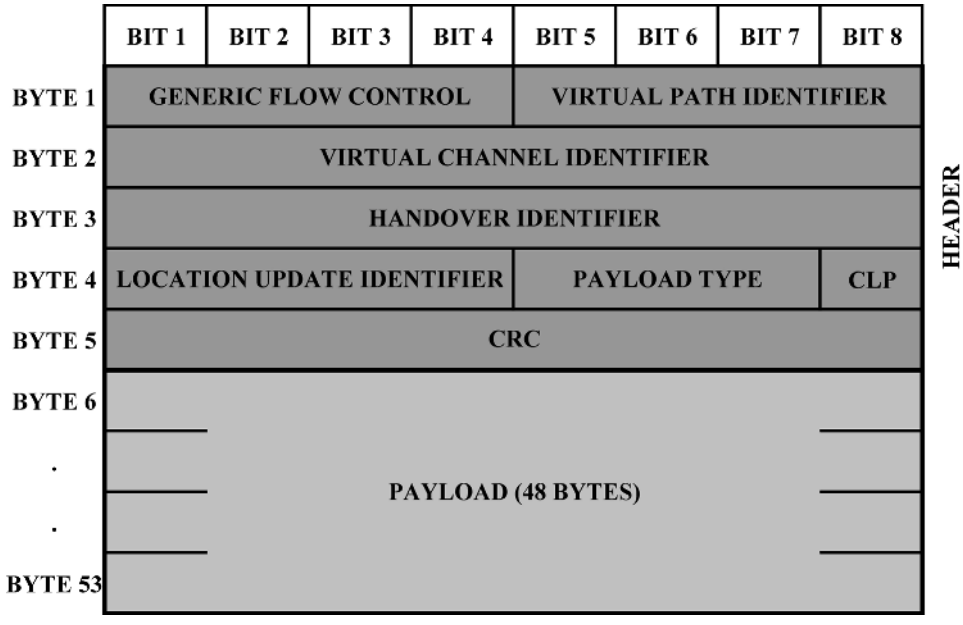
Figure 5. Future mobile network architecture with different technologies (2G/3G/4G) engaging a multi-layer ATM interconnection architecture

internet. However in geographical areas where traffic load is less, as in the coverage area of radio access controllers, the use of Medium ATM switches is preferred for cost reduction. In such a case Medium ATM switches are similar to the role of BSC (RNC) in GSM (UMTS) network, from a transmission node point of view, with the capability of supporting all different radio access technologies. They would be responsible for the radio part and the header of the ATM-cell will be slightly modified in order to accommodate the cellular functionalities of handover and location updating as mentioned before. In geographical areas of intensive traffic load, as a city, the concentration of cells per km will be large. Especially in train stations, airports or inside company buildings the coverage pico cells will be of the order of 50-200 m. Hence, in the center of a city with a lot of buildings the cell concentration might be 20 pico cells per km. The use of Medium ATM switches in such a case might be 10 switches per km, which is extremely expensive. The solution is the use of Small ATM switches or multiplexers for massive wireless links transmission rate and concentration of multiple radio AP traffic towards the core network (Medium ATM switch). The cost is decreased and the flexibility in design and planning features is increased.

Contribution of Modified ATM Cell to Future Cellular Network Architecture

Medium ATM switches are connected straight or through ATM multiplexers towards the radio APs (BTS's, NodeB's, etc.) by modified ATM cells in order to include HOI and LUI. These ATM contain 48 bytes of payload but the five bytes of the header are modified as shown in Figure 6. It is obvious that the field of VCI and VPI are compressed from three bytes to two bytes. The usage of 12 bits for VCI is enough to encode approximately 8000 virtual channels that can exist in a VP. This is not enough for the 65000 different VC in one VP for regular ATM cells, but keep in mind that the ATM switch is of medium or small switching capability. The usage of four bits for VPI is enough for the switch to accommodate 16 different virtual paths. The HOI is an eight-bit field to inform the system for the priority of 256 different handover requests simultaneously. The field of Location Updating Identifier (LUI) is accommodated only in four bits. LUI is smaller than HOI since in a pico or micro cellular system, handover attempts are more frequent than location updating attempts. Hence 16 different location update attempts are enough for the radio access network supported by the Medium ATM switch. The same ATM cell format can be used at Small ATM switches (multiplexers) towards the radio APs as well.

Figure 6. Proposed ATM cell structure for use at medium and small ATM switches



Transmission Solutions for the Multi-Layer Architecture

According to the proposed ATM solution, the radio APs are interconnected to the Small or Medium ATM switches through wired or wireless physical layer with wired and wireless means. In geographical areas where multiple radio APs are indoors (like shopping centers, train stations, or in-building coverage), the optimum transmission network could be an optical one with POF media and a general architecture according to previous section. The characteristics of the plastic materials provide flexibility to the indoor area architecture and also perfect response for short distance links. The number of base stations provides transmission interconnection links that are cost effective with high bit rates. ATM modified cell of Figure 6 is transmitted through the physical layer

In geographical areas of outdoor coverage (city centers or crowded areas) the wireless ATM solution through a certain radio link would be preferable. Regarding the interconnection of Medium ATM switches with Large ATM switches or between each other, SDH or SONET solutions will be preferable because of high capacity and enhanced error protection.

Conclusion

Node interconnection architecture of future mobile and wireless networks is considered to be crucial in the near future because of high capacity demand, high speed switching and high cost of investment. Generally speaking, UMTS, WLAN's, and so on are a revolution in telecommunications and telephony. The use of ATM switches and packet switched telephony is already a mature idea and telecommunication engineers are quite familiar. Telecommunication companies have invested billions on SPC technology and telephony networks and are already under investments for 3G networks. Beyond future mobile technology elements (Node B's, RNC's core network elements, routers, etc.), which present at a first glance the major cost of installation and application of future mobile communications, the realization of such a network supporting a huge number of subscribers with extremely high (for today's point of view) bit rates will result in a great investment regarding the backbone interconnection network needed to support the promised services and QoS. Thus, although the air interface (radio AP towards the end user) seems nowadays to be the basic restricting factor in the capacity of future mobile communications networks, a thorough examination should take place on the backbone network, in terms of capacity and cost. A simple example will clarify the above consideration: 2G BTS's in mature (with respect to radio coverage and capacity) companies seem to be adequately supported by one or two E1/T1 lines for interconnection towards digital cross connectors or directly to the BSC in order to support over 150 voice users. This would not be the case for UMTS. A maximum of three packet switched users with a rate of 384Kb/s can be nowadays supported through a single E1. For a mature 3G network with high penetration, multiple E1's will be needed to support the interconnection of Node B's to the RNC (it is not by chance that Node B's are already disposed with eight to 16 E1 integrated). The same is true for the interconnection of RNC's, where the normal interfaces are nowadays of the order of STM-1 (155Mb/s). This poses a certain problem for operators. The planning and management of a compact, multi-layer, easily adaptable and fast (in terms of switching) interconnection architecture, as the one proposed will be a major task for operator planners.

Next generation wireless networks are rapidly evolving based on several diverse planning techniques that have been proposed over the years. It is important to realize the weakness to implement extremely innovative architectures, since in most of the cases the basic modules to compose a network are the same and exist in several specifications. A safe way to implement a different architecture is to work with elementary modules. This chapter investigates innovative ideas, related to the elementary modules of wireless ATM and optical networks, concerning the interconnection backbone network of existing mobile networks.

From a theoretical point of view the proposed architecture and modifications in the existing fixed ATM network proves the flexibility of ATM protocol, the migration to wireless ATM access network and the possibility to incorporate it in the future specifications.

Future research will be conducted in wireless ATM networks. Several issues have to be considered. Traffic modeling for wireless ATM networks has not been examined a lot so far. Using these traffic models, a worst case analysis will be extracted to complement and design the parameters for non-congested wireless ATM networks. Cochannel interference is a critical parameter to consider in a dynamically frequency assigned wireless ATM network. IP over ATM or pure ATM wireless networks also have to be further investigated in order to decide to the most promising technique for data rates and integrability with other existing technologies

Concerning the POF optical link suggestion for indoor multi-layered cellular networks, the killer application will be the design of short distance links with the most available data bit-rates. POF are very flexible and nowadays, great interest has been presented from companies and industry to expand its applications to all available areas. POF links will be mostly used in LAN's, last mile connections, indoor connections and of course mobile applications. In conjuction with the pico-cellular architecture developed for next generation wireless cellular networks (UMTS,4G), the use of POF for interconnections among radio-access and core elements of network, in an indoor environment is a very good solution. Of course more research has to be conducted yet in several areas of POF technology. Better materials to lower the attenuation, to increase the data rate by decreasing the dispersion and to move the optical window towards the semiconductor lasers optical output are under investigation in world-wide industry. Better models have to be developed, including bending of fiber, and several non-ideal optical sources.

References

Cox, D.C. (1995). Wireless personal communications: What is it? *IEEE Personal Communications Magazine, 2*(2), pp. 2-35.

Daum, W., Krauser, J., Zamzow, P.E. & Ziemann, O. (2002). *POF-polymer optical fibers for data communications*. Springer-Verlag.

Esmailzadeh, R., Nakagawa, M. & Jones, A. (2003). TDD-CDMA for the 4th generation of wireless commnications. *IEEE Wireless Communications, 10*(4), 8-15.

Houssos, N., Alonistioti, A., Merakos, L., Mohyeldin, E., Dillinger, M., Fahrmair, M. & Schoenmakers, M. (2003). Advanced adaptability and profile management framework for the support of flexible mobile service provision. *IEEE Wireless Communications, 10*(4), 52-61.

Ishigure, T., Satoh, M., Takanashi, O., Nihei, E., Nyu, T., Yamazaki, S. & Koike, Y. (1997). Formation of the refractive index profile in the graded index polymer optical fiber for gigabit data transmission. *Journal of Lightwave Technology, 15*(11), pp. 2095-2100.

ITU Q.2931 ATM Network Signaling Specification.

Lach, H. Y., Janneteau, C. & Petrescu, A. (2003). Network mobility in beyond-3G systems. *IEEE Communications Magazine, 41*(7), 52-57.

Li, W., Khoe, G., van der Boom, H., Yabre, G., de Waardt, H., Koike, Y., Yamazaki, S., Nakamura, K. & Kawahadara, Y. (1999). 2.5 Gbit/s transmission experiment over 200 m PMMA graded index polymer optical fiber using 645 nm narrow spectrum laser and a silicon APD. *Microwave Optics Technology Letters, 20*, 163-166.

Louvros, S., Iossifides, A.C., Economou, G., Karagiannidis, G.K., Kotsopoulos, S.A. & Zevgolis, D. (2004). Time domain modeling and characterization of polymer optical fibers. *IEEE Photonics Technology Letters, 16*(2), 455-457.

Rajagopolan, B. (1995). Mobility management in integrated wireless ATM networks. *Proceedings of Mobicom 1995*, Berkeley, CA.

Siegmund, H., Redl, S.H., Weber, M. K. & Oliphant, M. W. (1995). *An introduction to GSM*. Boston: Artech House.

Simoens, S., Pellati, P., Gosteau, J., Gosse, K. & Ware, C. (2003). The evolution of 5 GHz WLAN toward higher throughputs. *IEEE Wireless Communications, 10*(6), 6-13.

Yabre, G. (2000a). Comprehensive theory of dispersion in graded-index optical fibers. *Journal of Lightwave Technology, 18*(2), 166-177.

Yabre, G. (2000b). Influence of core diameter on the 3-dB bandwidth of graded-index optical fibers. *Journal of Lightwave Technology, 18*(5), 668-676.

Yabre, G. (2000c). Theoretical Investigation on the Dispersion of Graded-Index Polymer Optical fibers. *Journal of Lightwave Technology, 18*(6), 869-877.

Weinert, A. (1999). *Plastic optical fibers: Principles, components, installation.* Publicis MCD Verlag.

Appendix - POF Channel Modeling

Consider the class of circular symmetric multimode graded-index polymer fibers with refractive index profile according to the *a*-profile grading. The multimode nature of the medium comes out of the Maxwell's equation solutions. Each particular solution corresponds to a particular mode propagating at its own velocity. Modes are clustered into groups in which each mode has nearly the same propagation constant. Dispersion in POF occurs mainly due to two different factors, time (modal) dispersion and material (chromatic) dispersion. The modal delay per unit length may be expressed as:

$$T_m(\lambda) = -\frac{\lambda^2}{2\pi c}\frac{d\beta(\lambda)}{d\lambda} = \frac{N_1(\lambda)}{c}\left[1 - \frac{\Delta(\lambda)[4+\varepsilon(\lambda)]}{\alpha+2}\left(\frac{m}{M(\lambda)}\right)^{2\alpha/(\alpha+2)}\right]\cdot$$

$$\cdot\left[1 - 2\Delta(\lambda)\left(\frac{m}{M(\lambda)}\right)^{2\alpha/(\alpha+2)}\right]^{-1/2},$$

where c is the speed of light in vacuum, $N_1(\lambda) = n_1(\lambda) - \lambda \cdot dn_1(\lambda)/d\lambda$ is the group index and $\mu(\lambda)$ is the profile dispersion parameter, expressed as

$$\varepsilon(\lambda) = \frac{-2\lambda n_1(\lambda)}{N_1(\lambda)\Delta(\lambda)}\frac{d\Delta(\lambda)}{d\lambda}.$$

Regarding material dispersion, the refractive index coefficients $n_1(\lambda)$ and $n_2(\lambda)$ of the core and the cladding respectively, follow a three-term Sellmeier function of wavelength Modal attenuation $A_m(\lambda, z)$ originating from conventional loss mechanisms, such as absorption, scattering (Rayleigh) and reflection, may be described in the simple form:

$$A_m(\lambda, z) = e^{-\gamma_m(\lambda)z},$$

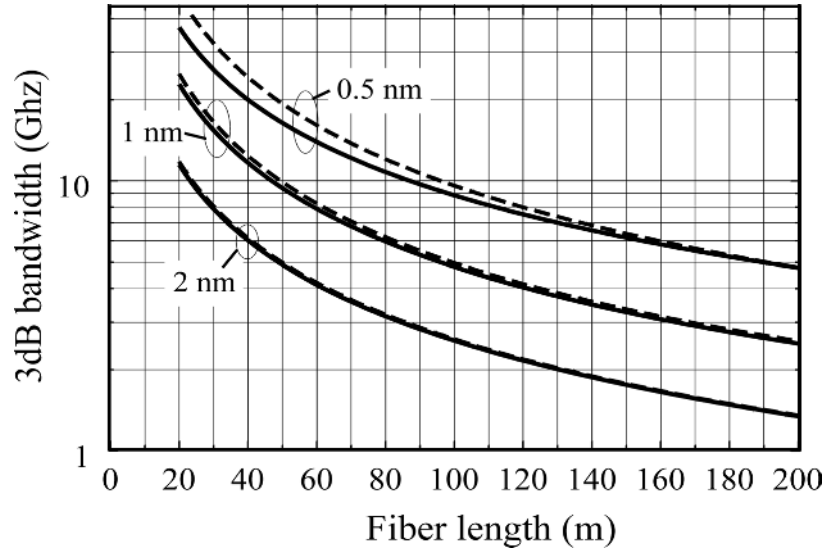
where z denotes the length of the fiber and $\gamma_m(\lambda)$ the mode attenuation coefficient. Since mechanism losses affect each mode in a different way the attenuation coefficient varies from mode to mode (the so-called Differential Mode Attenuation, DMA).

The channel model under consideration is based on formula:

$$h(t,z) = \sum_{m=1}^{M(\lambda)} \int_0^{\infty} A_m(\lambda, z)\delta[t - zT_m(\lambda)]d\lambda,$$

where $A_m(\lambda, z)$ is attenuation introduced by the medium for mode m, $T_m(\lambda)$ is its corresponding arriving delay (due to both chromatic and mode dispersion effects) at wavelength λ and $\delta(t)$ is the Dirac delta function. Since we are mainly interested for POF's response in a specific wavelength region (wide enough to include the optical excitation bandwidth of LED) we restrict, without loss of generality, the analysis and simulation between 620 nm to 680 nm. A sufficient number $N_\lambda = 600$ of equally spaced samples are taken in the wavelength domain, in the region of interest, that is, between 620nm and 680nm, leading to a wavelength resolution of 0.1nm (resolution may be increased if needed). In Figure 7 it is obvious that the resulted bit rates are very close to the predicted ones from previous stated experiments and standards.

Figure 7. 3dB bandwidth of PMMA POF with different source spectral widths and for index exponents of a = 2.1 (——) and a = 2.3 (– – –)



Chapter VIII

Packet Level Performance Measurement Schemes and their Limitations

John Schormans, Queen Mary University of London, UK

Chi Ming Leung, Queen Mary University of London, UK

Abstract

New business opportunities for mobile, wireless, and fixed networks are going to require managed packet based services; this requires SLAs that relate to the level of QoS purchased, and the measurement (monitoring) of information loss and delay at the packet level. In this chapter we investigate the two available measurement techniques: passive and active monitoring. We show that passive monitoring techniques can provide excellent accuracy with minimal computational overhead. However, it also has the disadvantage that it is necessary to have access to all the routers in all measured end to end paths, so severely limiting scalability. Alternatively active monitoring techniques can provide global reach; however it is critical that we go on to show that this technique has the disadvantage that (under many

circumstances) the measured results may be very inaccurate. Finally we propose some ideas which may enhance accuracy.

Introduction

In order to best manage new business opportunities, many organizations are employing packet based networks, for example, managed IP networks, for all data transfer. IP and MPLS packet networks, both mobile/wireless and fixed, are carrying a heterogeneous mix of traffic, with widely differing Quality of Service (QoS) requirements. The service model of emerging multiservice packet networks is based on the network's ability to guarantee QoS to user applications. It is clear from the origin and nature of much of the traffic that information loss and delay will be very important, and particularly so in view of the fact that a significant proportion of these organizations will be making use of mobile and wireless network systems from 2.5G and 3G eventually to 4G. These systems will, by their nature, tend to introduce a larger element of fixed delay greater than that found in land-based networks, and, as a result of fading effects, will also tend to exhibit higher loss rates.

Business plans associated with 3G and beyond are based on the idea that networks will provide services to mobile and wireless users that are potentially both data intensive and real-time in nature. For example a user requiring to know the location of a restaurant in a city will want to receive both enough data to make the location clear—a map perhaps—and receive it in time (and sufficiently uncorrupted by data loss) to be useful. Such a 3G system, to be commercially viable, cannot operate in the slightly hit and miss fashion that, for example, current 2G text mailing works. Here delays of up to days can be encountered especially when messages are passed between operators; this would be totally unacceptable.

Packet delay will consist of two components: the deterministic component and the stochastic component. The propagation delay is intrinsic and fixed (therefore deterministic) for a specific path. The stochastic component comprises many elements, among which the queuing delay is by far the most significant in packet networks like IP. In many cases, where the transferred information is commercially valuable or organizationally significant, transactional data will often need to achieve mean end to end delays that are nearly real-time. In addition to desired bounds on average delays, tight limits may well be needed on the proportion of data that is:

- Delayed by longer than an agreed time (i.e. bounds on the delay jitter), perhaps of the order of 100's of milliseconds.
- Lost as a result of e.g. buffer overflow.

Service Level Agreements (SLAs) (Cisco, 2001; Verma, 2000), will therefore be in place between organizations, or even different divisions within an organization, and these will be used to define the expected quality of service provided, including these bounds on information loss and delay.

In order to ensure that packet delay and loss targets featured in these SLAs are being met, organizations will manage the situation by using packet level monitoring to provide measurements from which guarantees can be checked. The type of measurement strategy employed depends on a number of factors, of which ownership and management responsibility are key. The two basic approaches to measuring end to end QoS are passive/non-intrusive and active/intrusive. In the former, where an organization has access to all of the network, so called passive measurements/monitoring, of the routers can be used. In passive monitoring/measurements of the level of packets in the buffers can be obtained from signalling information internal to the routers. From this information relating to the packet queue level in buffers, the end to end probability distribution of packet delay can be re-created to a very high degree of accuracy. An example of this process we develop in more detail in this chapter.

However, a very different challenge for managing performance is created when end to end communication is provided over a number of separately owned and operated networks because the customer's "footprint" does not match that of any service/network provider. This is likely to be very common in an era of growing globalization of business opportunities. In this case access to all the relevant equipment's internal measurements is not possible. A network operator will normally only provide SLAs for "on net" traffic, since it is unlikely to take the risk of guaranteeing performance for network segments over which it has no control. This means passive monitoring is no longer possible and active monitoring, using so-called packet probing must be employed to determine the end to end performance. Later in this chapter we discuss active monitoring in more detail and provide examples of the limitations on the accuracy achieved.

For corporations and commercial organizations needing to take advantage of managed IP provision, a significant difficulty may be that IP networks are evolving in a manner that is essentially heterogeneous. There is no single global network covering the world, rather an interconnected collection of different networks with different owners. Network heterogeneity is matched by traffic heterogeneity: the growth in user applications from VoIP and picture messaging to file transfer - both real time and non-real time - all cause different patterns of

traffic to appear in packet networks. However, despite this apparent complexity, queueing theory has shown that there are typical cases which have been found to be ubiquitous (Pitts & Schormans, 2000; Roberts, 1991; Schormans & Pitts, 2001). While the details are beyond the scope of this chapter, we are able to use this prior experience and knowledge as a fundamental input to any measurement/monitoring scheme.

The objectives of this chapter are therefore to:

- Discuss the importance of measuring packet level performance in broadband multi-service packet networks, and relate this to the business case;
- Schematically describe the functioning and limitations of both passive and active measurement techniques, providing mathematical details in self-contained sub-sections (that may be left by the reader, or read as required);
- Provide quantitative examples illustrating the accuracy of passive measurements (where passive monitoring is possible) for realistic networking scenarios; and
- Provide quantitative examples illustrating the accuracy of active probing (measurements), and discuss the reasons that such schemes are necessarily limited in accuracy.

Specifically we show that:

1) Where passive monitoring is possible, an organization can be provided with very accurate predictions for the end to end performance, and
2) Where active probing is needed, the extra load added by the probing packets can become suddenly excessive, especially at high loads (i.e., at exactly the sort of loads at which accurate performance monitoring is most desirable). If the probing rate cannot be increased (e.g., with increasing load) this will mean that the accuracy of the returned samples will decrease markedly. In consequence managing performance by using active monitoring must be done carefully.

We provide discussion and examples that are intentionally generic, and therefore as widely applicable as possible. It should be noted however that, for the reasons stated earlier, what applies in packet level monitoring is most critical when used in conjunction with mobile and wireless networks: delays and loss are potentially higher, and bandwidth is usually significantly more constrained. Later we show how important this can be.

Business Case for Measurement and Expected Market Response

Ideally, networks are designed to provide enough resources (buffering and bandwidth) adequate to bound delays and losses. Particularly in the case of mobile and wireless networks this is not simple to bring about however, as network resources are scarce and therefore comparatively expensive. Partly for this reason mechanisms, such as queue scheduling, active queue management, and path management, have been developed that provide traffic aggregates or flows to experience different levels of loss and delay. Premier level data and other important information will, as we will see, receive expedited passage across the network(s). This sort of QoS differentiation has been established as a key selling point that underlies many corporations' business plans.

The need to manage effectively QoS differentiation and varied levels of QoS guarantees leads directly to the need for reliable and accurate measurement techniques. These are vital to ensure that network configuration and mechanisms are providing the intended and commercially agreed service. As an example of how this is already evolving we can consider the case of a large US carrier. This organization is confident it can deliver its contractual promises to business customers, and so it is planning to give customers (using its premier data services) a 100% credit on monthly charges if it fails to achieve any SLA metric (LightReading, 2003). Furthermore, it is claimed that most of this carrier's rivals offer only a 10% refund if their service fail to match their SLA. It is also reported that the carrier in question claims already to be meeting its SLAs more than 99% of the time, and so is confident that to offer this level of service is good business. It can be seen that measurement are key to informing the business case at each stage.

It is further reported (LightReading, 2003) that the carrier in question is set to offer a "Jitter SLA" that will "give customers the confidence that IP networks can deliver quality of service for services such as voice or video over IP". It is claimed that jitter levels can be reduced down to two milliseconds on its U.S. network, and the plan now is to begin deploying the probes needed, in its international POPs, to measure the IP network's performance internationally. As a result, it will be some time later in 2004 before it will be known what sort of guarantees can be made for jitter levels on international connections.

Market analysts have forecasted (Analysys, 2003) that the global SLA-based WAN market will grow from $29 billion in 2003 to $52 billion in 2006, driven by applications such as intranets and remote access for mobile staff, making wireless and mobile services a particular driver for this growth. They go further to note that multiple networks, technologies and (often) service providers may

have to be combined to offer "a complete service based solution to multi-site companies". The problem then becomes one of SLA based service assurance over multiple scenarios, and this requires measurement. As reported in (Yankee, 2003) users face the problem of "meaningless", that is un-guaranteed, SLAs; their problems also include:

- What metrics should be measured (availability, loss delay, jitter etc.)?
- How are terms like loss and delay to be most meaningfully defined?
- Is measured performance averaged (over an hour, day week, month etc.)?
- How will CoS affect matters?

The solutions to these problems lies in part with (IT and network) managers obtaining a better understanding of QoS guarantees by measurement, and in part with properly monitored and measured networks becoming fully commercially available.

Commercially available networks that are fully monitored are becoming a reality, indeed (Nexagent, 2003), future business in global networking will be absolutely dependent on accurate measurements.

Correlation with the Standardization Process

The work reported in this chapter correlates well with the (considerable) on-going standardizations work, which is being carried out for and by the Internet Engineering Task Force (IETF), and others. The importance of the IETF is that it represents the interests of very important groupings of suppliers and consumers within the packet networking community. IETF document RFC3393 is entitled: IP Packet Delay Variation Metric for IP Performance Metrics (IPPM). This paper (RFC3393) utilises a delay jitter measurement technique that aggregates instances of packet delay, thereby, over time, formulating an estimate of delay distribution. This RFC is based on prior (IETF) work reported in RFC2679.

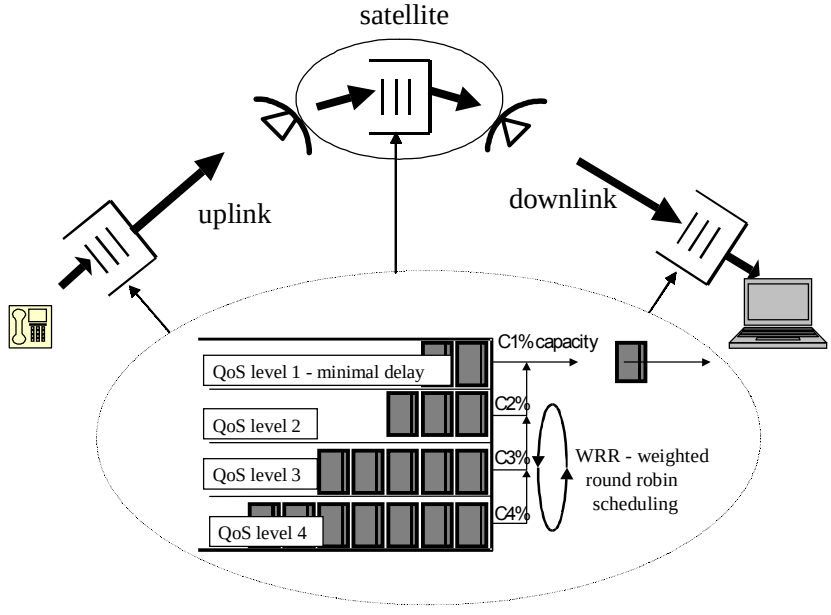
There are two important conclusions to be drawn from an examination of this (and similar) work: 1) the cluster of interest groups that form the IETF consider packet level measurement to be very important; 2) the standardization process is concentrating almost exclusively on technological issues associated with how packet probing protocols and techniques can be developed. The topics covered in this chapter are both consistent with, and complementary to, the issues

considered in the standardizations process. In this chapter we provide a quantitative investigation into both the schemes proposed for packet level measurements; this allows us critically to move beyond the protocols, presenting results which make clear how accurate the schemes are, and where are their limitations. Therefore our results are critical for IT managers: there is little point investing in monitoring protocols and equipment if it is not capable of achieving the desired objectives of accuracy coupled with scalability and minimal network overhead. Furthermore, once the investment has been made it is financially imperative to make best use of it. This implies the need for an understanding of what can be achieved. By addressing the accuracy/scalability of passive monitoring and the network overhead/accuracy trade-off in active monitoring we provide just the right insights; IT managers will know what their equipment is, in the limit, capable of, and what is it not capable of, for example, they will know (be able to calculate) what sampling rate to set up for packet probing given the desired level of accuracy.

Table 1. Gives a summary of the relevant work in the standardization bodies; see Morton (2004).

	IETF IPPM RFC's	ITU-T Rec's
Framework	2330	Y.1540 cl 1 thru' 5
Sampling	2330 Poisson 3432 Periodic	(for future work in SG 4)
Loss	2680	Y.1540 cl 5.5.6
Delay	2679 (1-way) 2681 (Round-trip)	Y.1540 cl 6.2
Delay variation	3393	Y.1540 cl 6.2.2 G.1020 (short term)
Availability	2678	Y.1540 cl 7
Bulk transfer capacity	3148	
Loss patterns	3357	Some in G.1020

Figure 1. Diagrammatic representation of packet network, and associated buffer scheduler



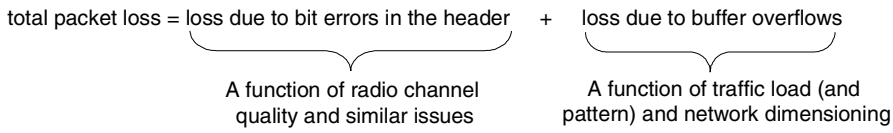
The Measurement Schemes and their Limitations

Delays and loss are the natural results of packet transmission, switching and buffering over the network. Figure 1 shows a schematic diagram illustrating how the network links may well be part of a wireless (perhaps mobile) network. This is potentially very important as these networks tend to feature limited bandwidth and buffer space. Also illustrated is the scheduler, which will divide the available link bandwidth among the different levels of QoS classes in a fashion that is in accord with the inbuilt design decisions.

The fact of mobile and wireless transmission has considerable importance here. Packet delays consist of the sum of both fixed (deterministic) and variable (stochastic) components, essentially:

total packet delay = packetisation delay + transmission delays + queueing delays

fixed (for fixed packet length) variable



There are other aspects, but these are of comparatively marginal importance. The packetisation delay is the time it takes to assemble the information bits into a packet, and is therefore a function of the packet size and the rate at which the data bits are arriving. So for voice over IP (VoIP) with packet lengths of 80 bytes (of user data) the packetisation delay would be relatively small and is known and fixed, and therefore does not need to be measured. Equally the transmission delay over a path is fixed: it is simply the time taken for the information to propagate along the full length of path, end to end. While this may vary, path to path, it is more or less fixed for any path, and therefore will appear in any measurement as a fixed value; this means it doesn't cause difficulties for any measurement scheme.

This is the justification for our concentration on the queueing delay part of the total delay: it is often the largest (although not necessarily in mobile and wireless networks) and it's always liable to be the hardest to measure as it will tend to exhibit the greatest variability. We return to this when we discuss the two main measurement methods in more detail.

It is important that, despite the apparently confusing array of different applications using these networks, and the extreme heterogeneity of the traffic patterns produced, recent results in queueing theory have led to an understanding that delay distributions will tend to exhibit essentially two types of queueing behaviour: short-term and long-term. These are sometimes called packet scale and burst scale respectively (Pitts & Schormans 2000; Roberts 1991; Schormans and Pitts 2001), and these are the terms we use in this chapter. The packet scale is associated with the random, phased, arrivals of packets, while the burst scale is associated with significant periods during which the aggregate of arrivals exceeds the service rate of the buffer. Later in this chapter we use this understanding further.

Passive Measuring

Passive monitoring works by monitoring the packet level, for example, buffer fill, statistics from the local routers, and using this information to infer overall end to end delays by the application of an algorithm such as the one we present in this

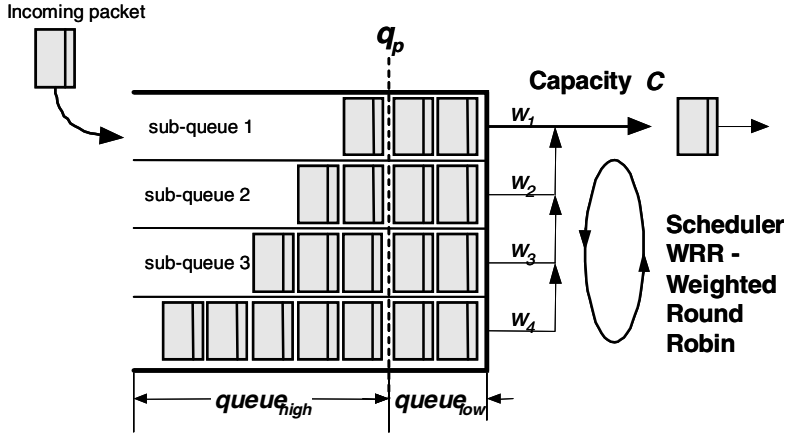
chapter. These schemes can use a variety of measurement techniques, for example, Polling MIB (Cisco, 2003), Packet Sniffing and Queue length monitoring. In fact queue length monitoring must be employed in any buffer management scheme using RED (Floyd & Jacobson, 1993) to determine the decision (i.e., drop/not_drop) for any incoming packet. And RED (or WRED) is extremely widely used in IP networks! However, regardless of which scheme is employed in detail, it is noted that passive monitoring is performed in the local network element(s) only. When the interior network elements are accessible, for example, in a network domain owned entirely by a single network operator, it is possible to infer the end to end QoS performance by using passive monitoring, otherwise active monitoring must be used.

Description of the Technique of Passive Measurement

In this section we describe a particular passive monitoring scheme in detail (Leung & Schormans 2002a), with the objectives of illustrating both what any passive monitoring scheme needs to be able to do (and therefore indicating the computational complexity involved in using one), and showing the level of accuracy that can be achieved.

Figure 2 depicts the passive queue monitoring scheme. Each sub-queue is logically partitioned into two regions via a partition point q_p. $queue_{high}$ refers to the region in which the queue length is larger than the partition point, whereas $queue_{low}$ is the region in which the queue length is smaller than (or equal to) the

Figure 2. Passive queue monitoring scheme



partition point. The purpose of this partition point is to isolate the burst-scale queuing region from the packet-scale queuing region. This partition point should be large enough to be in the burst-scale region, but it has been shown that the measurement accuracy is largely insensitive to the location of the partition point (Leung & Schormans, 2002a).

There are four measurement data for each sub-queue: $freq_{low}$; $freq_{high}$; q_{low} and q_{high}. During each measurement period, the current queue length, q, seen by an incoming packet is compared with the partition point. If q is larger than q_p, then $freq_{high}$ is incremented by one and q will be added to q_{high}. The same process is carried out for $freq_{low}$ and q_{low} when q is smaller or equal to q_p. The measurement scheme consists essentially of just comparison and addition, minimizing computational complexity and overhead. After measurements, the data for the sub-queue is stored (and is retrievable) at (from) the network operating centre for delay distribution re-construction. As there are only four types of measurement data for each sub-queue, this minimizes the extra traffic between the local nodes and the network operating centre.

Mathematics of the Passive Measurement Scheme

As previously described, measurement results will be used to estimate the following parameters:

p_{low} = Prob(an arriving packet sees queue $\leq q_p$) = $freq_{low}/(freq_{low} + freq_{high})$

p_{high} = Prob(an arriving packet sees queue $> q_p$) = $freq_{high}/(freq_{low} + freq_{high})$

$\bar{q}_{low}$ = the mean queue length conditional on being in $queue_{low}$ region = $q_{low}/freq_{low}$

$\bar{q}_{high}$ = the mean queue length conditional on being in $queue_{lhigh}$ region = $q_{high}/freq_{high}$

$\bar{q}$ = the mean queue length = $(q_{low}+q_{high})/(freq_{low}+freq_{high})$

With these parameters, we can obtain the Maximum Likelihood Estimate of the burst-scale decay rate (Leung, 2003) (this is fully developed in Appendix 1).

$$\eta_b = 1 - \frac{1}{\bar{q}_{high} - q_p}$$

Since the p_{high} represents the probability of an arriving packet seeing the $queue_{high}$ region, p_{high} can be expressed by relating it with the burst-scale queuing model.

$$p_{high} = \sum_{k=q_p+1}^{\infty} c_b \eta_b{}^k = c_b \frac{\eta_b{}^{q_p+1}}{1-\eta_b}$$

Therefore, the burst-scale decay constant is given as:

$$c_b = \frac{1-\eta_b}{\eta_b{}^{q_p+1}} p_{high}$$

With the burst-scale decay constant and the decay rate estimate, the per-hop queue length distribution, $Q(\cdot)$, can be re-constructed. For the $queue_{high}$ region, $(x > q_p)$, the queue length distribution is represented as $c_b \eta_b{}^x$. Since the burst-scale queuing is more significant than the packet-scale queuing, the $queue_{low}$ region is simply represented by a single point at $\bar{q}_{low}$ with probability p_{low}. This completes the re-construction of the per-hop queue length distribution $Q(\cdot)$.

Mathematics for Calculating the Delay Distribution Using Passive Measurements

Queue length x at sub-queue k with queue service rate C_k and mean packet size e_k will cause a queuing delay of $x \times e_k/C_k$. Likewise, the mean queue length $\bar{q}$ gives the mean delay time = $\bar{q} \cdot e_k / C_k$ at this hop. The end to end packet delay is the sum of per-hop packet delays along the path, experimental results revealed that the packet delay on the successive links can be safely assumed to be independent in a WAN (Vleeschauwer, 1995). Therefore the mean end to end packet delay is the sum of the per-hop mean packet delays, and the end to end delay distribution can be obtained by convolving the per-hop queue length distributions:

$$Q_{\text{end-to-end}}(\cdot) = Q_1(\cdot) \otimes Q_2 \otimes (\cdot) \ldots Q_n(\cdot)$$

This simple convolution equation is applicable when the service rates of the queues along the path are identical, as the queue length x will correspond to the same delay time at all nodes. In any generalized mobile, wireless or even fixed

network this may not be valid (there will be different queue service rates along the path e.g. different link bandwidth or different assigned scheduler's weights). To account for this, a normalization process is necessary to adjust the queue length distribution and thereby reflect the true delay time with respect to a specific reference service rate C_r.

As discussed, a queue length of x packets at queue k, which has a service rate C_k, will cause $x \cdot e_k / C_k$ delay. Assuming this queue is served at a rate of C_r (the reference service rate) instead, then, in order to have the same delay effect, the corresponding queue length x' packets can be determined by multiplying with a modifier (C_r/C_k) on x. Or, $x = (C_k/C_r)x'$. By substituting this into the burst-scale queuing model, it shows that the burst-scale decay rate should be modified to $\eta_{bk}' = \eta_{bk}\frac{C_k}{C_r}$ from the original estimate η_{bk} obtained above. With respect to the reference service rate C_r, the point $\bar{q}_{lowk}$, q_p should be modified by (C_r/C_k), and therefore the point representing the $queue_{low}$ region is shifted to $(C_r / C_k)\bar{q}_{lowk}$ with the probability p_{lowk}, whereas $queue_{high}$ region starts at $(C_r/C_k)q_p$. Since the probability of an arriving packet $queue_{high}$ region is equal to p_{highk}, the burst-scale decay constant c_{bk} at node k is modified to c_{bk}' as follows:

$$c_{bk}' = \frac{1-\eta_{bk}\frac{C_k}{C_r}}{\eta_{bk}^{(q_p+1)}} p_{highk}$$

With the parameters, q_{lowk}', q_p', η_{bk}' and c_{bk}', we can re-construct the normalized queue length distribution $Q_k'(\cdot)$. Based on the normalized queue length distributions, the end to end delay distribution with respect to the reference service rate can be obtained by using convolution.

$$Q_{\text{end-to-end}}(\cdot) = Q'_1(\cdot) \otimes Q'_2(\cdot) \otimes \ldots Q'_n(\cdot)$$

Mathematics for Calculating the Packet Loss Probability Using Passive Measurements

The per-hop packet loss probability (PLP) can be passively measured by simply counting the number of packets received and dropped at the local node.

$$PLP = number\ of\ packet\ dropped\ /\ number\ of\ packet\ received$$

Based on the link independence assumption, the end-to-end packet loss probability of a specific path can be estimated as follows (Verma 2000):

$$\text{PLP(total)} \approx 1\text{-}\prod_{j=1}^{N}(1-PLP_j)$$

Where PLP_j is the packet loss probability measured at node 'j'.

Accuracy Achieved by Passive Measurements in Simulations of Real Traffic Examples

Simulations of a typical delay sensitive voice over IP (VoIP) traffic scenario (as might be typical of an IP network) are now used to test the effectiveness of the passive monitoring scheme. This is as illustrated in Figure 3, where the link bandwidth is distributed by a scheduler with logical sub-queues. Each sub-queue multiplexes the traffic of interest, here called the foreground traffic, which is interfered by ambient "background" traffic. The foreground traffic traverses *N*

Figure 3. Representation of foreground and background traffic mixed in the simulated scenario

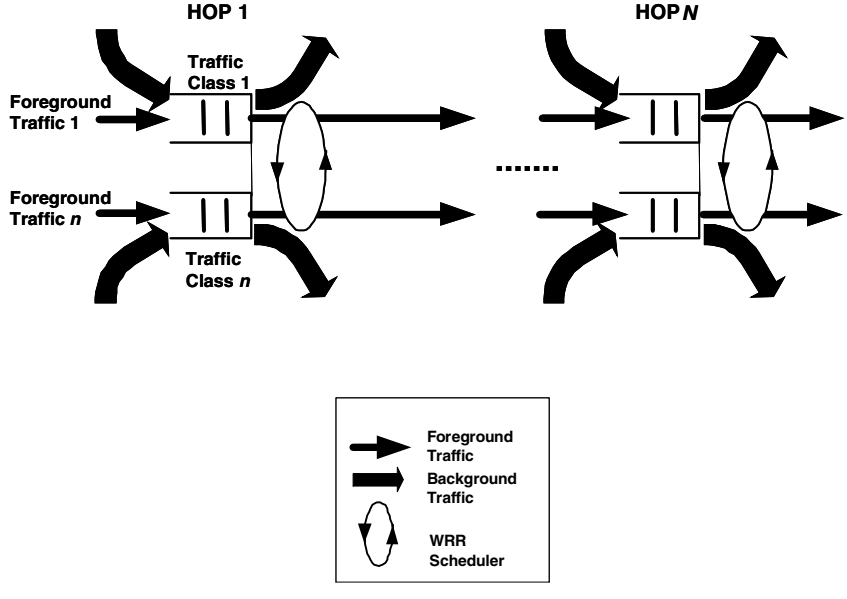
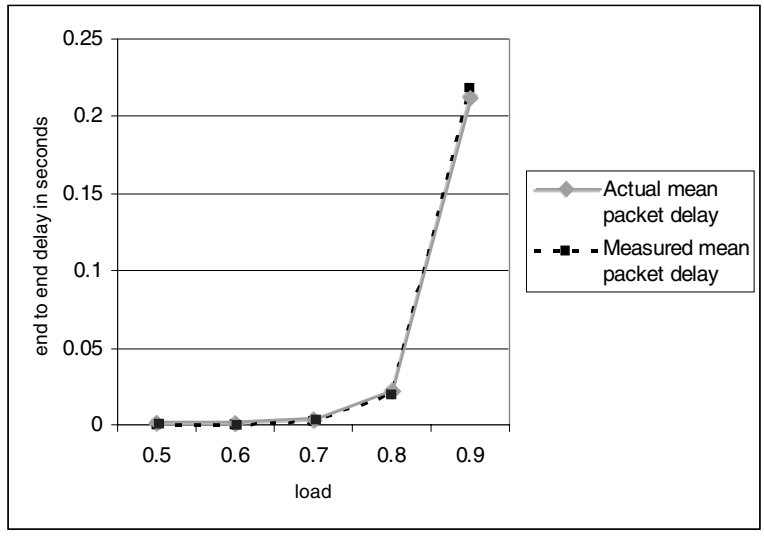


Figure 4. Comparison between the actual and measured end to end packet delay



hops before reaching the receiving end, whereas, at each node the background traffic is removed from the system (Liu, 2002; Stewart, 2002).

In this first example we use a VoIP traffic model only, and therefore only a single sub-queue per buffer. The traffic model of VoIP sources used is the usual one: mean ON (source active) time = 0.96 second, mean OFF (source silent) time = 1.69 second; both active and inactive periods of each VoIP source having an exponential distribution. In this test scenario only one traffic class is considered. There are four hops, each with non-negligible queuing delay, and the link bandwidths are all set to 2.048Mbps.

The mean queue lengths measured in this passive monitoring scheme are converted to the delay time by using the formula $\bar{q} \cdot e_k / C_k$ as discussed previously. The mean end to end delay time can then be estimated by summing all the per-hop measured mean delay times. In Figure 4 this is compared with the actual mean delay time, for various load conditions as shown. It can be seen that:

- The mean end to end packet delay time increases rapidly when the load is beyond 0.8 (a result of significance particularly for mobile and wireless operators, in whose networks bandwidth may be lower and hence loads higher).
- The passive monitoring technique measures the delay performance with excellent accuracy.

In the previous example, there was only one traffic class. Another sub-queue is now added in each node, and this sub-queue is used by the data traffic (the model of this is also of the ON-OFF type) which is set to generate traffic patterns that are 10 times as bursty as the VoIP traffic patterns. While there are still four hops, we also consider different link bandwidths in this simulation. The link bandwidths at odd-numbered hops are still 2.048Mbps, whereas the link bandwidth at even-numbered hops is twice that of the odd-numbered hops. The link bandwidth is distributed among the two sub-queues by a Weighted Round Robin (WRR) scheduler. The scheduler's weights are configured in such a way that one-tenth of link bandwidth is allocated to VoIP, whereas, the rest is allocated to the second traffic class bursty Markovian ON-OFF traffic. The load at all sub-queues is equal to 0.8 to account for potentially highly-loaded ingress and egress links.

At each sub-queue the data is collected for the per-hop queue length distribution, and from this the end to end delay distribution is re-constructed. We selected the sub-queues' service rates at odd-numbered nodes as the reference service rates, that is, 0.2048Mbps for VoIP traffic class and 1.8432Mbps for bursty Markovian ON-OFF traffic class. The actual bandwidth received by a sub-queue is lower-bounded by $w_i C$, where w_i is the scheduler's weight to a sub-queue and C is the bandwidth. The unused bandwidth of a sub-queue is shared by the other non-empty sub-queues, which is known as bandwidth stealing (Kuzmanovic & Knightly, 2001). For this reason the actual received bandwidth depends on how busy the other sub-queues are (so the bandwidth stealing effect is not significant under high load conditions). Leung (2003) proposed an algorithm to measure the actual received bandwidth of a sub-queue, so it therefore becomes possible to incorporate this into the passive monitoring scheme.

After the per-hop queue length distributions at the even-numbered hops are normalized with respect to the reference service rates, the end to end delay distributions are obtained by convolving the normalized per-hop queue length distributions along the path.

Figure 5 shows a comparison between the actual and estimated end to end delay distributions of VoIP and the bursty data ON-OFF traffic. It shows that the bursty data traffic tends to produce much longer delays for the same load. It can be seen that a very high level of accuracy is being achieved by the passive monitoring scheme. There is evidence that this accuracy is consistent across a wide range of traffic scenarios, including the multiplexing of self-similar traffic models, see Leung & Schormans (2002b).

Turning to the situation regarding passive measurement for end to end packet loss, a comparison of the actual and measured packet loss probabilities for VoIP traffic is shown in Figure 6. As discussed before, the per-hop packet loss ratio is obtained by locally counting the number of packets dropped and received, and the end to end packet loss ratio are then estimated based on the per-hop packet loss ratio measurement results. The VoIP data packets have traversed four

Figure 5. End to end delay distribution

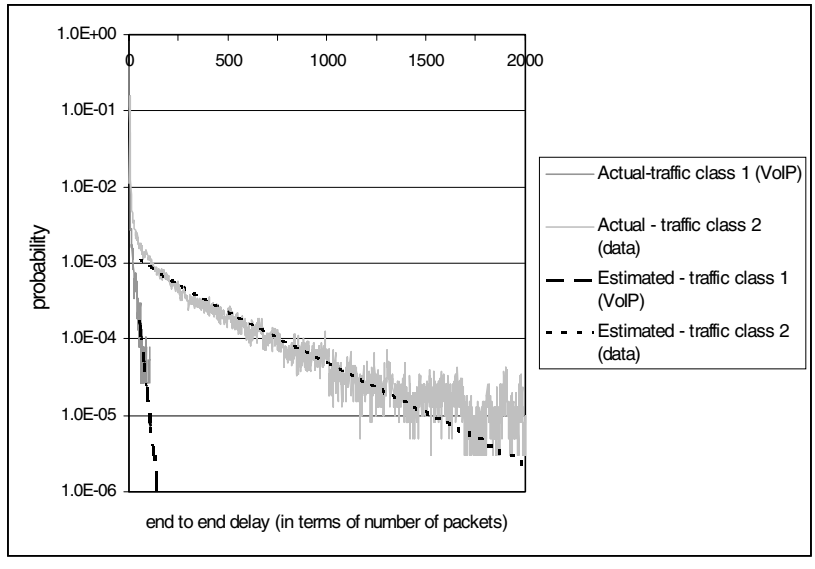
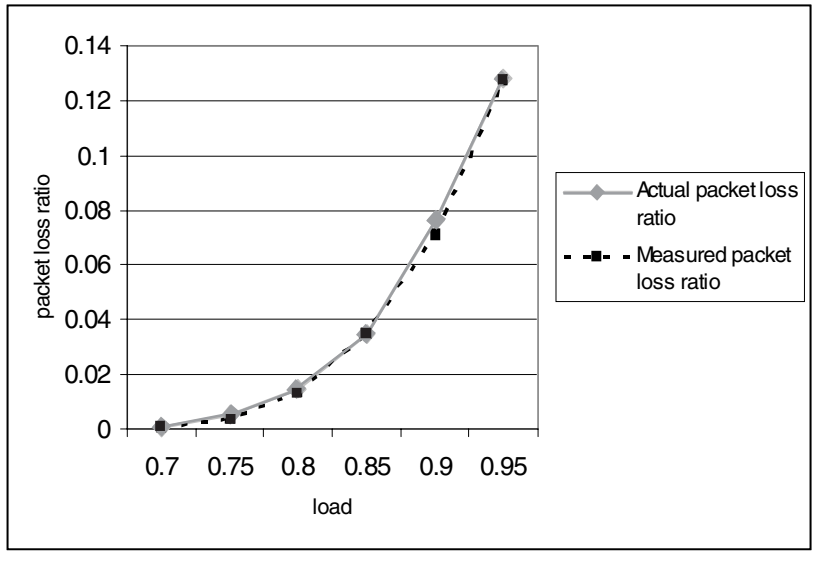


Figure 6. Comparison between the actual and measured end to end packet loss probability



nodes. We compare the end to end packet loss with the measurement results by using passive measurement as discussed. With reference to Figure 6, it can be seen that the measurement results shows good agreement with the actual loss proportions under various load condition.

Limitations of Passive Measurements

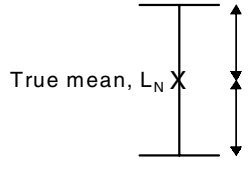
The effectiveness of passive monitoring in estimating end to end latency and loss probability performance has been demonstrated. The scheme provides excellent accuracy without intruding on the network. The main drawback is that the measurement data is only available in the interior network elements. Under circumstances involving global networking therefore, active monitoring becomes the natural choice for packet level measurement, and it is this we now consider.

Active Measuring

Most new business opportunities based on networks are, or are potentially, global in reach if not in scale (and it is normal to have ambition towards global scale too). As a result, it is hard to foresee that many business plans can be based on passive monitoring, as global, international and even national networking tends to imply inter-networking between organizations. Indeed this fact itself has become a driving force behind certain new opportunities in managing QoS aware inter-

Box 1. Standard error in measuring delays through packet networks as a sampling error

δ = Absolute error in measurements



100(1-α)% confidence interval

$$L_N \pm t_{N-1,1-\frac{\alpha}{2}} \cdot \frac{S_N}{\sqrt{N}}$$

For the sampled mean, L_N, to be within the desired absolute error of ± δ of the true mean, N must be such that:

$$t_{N-1,1-\frac{\alpha}{2}} \cdot \frac{S_N}{\sqrt{N}} \leq \delta$$

S_N = standard deviation of the measurements
N = the number of samples (probes) in the period

organizational networking. This implies that the only viable option for measurements is active monitoring. It is essentially a quality sampling system, as are many commercial / industrial quality control systems. The trade-offs are therefore similar: essentially accuracy increases with the number of samples taken (probes injected), but the working overhead (in this case bandwidth used) increases too. The most important limitation of active monitoring turns out to be that, under certain circumstances, the bandwidth available for probing may be so small that there are not enough probes available per measured hour to provide a level of accuracy that makes monitoring worthwhile; see Box 1. Networking managers have to be aware of this in order to create meaningful SLAs.

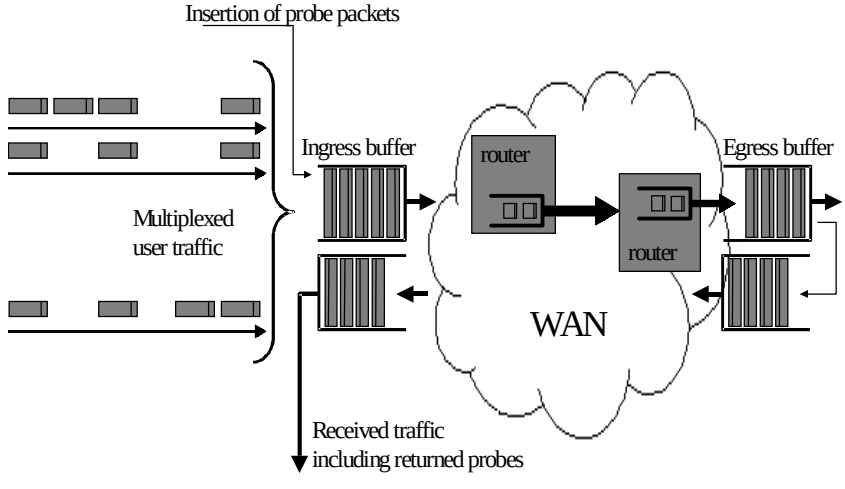
As with passive monitoring again measurements are needed for:

- Probing for Delay, in which the main limitation is on the accuracy that is likely to result from having a limited amount of bandwidth at access links that can be dedicated to probing.
- Probing For Loss Probability, in which the main limitation is the mean time until a probe encounters an overflowing buffer, on the understanding that for measurements to be valid there must be a few probes that do this in every measured hour.

Description of Active Measuring

The model we have adopted is shown schematically in Figure 7.

Figure 7. Schematic network model for probing delay



In this model, aggregated user traffic is inserted into the WAN via an ingress buffer, and, as this happens, probe packets are added into the aggregated stream. At the receiving end the probes are re-inserted into the returning traffic stream, and pass back across the WAN before being buffered again at the receiving (originally the transmitting) end. A variation on this would not reflect the probe packets back, but would instead measure their one-way delay (or loss) by means of external (out of band) signalling (i.e. of clock timing in the case of measuring delay). It is known from sampling theory that the greater the variability of the sampled data the more samples are needed for accurate estimation, even for mean values of the distribution only. This is illustrated schematically in Figure 8. In this case (the case of actively probing packet networks) the variability is highly dependent on two main factors: the load on the network, and the type of traffic being carried. Research has shown (Willinger, Taqqu, Scherman and Wilson, 1997) that highly bursty traffic is frequently found in packet networks and this results in very large variances associated with the number of packets in queues, and hence the packet delays, (Pitts and Schormans, 2000; Schormans and Pitts 2001).

There are different patterns by which the probes can be inserted into the network, and these are illustrated in Figure 9. Precise details are outside the scope of this chapter. Essentially, the use of any active monitoring scheme results in a probing bandwidth vs. accuracy trade-off, and this is illustrated in Figure 10. The upper part of Figure 10 shows how, for a constant desired

Figure 8. Schematic diagram illustrating the conceptual effect of highly variable network traffic

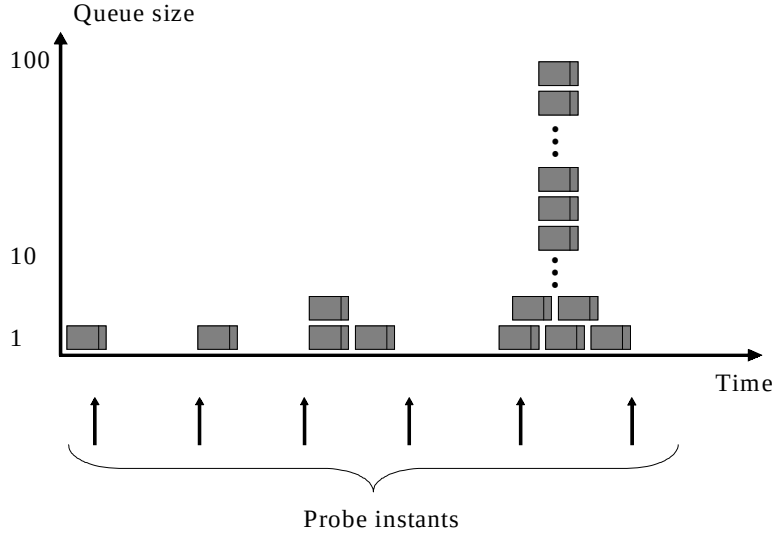


Figure 9. Different types of sampling in active monitoring

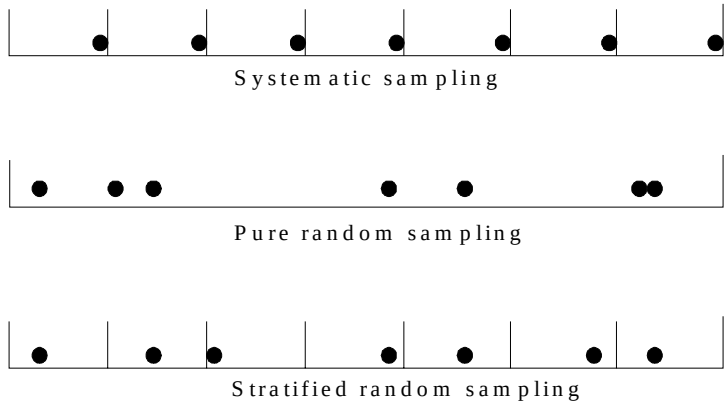
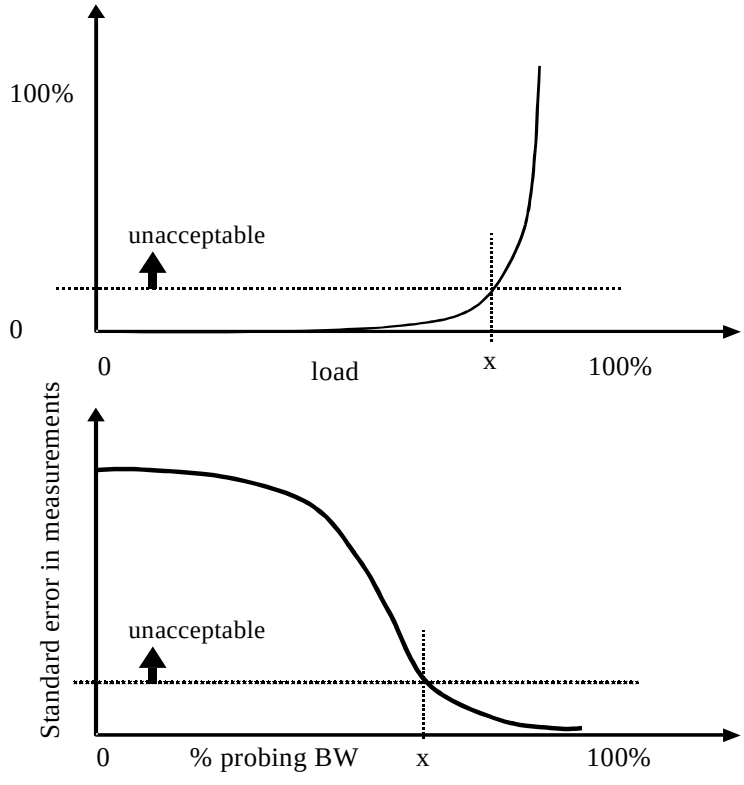


Figure 10. Schematic representation of the BW/accuracy trade-off in active monitoring



accuracy, the probing rate (bandwidth) will have to increase with traffic load; the lower part shows how the accuracy falls away with increasing traffic load if the probing rate (bandwidth) is kept constant. One problem with active monitoring is to establish the value of X in advance.

Mathematics Associated with Active Monitoring Schemes

This section presents the main parts of the analysis that quantify the levels of accuracy that can be expected from any active monitoring implementation. This analysis is based on two areas of theoretical knowledge: queueing theory and sampling theory. The former is needed to provide the variance of the queue level in the buffers, that is, an indication of how variable is the data that packet probing has to measure; the latter is needed to translate that variability into limits on the accuracy thereby obtained.

Queueing theory has determined that the "envelope" bounding the probability distribution of the delay of packets passing through a buffer is known to keep largely the same shape (packet scale and burst scale) over a very general multiplex of traffic types carried by IP, as used earlier. This packet scale + burst scale envelope is a function as follows:

$$\text{Envelope} = f(\rho, T_{on}, R)$$

Where:

- ρ = the load on the buffer
- T_{on} = mean ON (active) time of a traffic source(s)
- R = mean rate at which packets are generated by a traffic source when it is on.

To determine how many probe packets are required to estimate the mean number of queueing packets, L, delaying an arriving packet, define:

- N = the number of probe packets (measurements) required
- $t_{N-1, 1-\alpha/2}$ = the Student t-distribution value for N-1 degrees of freedom (i.e. sample size = N), and (100-±)% confidence interval for the estimate of the mean queue size (L)
- δ = the chosen value of absolute error in the measurements

L_N = mean number of packets in a queue delaying an arriving probing packet, estimated from N measurements

S_N = standard deviation of the measurements

S^2_N = variance of the measurements

For the number of probes needed to estimate the mean L_N, the 100(1-α)% confidence interval for the true mean is given by sampling theory as:

$$L_N \pm t_{N-1,1-\frac{\alpha}{2}} \cdot \frac{S_N}{\sqrt{N}}$$

For the sampled mean L_N to be within the error ±δ of the true mean, the number of samples taken (N) must be such that:

$$t_{N-1,1-\frac{\alpha}{2}} \cdot \frac{S_N}{\sqrt{N}} \leq \delta$$

And:

$$\left(t^2_{N-1,1-\frac{\alpha}{2}} \cdot \frac{S^2_N}{\delta^2} \right) \leq N$$

Omitting all the details (available in (Timotijevic and Schormans 2003a)) it turns out that, in the case of packet scale and burst scale queueing, we must account for the fact that the variance (S_N^2) is itself a function:

$S_N^2 = f(r, T_{on}, R, C)$

Where C = the channel capacity.

From these results it becomes possible to find the variance of the distribution of the number of delaying packets in a buffer. For a network of 'K' buffers the variance is approximately K times the individual variances (we use this in our results in the next sub-section).

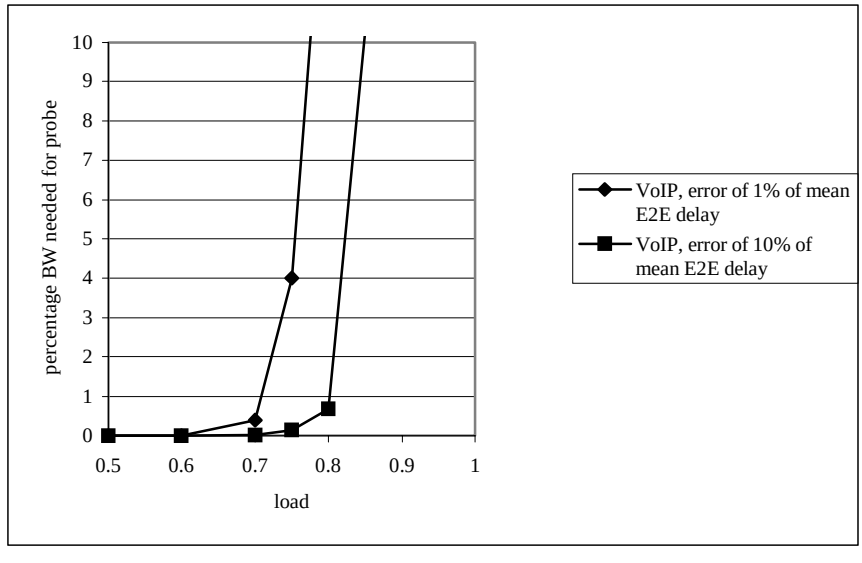
Accuracy Achieved by Active Measurement for Real Traffic Examples

From the results summarized in the previous section it is possible to quantify the accuracy of active monitoring, (Schormans & Timotijevic 2003; Timotijevic, Leung & Schormans, 2003; Timotijevic & Schormans, 2003a; Timotijevic & Schormans, 2003b). In this section we now provide schematic guidelines as to the main results.

In order to show the importance of our work for managing IP dependent business we now consider some practical networking examples of general interest. Our examples are for VoIP traffic, and a generic model of data. VoIP is a form of traffic that is expected to grow considerably in years to come, and therefore can be expected to form a large proportion of the revenue stream of many network and service providers. The "data" model is not a particularly bursty one, and so in effect forms an upper bound on the level of accuracy we can expect (the more bursty the traffic, the more variable the queueing distributions, and hence the less accurate the results of the probing will be).

Figure 11 shows the results of probing, using 100 byte probes for VoIP traffic which also use 100 byte packets. Bandwidth required by the probes (to achieve the desired level of accuracy) is plotted against load; this is for probing the mean

Figure 11. Active monitoring results for 100 byte probing packets with VoIP

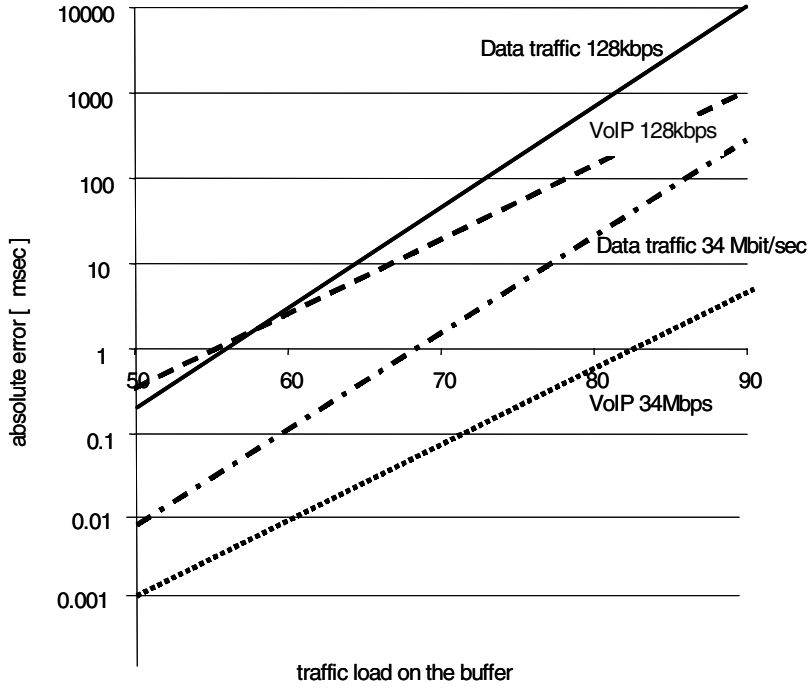


delay only. It can be seen that the probing overhead (required to keep a constant measurement accuracy) increases dramatically as the overall load increases above 50%.

However, networking and IT managers will be aware that it will, in general, not be possible to increase significantly the probing rate; rather the rate will be limited and this will have the result that accuracy will be degraded instead. We now investigate measurement accuracy by using the mathematical techniques developed in the previous section. Specifically we find the absolute measurement error (in measuring the mean delay only) against the traffic load on the buffer for the following traffic scenarios (see Figure 12):

- VoIP using 100 byte packets over 128 kbps access link
- VoIP using 100 byte packets over 34Mbps access link
- A generic data using 1000 byte packets model over 128 kbps access link
- A generic data model using 1000 byte packets over 34Mbps access link

Figure 12. Active monitoring results for the four traffic access link scenarios studied



In generating some representative results we have concentrated here on the access link, as the access bottleneck is the key element in limiting probing accuracy across a WAN. Our data packets are different sizes to represent the different traffic: 100 bytes for VoIP and 1000 bytes for data. We use probing packets that are all 100 bytes long; however, in a practical situation, where probes are being used also to determine the loss probability, probes may have to be much the same size as the informational packets: prior work has clearly shown that probing packets have to be the about the same size as the informational packets--if this is not the case then the measured *loss probability* will be wrong by many orders of magnitude, (Schormans & Timotijevic, 2003). Since our concern in this set of results is only delay we do not have to worry about this.

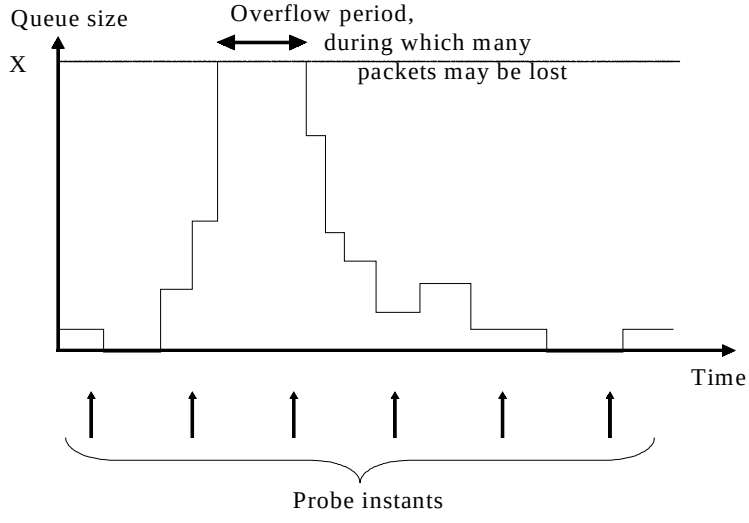
Overall, these results (in Figure 12) indicate that the accuracy achieved by active monitoring falls away dramatically as the traffic load on the access buffer increases. For loads greater than 80% (quite possible levels of load on bandwidth constrained access links in fixed networks, and so even more likely in mobile and wireless networks), the absolute error in measuring the means of the packet delays increases to values around one second! This, when seen in the context that real-time traffic like VoIP will have to be guaranteed mean delays around 50 milliseconds, is a huge figure! Furthermore, this only considers *mean* delay, and over a single buffer. The relative performance of active monitoring is worse when used it is used determine delay jitter or packet loss probability.

Turning to the situation regarding probing for packet loss, the effectiveness of probing may be severely degraded by the fact that events involving lost packets—buffer overflow events—are relatively rare but involve the loss of a very large number of packets (on average) per overflow event. This is very significant with respect to the measurement of packet loss probability for the following reason: while buffer overflow events are roughly randomly distributed over time, the packet loss events are not—they will be highly correlated. This is shown schematically in Figure 13.

Results for the time taken to reach measurable (buffer overflow) events for multiplexed VoIP traffic have been obtained. These were generated for probing with a moderate packet loss probability, 10^{-4}, which is probably about right for most future real-time (e.g. VoIP) traffic, and follows IETF recommendations in Y.1541. The key point is that as the required packet loss probability (that must be evaluated by probing) becomes lower it becomes less and less possible to measure it accurately. These results have again shown that performance falls away dramatically as bandwidth decreases, as traffic burstiness increases and as load increases.

Our results show clearly that for low speed access links, for example, 128kbps, there are not enough buffer overflow events per measured hour (on average) for reliable measurements. This implies that such low speed links may not usefully

Figure 13. Schematic illustration of the relationship between probe instants and the actual queue size in a packet buffer, showing the difficulties of loss probing for bursty traffic



ever be probed for packet loss probability, even when the traffic is not very bursty. At the higher link access rates, 34Mbps and 150Mbps, there can be expected to be enough events, but only at relatively low utilisations. This has critical implications particularly for mobile and wireless networks.

Limitations of Active Measurement Schemes

It has been found that it may be very hard to measure network performance accurately using active monitoring. Measurement of mean delays should often work, but even here accuracy will be constrained by load and traffic burstiness with respect to the available bandwidth (see Figure 13). Use of active monitoring is further complicated by the fact that any monitoring measurements must be done over a period of time that reflects users' experiences—to see the Cisco Systems recommendations on this see (Cisco 2003). Furthermore the available bandwidth is likely to be cut up in ratios like 10%, 40%, 30%, 20% to support each of the CoS classes in the network (see Figure 1). Voice will likely be in the 10%, Web browsing into the 40%, e-mail into the 30%, and other low priority traffic (perhaps maybe ftp or telnet) in the 20%. And probing will have to be done

separately for each different CoS class, as well as (probably) for each different mean packet length. Also there are basically three methods of making active measurements: Cisco Service Assurance Agent (SAA), UDP echo and ICMP Ping, and the way in which these various active tests are handled by the router processor/OS in any Customer Premises Equipment (CPE) can vary, with the following effects:

- SAA is likely to offer relatively stable performance, i.e. not too much variation, as load on the router increases;
- UDP echo may also not be as good as SAA;
- A router is likely to treat ICMP as a relatively low-priority activity, and so this measurement technique may be disadvantaged, and so may not be very stable or accurate in measurement of performance at all. Additional instability or variance in these measurements may severely affect the value of active measurements.

Solutions and Recommendations

We have seen that passive monitoring schemes can provide excellent accuracy and minimal computational overhead; however they lack the global reach to support globally significant new services at 3G and beyond. Active monitoring schemes will suffer from inaccuracy when used across bandwidth bottlenecks (it is likely that, particularly mobile and wireless networks, will suffer from exactly these bottlenecks even when fixed-wire networks do not, and this may not be avoidable at this time). There are a number of possible paths to solving these drawbacks, and in this section we discuss these with a view to making the various possibilities more transparent to IT and networking managers, essentially so that managers can chose which option is the most economical for them, given their requirements and infrastructure.

These possible solutions fall into one of the following three categories:

1) Greater investment in larger bandwidth links, fully end to end including the access lines;
2) Adjusting the focus of the QoS metrics guaranteed under the SLAs; and
3) The development of more ‘intelligent’ measurement schemes.

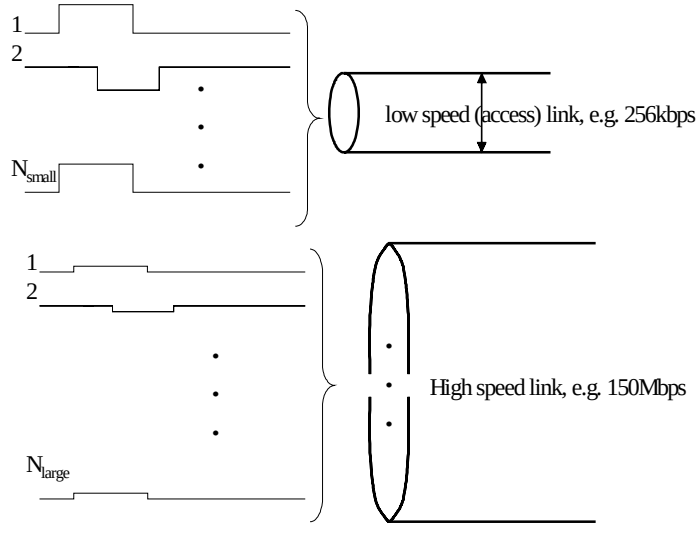
Buying More Bandwidth

One of the main reason buying more bandwidth is a possible effective solution is that more bandwidth diminishes the unfortunate effects of traffic burstiness. And very bursty traffic has been found, experimentally, on the current internet, for example see Willinger, et al. (1997) and Leland, Taqqu, Willinger & Wilson (1994). Prior work on measurements (Timotijevic et al., 2004) has strongly indicated that high burstiness always degrades measurement accuracy. This aspect of the problem is indicated diagrammatically in Figure 14: the greater the individual traffic source burstiness with respect to the channel capacity the worse the effect on the measurement accuracy.

Note that there are many definitions of "traffic burstiness"; some of these are quite complex in a technical sense, and rely on aspects of traffic theory that in general IT and networking managers will not have time and energy to investigate. Therefore we propose to use the simplest definition:

$$\text{Traffic burstiness} = \frac{\text{PEAK\_packet\_rate}}{\text{MEAN\_packet\_rate}}$$

Figure 14. Diagrammatic representation of the minimizing effect on traffic burstiness of higher bandwidth



which is generally fine, and is also easily understood intuitively. It can be applied to individual traffic sources, for example, a single voice connection, or equally well to a multiplex of many sources, for example, as may emanate from a large corporate headquarters.

The first solution then is to buy more bandwidth (where this is possible, and we emphasize that it may not be). This can be achieved by either over-dimensioning, or by the more sophisticated approach of greater traffic aggregation; by greater traffic aggregation we mean sharing a larger amount of bandwidth between a larger number of traffic sources. When implemented, either would have the effect of reducing the magnitude of the burstiness of the individual traffic sources with respect to the overall capacity of the links. Queueing theory tells us that the result will then be lower queue level fluctuations (packet scale queueing only) in the network buffers. These lower variance fluctuations can be much more accurately measured using a smaller overhead in active probing systems (and with less stored information in passive monitoring systems).

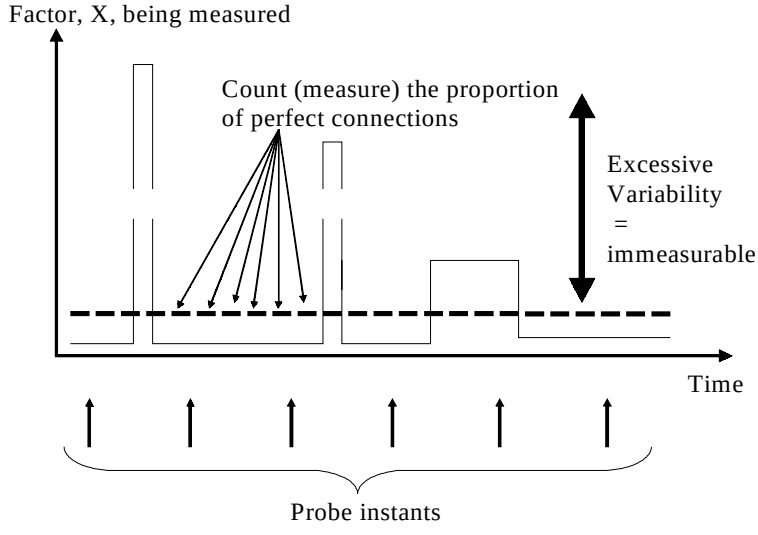
Adjusting the Focus of the QoS Metrics Guaranteed in the SLA

The second possible solution is in two parts: 1) avoid attempts to guarantee the twin aspects of network performance most usually guaranteed, mean packet delay and packet loss probability; 2) provide an actively measured guarantee of the proportion of connections (of the given type that are being measured) that will be either "perfect" or "imperfect" (where perfect means packet delays <T, and proportion of packets lost <p).

This is represented diagrammatically in Figure 15. "Perfect" for a connection will therefore imply delays, and loss probability, that is on average below the specified limits for both; it makes sense for these time units to have duration equal to the mean of the connections of interest. The counting process implies that each time unit will be either perfect or imperfect.

This may make better sense (than option 1) for mobile and wireless operators, where to over-dimension bandwidth may not be possible. This method of evaluating the proportion(s) could be used either with passive monitoring or active monitoring, but due to the worldwide nature of managed IP traffic, and the fact that this trend will continue with the growth of mobile and wireless systems beyond 3G, it seems most likely that it will have to be paired with (active) packet probing.

Figure 15. Schematic representation of second possible solution for inaccuracy in active monitoring



The Development of More 'Intelligent' Measurement Schemes

This is part of the (near) future. University (and other) researchers are working on schemes that will be scalable and require minimal overhead so that they can be used to optimally manage the measurement and monitoring aspects of IT systems. These will be particularly important for mobile and wireless (for 3G and beyond) as they will make smaller demands on limited bandwidth, and also will be designed to push complexity to the edges of the network (vital in mobile and satellite systems). It is foreseen that these should combine the best aspects of both passive and active monitoring; one (Schormans, 2004) currently at the proposal stage, would integrate passive monitoring at each node with probing in such as a fashion as to significantly minimize the need to use bandwidth (which is still likely be a very valuable commodity in mobile and wireless networks) for probes.

Conclusion

New business opportunities for mobile, wireless and fixed networks are going to grow to be worth in excess of USD 50billion by 2006. However these opportunities are for *managed* packet based services; this means that customers will require SLAs that are genuine, monitored and supported by enforceable guarantees. These SLAs will define (among other things) the QoS level that is being bought and sold, in terms of information loss and delay at the packet level. This means that managing new business opportunities (whether mobile, wireless or fixed networks) for services at 3G and beyond is going to require that the available measurement techniques, and their limitations, are fully understood.

To this end, we have investigated the two available techniques: passive and active monitoring. In this chapter we have shown that passive monitoring techniques have the advantage that:

- They would, if properly implemented provide excellent accuracy and minimal computational overhead.

However, these techniques also have the disadvantage that:

- It is necessary to have access to all the routers in all measured end to end paths, and this limits the scalability of any such system. The IT manager at any network cannot usually have access to delay/loss measurements at any other network, and this severely limits the ability to implement a meaningful business plan in a globalized world.

We have shown that active monitoring techniques have the advantage that:

- They can provide global reach, thus (theoretically anyway) the IT manager at any network can have access to delay/loss measurements right across the world. This would, if done well, allow such managers to implement a meaningful business plan in a globally significant way.

However, in this chapter we have shown that active probing also has the disadvantage that:

- Under many circumstances measured results will lack accuracy, meaning that guarantees are potentially being put into a legal framework that cannot

be properly supported. This leaves the way open for costly legal disputes about the level of performance that is actually being provided as opposed to that which is supposed to be being provided.

The first step to having a solid strategy for dealing with these potential difficulties is to understand the problem--this understanding is what we have tried to achieve in our in-depth discussions of passive and active monitoring. The second step is to propose methods whereby these difficulties, once understood, can be circumvented; we have proposed finally a number of methods which would allow accurate monitoring.

References

Analysys, (2003). Multi-carrier, multi-technology, SLA-based WAN services - a USD12 billion opportunity by 2006.

Cisco, (2003a). Measuring delay, jitter, and packet loss with Cisco IOS SAA and RTTON. From *http://www.cisco.com/warp/public/126/saa.pdf*.

Cisco (2003b). Service-Level Management: Defining and monitoring service levels in the enterprise. *Cisco White Paper*, available from: *http://www.cisco.com/en/US/products/sw/cscowork/ps2428/products_white_paper09186a0080091ba5.shtml*.

Cisco (2001). Support services – Service level agreements. Available at *http://www.uu.net/us/support/sla/*.

Cole, R.G., Kalbfleish, C. & Romascanu, D. (2000). A framework for active probes for performance monitoring. *Internet Draft*, Sept. 2000.

Floyd, S. & Jacobson, V. (1993). Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on networking*. 1(4), 397-413. 1993.

Kuzmanovic, A. & Knightly, E. (2001). Measuring service in multi-class networks. *Proc. of IEEE INFOCOM'01*.

Leland, W.E., Taqqu, M.S., Willinger, W. & Wilson, D.V. (1994). On the self-similar nature of Ethernet traffic. *IEEE/ACM Trans. on Networking*, *2*(1), 1-15.

Leung, C.M. (2003). Non-intrusive measurement in packet networks and its application. PhD thesis. University of London.

Leung, C.M. & Schormans, J.A. (2002a). Measurement-based end to end latency performance prediction for SLA verification. *Eleventh Interna-*

tional Conference on Computer communications and networks. 412-417.

Leung C.M. & Schormans, J.A. (2002b). Measurement-based queue length distribution estimation for power-law traffic. *Electronic letters, 38*(24), 1608-1610.

LightReading (2003). LightReading ITU Telecom World News Analysis, "Sprint Doubles Down on MPLS", Oct 14, 2003, *http://www.lightreading.com/document.asp?doc_id=41888&site=itu*.

Liu, E. (2002). A hybrid queueing model for fast broadband networking simulation. PhD thesis. University of London.

Morton (2004). IP/MPLS Network Performance and QoS. *IEE Telecommunications Quality of Service (QoS2004)*, Savoy Place, London, March 2004.

Nexagent (2003) *http://www.nexagent.com*.

Pitts, J.M. & Schormans, J.A. (2000). *Introduction to IP and ATM design and performance*. John Wiley & Sons.

Ribeiro, V.J., Riedi, R.H., Richard, G., Baraniuk, G., Navratil, J. & Cottrell, L. (2003). PathChirp: Efficient available bandwidth estimation for network paths. *Passive and Active measurement workshop PAM2003.*

Roberts, J.W. (1991). Information technologies and sciences: report of COST project 224. *Commission for the European Communities*, 1991.

Shalunov, S., Teitelbaum B. & Zebauskas M. (2001). A One-way Delay Measurement Protocol. *Internet Draft*, Feb. 2001.

Schormans, J.A. (2004). A proposal for combined passive and active measurement in packet networks. QMUL internal report, available from john.schormans@elec.qmul.ac.uk.

Schormans, J.A. & Timotijevic, T. (2003). Evaluating the Accuracy of Active Measurement of Delay and Loss in Packet Networks. *6th IFIP/IEEE International Conference on Management of Multimedia Networks and Services,* September 2003, Queens University, Belfast.

Schormans, J.A. & Pitts, J.M. (2001). From Erlangs to Excess Rate. Journal of the IBTE, Oct-Dec 2001.

Stewart, R.A. (2002). End-to-end delay analysis for small/medium scale IP networks. PhD thesis, University of London.

Timotijevic, T. & Schormans, J.A. (2003a). Bandwidth overhead of probe technology guaranteeing QoS in packet networks. *IEE Electronics Letters*, *39*(10), 816-818.

Timotijevic, T. & Schormans, J.A. (2003b). Bandwidth overhead of measurement technology used for guaranteeing QoS in packet networks. *Interna-*

tional Network Optimization Conference (INOC), Paris, France, October 2003.

Timotijevic, T., Leung, C.M. & Schormans, J.A. (2004). Accuracy of measurement techniques supporting QoS in packet based Intranet and Extranet VPNs. To appear in IEE Proceedings Communications *Special Edition on VPNs and Broadband Technology*.

Verma (2000). *Supporting service level agreements on IP networks*. MacMillan Technical Publishing.

Vleeschauwer, D. (1995). Experimental verification of the statistical independence of cell delays introduced in consecutive switches. *B-ISDN teletraffic modelling symposium*. 105-116.

Willinger, W., Taqqu, M.S., Scherman, R. & Wilson, D.V. (1997). Self-Similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking, 5*(1), 71-86, 1997.

Yankee (2003). Off the beaten pipe: Achieving end-to-end quality guarantees. *Convergent Communications Europe*, Graham Finnie.

Yates, D.J., Kurose, J.F., Towsley, D. & Hluchyj, M.G. (1993). On per-session end-to-end delay distributions and the call admission problem for real-time applications with QoS requirements. *ACM SIGCOMM symposium on communications architectures and protocols.* 2-12.

Appendix

Maximum likelihood Estimates of the value of the burst-scale decay rate using passive measurements.

Let x_1, x_2, …., x_n are n samples of the queue length seen by the packet arrival and the queue length is greater than the partition point. Therefore:

$$\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \bar{q}_{high}.$$

By assuming the queue length is continuous, the conditional probability of queue tail probability provided that it is greater than the partition point is given as:

$$Q(x \mid x > q_p) = \frac{c_b \eta_b{}^x}{\int_{q_p+1}^{\infty} c_b \eta_b{}^y dy} = \theta e^{-\theta(x-(q_p+1))}$$

where $\eta_b = e^{-\theta}$. Then, the likelihood function $L(\theta)$ of the n measurement samples is defined as:

$$L(\theta) = \prod_{i=1}^{n} \theta e^{-\theta(x_i-(q_p+1))}$$

By differentiating the above equation, the maximum-likelihood estimator θ_{ML} is given as:

$$\left.\frac{\partial \ln L(\theta)}{\partial \theta}\right|_{\theta=\theta_{ML}} = 0$$

Since:

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta} - \left(\sum_{=1}^{n} x_i - n(q_p + 1) \right),$$

therefore,

$$\theta_{ML} = \frac{1}{\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) - (q_p + 1)} = \frac{1}{\overline{q}_{high} - (q_p + 1)}$$

As $\eta_b = e^{-\theta}$, therefore, the maximum likelihood estimator of η_b is given as:

$$\eta_b = e^{\frac{-1}{\overline{q}_{high} - (q_p+1)}}$$

The expression is expanded by Taylor's series. By taking the terms up to the first order, we reach:

$$\eta_b \approx 1 - \frac{1}{\bar{q}_{high} - (q_p + 1)} \approx 1 - \frac{1}{\bar{q}_{high} - q_p}.$$

Section V

Security Issues

Chapter IX

The Smart Card in Mobile Communications: Enabler of Next-Generation (NG) Services

Claus Dietze, The EuropeanTelecommunications Standards Institute (ETSI), France

Abstract

This chapter gives an introduction into the smart card technology and its history by outlining the role of the smart card in mobile communications systems. The role of the smart card as a key enabler for services requiring or utilizing unambiguous user-identification is outlined. These services include multimedia and high-security services such as mobile commerce or mobile banking. Smart cards containing the described mechanisms provide the user with privacy and the capabilities to use information, personalized according to his needs, in a wide-spread system with a virtually unlimited number of services. Furthermore, the capabilities of the smart card to enhance services, to secure the issuers' revenues and to increase the usage of the services by providing a trustful platform for the user are described. Future evolutions and further developments of the smart card are illustrated, including how they pave towards new types of applications and services.

Introduction

The smart card in mobile communications is used both as a service platform and as a marketing instrument for the network operator. The (Universal) Subscriber Identity Module-(U)SIM—is the network operator's "business card" that is handed out to the end-user. The design of the artwork printed on the smart card, the packaging, and the functionality directly influence the positioning of the operator's brand in the market. The smart card as used in mobile communications enjoys a high reputation and is very important for the network operators. It does not only provide security and trust thus securing the revenues of the network operator, but is also a platform for value added services. Its importance for the network operator is impressively expressed by one of the world-leading network operators: they included the shape of the SIM into their corporate identity and use it within their logo and advertisement. Why this is absolutely justifiable will be outlined in the following chapter.

This chapter is divided into the following seven sections:

- The first section gives a brief introduction into the structure of the chapter and subject;
- The following section derives a dedicated definition for the term "smart card in mobile communications" to create a common understanding for the remainder of the chapter;
- The next section briefly lists and describes the main different specifications for smart cards used in today's mobile communications systems;
- The next section describes the technological and commercial evolution of the early SIM towards the next generation smart card (UICC, USIM, ISIM) used for 3G and further generations. Issues such as the technological constraints as well as the enhancements of the smart card are described and their impact on the market is highlighted;
- We then illustrate the role of standardizing organizations and explain the importance of standards for the success of a mobile communications system and the smart card in particular;
- The following section details the key capabilities of current and future smart cards and describes their importance for the creation of successful mobile services;
- And finally, we give an outlook on future evolutions of the smart card in mobile communications.

Defining the "Smart Card in Mobile Communications"

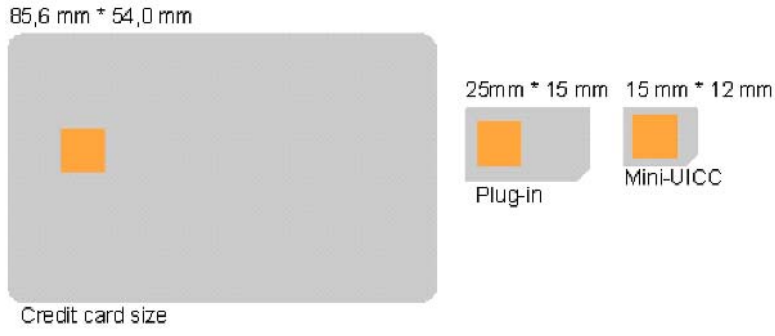
When searching the internet or other technical literature for definitions and explanations of the term "smart card", the following can be found: "The smart card is a credit-card size plastic card containing a micro-processor". Please also refer to the Smart Card Handbook for further information on smart card technology in general. For the context of this chapter and for the usage of the smart card in mobile communications, this definition is only to some extent true and need to be modified. A more appropriate definition of what the smart card in mobile communications actually is, is developed below by examining the features and applications implemented on and executed by it. The first indication on the purpose of a particular product may in many cases be derived from its name. This also holds for a smart card in mobile communications. As, of course, everybody using a Global System for Mobile communications (GSM™) phone knows, it has been called the Subscriber Identity Module or simply the "SIM". In fact, the capability to uniquely and securely identify one single user within the network has been one of the key features for the SIM since the beginning. How this feature was extended during the evolution of the SIM will be outlined later in this chapter.

Coming back to the above cited definition of the "smart card", the following precision are made below that focus on the use of the smart card in the area of mobile communications. The first precision concerns the first part of the definition, that is, "The smart card is a credit-card size...".

Simply looking at a SIM reveals that the actual size is much smaller than the size of a regular credit card. This reduction in size was felt necessary already at a very early stage in order to allow the smart card to be inserted in smaller and smaller devices, i.e. mobile terminals. In the respective specifications and standards this small size SIM is called Plug-in or ID-000. A further reduction in size was introduced into the standards in the beginning of 2004 (Mini-UICC) and show that the size of the smart card should not be part of the definition.

Another physical characteristic of the SIM or the next generation smart card for telecommunication is even more important. It was one of the crucial factors for the success of the SIM. The SIM is a token that can be removed from one terminal and easily put into another one. This allows the user to transport all personal as well as end user subscription related data from one terminal to another, for example when buying a new terminal. Even in the days of tri-band terminals allowing end-users to perform calls in almost every part in the world, new access technologies arise that again benefit from the "removableness" of the SIM. Wireless Local Area Networks (WLANs) could be mentioned as just one example of where the smart card may need to be removed from the mobile

Figure 1. Size reduction of SIM card



terminal and put into a WLAN device. Another solution will allow the smart card in the mobile terminal to be used for the authorization of the WLAN session that runs on a different piece of hardware.

The second modification of the definition is related to the part "... size plastic card...". Due to the reduced size of the SIM (see above) only a small piece of plastic is used to hold the module containing the micro-processor. From this point of view the material that is used to hold the module should not have any significance in the definition and could be left out.

The smart card is also described as "...containing a micro-processor" for the execution of functions implemented on it. Even though the micro-processor of the smart card is its heart, the soul of the smart card is or are the applications implemented on it. The applications characterize the smart card and make it useable in dedicated markets. In addition to the micro-processor, more and more memory capacity is required in the smart card. This memory is needed in order to contain multiple applications and value added services as well as complex configuration and provisioning parameters for services such as Multimedia Messaging Service, General Packet Radio Service (GPRS) connectivity or others. Rather than defining a smart card through its possession of a micro-processor, the smart card in mobile communications should be defined through its capabilities to execute applications and to manage specific types of data such as data which is personalized according to the individual users' needs.

As a result of the above observations the following is offered as a more accurate definition of the "smart card used in mobile communications": "The smart card in mobile communications is an individually personalized and removable authentication token. It is used to execute dedicated applications and manages specific data within the mobile communications system."

Smart Card in Mobile Communications Systems

Having defined the "smart card in mobile communications", we may now consider its role and the respective specifications for different mobile communications systems.

Besides the already frequently mentioned SIM used in GSM/EDGE Radio Access Network (GERAN) or USIM/UICC used in the Universal Terrestrial Radio Access Network (UTRAN), smart cards are also specified for other mobile communications systems—with the difference that in these systems the smart card is an optional component whereas it is mandatory in GERAN and UTRAN systems. The SIM specification, GSM TS 11.11, is the mother of almost every specification that was developed for other mobile communications systems. It was used as a basis for a smart card used in Terrestrial Trunked Radio (TETRA) systems that focus on emergency services as well as for the smart card used in the Digital Enhanced Cordless Telecommunications (DECT™) system and the Code Division Multiple Access (CDMA) system, just to mention three.

System Architecture

The smart card in mobile communications represents one crucial component in the network infrastructure. It mainly plays two different roles: to provide secure access and provisioning to the network and to provide additional value added services for the end user and/or the network operator. Both roles are outlined in the following two subsections.

Network System Component

The smart card as a network system component is primarily used to authenticate the subscriber to the network as in GSM or to mutually authenticate both the subscriber and the network as in UTRAN. The following Figure 2 shows the simplified authentication procedure in a UTRAN system and illustrates the role of the smart card. All authentication relevant computations on the user side are executed inside the smart card. The secret key K used for the computations never leaves the smart card and is safely stored in the secure memory area of the chip. The smart cards' counterpart for the authentication in the network is the Authentication Centre (AuC). The AuC also possesses the secret key K and is therefore able to calculate the expected response of the smart card (RES). By

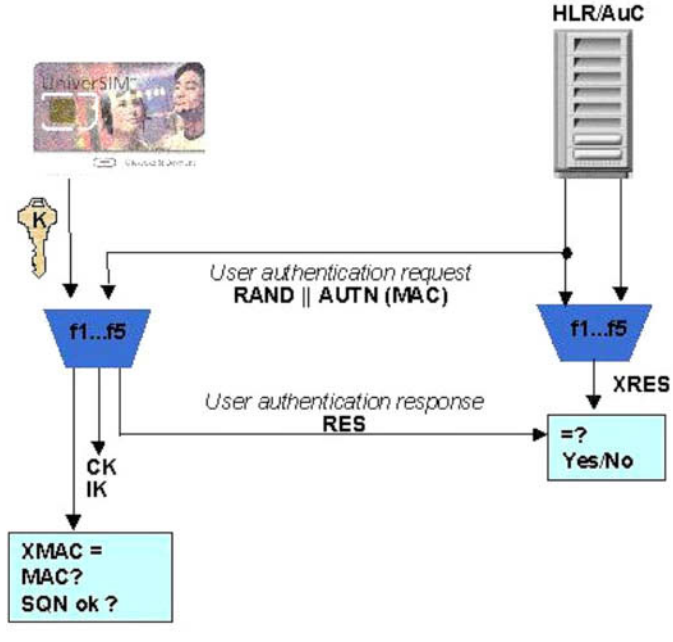
performing the respective calculations (using functions f1 to f5) on both the user and the network side and by comparing the results of the calculations with the values submitted by the respective counterpart, a decision on the network access can be made by both parties.

Connectivity parameters such as Multimedia Message Service (MMS) parameters and service related information such as preferred network identities are stored on the smart card. This information is used by the handset to access the appropriate networks and services. Putting the connectivity parameters on the smart card allows the user to access the network operators' services independent of the terminal used.

Value Added Services System Component

The Card Application Toolkit (CAT) provides a common set of commands that was derived from the SIM Application Toolkit. This framework enabled the development of additional applications that can be put on the smart card. These applications are in general linked to a network entity in the background server system. This server system is not necessarily located at the network operators premises and could also be operated and maintained by third party service providers such as news content providers or banks. An end-to-end (server-to-

Figure 2. Authentication procedure



card) communication channel for applications stored and executed on the smart card can be established. Dedicated security mechanisms that are defined in the respective specifications (see TS 03.48 and TS 102 127 for further information) provide a secure channel between the smart card and the network entity. Application relevant data can be encrypted and will only be available for the application server and the application in the smart card.

Figure 3 illustrates today's system architecture for SMS based services in a 2G network. The SIM issues a short message that contains dedicated and application specific information. This information can optionally be encrypted. The data part contained in the short message is then routed via the Short Message Service Center (SMSC) to the application server. The application server (optionally decrypts and) interprets the contents of the message and triggers the relevant behaviour such as downloading further information to the SIM.

GSM/EDGE Radio Access Network (GERAN) and Universal Terrestrial Radio Access Network (UTRAN)

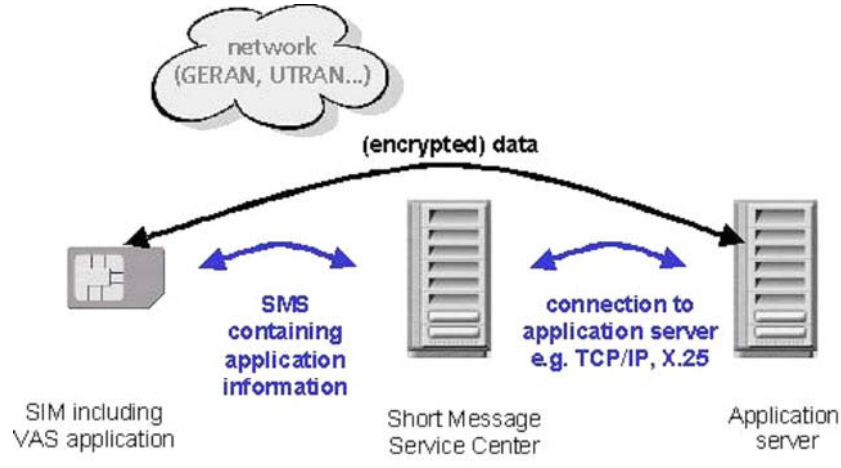
The main (and most successful) mobile communications system that currently involves a smart card is certainly GERAN (GSM). The Subscriber Identity Module—the SIM—is specified as a mandatory component in the whole system. Therefore, all mobile terminals have to include a smart card reader that is able to contain the smart card. The technical core specifications that are being maintained and further developed by the respective standards organizations are:

- TS 11.11 the "Specification of the SIM-Mobile Equipment Interface" that started in the early days and was developed further until Release 99; and
- TS 51.011 the "Specification of the SIM- Mobile Equipment Interface" which is the corresponding document that only exists in Release 4.

The members of the 3rd Generation Partnership Project (3GPP™ concluded to freeze the SIM specifications at Release 4 and to include further enhancements to the SIM into the evolutional counterpart of the SIM in UTRAN—the Universal Subscriber Identity Module (USIM). The main specifications for the USIM are:

- TS 31.101 defining the "UICC-Terminal Interface; Physical and Logical Characteristics" of the smart card from Release 99 onwards; and
- TS 31.102 specifying the "Characteristics of the USIM Application" from Release 99 onwards.

Figure 3. Simplified value added services architecture



A comprehensive overview of how the two smart card applications, the SIM and the USIM, interwork and how they could be combined on one single smart card that could be used for GERAN as well as UTRAN is also part of these specifications. A related technical report, TR 31.900 "SIM/USIM Internal and External Interworking Aspects", describes these aspects in detail. It outlines further the key role that the smart card plays for a network operator that migrates from 2nd Generation to 3rd Generation networks.

Code Division Multiple Access (CDMA)

The second major mobile communications player in the industry is CDMA. 3GPP2, the equivalent to 3GPP for the specification of CDMA, is responsible for the definition and maintenance of the specifications for CDMA2000. 3GPP2 is a partnership project consisting of the following partners: Association of Radio Industries and Businesses (ARIB-Japan), China Communications Standards Association (CCSA-China), Telecommunications Industry Association (TIA-North America), Telecommunications Technology Association (TTA-Korea) and The Telecommunication Technology Committee (TTC-Japan).

In CDMA networks the smart card is optional, that is, all parameters such as subscription data, network settings, and security functions are stored and

personalized into the handset. Due to the absence of the smart card CDMA has been facing the following issues:

- No roaming to GSM/GERAN networks for the subscriber, which means a new card and a new terminal is required when travelling abroad;
- Difficult handset exchange due to difficult transfer of personal and subscription related data from the existing to the new terminal;
- Difficult manufacturing process for terminals due to personalization of each of the terminals with user individual data; and
- No SIM Application Toolkit based services and applications available that can be easily transferred from one terminal to another.

These limitations lead to strong requests for a smart card. This request was also supported by the CDMA Development Group (CDG). Therefore 3GPP2 specified--also based on the SIM specification in TS 11.11—the requirements for the Removable User Identity Module (R-UIM) in technical specification C.S0023-0. The R-UIM is an extension of the Subscriber Identity Module (SIM) capabilities, to enable operation in a radiotelephone environment. Examples of this environment include, but are not limited to, analogue CDMA. The specification is based on the SIM specification and includes additional commands and responses necessary within the context of CDMA. The introduction of the R-UIM allows subscriber to "plastic roam" (by switching the smart card) between CDMA and GSM networks.

Digital Enhanced Cordless Telecommunications (DECT)

The European Telecommunications Standards Institute (ETSI) has developed a total of more than 30 publications (technical specifications, technical reports or technical base for regulation) for the Digital Enhanced Cordless Telephony (DECT). The first system became operative in 1992. The DECT Authentication Module (DAM), based on the SIM specification TS 11.11, was specified in the early 1990's. This enabled a smart card to be used as an authentication token for the end user to be introduced and several specifications were approved:

- ETSI ETS 300 331 ed.1 (1995-11): Digital Enhanced Cordless Telecommunications (DECT); DECT Authentication Module (DAM); and
- ETSI ETS 300 825 ed.1 (1997-10): Digital Enhanced Cordless Telecommunications (DECT); 3 Volt DECT Authentication Module (DAM).

Mobile network operators that also operate as fixed line operators have experienced the advantages of a smart card in a mobile communications system and are seeking to adapt services that are being used in the mobile world also to the home and fixed line environment. This is an interesting phenomenon. Even though the fixed line telephony has been used for far longer than mobile telephony, the standards for the features and the look and feel of terminals are being set by the mobile industry. This trend will also impact services being used in the mobile world that are going to be introduced in the fixed line environment. Features such as short messages and multimedia messages are being introduced into the "regular" phones, which are mainly DECT phones. The inclusion of a smart card adding further advantages to them seems to be a logical evolution of today's DECT phones. A DECT Local Area Network (DECT LAN) also appears to be an area where secure user authentication and therefore the DAM could be essential.

TErrestrial Trunked RAdio (TETRA)

TETRA is an open digital standard developed by ETSI that describes a common mobile radio communications infrastructure. This infrastructure is targeted primarily at the requirements and needs of public safety groups such as police and fire departments. The requirements comprise the need to rely on fast and accurate file communication even if no network coverage is given. These groups have been high-end users of private/professional mobile radio (PMR) or public access mobile radio (PAMR) technology. Based on digital, trunked radio technology, TETRA is targeted to be the next-generation architecture and standard for current, analogue PMR and PAMR markets. As TETRA is targeted to public safety groups, privacy and confidentiality of the data and voice communication is essential.

Again based on the SIM specification TS 11.11, a smart card, the TETRA SIM, was specified for the usage in the TETRA system. The TETRA SIM used for user authentication and storage of configuration data and phonebooks was specified by the respective ETSI technical body in the mid 1990s in:

- ETR 295 "Radio Equipment and Systems (RES); Trans-European Trunked Radio (TETRA); User requirements for Subscriber Identity Module (SIM)". This Technical Report describes the high level requirements that have to be fulfilled by the TETRA SIM. ETR 295 was published in 1996 and indicates that the SIM is an optional device within TETRA Mobile Stations (MS): thus this ETR does not preclude the implementation of MS without a SIM.

- ETS 300 812 "Terrestrial Trunked Radio (TETRA); Security aspects; Subscriber Identity Module to Mobile Equipment (SIM - ME) interface", Edition 1, was the first version of the TETRA SIM specification. It was published in 1998.
- ETS 300 812 Edition 1 was revised to EN 300 812 version 2.1.1, published in 2001.
- ES 200 812 part 1 (Physical and logical characteristics) and part 2 (Characteristics of the TSIM application), published in 2002, are the Edition2/Release 1 versions of TS 100 812-1 and TS 100 812-2 (see chapter 3.2) and are technically identical. They are intended to ensure a smooth transition to TETRA Release 2.

In line with the evolution of the smart card to a multi-application platform, the TETRA SIM specification in TETRA Release 2 split off the physical characteristics and concentrated on the definition of a TETRA SIM application. The aim was to bring convergence with the Universal SIM (USIM), to meet the needs for TETRA specific services whilst gaining the benefits of interworking and roaming with public mobile networks such as GSM, GPRS and UMTS™ TETRA Release 2 started in September 2000 in order to enhance the services and facilities of TETRA. Release 2 of the TETRA SIMs (TSIM) aligns the TETRA SIM application with 3GPP. Release 2 was previously known as Edition 3 of the TETRA SIM specifications.

Evolution of the Smart Card in Mobile Communications

Evolution Of Smart Card Hardware

In 1988 the idea of introducing a smart card in mobile communications led to the first conception of the SIM. In these early days microprocessor chips that could be embedded in a smart card had only a very limited amount of memory that did not allow storage of large data sets. Therefore the first functionality implemented on the SIM focussed on the authentication algorithm. From the beginning the smart card was used to provide the end user with a secure token that enabled her/him to access the GSM service. For network operators the SIM allowed from the very beginning to manage billing and other information about their subscribers. The security features of the smart card in mobile communications were constantly enhanced and developed.

Data storage capability was discussed as early as May 1988. It was introduced for the provision of services including Short Message Service, Advice of Charge, Abbreviated Dialling Numbers and Public Land Mobile Network (PLMN) selection, but memory was a major constraint. Initially, the total memory capacity of a SIM was about 10 kB for both the operating system (read only memory—ROM) and data storage (programmable memory—EEPROM). Only in the mid-1990's did larger chips with 8 kB of programmable memory become available. Today, chips provide 128 kB to accommodate (programmable) data and applications, and at least the same amount for the operating system and other ROM-based applications. New technologies such as flash or floating EEPROM technology will soon enable the mass deployment of even larger chips, with 1 MB or more of programmable memory—today's expectations with today's available technology reach to an estimated maximum of about 16MB memory within the today's (U)SIM.

Considering the fact that comparing the 128KB with a 1MB smart card is already an increase in available memory of about 800%, the question arises how to actually use up all memory. The answer is given by the network operators: The SIM is the property and under full control of the network operator. Therefore all important information and network connectivity parameters as well as service related information should be stored and managed by the SIM.

Also, the separation of programmable and read only memory could be dissolved in favor of the programmable memory. New technologies allow the storage of the operating system in a special one-time programmable memory that can be loaded onto the smart card during the production process. This has the advantage that packages of features, tailored to the network operators' needs, could be loaded onto the smart card rather than burning the complete set of features into the ROM part of the chip. Figure 4 illustrates the traditional split of the memory into ROM and EEPROM. It shows the separation of the memory into an operating system area (ROM) and the applications and data area (EEPROM). The provisioning of the memory is fix and cannot be changed after the production of the chip. New technologies such as flash or "floating EEPROM", as shown in Figure 5, provide a more flexible memory management. The operating system, applications and data share one common memory pool. This pool is not split and can be managed according to the network operators needs.

Features that resided in ROM by default and that are not required by the network operator could be removed, freeing more memory for additional applications that are requested by the network operator or their customers. Based on this technology development cycles and thus time to market will be reduced. Providing new operating systems and features in the ROM mask of a chip is a time consuming task. From the finalization of the development by the smart card manufacturer until the reception of first commercial samples of the chip takes

between three and six months. Making the development of these components independent of the silicon manufacturer saves valuable time for the introduction of new features and enhancements for the network operator.

Along with the evolution of the memory, SIM processors have followed a similar growth pattern. From the first eight bit processors to today's chips which have the computing power of 16 bit and 32 bit processors. Dedicated crypto co-processors allow the powerful execution of asymmetric crypto-algorithms. Asymmetric crypto-algorithms enable the smart card to play a major role in the application of Public Key Infrastructures (PKI) to regulate the use of certificates for authentication in e-transactions.

Evolution of Software and Operating System Architecture

In parallel to the evolution of the chip hardware the operating system as well as the capabilities of the operating system evolved. SIM development reached a major milestone in 1996 when ETSI approved the first technical specification for the SIM Application Toolkit (Technical Specification TS 11.14). This specification defined a set of commands and procedures to enable the card to contain applications specific to the issuer (the network operator), allowing the operator to introduce a wide range of new services including information and location based services, banking and Internet access. Today, the smart card in mobile

Figure 4. Traditional memory separation scenario

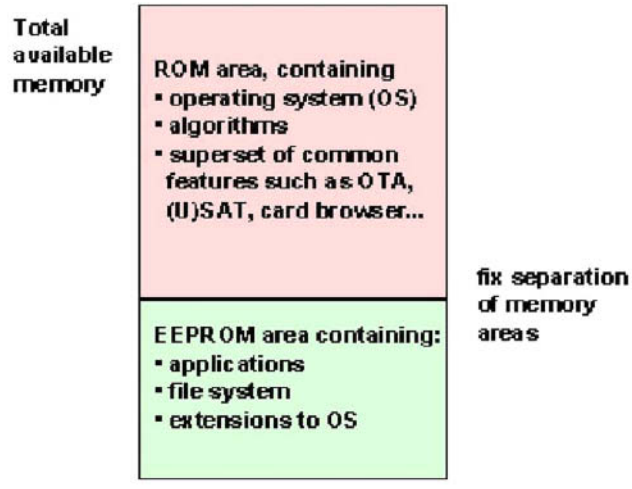
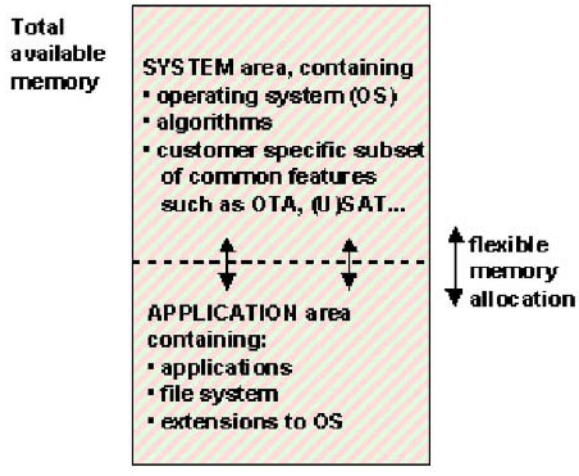


Figure 5. Flexible memory separation scenario



devices operates and manipulates menus and services and authenticates users for service access. The high level of security which the SIM offers means it can secure financial transactions over a mobile phone, enabling mobile commerce. Work continues to introduce additional features. The introduction of the SIM Application Toolkit was the first step towards the opening of the SIM for additional and customer, that is network operator-specific applications. These applications had to be developed by specialists that had access to the operating system of the card and that were able to low level code the application (hard code) into the EEPROM of the SIM card. Therefore development was time consuming and expensive. Only smart card manufacturers were able to implement such applications on their cards. As different smart card vendors had their own (different) smart card operating system the development had to be done as often as the network operators have smart card suppliers.

A need for an Application Programming Interface (API) arose that, as for other computer systems, allowed applications to be developed by almost anybody and rapidly. It should be possible to run these applications interoperable on any smart card, independently of the smart card vendor. To fulfil this request, ETSI approved a high level requirements specification to introduce such an API in TS 02.19. The first API fulfilling the requirements of TS 02.19 was approved in 1998 (TS 03.19) and was based on the Java Card™ specifications developed by the Java Card™ forum. The new Java Cards™ were supposed to allow development of applets in a quick and cost efficient way. However, reality showed that even though Java™ seemed to be the right way to go, all aims could actually not be

achieved in the beginning. Interoperability on a 100% technical level needed to be slowly established and proven in the market. Using Java™ technology does not automatically guarantee interoperability.

Smart card vendors offer development kits for their Java Card™ products (such as Sm@rtCafe Professional of Java™ Mobile Application Designer (JMAD) from Giesecke & Devrient, Cyberflex from axalto or GemXPlore from Gemplus). New services and applications can comparatively easy be developed, leading to an increased demand of the smart card vendors' Java Cards™. Actually, only very few application developers could profitably enter the market and offer new smart card applications to network operators. The reason is rather simple: network operators have been used to getting applications almost for free from the smart card vendor. Due to the history of application development on SIM cards and the competitive situation in the market, most of the smart card vendors continued to offer the applications for free or at least for a far less cost than an independent application developer. Nevertheless Java™ proved to be very important within the industry especially with regards to time to market and flexibility for the network operator. It allows the one time development of services utilizing SIM Application Toolkit commands. After perambulating the learning curve and due to the help of external organizations such as SIMalliance and ETSI, the major smart card vendors managed to provide truly interoperable Java Cards™. That means that applets that have been developed by one party should run equally on cards of any other party that also followed the appropriate specifications for the development of the applet and the Java Card™.

With the advent of 3G, the SIM has evolved to become the "USIM" (the Universal Subscriber Identity Module). Whereas the SIM is the definition of a complete smart card including the physical and logical characteristics (i.e., the plastic and the chip), the USIM is defined as being an application. The USIM resides on a smart card that is to be implemented according to the technical specifications for the smart card platform, the UICC. Figures 6 and 7 below illustrate the different concepts. Figure 6 shows the traditional SIM architecture whereas Figure 7 illustrates the new modular concept. A smart card for 3G mobile communications consists of the UICC containing at least one USIM application.

This new approach of separating the physical and logical characteristics from the functions and applications enables the smart card to become multi-application capable. It is like having a PC with a basic operating system (being the UICC) and the Internet explorer managing the access to the network (being the USIM). The USIM provides features which equip it to play a key role in crucial aspects of 3G such as managing security access, virus intrusion, customer profiles, mutual authentication, downloading, and a new phonebook allowing the management of additional information such as fax numbers and e-mail addresses. The

USIM also has the ability to store applications for network services, offering, for example, pre-paid service activation and control, information services, directory services, mobile banking, and ticketing. See later in this chapter for further information.

The UICC allows users access to global roaming by means of their smart card, irrespective of the radio access technology used. It is able to contain multiple applications, allowing smooth roaming and interworking between different services and networks, whether GSM, the new Wideband Code Division Multiple Access (W-CDMA) or other networks; the handset will be able to access a portfolio of services and applications available to users via their user profiles.

The UICC's revolutionary ability to handle true multi-applications, providing the platform for independent applications which can even run in parallel, present an interesting test of both the ingenuity of marketing experts and the ability of different market sectors and manufacturers to co-operate in the deployment of services. For instance, telecommunication operators are able to issue UICCs containing both a USIM and an electronic purse.

The UICC also contains new features such as enhanced security, further Application Programming Interfaces (APIs), new form factors, enhancement of the interface speed, access to shared multimedia sessions through the Internet Protocol Multimedia Subsystem (IMS), and, of course, backwards compatibility for network operators, allowing them a smooth transition from 2G to 3G.

Figure 6. Single application smart card

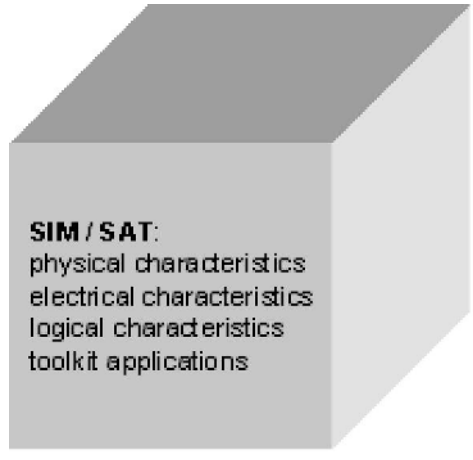
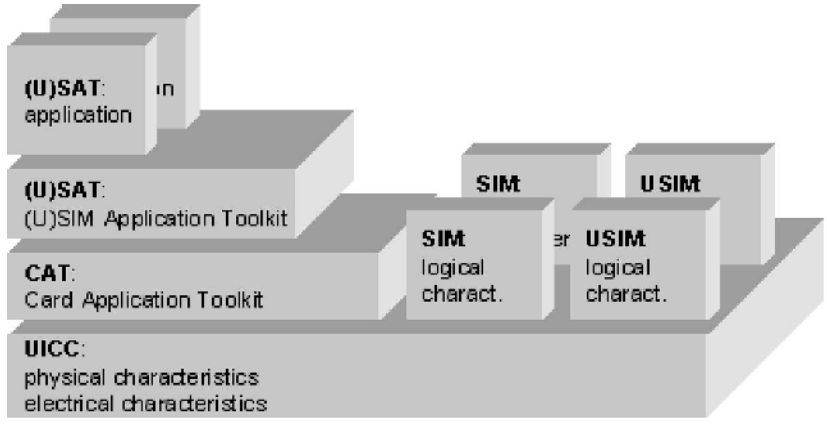


Figure 7. Multiapplication smart card



Evolution Milestones

Figure 8 provides an overview on major steps achieved during the evolution and development of the smart card in mobile communications, in particular the SIM and USIM.

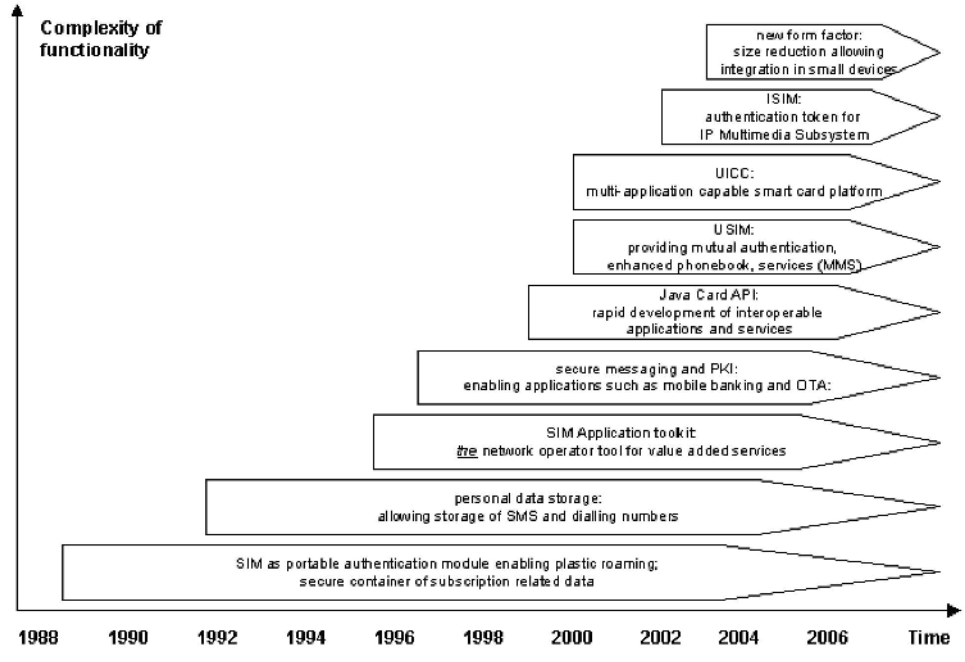
Supplementary information on the evolution of the SIM can be found in the article "The Subscriber Identity Module, Past, Present and Future" of Dr. Klaus Vedder in [1].

Further evolution of the smart card in mobile communications is certain as the market for GERAN and UTRAN and therefore the market for the smart card used in such systems continues to grow. To give an indication on the market size: from its first delivery to the present, far more than 2 billion SIM cards have been delivered.

Standardization

When talking about markets, features, services and business one of the main drivers for the business is in most of the cases underrepresented: standardization. Standardization is a subject that is not very often discussed in connection with services, marketing, and business but is actually one of the most important

Figure 8. Evolution of the SIM



factors in the game. Standards aim to establish systems that make interoperable solutions among different manufacturers possible and therefore also reduce cost. This drives the acceptance of the system in the market. Without standardization systems like mobile communications could not have become reality on such a large scale. One can without doubt say that standardization has been the driver for the success of GSM and the SIM.

Today's purchasing decisions are more and more based on whether the solution or product is implemented according to a particular set of standards. The availability of specifications and standards was a key factor for the success of the SIM as well as the ability to provide the network operator with a standardized subscriber authentication method. The revenue of the network operator depends to a large extent on the security of the authentication and thus the billing system. It is crucial to rely on a defined method for the subscriber authentication and in addition a set of harmonized and standardized security features. The production of the relevant standards has been undertaken by ETSI.

ETSI is the recognized European Standardization Organization for telecommunications and related fields of broadcasting and information technology. From its inception in 1988, the Institute has been at the leading edge in setting security standards. It achieved an outstanding success with the standardization of the Global System for Mobile communications (GSM), which included authentication, anonymity and customer privacy. This represented the first full, worldwide, commercial deployment of encryption and smart cards, and ETSI's standardization of the SIM for GSM has helped make it the most widely deployed smart card ever.

With the closure of ETSI committee SMG9 (Special Mobile Group 9) in the year 2000, which was responsible for specifying the SIM, and the establishment in December 1998 of the 3rd Generation Partnership Project (3GPP), of which ETSI is a founding partner, ETSI's work on the SIM application, i.e. the non-platform and non-generic part, was transferred to 3GPP's Technical Specification Group TSG-T3. T3's task is to further evolve the SIM application to meet the demands of the new 3rd generation (3G) mobile network.

Further smart card-related work continues within ETSI's Smart Card Platform Project (EP SCP), which was founded in 2000 as the successor of SMG9. EP SCP is the focal point in ETSI for the standardization of the common Integrated Circuit (IC) card platform for 2G (e.g. GSM) and 3G mobile communications systems, the UICC. (As described earlier, the UICC comprises the platform specifications implemented on the smart card, together with all resident applications based on them. It also contains the USIM as an application for access to the 3GPP system, and/or the R-UIM application for access to the 3GPP2 system.) The work of EP SCP provides a common platform on which others, including organizations from the financial sector, can base their system-specific applications.

In addition, EP SCP has worked to make the specifications for GSM independent of the bearer network and, as part of the process, new deliverables have been approved. These specifications provide standardized security mechanisms for the interface between a network entity (e.g. a toolkit application) and an entity in the UICC. They also make available a standardized method for the secure, remote management of files and applications on the UICC. A requirements specification for a generic API, the UICC-API, was approved in 2002, and the work on a corresponding functional and architectural specification was completed in May 2003. This allows the rapid development of interoperable card-based applications (applets).

The upshot of these developments is that, while the SIM retains its original function of authentication, it has evolved to become both a service platform offering multiple value added services and a multi-application platform providing interoperability and interworking between different access technologies.

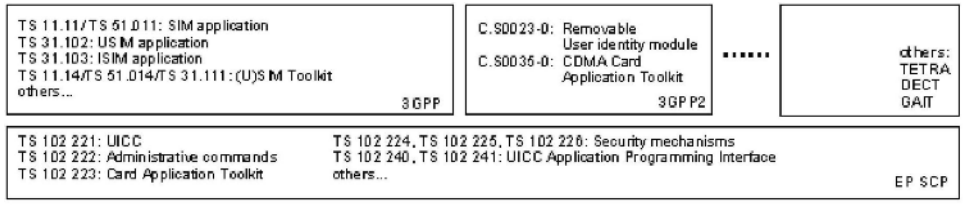
Further advancements of the UICC platform and the applications based on it will create new possibilities. For example, the new form factor specified by EP SCP allows the development of smaller devices, for example, for data transmission only, and offer additional communication and financial applications. New chip technology and the continuation of standardization work will advance the capabilities of smart cards to the point where they are really personal mini-computers. This development in turn will enable new applications, which will drive the growth of other new technologies, particularly 3G mobile.

The following Figure 9 illustrates the responsibility and interdependencies of the different standards producing bodies. ETSI project Smart Card Platform (EP SCP) is the initiator of common and application independent platform specifications. It is responsible for the creation and maintenance of these platform specifications on which other committees such as 3GPP and 3GPP2 base their system and application specifications such as the SIM, USIM and CDMA Card Application Toolkit.

Institutes such as ETSI, and partnership projects (such as 3GPP and 3GPP2) between different institutions, are important for the generation of specifications and standards. It is very easy to imagine that a set of specifications for a system such as 3G is a huge effort and that the different parties that are going to base their products and services on these specifications heavily contribute to such standards. Hundreds of member companies from all over the world develop their products and services based on the standards they produce under the umbrella of the institutes and partnership projects.

Clearly, decision-making and Intellectual Property Rights (IPRs) are or could become quite an issue among the member companies. For these reasons ETSI and 3GPP respectively, created working procedures and established an IPR policy under which the specifications are produced, with the aim of minimizing

Figure 9. Organization of standards and responsibility of standards bodies



these issues. This framework is accepted by the different parties when they become a member and sign the membership agreement of the institute.

The development of specifications/deliverables is consensus driven. The way decisions are made is well defined in the rules and working procedures. Given the diversity of interests, it is remarkable that, in the long history of standardization, the use of the ultimate decision-making tool where no consensus can be reached, — the vote — is very rare. In the case of standardization for smart cards in telecommunication, up to now only one single vote has been; this was in December 2003 in ETSI Project Smart Card Platform, and concerned the introduction of the new form factor for the smart card.

Last but not least, the fruitful collaboration between the different standardization organizations and their collaboration with the industry partners ensures that the standards meet market requirements.

2G, 3G and NG Services Based on the Smart Card

Smart card technology has consistently and reliably provided solutions to current and future requirements and challenges. The smart card evolution cycles become ever shorter, whilst the capabilities of the smart card increase as dramatically as for any other product in the information technology domain. Even though those capabilities increase, the key features of the smart card in mobile communications remain the same, yet are subject to enhancements and evolution. This section explores those key features and discusses the services and applications that utilize these key features. It concludes with a review of some of the work areas that are currently being considered by the smart card specification groups. They are listed and briefly described to illustrate just some example areas of further enhancements and future trends.

Security Mechanisms

One of the main features of the smart card is to provide security: security when authenticating the subscriber to the network (and also the network to the subscriber) by performing cryptographic operations as well as security mechanisms to protect security relevant data from unauthorized access. Security means trust and the level of trust depends on the relevance of the performed transaction (and the connected value). The end user need to trust the stability and

security of the system when performing dedicated operations that in particular have some financial impact. These services include: digital signature to sign commercial transactions; mutual authentication of subscribers and service providers to get access to special (with cost) services; storage of secret user data such as keys or PINs to grant access to the secret or personal data on the smart card; secure authentication to access Virtual Private Networks for company subscriptions; providing features to secure copyrights (Digital Rights Management) when for example transferring valuable content from one terminal to another terminal; preventing theft of mobile terminals by connecting the smart card and some secret data to the terminal; and many more.

As the smart card is the token which uniquely identifies a subscriber in the network it is also an ideal container for all subscriber individual data that is not related to the authentication procedure. Such data could either be personal data stored on the smart card by the end user or data that is used by specific applications (e.g. application toolkit services) running on the smart card and managed by the card or application issuer. Personalized applications that behave according to the individual needs or characteristics of every single end user can be envisaged. Personalized applications are commonly used within the Internet. Once subscribed to a bookstore and as soon as some articles have been purchased, an individual customer profile is created. This customer profile will then initiate the application to behave in a personalized way when visiting the bookstore again. Starting with a personal greeting reaching to a tailored product offering for the customer to more easily find the appropriate book. Services of this kind are also introduced to the mobile world. The creation of Web pages for the terminal that can be displayed on the screen mirror the Internet behaviour on the mobile world.

As smart card applications are used for billing services the security of the smart card has also been subject to attacks. Appropriate countermeasures were developed by the industry to fend off such attacks, for instance, power analysis on the chip and measurement of the required power consumption during dedicated operations could be used to determine security relevant data. Random waiting periods to the processor and the use of processors with constant and steady power consumption were some of the introduced countermeasures.

Personal Identity and Data Management

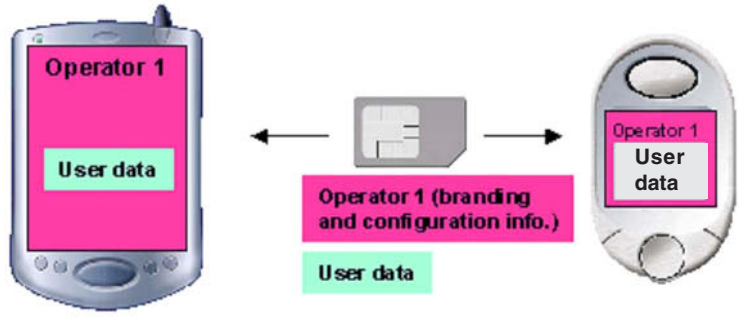
The smart card is a removable device allowing an easy transfer of personal data (such as phone numbers) from one mobile terminal to another and of course to allow plastic roaming (in different networks, see also sections above).

Life cycles of terminals and smart cards, and the necessity to change the terminal from time to time for reasons of functionality, forces the end user to also transfer all personal data from the existing terminal to the new terminal. This has to be done whilst keeping all relevant authentication parameters and mechanisms, as access to the network still needs to be provided. In order to ensure a smooth migration from one terminal to another without any tools, the smart card and all personal data stored on it can be easily ported from one terminal to another by simply taking it out of the reader and putting it into the reader of the new terminal. With the increasing number of services and a resulting increasing amount of personal data to be stored, the migration became more and more inconvenient and complex for the user if the data was stored in the terminal. Dedicated connectivity parameters and settings for the network and some services can easily be ported as well as additional keys and PINs for high security applications such as mobile stock trading and virtual private network access. The same applies to personal data such as phone numbers, email addresses and templates stored on the smart card. Equally, applications that are stored on the smart card remain available to the end user when the exiting terminal is replaced by a new one. For the network operator, branding information and connectivity parameters that are stored on the smart card are available on any terminal.

Value Added Service Platform

The smart card is a platform for value added services. As mentioned above, a standardized framework to develop services exists for the smart card. This framework is supported by virtually all terminals in an harmonized way. The framework comprises the Card Application Toolkit (SIM application toolkit in 2G and USIM application toolkit in 3G) allowing the creation of applications that utilize the capabilities of the terminal with regards to the user interface (keyboard, display...) and the radio interface and bearers (set up call, set up GPRS connections...). Value added services starting with simple information services such as weather information or horoscope can be rolled out. But also more complex applications involving 3rd parties (such as a financial institution or a WLAN operator) are generated and can be easily put on the smart card. The SIM Application Toolkit and/or USIM Application Toolkit provide mechanisms to also retrieve location information and allow the development of location-based services. Dedicated routines and procedures on the smart card are defined for the execution of call control features that allow the restricted or dedicated use of subscriptions related to the calls that can be set up. In order to offer flexible solutions, a dynamic approach exists for the interpretation or execution of applications on the smart card. A smart card browser was defined that allows development of applications that are stored on a Web server and that are

Figure 10. Personalization of terminal by smart card



dynamically downloaded to the smart card and interpreted accordingly. Besides the evolution of the smart card hardware, the toolkit is one of the areas that has developed most rapidly in order to reflect the new technologies and make them available from and for smart card applications. All major smart card vendors offer development tools to generate applications based on the card application toolkit standards by even providing a graphic interface to the developer. Adding the Java Card™ API to the smart card allowed an even faster and easier way to develop applications for the smart card (see also section 4.2).

Today's smart cards offer true multi-application capabilities. New generation smart cards are multi-application capable. This means that completely independent applications can run on the same physical smart card. Whereas in the "old" days the smart cards contained one application and the SIM was a specification of both the logical and the physical characteristics, the new concept is to separate the physical aspects from the application specific characteristics. This allows the combination of completely independent applications on the same smart card platform. For example, a credit card application could be combined with a USIM application on the same smart card. These applications are completely independent from each other. That means that the credit card application (e.g. EMV) does not rely on the mechanisms provided by the USIM application. Whereas today a banking application uses the bearers offered by the mobile communication system, an infrared interface could be used for the payment transaction if the credit card application was separate. The SIM application (remember, the SIM is now an application and not a complete smart card anymore) or any other application may be located on the same card. The maximum number of applications running on the same smart card is only limited by the available

memory space. The multi-application capability of the smart card enables network operators to prepare for a smooth transition from 2G to 3G by providing both the 2G and the 3G application on the same smart card. This concept will permit the network operators to also smoothly migrate from 3G to any further generation system. Considering the technical capabilities and the increase in memory available on the smart card, concepts are developed to rent out a specific part of the memory on the card to a third party. This allows them to offer their services to the customer base of the network operator. Any third party could provide applications such as ticketing or banking in case the network operator itself decided to concentrate on other core business rather than offering all services and applications by himself.

Service Deployment and Management

The smart card is owned by the network operator. Even though the smart card is physically possessed by the end user, it is still the property of the network operator. As the smart card is the only token being under full control of the network operator, the network operator is independent from the type and model of mobile terminal used by the customer. The smart card enables the network operator or service provider to offer applications according to standardized methods on all different types of mobile terminals. These services will then be available to practically all customers. Operators are putting one generic service enabling application on the smart card that can be managed and enhanced remotely. This generic service application is a flexible tool that adjusts itself depending on the terminal in use and depending on the subscribers needs. It is a platform for specific applications such as info-services, banking or any other type of application. This ensures that the network operator can base at least some or parts of their most important services on the smart card. So, as more and more smart terminals and personal digital assistants (PDAs) with mobile communications capabilities enter the market, the smart card with its features will still be the property of the network operator and remains independent of the terminal.

Services are easily deployed. Service deployment need to be fast and cost efficient. Bringing new applications and services into the market is essential and has direct impact on revenue. The Java Card™ concept allows different parties to develop applications that can be placed on the smart card for mobile communications. Rapid development times represent one important factor for the rapid supply of new services. More complex, however, is the provisioning and management of these services in the field. Services need to be brought to the customer (see also section 4.2). Three principles for the deployment of services are:

- Delivery of smart cards already containing the new application;
- Update of the smart card in the shop / at the point of sale; this update is done either offline (by an autonomous PC system) or online (connected to a background server system); and
- Update of the smart card by using over the air mechanisms.

Logistics for the shipment of smart cards is well established in the industry. In general, network operators order only a limited number of different smart card types. This makes the deployment of applications based on the smart card rather simple. It can be done by using the established way of delivering the smart cards to the end user via the network operator's shop or by fulfilment services offered by the smart card manufacturer. Application could be loaded onto the smart card already during the production phase and delivered to the end user already containing all relevant services. In case new applications need to be brought onto cards that are already deployed in the field, they can be updated and applications can be uploaded on the smart card on the air interface (over the air — OTA) by using standardized mechanisms (e.g. TS 03.48/TS 23.048) and available "bearers" such as SMS or GPRS. The configuration of the smart card can also be done at the point of sale using dedicated tools to update the card in a card reader. At the point of sale the smart card is removed from the terminal and inserted into a card reader connected to a personal computer. The employee in the network operators' shop is then able to configure the card according to the subscribers' requirements and the available new applications.

OTA capabilities of the smart card include the possibility to remotely manage the file system and the applications of the smart card. By means of existing bearers (such as SMS and GPRS) an OTA server is able to create, delete and modify files on the smart card. The smart card contents can be read out and sent to the background server, whilst new file contents can be written on the smart card from the background server. The OTA mechanisms can be utilized to update provisioning information for services such as MMS and to perform adjustment of parameters for bearers such as GPRS. Additionally, in the area of user equipment management (UEM) some relevant terminal related information could be stored and maintained on the smart card. The OTA capabilities even permit entire applications to be downloaded onto, and managed on, the smart card. Management of applications include provision of the application to the customer, update of the application or the application status, activation or deactivation due to business or technical reasons, re-activation, and finally deletion from the card. The OTA server system is capable of managing each individual card by storing information on the applications and services loaded, the memory space available for further applications any many more. This provides

the marketing departments of the network operator a powerful tool to determine which kind of service is accepted, how frequently it is being used and what the effort was to deploy a new service to which category of subscriber.

Interworking Aspects

Platform for the interworking of different access technologies such as WLAN and 3GPP networks. The smart card offers a secure authentication platform to access networks. As wireless communication is a huge success in the market new types of wireless communication systems have been developed and further technologies may exist in the future. WiFi hotspots in public areas require secure authentication just as it is required in 2G and 3G networks. Billing for the usage of a service is only possible if there is a way to secure the network against unauthorized access. As the smart card is already used for authentication of the subscriber by the 2G or 3G operator, an enhancement of either the smart card, the authentication system or both is a natural step. This enhancement could then be used to authenticate the user and to grant him access to the WLAN access network. Combining the authentication functionality of different systems on one single smart card and using the same terminal to access different access technologies (e.g. whichever is cheaper to fulfil the end users needs) makes the smart card an interworking token. Furthermore (as mentioned previously) the smart card enables the issuer to prepare for a smooth transition from 2G to 3G and beyond to nG systems by providing the related authentication application (WLAN, 2G, 3G...) on the smart card. Interworking aspects between the WLAN systems and the existing 2G/3G networks are currently investigated by standardization groups. For the smart card two options are considered. Firstly, to use the existing SIM and/or USIM application to provide secure access to WLANs, for example, by implementing the IETF standards Extensible Authentication Protocol EAP-SIM or EAP-AKA. Secondly, to develop a new independent application that reside in parallel to the SIM and the USIM application on the UICC, the smart card platform. Which ever way will be decided, the smart will be an integral part within the system.

IP Multimedia Systems as specified by 3GPP allow IP based services to be used within the 3G context in the mobile world. The authentication to this kind of IP based services such as multimedia video streaming is being done by means of an application based on the smart card. The IMS Subscriber Identity Module (ISIM), as defined in TS 31.103 “Characteristics of the ISIM application” allows secure access to IM services. This opened the way of the smart card into the authentication for IP based services and underlines the importance of the smart card for secure authentication of both subscribers and networks.

Last but not least, the smart card is fully standardized and harmonized. All required specifications for the implementation of services based on the smart card and for the smart card itself are available and mature. This allows an interoperable and harmonized implementation of services. Harmonization can be interpreted as at least twofold. The first interpretation is the harmonization of applications within the network operator group. Core applications can be defined for each subsidiary of the network operator that can then be easily adjusted according to individual local needs. The second interpretation is the harmonization of services among terminals. Different terminals (and PDAs etc.) behave in the same specified way and execute the applications stored on the smart card in the same way. As there is a much broader diversification of terminal operating systems compared to only one smart card operating system, network operators would have to develop their services for each different type of terminal operating system.

Current Work Areas and Next Generation Smart Card

The above described set of features is not exhaustive, it covers the most distinctive ones, allowing the derivation of an unlimited number of services and applications. In addition to the features described so far, some concrete enhancements are currently being discussed in the respective standardization groups, and are briefly outlined below:

- *Multimedia Broadcast/Multicast Service (MBMS) security*: In order to protect the Multimedia Broadcast/Multicast Service some security mechanisms need to be implemented to prevent unauthorized users to get access to the MBMS service. As the smart card is a proven token for containing security functionality, these mechanisms and features are being enhanced in order to also provide the requested level of security for the new service.
- *Voice Group Call Services*: The one-to-one communication channel within the 2G and 3G network need to be extended to a many-to-many communication. The related network authentication of the multiple members of the group that want to communicate to each other at the same time as well as the authentication of the individual members within the group is crucial. The classic authentication procedure based on the smart card needs to be enhanced to allow users to join or perform a group call. This service is especially important for emergency services and is already available in systems other than 2G or 3G, such as TETRA or TETRApol.
- *UICC Security Services Module (USSM)*: The UICC may contain several applications, each dealing with keys and realizing crypto-services. Some

keys might be shareable even though there are no standardized mechanisms to share keys and indicate allowed functions to authorized applications. To allow applications on the UICC to use shared security objects, it is essential to introduce standardize mechanisms to administer and to use these shared objects on the UICC. The USSM will consist of security objects (keys, PINs, etc.) including information on allowed functions and authorized applications, an API for administrative functions to administer objects of the USSM, and an API for cryptographic functions (non-administrative) to be used by applications.

- *Advanced Communication*: The demands on the classic smart card communication channel are driving it beyond its design and intents. Bearer independent protocols allow UICC applications to access communication channels whose native speed is greater than needed for classic terminal/UICC traffic. Higher level protocols such as network and transport protocols are starting to appear on some bearers. The channel is multiplexed between multiple applications using a number of different techniques. This work item considers the evolution of the smart card communication channel with respect to transfer rate, size and protocol.
- *Large Files*: Applications such as multimedia or identification applications require data storage capabilities that are reaching the current file size maximum of 65,535 bytes. Increasing the maximum file size beyond this limit impacts a wide range of size fields, parameters and commands within the standards. This work item will upgrade the standard in a synchronized and harmonized manner by providing backward compatibility.
- *Reduced voltage class*: The aim of this work item is to respect the requirement to prolong the lifetime of the terminal's battery. In order to address this need the electrical characteristics of a new (1.2V) UICC-ME interface (the Class D operating conditions) has been defined and is awaiting final agreement.
- *Next Generation UICC*: This work item identifies and evaluates commercially-viable hardware and software technologies needed to define a next generation smart card platform. The scope of the work item includes, but is not limited to, the possible role of memory management units, ASIC co-processors, proof-carrying code, new memory architectures, natural clocks, multi-tasking operating systems, embedded electrical sources, free-running oscillators, integrated biometrics sensors, universal byte codes, alternative form factors, new chip carriers, and high-speed communication channels. One of the essential characteristics of these new technologies that will be catalogued is their impact—positive and negative—on the security of the UICC platform.

Outlook

In Information Technology memory space and processing power has never been enough. This holds for the personal computer as well as for almost every component in the system. Displays have been too small, the resolution not good enough, battery lifetime was too short and the size of the battery too large. The smart card was therefore also seen as offering too little memory and not enough processing power. Today's multimedia applications and services demand more and more memory space for the storage of application or service related data such as pictures, movies and configuration data. But one thing needs to be remembered when demanding more memory, more processing power or more whatever: the smart card today already has the processing power of the early personal computers (PCs) concentrated on just a few square millimeters of silicon. And it is not the size but the effectiveness and the clever design of applications and services that make the service successful (and, of course, its market relevance).

The capabilities of the smart card will increase both physically (hardware) and logically (features implemented as software). From a hardware point of view, which has been the main constraining factor for the smart card, new technologies will enable a noticeable increase of memory capacity. Memory sizes of megabytes seem to be possible in near term and continued chip development appears to promise even multiple megabyte capacity in the mid term. The development of memory cards for digital and video cameras, together with the size reduction that is envisaged for these components, gives a good indication of the potential for smart cards.

As in the PC world, chip technologies and processor capacities continue to evolve. We can foresee smart cards with the capability to execute complex security calculations such as asymmetric en- and decryption. Equally, the development costs for new high-end smart cards can be compared with those in the PC area. The prices of the PCs are more or less always stable, whilst the capabilities of the system, such as memory, processing power and software packages included, steadily increase.

Much interest is shown in smart cards with a contact-less interface for areas such as public transportation. The combination of contact and contact-less technology in mobile communications devices would also enable future services to be deployed by means of the mobile terminal. New standards such as near field communication (NFC) where an active component in the terminal could act as a card reader for contact-less cards could dramatically enhance the uses of the smart card.

Where the hardware capabilities of the smart card permit, and suitable opportunities exist, merging markets are foreseen. Transferring applications that reside on physically different smart cards onto the new multi application smart card makes sense for a number of applications, especially those that require the possibility to communicate with external entities or a personal card reader. Such applications are the ideal candidates to be incorporated into the smart card for mobile communications.

Mobile communication is a growing market: new applications and new areas of use are established regularly. Niche markets or markets that traditionally could not be served due to physical limitations of current technology will be served by new products. Telematics, just to give one example, is one of the areas that could not be served with a full range capability due to the limitation of the temperature range of existing hardware and chip products. The continuous development of these components is expected to allow increases in the temperature range to a degree that is acceptable in telematic systems.

Further reductions in size or even a completely new design and architecture of the smart card is part of the investigations carried out by the ETSI Project Smart Card Platform. The UICC next generation is a work item that has been set up by the committee to search for a smart card solution and architecture that will meet the future requirements. As well as possible size reductions, the definition of new communication protocols, new file systems and other enhancements are being considered as required by future markets. A new file system that reflects the capabilities of the smart card to store megabytes of data will enable the smart card to ease the management of data storage on the card. This will make it much easier for the issuer of the card to maintain and to manage the data related to the services on the card.

The evolution from a one-application card to a multi-application card is paralleled by evolution of the operating system. Multi-tasking operating systems and other state of the art personal computer features are required and may be added one by one to the smart card. This allows the addition of further applications that run in parallel or in background mode.

A further trend concerns the existence of multiple different communications systems such as 2G, 3G, WLAN, DECT and others. These will lead to a merger of the applications for the end user, and thus for end user devices. The token containing the authentication data and performing the trusted operations in the network for the end user will contain different authentication and network access applications. These applications may share some of the information stored on the smart card. This puts the smart card in a role of providing a smooth roaming from one access technology to another. The smart card acts as the medium that keeps the subscriber connected and that provides the means of interworking between the different networks. Smart card ownership in such a

scenario is crucial as it means that the churn of subscribers could be reduced: the subscriber could be more easily added as a customer for a new access technology by simply putting the required information on the smart card to access the new network. The network operator issuing the SIM and also operating WiFi hotspots could very easily enable the subscriber to access the WLAN at the airport by simply re-using the SIM card for the authentication. This makes it more difficult for new providers to offer such a service.

The pressure in the industry to generate additional revenues is extremely high. Companies have invested massively in the purchase of licenses and the setting up of new generation networks. Further enhancements to the network infrastructure and the introduction of further generations of communication systems will be equally expensive, so even more revenue has to be generated in order to be prepared for those future systems. This revenue will be generated by those companies who succeed in providing the appropriate services that are accepted by the consumers. Including the power of the smart card in the concept of such services may well be one major step towards achieving that goal.

References

SCP-010141 work item description on “Advanced communication”

SCP-010142 work item description on “Large files”

SCP-010265 work item description on “Introduction of a new voltage class”

SCP-020185 work item description on “UICC next generation”

SCP-030281 work item description on “USSM (UICC Security Services Module)”

TS 11.11 “ Specification of the SIM-ME Interface”

TS 31.101 “UICC-Terminal Interface; Physical and Logical Characteristics”

TS 31.102 “Characteristics of the USIM Application”

TR 31.900 “SIM/USIM internal and external interworking”

Klaus, V. (2001). In Hillebrand, F. (Ed.). *GSM and UMTS: The Creation of Global Mobile Communication*. Wiley Europe.

Rankl, W. & Effing, W. (2003). *Smart Card Handbook*. London: John Wiley & Sons.

Endnotes

Trademark Information

- **DECT™**, **TIPHON™** and **UMTS™** are trade marks of ETSI registered for the benefit of its Members.
- **3GPP™** is a trade mark of ETSI registered for the benefit of the 3GPP Organizational Partners.
- **GSM™**and **Global System for Mobile Communication** are registered trade marks of the GSM Association.
- **Java** and all Java-based marks are trade marks or registered trademarks of Sun Microsystems, Inc. in the US and other countries.

Chapter X

Recent Developments in WLAN Security

Göran Pulkkis, Arcada Polytechnic, Finland

Kaj J. Grahn, Arcada Polytechnic, Finland

Jonny Karlsson, Arcada Polytechnic, Finland

Mikko Martikainen, Arcada Polytechnic, Finland

Daniel Escartin Daniel,
Escuela Universitaria Politecnica de Teruel, Spain

Abstract

This chapter is a topical overview of wireless local area network security evolution. WLAN security threats are surveyed. Covered standards include 802.11/WEP, 802.1X/EAP, 802.11i/WPA and 802.11i/WPA2. Special attention is given to user authentication schemes and to protection of data communication. WPA is also compared with the present WLAN security protocol WEP. Other covered issues are available WPA supported technology and open source WLAN security software. A WPA secured WLAN test

network is described and tested. A WLAN designed according to the new security standards is a real alternative to a secure enterprise LAN and also a reliable platform for e-commerce. Finally, WLAN security management and current research related to WLAN security are surveyed.

Introduction

Security of wired local area networks (LAN) is:

- Internally based on:
  - The use of switches for network unit interconnection
  - Challenge response user authentication
  - User dependent access rights
- Externally based on:
  - Gateway protection by firewalls
  - Intrusion detection hardware/software
  - Physical protection of networking hardware and data transmission media

For protection of data communication security standards like IPSec, SSH and TLS/SSL on OSI layer 3 and higher are available.

Attacks against the physical layer and the data link layer can be avoided in traditional LANs by preventing unauthorized access to the neighborhood of networking hardware and data transmission media. Propagation of electrical signals in metal wires can be eavesdropped upon only by closely placed sensible monitors of electromagnetic radiation. Propagation of optical signals in optical fibers creates no electromagnetic radiation and thus offers no possibility for eavesdropping.

The immergence of wireless local area networks (WLAN), however, also has opened the OSI layers 1 and 2 for eavesdropping, intrusion and information content manipulation attacks. Data transmission media is far from easily protected in a WLAN. The radio communication of a WLAN cannot in practice be encapsulated by a wall, through which no electromagnetic radiation leaks. The data transmission between WLAN units can be eavesdropped from quite a long distance with sensitive radio communication receivers.

WLAN security thus requires:

- Reliable protection of data communication on OSI layer 2 (=the data communication between WLAN units) and
- User authentication protocols with reliable protection on OSI layers 1 and 2.

The main purpose of this chapter is to present the evolution of WLAN security with emphasis on the role of the proposed new WLAN security standard WPA (Wi-Fi Protected Access).

WLAN Security Evolution

A WLAN consists of station computers and access points (AP). A WLAN architecture with dedicated station computers and a dedicated access point is called a Basic Service Set (BSS). A WLAN architecture with dedicated station computers and many dedicated access points is called an Extended Service Set (ESS). An access point is a wireless hub for the station computers in a BSS. A WLAN architecture without dedicated access points is called an Independent Basic Service Set (IBSS). In an IBSS WLAN each station computer also is an access point for other station computers. An IBSS WLAN is also called an *ad-hoc network*. In the following the abbreviation WLAN will be used as a synonym for a wireless local area network with one or several dedicated access points.

Security Threats and Vulnerabilities

A radio interface is by nature easy to access. The number of security threats is large, requiring ongoing security monitoring. Security threats are either passive or active attacks. Active attacks involve altering data streams. Passive attacks, on the other hand, include snooping on transmission (eavesdropping, sniffing a wireless network). These attack categories can further be divided into subcategories (Nicopolitidis, Obaidat, Papadimitriou & Pomportsis, 2001).

Active attacks are:

- *Masquerade:* one entity pretends to be another entity
- *Reply:* passive capture of a data unit and the construction of unwanted access

- *Modification:* some portion of a genuine message is changed
- *Denial-of-service:* attempt to prevent legitimate users of a service from using that service

Passive attacks are:

- *Release of message contents:* the attacker reaches information being transferred
- *Traffic analysis:* reveal useful information in guessing the nature of the exchanged information

The four main security threat types for a WLAN are eavesdropping, identity theft, denial-of-service (DoS), and internal threats (Potter & Fleck, 2003).

Eavesdropping

By listening to airwaves, a NIC (Network Interface Card) can easily pick up encrypted messages. The intruder receives all packets within its range, which is optimized by using a high gain directional antenna and highly sensitive receivers.

Hackers normally use their mobile device to access a network. An incorrectly configured access point (AP) may broadcast its SSID (Service Set Identifier). This makes it possible for the intruder to associate and share the network's connection. Moreover, a large number of APs are not WEP enabled.

Sniffing software developed to locate APs, crack WEP keys, etc. can easily be found on the Internet. Examples of such sniffing programs are: Airsnort (AirSnort Portal, 2004), Ethereal (Ethereal Portal, 2004), Kismet (Kismet Portal, 2004), THC-Wardrive (The Hacker's Choice, 2004), War-Linux (WarLinux, 2004), Wellenreiter (Wellenreiter, 2004), Wepcrack (WEPCrack, 2004), and Winairsnort (Seattle Wireless, 2004).

Identity Theft

SSID, WEP authentication and Media Access Control (MAC) addresses are used to authorize clients to connect to an access point. Approved SSIDs and MAC addresses can be picked up by the intruder. The WEP key can also be retrieved by an intruder some hours after the intrusion has occurred.

Denial-of-Service (DoS)

WLANs are vulnerable to both wired and wireless DoS attacks. There are three basic DoS attack types (Cert, 2004):

- Consumption of scarce, limited, or non-renewable resources;
- Destruction or alteration of configuration information; and
- Physical destruction or alteration of network components.

Typical DoS attacks are (The Hacker's Choice, 2004):

- *Channel flooding (jamming)*: static noise or channel overlapping cause WLAN signals to collide and produce CRC errors
- *Disassociation attack:* amplified transmission of disassociate management frames is used
- *Michael vulnerability:* vast quantities of unauthorized data is sent to the network, the system assumes it is under attack and shuts itself down
- *Man-In-The-Middle:* spoofed beacon frames are transmitted by the intruder at shorter intervals and at a greater signal strength causing the mobile nodes to select the intruder as the best choice AP

Internal Threats

Internal threats include unauthorized APs and improper AP setup:

- *Rogue APs:* employees are deploying unauthorized WLANs to the corporate network(s)
- *Incorrectly configured APs:* initially many systems are shipped without any security

WLAN Security Architecture

A LAN architecture can be a single network segment layout or several network segments interconnected by bridges. An AP for a WLAN consisting of a single

network segment must also have network gateway functionality, if this WLAN is part of a WAN or of a MAN. An AP in a WLAN, which is one of several network segments in the same LAN, must also have network bridge functionality. To minimize the risks of OSI layer 2 security incidents it is recommended to implement WLAN bridging on a LAN gateway device and to avoid pure AP bridging devices (Potter & Fleck, 2003).

OSI layer 1 and 2 security is implemented by the hardware and the software of the network interface, for example by the hardware and software driver of a wireless NIC (Network Interface Card).

WLAN Security Standards

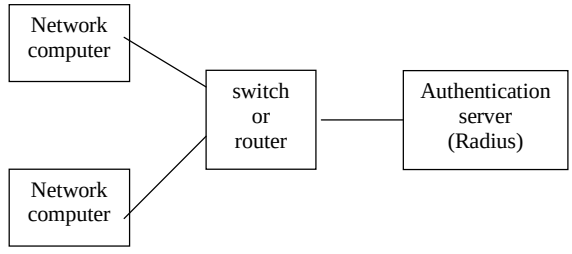
WLAN standards are introduced by three major standardization organizations, IEEE (IEEE Standards, 2003), Wi-Fi Alliance (Wi-Fi, 2003), and IETF (IETF, 2003). Most of the standards are issued by IEEE. Wi-Fi Alliance handles the practical implementation of these standards through interoperability testing and certification, while IETF is engaged in the evolution of Internet architecture.

The 802.11 standard launched practical WLAN solutions for the home and for most small offices. The lack of a centralized method for authentication limited the usability of these systems. The data encryption solution WEP was a weak implementation of the RC4 algorithm. This encryption was only a protection against casual eavesdropping.

IEEE 802.1X introduced a distributed architecture to increase scalability and to address the port based authentication function. This standard defines a client-server-based access control and an authentication protocol, which prevents unauthorized clients from connecting to a LAN through publicly accessible ports. The principle of the IEEE 802.1X standard is shown in Figure 1 for a wired network. Wired connections of network computers to switch or router ports must be approved by an authentication server. An example of practical port based authentication is the 802.1X configuration guide for Cisco Catalyst routers and switches (Catalyst, 2003).

In a WLAN, the access point has the role of the router or switch, and approved wireless connections have the role of approved port connections. IEEE 802.1X is also included in the new WLAN security standard Wi-Fi Protected Access (WPA). TKIP, a stronger implementation of the RC4 algorithm, is incorporated in WPA for protection of the data communication between workstation computers and access points. Data encryption and authentication functions are coupled together through a sophisticated keying system.

Figure 1. Port based IEEE 802.1X authentication



The IEEE and the Wi-Fi Alliance introduce a single unified standard built on the pending IEEE 802.11i and WPA2 standards. New features are Advanced Encryption Standards (AES), message integrity and fast-roaming support (Burns and Hill, 2003).

802.11/WEP

Wired Equivalent Policy (WEP) based on the RC4 encryption method uses keys that are both static and known by the WLAN stations. As a result, the academic community has demonstrated that WEP alone does not provide adequate security (Borisov, Goldberg and Wagner, 2001). The WEP's weaknesses include (Burns and Hill, 2003):

- Station identification addresses can be easily captured;
- Static user keys are rarely changed;
- Keys are duplicated on client stations;
- The RC4 encryption algorithm is weakly implemented;
- The Initialization Vector (IV) is short and will be reused; and
- Management messages are not authenticated.

Freely available packages that allow attackers to discover the WEP key can be found on the Web (Sourceforge Project wepcrack, 2001).

IPSec VPN

Using WEP is more secure compared to using no security at all and still remains useful for protecting small office and home SOHO WLANs were network

traffic is light. Nevertheless, WEP is not sufficient for protecting enterprise WLANs due to its serious security flaws. Many large companies therefore started to strengthen their WEP protected WLANs with other third-party solution like Virtual Private Networks (VPNs) (Wi-Fi Protected Access, 2003).

VPN is based on the IP Security Protocol (IPSec) which is developed by the Internet Engineering Task Force (IETF). IPSec related Internet drafts are found at www.ietf.org (IP Security Protocol, 2004). VPN was developed to enable clients to securely connect to servers over the public Internet. VPN employs strong authentication and encryption algorithms. With VPN, a secure tunnel between two endpoints is created, protecting against intrusion for packets transferred over the Internet.

Internet Key Exchange (IKE) is a protocol within IPSec which provides authentication methods for protecting data and communications. These authentication methods are a pre-shared secret key between two parties or a standard public key authentication including support for digital certificates. Using IPSec VPN, in tunnel mode, both the header and payload data in packets transferred between two endpoints are protected using key encryption.

VPN can be utilized for securing an 802.11b network by establishing one-to-one secure connections between wireless clients and a VPN gateway, which is located behind the wireless access point. This means that every client in the wireless network communicates with its wireless access point over a secure VPN tunnel.

802.1X/EAP

802.1X uses the Extensible Authentication Protocol (EAP) and was adopted as a standard by IEEE in August 2001. EAP was originally designed for PPP based connections. The standard was further developed for port-level authentication for 802 based wireless and wired networks. Support for very large deployments at low cost was one of the considerations.

In an unprotected LAN supporting DHCP it is very simple for a user to connect to and use the services of the network by physically connecting the computer to a network switch. In a WLAN it is even simpler because neither a wire nor a physical contact to the network is needed. When 802.1X security is implemented and supported by the wireless access point (AP) or the switch, the access to any network services are denied until successful authentication of a client has been performed.

The 802.1X standard uses a 3-element WLAN configuration. The elements are called supplicant (station, client), authenticator (wireless access point) and authentication server. A suppliant is client software supporting 802.1X authen-

tication. The authenticator is in a LAN the network switch. In a WLAN, the authenticator is the access point which:

- Listens to connection requests from clients
- Delivers EAP messages between the client and the authentication server, i.e. RADIUS with 802.1X and EAP support
- Accepts or rejects access to the network services whether the authentication succeeded or not.

During the authentication process the authenticator remains in uncontrolled state allowing only EAP messages between the suppliant and the authentication server and switches to *controlled state* which allows full access to the network when authentication is successfully performed. (Potter & Fleck, 2003)

EAP, IETF standard RFC 2284 (IETF, 2003), can be extended to run over any transport mechanism and verification can be handled by any crypto system. With 802.1X in wireless networks, EAP is used to pass authentication information messages known as EAP over LAN (EAPOL) messages between the suppliant and the authentication server, and to handle the presentation of user's credentials in form of a unique username and passwords or digital certificates. By using EAP, a number of specific authentication schemes called EAP types can be used. An authentication, authorization and accounting (AAA) access point that supports EAP need no understanding of the specific EAP type (Extensible, 2002).

For a WPA protected WLAN, at least the following EAP types (options) are supported:

- *EAP-TLS:* EAP-Transport Layer Security (TLS) provides certificate based and mutual authentication of the supplicant and the server. This requires both client-side and server-side certificates for authentication. Dynamic session key generation is also included. EAP-TLS relies on Public Key Infrastructure (PKI) concepts. A more detailed description of PKI including the concepts of certificates, certificate authorization, validation of user identity and hardware tokens will be given later in this chapter.
- *EAP-TTLS:* Tunneled Transport Layer Security (TTLS) was developed as an extension of EAP-TLS. Certificate-based, mutual authentication of the client and the network through an encrypted channel/tunnel is provided. This method unlike TLS only requires server-side certificates. Dynamic WEP keys are also provided.

Table 1. Features and benefits of 802.1X EAP types (Networking, 2003)

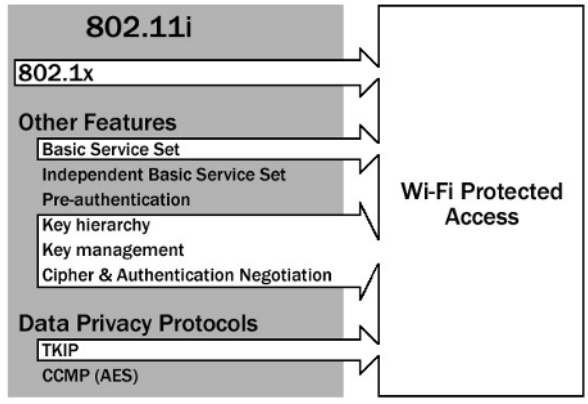
802.1X EAP	TLS	TTLS	PEAP	LEAP
Feature/Benefit	**Transport Level Security**	**Tunneled Transport Level Security**	**Protected Transport Level Security**	**Lightweight Extensible Authentication Protocol**
Client side certificate required	Yes	No	No	No
Server side certificate required	Yes	Yes	Yes	No
WEP key management	Yes	Yes	Yes	Yes
Rouge AP detection	No	No	No	Yes
Authentication attributes	Mutual	Mutual	Mutual	Mutual
Deployment difficulty	Difficult	Moderate	Moderate	Moderate
Wireless Security	Highest	High	High	High

- *PEAP:* Protected Extensible Authentication Protocol (PEAP) is like TTLS a method to securely transport authentication data via 802.11 wireless networks. The PEAP authentication process occurs in two phases. The first phase creates an encrypted channel between the supplicant and the authentication server using TLS. This phase requires a server side certificate. The second phase uses a different EAP type, such as MS-CHAP v2, to authenticate the users. MS-CHAP is a password-based, challenge-response, mutual authentication protocol that uses Message Digest 4 (MD4) and Data Encryption Standard (DES) algorithms to encrypt responses.
- *LEAP:* Lightweight Extensible Authentication Protocol (LEAP) is an EAP authentication type. It supports mutual authentication and uses the username-password model. Dynamically generated WEP keys are used for data encryption.

802.11i/WPA

Formerly known as Safe Secure Network (SSN), Wi-Fi Protected Access (WPA) now offers a strong security solution. The WPA technology is intended to address the WEP vulnerabilities and to be (Disabato, 2003):

Figure 2. A comparison between WPA and 802.11i



- A software/firmware upgrade to existing access points and NIC's;
- Inexpensive in time and cost to implement;
- Cross-vendor compatible; and
- Suitable for enterprise, small sites and home networks.

WPA performs operations that encompass 802.1X/EAP authentication, sophisticated key management, and encryption techniques (Burns & Hill, 2003). WPA also runs in enterprise or pre-shared key (PSK) mode. WPA is a subset of the 802.11i draft standard and is also expected to be forward compatible with the standard. A comparison between WPA and 802.11i is shown in Figure 2. More details concerning WPA will be given later in this chapter.

802.11i/WPA2

Further addressing of security issues of the WLAN is handled by an IEEE Task Force, expected 802.11i ratification is the first quarter of 2004. Wi-Fi Protected Access version 2 (WPA2) includes full 802.11i support, while WPA2 will replace RC4 with AES. WPA2 will also include the CCM protocol, which combines CBR (Counter Mode Encryption) and CBC/MAC (Cipher Block Chaining/Message Authentication Code). The new standard implementation is hardware accelerated and will require replacement of most access points and some NIC's (Network Interface Cards). WEP, WPA and WPA2 are compared in Table 2 (Disabato, 2003).

Table 2. Comparison between WEP, WPA and WPA2

	WEP	WPA	WPA2
Cipher	RC4	RC4	AES
Key Size	40 bits	128 bits encryption 64 bits authentication	128 bits
Key Life	24-bit IV	48-bit IV	48-bit IV
Packet Key	Concatenated	Mixing Function	Not Needed
Data Integrity	CRC-32	Michael	CCM
Header Integrity	None	Michael	CCM
Replay Attack	None	IV Sequence	IV Sequence
Key Management	None	EAP-based	EAP-based

Future Trends

It is interesting to analyze the future in wireless networking. The wireless landscape has seen many changes in the past two years, improving considerably in security and availability. It is reasonable to assume that the next step in wireless would be to develop larger networks. Technologies like WWAN (Wireless Wide Area Network) and WMAN (Wireless Metropolitan Area Network) might become household names in the next few years. With these technologies there will be a cost effective way to connect smaller LANs (wireless or wired) into larger networks.

WWAN Example: 802.16a/WiMAX

WiMAX supports a new technology that provides broadband Internet access for wireless communication (802.16a/WiMAX, 2004). WiMAX will improve the performance and probably offer cheaper products. The following step after Wi-Fi is considered to be WiMAX.

Some of the advantages are for example higher Quality-of-Service, enhanced security, higher data rates and efficient use of the radio frequency spectrum. It is important to emphasize that WiMAX and Wi-Fi are considered to be complementary because they are addressed to different segments of the market.

Wi-Fi is oriented towards local networking and WiMAX towards metropolitan area networking.

Wi-Fi Protected Access

Hitherto, WLAN has not been a reliable implementation alternative to a secure local area network due to the weakness and flaws of the existing wireless security standard, WEP. IEEE and Wi-Fi Alliance members have therefore started the development of a new security standard for both home and enterprise WLANs. This new security standard, WPA, is derived from and will be forward compatible with the presently proposed IEEE 802.11i standard. WPA presents secure solutions to all known flaws of the WEP protocol. All attempts to find WPA vulnerabilities have as yet been unsuccessful.

Wi-Fi certified the first products with WPA version 1 support for testing purposes in early 2003. Today there are already a number of such Wi-Fi certified products on the market. During 2004 Wi-Fi intends to completely replace WEP with WPA in certified products. All new security features will be included in WPA version 2. This new WPA version will include full IEEE 802.11i support and is expected to be released by late 2004.

WPA has a number of design goals:

- To be a strong, interoperable security replacement for WEP,
- To be software upgradeable to existing WI-FI certified products, and
- To provide real security for both home and enterprise networks.

To obtain these goals, two primary security enhancements are necessary, *improved data encryption* and *unambiguous user authentication*. Encryption is weakly implemented and user authentication is missing in WEP.

Specification and Requirements

The present specification documentation of WPA consists of:

Table 3a. Checklist of features required by WPA (WPA, 2003)

Function	Status
48 bit TKIP (including phase 1 and 2)	Required
Fragmentation of TKIP data packets. *Note: The station will not be able to send full size 802.11 MPDUs if fragmentation is not supported*	Optional
De-fragmentation of TKIP data packets	Required
Use of integrity check and IV for replay protection	Required
Michael	Required
Michael counter measures	Required
WPA information element in beacon, probe response, association/re-association request	Required
Privacy bit set in capability information element Beacon/probe response/association/re-association request	Required
4-way handshake	Required
Validation of WPA IE in beacon, probe response, association/re-association request with WPA IE in 4-way handshake	Required
Group key update	Required
Pairwise Request (with or without error)	Required
Group Request (with or without error)	Required
Encryption of 802.1X messages with Pairwise Keys	Required
802.1X messages not encrypted with Group Keys	Required
WPA authentication mode	Required
WPA-PSK authentication mode	Required
WPA-None authentication mode	Optional for NIC
Open 802.11 MAC authentication for all WPA authentication modes	Required
WPA-PSK ASCII pass phrase hash	Required
WPA-PSK 256 bit key	Recommended
Non-WPA support	Recommended
Non-WPA and WPA mixed mode	Recommended
Group key cipher	Required
Pairwise key cipher	Required for NIC Optional for AP

Table 3b. Checklist of features required by WPA (2003)

Function	Status
No sending of non-802.1X data packets until the correct key is installed. (Note: This is Group key for multicast/broadcast from AP or from station if Pairwise key is not installed. This is Pairwise key for unicast from AP or all traffic from station if a PW key is installed	Required
Queuing of EAPOL-Key messages when in power save	Required
Saving of IBSS IV	Required
Support for RADIUS	Required for AP
Group Key Update on a time interval	Recommended
Group key update on a disassociation of a authenticated station	Optional
Use of PRF for Pairwise key generation	Required
Use of PRF for Group Key generation	Required
Use of random number on AP for master key for Group Key generation	Required
Initialization of Key Counter	Required
Initialization of EAPOL IV from Key Counter	Required

- The IEEE Standard Draft 802.11i/D3.0 with specifications for enhanced security on the MAC and physical layers of a WLAN (Draft Supplement, 2002) and
- The document "Wi-Fi Protected Access (WPA)" published in April 2003 by the Wi-Fi Alliance, this document captures those clauses of the IEEE draft that comprise the enhanced security implementation called WPA for the standard 802.11i. (WPA, 2003).

Both documents can be purchased from the publishers, see Tables 3a and 3b.

WPA can be seen as a snapshot of the 802.11i standard including 802.1X, BSS, key hierarchy, key management, cipher and authentication negotiation, and TKIP. A robust security network (RSN) relies on the 802.1X entity above IEEE 802.11 to provide authentication and key management services.

User Authentication

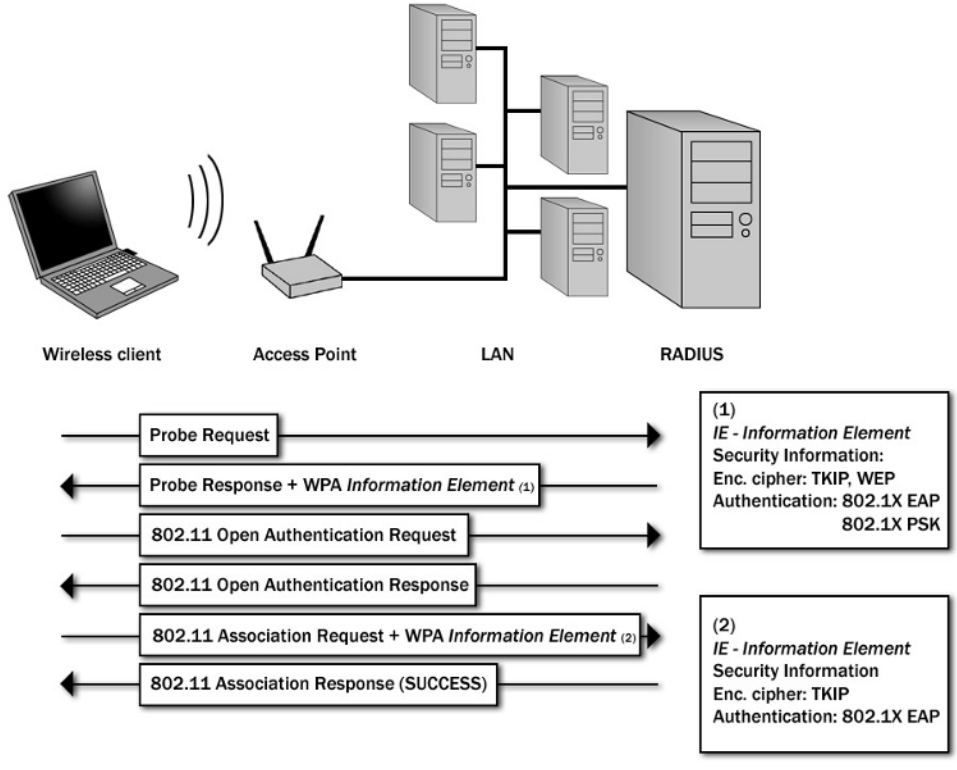
WPA provides strong user authentication by implementing the IEEE 802.1X standard and the EAP. This framework uses an AAA server, such as RADIUS, to authenticate each user in the wireless network.

With WPA, mutual authentication is required, meaning that the AAA server also must be authenticated on the wireless client before the authentication procedure is complete. This prevents accidental user access to a rogue network, which might steal network credentials from such a user (Overview Wi-Fi, 2003).

WPA also provides the opportunity of using pre-shared key (PSK) authentication for users in small office or home office WLAN environments where AAA servers are unavailable. The difference is that a password is manually entered on the client and the authentication succeeds if the password matches the one entered on the AP (Wi-Fi Protected, 2003).

Before the actual user authentication can start a wireless client must discover an AP and associate to it. During the *association* process (defined in Draft Supplement to ISO/IEC 8802-11, 2002, p. 13), the security policy is negotiated. The client starts the association process by sending a *probe request* to the AP. After receiving this request, an AP advertises its support of WPA by sending a beacon frame with a new 802.11 WPA information element containing the wireless AP's security configuration (Microsoft Knowledge, 2003). Figure 3 shows how a wireless client discovers an AP and negotiates the security policy.

Figure 3. Security association



802.1X EAP

As a continuation of successful association, the authentication exchange can begin using the authentication method negotiated during association. In case of 802.1X/EAP, authentication information is exchanged between the supplicant and the authenticator using EAP over LAN (EAPOL) messages. In this procedure the access point works as an intermediary between the client and the authentication server. The EAP authentication process is shown in Figure 4 (Goransson, 2003).

Pre-Shared Key

For smaller wireless networks, like home networks with no available authentications server, WPA offers the use of Pre-Shared Key (PSK) authentication. The PSK is, on the client and the AP, manually entered as an ASCII or hexadecimal

Figure 4. EAP authentication

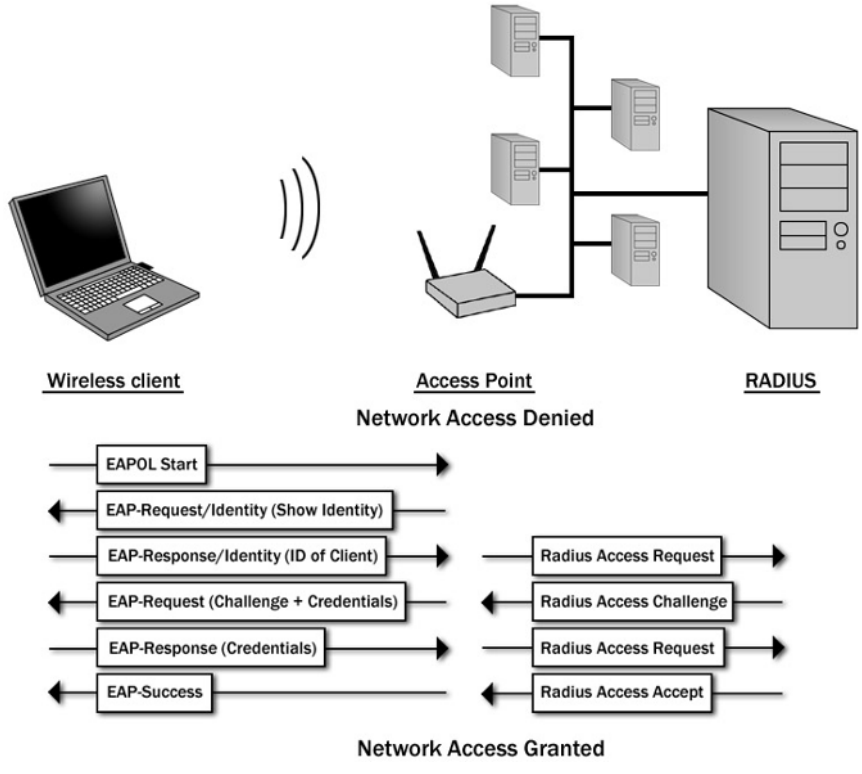
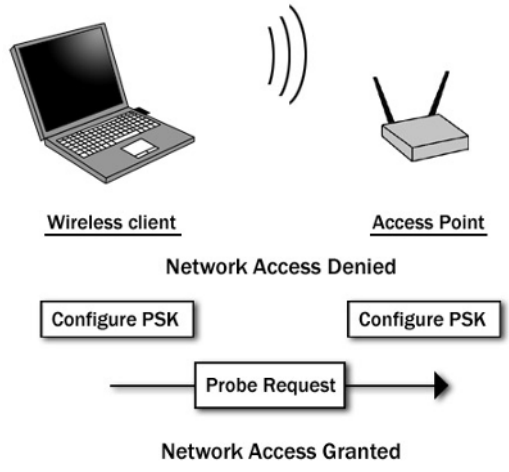


Figure 5. Pre-shared key authentication



pass phrase with a suitable length (depends on the implementation). The system generates a 256 bit key of the ASCII pass phrase which then is used as the actual Pre-Shared Key. The PSK authentication is performed as shown in Figure 5 (Goransson, 2003).

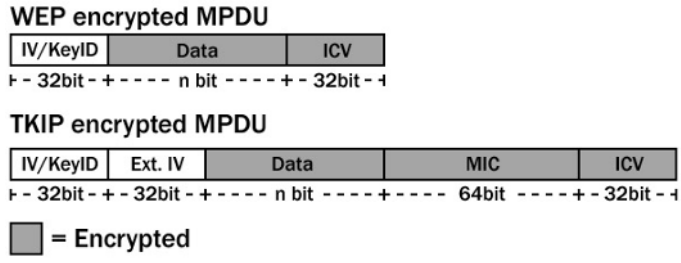
Protection of Data Communication

WPA greatly improves the WEP encryption using the Temporal Key Integrity Protocol (TKIP) in conjunction with the 802.1X framework and WPA 802.1X key management. This key hierarchy and key management methodology removes the predictability intruders relied upon to exploit the WEP key (Wi-Fi Protected, 2003).

TKIP

The purpose of the Temporal Key Integrity Protocol (TKIP), (Draft Supplement, 2002, p. 35-47; WPA, 2003, p. 21-31), was to fix all security flaws in WEP. TKIP uses keys that are dynamically generated and distributed by the authentication

Figure 6. WEP—and TKIP encrypted MPDU formats



server. These encryption keys are regularly changed and rotated in a way that the same encryption key is never used more than once. All this happens automatically in the background, invisible to the user.

TKIP is, like WEP, based on the RC4 stream cipher algorithm but provides much stronger data encryption by adding four new algorithms:

- Message Integrity Code (MIC) named Michael, to defeat forgeries
- 48-bit Initialization Vector (IV) and an IV sequence counter, to prevent replay attacks
- Per-packet key mixing function, to de-correlate the public IVs from weak keys
- Re-keying mechanism, to provide fresh encryption and integrity keys

TKIP increases the length of a WEP encrypted Mac Protocol Data Unit (MPDU) by 96 bit, 32 bit for the extended IV information, and 64 bit for the Message Integrity Code (MIC), see Figure 6.

Pair-Wise Key Hierarchy

The Pair-wise key hierarchy, (Draft Supplement, 2002, p. 88-90), is used to derive session specific session keys needed by TKIP for encrypting unicast data packets. By the end of a successful mutual EAP authentication an EAP master key is generated by the supplicant and the authentication server. The generation of this key is authentication type dependent, but if e.g. the EAP - TLS authentication protocol (Adoba & Simon, 1999) is defined, then the key (here

called TLS master secret) is generated through the TLS handshake protocol. The TLS master secret is generated from a random number (pre-master secret) by the supplicant during the EAP authentication process. The pre-master secret is delivered by the supplicant to the authentication server encrypted by the server's public key and is thus only known by the supplicant and the authentication server. From the TLS master secret, a 256 bit Pair-wise Master Key (PMK) is generated by both the supplicant and the authenticator. This PMK is delivered by the authentication server to the AP at the end of the EAP authentication procedure.

Alternatively, using other EAP types the PMK can be generated as a random value on the authentication server which then is transported to the AP and the client protected by the EAP master key. If the Pre-shared Key authentication method is defined, the 256 bit key derived from the manually configured pass phrase on the client and on the AP is used as PMK.

WPA implements a protocol, called the 4-way handshake (Draft Supplement, 2002, p. 97-104), as an addition to the 802.1X key management. Using dynamic WEP keys and EAP authentication, the PMK is directly used for encryption. WPA uses the PMK for generating Pair-wise Transient Keys (PTK) through the 4-way handshake protocol which are then used in encryption and hashing functions.

The 4-way handshake, shown in Figure 7, exchanges EAPOL-Key messages between the supplicant and the AP to:

- Confirm that both parties have the same PMK and that it is current;
- Derive the PTK from the PMK;
- Install the encryption and integrity keys into IEEE 802.11; and
- Confirm the installation of the keys.

The PTK is a value derived from the Pseudo Random Function (PRF), (Draft Supplement, 2002, p. 88), which hashes various inputs to derive a pseudorandom value. The output value (as shown in Figure 8) is divided into 4 different 128 bit keys:

- MK is used to integrity check an EAPOL-Key Message
- EK encrypts the Key Material field in an EAPOL-Key Message
- TK1 is a temporal key used by TKIP for encrypting data packets
- TK2 is a temporal key used by TKIP for decrypting data packets

Figure 7. The 4-way handshake

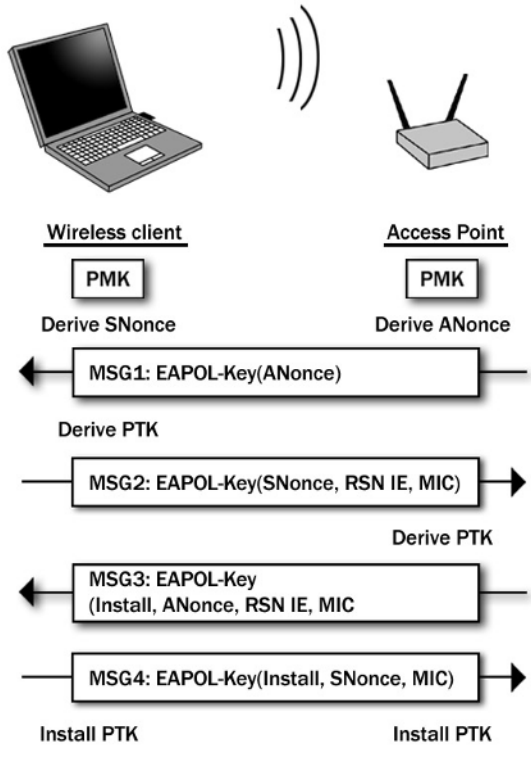


Figure 8. PTK derivation for the TKIP

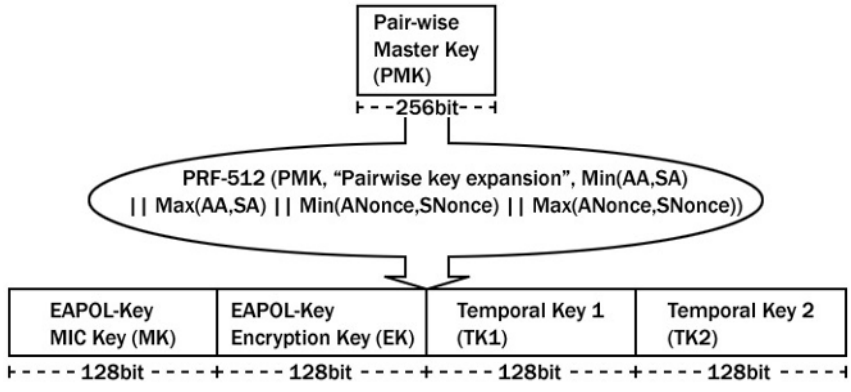
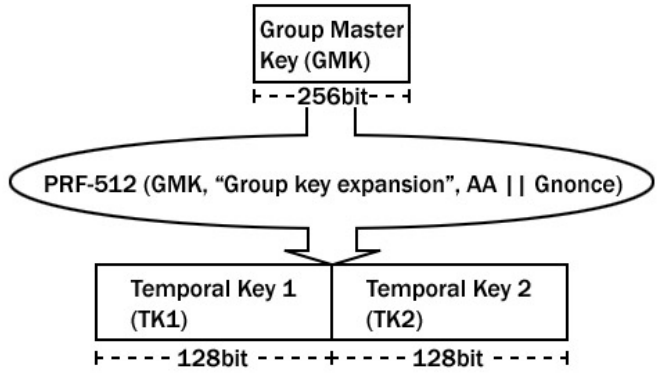


Figure 9. GTK derivation for the TKIP



Group Key Hierarchy

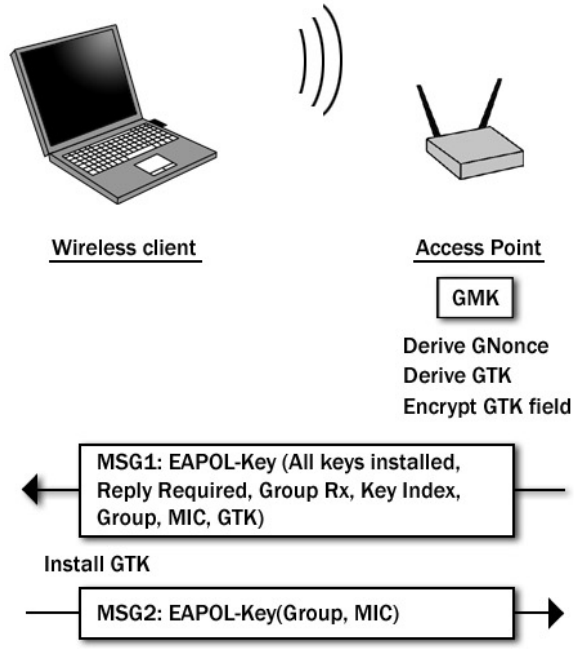
WPA implements the Group key hierarchy, (Draft Supplement, 2002, p. 90-92), to derive group keys, needed by TKIP for encrypting multicast data packets. Within this key hierarchy the authenticator utilizes the PRF to generate a Group Transient Key (GTK) from a Group Master Key (GMK), which is a 256 bit random value generated by the AP (See Figure 9).

The GTK is divided in two different keys:

- *TK1:* a temporal key used by TKIP for encryption
- *TK2:* a temporal key used by TKIP for decryption

Group keys are delivered by the AP to all its authenticated clients through a two EAPOL key message exchange called Group Key Handshake (Draft Supplement, 2002, p. 104-109). See Figure 10. During delivery GTKs are encrypted with each session's EK and digitally signed by each session's MK.

Figure 10. Group key handshake



Available WPA Supported Technology

Wi-Fi Certification

The Wi-Fi (Wireless Fidelity) Alliance (Wi-Fi, 2003) is an organization of great importance within the wireless networking scene. Wi-Fi encourages manufacturers to use standardized 802.11 technologies when producing networking products and markets these products to the consumers so that users stick to the same standard. Its main purpose though, is to test and certify Wi-Fi product interoperability. A Wi-Fi certified product guarantees interoperability with all other Wi-Fi certified products independently of brand.

For a company to have their product Wi-Fi certified, they must first apply for membership to the Wi-Fi Alliance. The company seeking to certify a product must then submit an online application for a test. A testing date is then agreed upon with one of many independent testing facilities around the world and a testing facility chosen by the Wi-Fi Alliance is then agreed upon (Agilent, 2000).

The testing phase consists of a series of tests designed to verify that the product can perform all major features from the 802.11a/b/g protocols with other Wi-Fi certified products. At the same time the product is tested for WPA compliance. When the test lab has verified compliance, the Wi-Fi alliance reviews the test results to check for inconsistencies and common application errors to further ensure the integrity of the product. After this, a Certification Authorization Letter is signed, and finally, the Wi-Fi Board of Directors will give their final approval (Wi-Fi, 2003).

WPA Certified WLAN Products

Since the first WPA certified products started to appear in early 2003, there has been a steady growth of the number of certified products. The Wi-Fi Alliance keeps a complete listing of all certified products on their Web page. However until now only hardware, such as (but not restricted to) access points, gateways, PCMCIA cards and USB adapters are being certified.

The Wi-Fi uses an easy to use labeling system called Capabilities labels for the certified products. On the label, included in all certified hardware, you can see what certificates the product has been tested for and has passed. The certificates are:

- 2.4 Ghz band 11 Mbps
- 2.4 Ghz band 54 Mbps
- 5.5 Ghz band 54 Mbps
- Wi-Fi Protected Access

Since the label includes everything you need to know about the product to determine which products are interoperable, it helps the average user (as well as the professional) to identify which products are needed to build a required WLAN. The label also includes a unique Certification ID number so that it can be identified if necessary. The label also can be found on the Certified Products Listing on the Wi-Fi Web Portal (Certified, 2003).

Wireless network cards and access points need to be upgraded with new firmware and drivers with WPA support. Most new hardware can be upgraded and many hardware vendors have already released WPA drivers and firmware for their WLAN products. Examples of such vendors are Cisco Systems, 3Com and Linksys.

Cisco Systems has actively taken part in the testing of the new WLAN security standard. Cisco released their first WPA supported software, IOS Software Release 12.2(11)JA, in June 2003. This software and later releases of it can be used for and provides WPA support to all 1100 and 1200 Series Cisco access points (Cisco Aironet 1100, 2003; Cisco Aironet 1200, 2003) and is available as a free download for all Cisco customers.

Linksys, which is part of Cisco Systems, has introduced both WLAN access points and network cards supporting WPA during the summer 2003. For example the Linksys WLAN PCMCIA adapter, Wireless-G Notebook Adapter Model WPC54G, which became WPA upgradeable after release of new firmware and drivers on June 24th 2003 (Linksys, 2003).

3Com is one of the most recent vendors to obtain Wi-Fi Protected Access certification. The company released a new product family of wireless access points based on the IEEE 802.11g standard (bandwidth support up to 54 M) with WPA support (3Com®, 2003). The 3Com Officeconnect Wireless 11g series products are available from August 2003.

Other WLAN Products with WPA Support

Software that offers WPA support is however not as common. Funk Software was the first company to release a WPA client (also called a *supplicant*), called Odyssey (Funk, 2003), which runs on Windows XP, 2000, 98, Me, Pocket PC, and Windows Mobile 2003. Presently other supplicant software is available, for example Meetinghouse's AEGIS (Meetinghouse, 2003). AEGIS supports all major authentication and encryption protocols. The server is available for Windows, Linux and Solaris, whereas the supplicant is available for Windows 98 and up, Linux, Solaris, Mac OSX, Pocket PC 2002 and Palm Tungsten.

OS Platform Requirements

Windows

Microsoft offers a free download of a WPA supplicant for Windows XP and Windows 2003. The WPA upgrade is not accessible through Windows update. However, it can be found as *Microsoft Knowledge Base Article 815485* (2003). In older Windows operating systems third party supplicant software is still needed to connect to a WPA WLAN, such as AEGIS or Odyssey (Higgins, 2003, p. 7).

Mac

To use WPA in an Apple Macintosh computer an AirPort Extreme base station and AirPort Extreme card is needed. Mac OS X 10.3 or later is also needed (AirPort, 2003). Of course, you will not need a base station if you are connecting to an existing WPA connection.

Unix/Linux

Since the market for Windows computers is larger than for Unix/Linux, there are at the time of writing no drivers with WPA support for any WLAN cards. However, Meetinghouse offers supplicant software for both Linux and Solaris (Meetinghouse, 2003).

Open Source WLAN Security Software

A large selection of Linux based open source WLAN software is presently available (Wireless LAN, 2003). For other open source UNIX implementations like FreeBSD and OpenBSD the selection of open source WLAN software is more restricted (Potter & Fleck, 2003). Also Mac OS based open source WLAN software is being developed (WirelessDriver, 2002).

Most of this open source software consists of drivers for different WLAN cards. Most drivers implement WLAN Client Mode. However, some open source drivers also implement Host Access Point (Host AP) mode.

Setup and configuration of pre-WPA WLAN interface security is described in (Potter & Fleck, 2003) for

- Station computers with Unix, Windows and Mac OS X operating system environments and
- Access point implementation in some open source UNIX based operating system environments.

Development Initiatives

Important open source development initiatives for a WLAN driver supporting Host AP mode and bridging the Linux HostAP, the FreeBSD HostAP and the OpenBSD HostAP drivers are for WLAN cards based on Intersil's Prism2/2.5/3 chipset (Host AP, 2003; Potter & Fleck, 2003).

The Linux HostAP WLAN driver is included in the recently launched Radionet Open Source Environment (ROSE) project, in which fully functional Linux based WLAN Access Point software is published for further open source development (Rose, 2003).

Security Features

Most open source WLAN drivers support WEP. Open source authentication software based on the IEEE 802.1X standard is being developed in the Open1x project (Open, 2003).

Open source IPSec protection of WLAN data communication (WAVElan, 2003) is part of the FreeS/WAN project for open source Linux based IPSec software development (Linux FreeS/WAN, 2003). Open source Linux based IPSec protection of WLAN data communication is also a development project for the National Institute of Standards (NIST) in the USA (NIST IPSec, 2001).

Open source WPA protection of WLAN data communication is being developed for version 2 of the Linux driver for WLAN cards based on Intersil's Prism2/2.5/3 chipset (Host AP, 2003)

Trusted WLAN Communication Related to Business

New WLAN technologies and standards are growing worldwide. Cheap and easy-to-use wireless products in LAN environments, and especially in Wi-Fi WLANs, expand both in enterprise environments as in home office environments.

Main features of WLANs such as flexibility, advisability, and cost savings are drivers. These features are tempting a lot of companies to implement this technology and integrate it in their current IT infrastructure. To obtain the maximum profit from this technology it is necessary to have a proper plan and to find the most appropriate products in order to cover needs and requirements established for a wireless system.

Today, most enterprises have built their IT infrastructure with wired networks, but this field is rapidly changing. WLAN systems are a flexible way of data communication and they are being implemented as an extension to wired networks or even as an alternative to them. Some of the advantages that WLANs provide are:

- Mobile workers have the possibility of communication and data collection from the point of activity.
- Connections running through walls and ceilings with minimum cable involved.
- Faster and less expensive installation and configuration compared to wired alternatives since these need complex infrastructure investments.
- Improved flexibility for expanding business with regards to company needs.

WLAN technology permits synchronization of information and merchandise flow, and permits input of information directly to the enterprise network from the activity point. In addition to this, WLANs give employees the freedom to work where and when they want, increasing their performance in different tasks.

WLAN equipment manufacturers like Intel and Cisco seriously recommend implementation of a security policy that defends the WLAN in case of attacks against it. Dave Juitt, CTO and chief security architect at Bluesocket, states in Vance (2004):

> *"Both WPA and 802.11i should provide reasonable (OSI) Layer 2 security, but any security professional with an ounce of sense will tell you that secure networks, be they wired or wireless, are based on layered security architecture. You begin with the underlying Layer 2 protocols and build up from there. In some deployments, a Layer 2 solution is enough. In others, you may want to add Layer 3 security, perhaps IPSec, on top of that."*

Wireless ISP (WISP)

- The first wireless internet service providers (WISPs) started offering their products in Finland around year 2000. Despite the evolution of WPA, VPN is still the most common way to solve the security issues concerning WLANs. VPN is probably still the primary security solution for wireless ISPs. 'Probably,' because most WISPs do not divulge on their homepages what kind of security solutions they offer. The reason for WISPs not offering WPA to their customers might stem from several reasons:
  - VPN is cheaper since it is software based.
  - Not all card manufacturers offer WPA drivers for their products.

- WISPs wanting to offer WPA know that WPA2 is on the way, and might decide to wait until WPA2 is ready for deployment.
- Security is not a big issue

So why should WISPs want to offer WPA instead of VPN (or other) solutions to customers? After all, WPA is only an interim solution until WPA2 is introduced. For home users, security is not as big an issue as for large enterprises, meaning that a single layer solution might be sufficient. WPA provides the easiest way to offer users a sufficient security package. If an ISP wanted to profile itself as a very secure alternative, this could provide an ample opportunity to carve a niche for themselves.

Wireless Networks Examples Oriented to Business

Two examples are presented in order to study different business opportunities offered by WLAN technology. Network security in small and large companies and in business is discussed generally. The number of people that are working, shopping, surfing and playing games on the Internet is increasing very quickly all over the world. Businesses may gain advantage by taking the security aspect into account, offering secure connections to their customers.

The first example is a Wireless LAN established in a small coffee shop, while the second refers to a large hotel, providing its costumers secure access to the Internet. In both examples, advantages and requirements of installing secure and effective services are analyzed.

Possible benefits from WLAN installations in these kinds of businesses are (WifiConsulting, 2004):

- Business differentiation;
- Catering to customers;
- Expanding clientele;
- Increasing benefit capability (charge for the extra service offered);
- Positive press and word of mouth; and
- Business image.

Requirements for a Small Coffee Shop and a Large Hotel WLAN

Compared to wired networks, the wireless networks provide high performance at a low cost. But on the other hand, there are still speed and bandwidth limitations and more security risks, since wireless networks do not have any physical layer protection. Therefore, just WLAN installation does not ensure success. A typical set of requirements that has to be taken into account is given below. Some of these requirements studied for the small coffee shop can be applied for the large hotel, but most of them have different shades and must be analyzed.

Number of Clients

To offer a good service to your customers it is advisable to take in account the maximum number of clients that the WLAN will support, since the "Bandwidth Rate" affects the quality of service.

Obviously, the number of clients in the hotel will be large compared to the coffee shop. Therefore, it will be more complicated to manage user accounts, to plan topology design, and the need for specialized staff that manages the WLAN must be taken into account.

Prices/Invoicing

The service provider generally considers two possibilities; to offer this service for free, as an extra service to attract customers or directly charge money for the service. Additionally, stemming from the possibility to charge, there are two alternatives: a download rate charge or a time charge. It is possible to consider the same alternatives outlined for the coffee shop.

Roaming (Range Area)

The most benefit is achieved when the whole local area is covered. For a small business like the coffee shop, it should be sufficient to cover the main customer area. On the other hand, to plan the range area for a WLAN in the hotel, where the clients are moving in different areas, a roaming policy must be established.

Clients Information

Helping the customer and offering them good information can be an important step to obtain a successful WLAN service.

Maintenance

To avoid unnecessary costs for the Wireless LAN service, it is advisable to secure a good maintenance service, or even to do it by yourselves. In a case, where the maintenance is more complicated, it is advisable to use outsourcing in order to avoid problems and unnecessary costs.

Selection of Wireless LAN Technology

Each business has different preferences when it comes to deciding which technology to choose. The list of standards offered by the IEEE organization (Wireless LANs, 2004) is a guideline for this business:

- *802.11a.* These products have a data transmission rate of 54 Mbps; covering a WLAN range between 23 and 50 meters. This technology is suitable for a large hotel where high data transmission rates are required.
- *802.11b.* This standard works in the 2.4 GHz band with a data transmission of 11 Mbps. It covers a range between 30 and 75 meters. It is thus ideal for the coffee shop, since the technology is inexpensive.
- *802.11g.* An alternative for both kinds of businesses in the future would be this standard, which will provide 54 Mbps working in 2.4 GHz and also offers compatibility with 802.11b.

Security

All the previously outlined issues are important for improving the success possibilities in your business. But, for avoiding possible legal problems, and for providing a good service to the customers, security is the most important feature to take into account. For example, if a customer needs to carry out an important transaction or to send confidential information and your WLAN security is cracked, you will suffer a security failure in your wireless network service. For this reason it is advisable to secure your wireless network, protecting it from possible attacks.

To protect the WLAN network, "Unique Authentication" and "Customer Data Privacy" must be provided to the customers. In the last years, "Temporal Passwords" have been used to secure "Unique Authentication" in all types of LANs. Currently, a new type of device has been developed to provide "Unique Authentication" to the users, the USB token. The advantage of these tokens is that they are relatively inexpensive and easy to operate. Working with these tokens, the coffee shop can act as a CA (Certificate Authority), since it will be able to enable and disable each USB token with specific control software for this task.

It is possible to get the "Customer Data Privacy" using WPA (Wi-Fi Protected Access) with USB tokens. As mentioned in this chapter, WPA provides authentication and encryption for wireless communication, improving the features offered by WEP. Therefore, for the moment, it is the most advisable protocol to apply in securing the WLAN communication. The same suggestions are offered for the coffee shop and the hotel.

WPA Secured WLAN Example

This section presents a WPA secured enterprise test WLAN set up in Arcada Polytechnic during the winter 2004. With this WLAN practical WPA studies were carried out, implementation and administration experience was gathered and analyzed. These experiences are essential for design, implementation and administration of WLANs for real use.

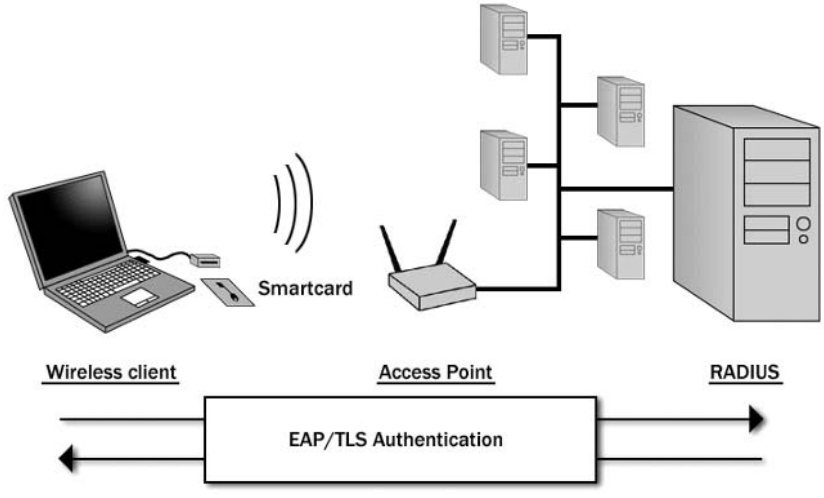
Test Network Overview

The test environment consists of a WLAN access point and Radius server connected to a wired LAN (see Figure 11). The WLAN client is connected to the wired LAN through the access point.

Notebook computers running both MS Windows and Linux operating system were tested as WLAN clients. The Windows client (Windows XP Professional) is patched with the Windows XP support patch for Wi-Fi Protected Access and Windows XP service pack 1 (required by WPA patch). Both patches are downloadable free from Microsoft's Web site (Microsoft, 2003).

WPA functionality in the Linux client, running Fedora Core 1, was achieved by installing and configuring the development versions 0.2.0 of Host AP driver and WPA supplicant. Both are products of the open source based project Host AP (Host AP, 2003). Furthermore a snapshot version of the XSupplicant open

Figure 11. Test network architecture



source based software (Open, 2003), prepatched with WPA support, was downloaded and installed to add support for WPA-EAP authentication.

Following is a list of WPA certified WLAN cards used in the test environment:

- Linksys WPC54G PCMCIA adapter
- Linksys WPC11 V3.0 PCMCIA adapter
- Intel Pro/Wireless LAN 2100 3B Mini PCI Adapter

A Cisco AIR-AP1230B-E-K9 controlled by the operating system 12.2(11)JA3 including WPA support was used in the test environment as a wireless access point. The Radius server is based on Linux open source software, freeRADIUS Version 0.9.0. freeRADIUS supports EAP authentication and is downloadable from FreeRadius (2003). This radius software however requires an EAP/TLS module and a special build before it can be used for PKI authentication, see (McKay, 2003).

User Authentication

The EAP-TLS protocol was used in the test environment for user authentication. PKI based authentication using X509 certificates between the client and the

Radius authentication server is thus defined. Both a soft token approach and hardware token approaches were tested for storing the user's certificate and private key.

Certificate Generation

The Linux based software, OpenSSL, was used for creating needed X509 certificates for the test environment. More information about the OpenSSL software and its latest downloadable versions are available at the OpenSSL portal (OpenSSL, 2003).

The following certificate types were created for the test WLAN:

- CA certificate
- User certificate
- Server certificate

The CA certificate is self signed, and used for issuing user and server certificates. Table 4 shows the main contents of this certificate.

The CA certificate is installed in both the client and the access point. Practically this means that on the wireless client running Windows XP it is installed in the store of "trusted root certification authorities". On the Linux client and Radius server it is stored in a file. On the Windows based client the CA certificate is presented in DER (.der) format while on the Linux based Radius server and wireless client it is presented in PEM (.pem) format.

Table 4. CA certificate

SUBJECT	CN = Test WLAN administrator OU = Research = Arcada L = Espoo S = Uusimaa C = FI
ISSUER	CN = Test WLAN administrator OU = Research = Arcada L = Espoo S = Uusimaa C = FI
PUBLIC KEY	RSA (1024 Bits)

Table 5. User certificate

SUBJECT	E = jonny.karlsson@arcada.fi CN = Jonny Karlsson S = Uusimaa L = Espoo OU = Research = Arcada C = FI
ISSUER	CN = Test WLAN administrator OU = Research = Arcada L = Espoo S = Uusimaa C = FI
PUBLIC KEY	RSA (1024 Bits)
ENHANCED KEY USAGE	Client Authentication (1.3.6.1.5.5.7.3.2) Smart Card Logon (1.3.6.1.4.1.311.20.2.2)
KEY USAGE	Digital Signature, Key Encipherment, Data Encipherment (b0)

The user certificate is digitally signed by the private key of the CA certificate and is used for authenticating a test user against the Radius server. Table 5 shows the main contents of this certificate.

Note that the field "Enhanced Key Usage" with the values "Client Authentication" and "Smart Card Logon" is added to the user certificate. The value "Client Authentication" is required by the EAP/TLS authentication type while "Smart Card Logon" is required by Windows XP when the certificate and private key is stored on a smartcard ore USB token.

The server certificate, shown in Table 6, is generated for the Radius server in purpose to authenticate it on a wireless client. Notable here is that the value of the CN (Common Name) field has to match the Radius server's host name and that the field "Enhanced Key Usage" with the value "Server Authentication" is added to match the EAP/TLS requirements. This certificate is, like the CA certificate, stored in a file on the Radius server in PEM format.

Table 6. Server certificate

SUBJECT	CN = test-radius OU = Research = Arcada L = Espoo S = Uusimaa C = FI
ISSUER	CN = Test WLAN administrator OU = Research = Arcada L = Espoo S = Uusimaa C = FI
PUBLIC KEY	RSA (1024 Bits)
ENHANCED KEY USAGE	Server Authentication (1.3.6.1.5.5.7.3.1)

Authentication Using Soft Tokens

There are different ways how user's X509 certified private keys can be stored. One way is to store them in soft tokens which means storing them in files, such as on a computers hard drive or floppy disk. The first step in the test network development process was to utilize this solution.

For the Windows XP client the OpenSSL software was used to create a PKCS12 (.p12) file of the earlier generated user certificate represented in PEM (.pem) format. The PKCS12 file including the private key was stored on the client computer's hard drive and the user certificate was imported to the "personal certificate store" of the Windows operating system.

On the Linux based client the user certificate and private key is represented in PEM format and stored in different files on the hard drive.

Authentication Using Hardware Tokens

EAP-TLS authentication using X509 certified private keys stored in hardware tokens is a more secure and practical solution than storing them in files. For the test WLAN two different kinds of hardware tokens were tried out with the

Windows XP operating system: Finnish Electronic Identity (FINEID) smart card (Finnish, 2003) and USB based smartcard Aladdin Etoken (Aladdin, 2004).

FINEID Smart Card

It was stated that a FINEID smart card cannot be used for authentication in WPA secured WLANs because it doesn't meet the EAP-TLS and MS Windows requirements for the user certificate. The field "Enhanced Key usage" with the required values "Client Authentication" and "Smart Card Login" is undefined.

This problem can be solved by creating a new certificate matching the EAP-TLS and MS Windows requirements for the existing key pair of the smartcard, as the FINEID smartcard has storage space for an additional certificate. For the test WLAN this was accomplished using the web-enrollment interface, which was developed by Tampere University of Technology in the "Electronic identification in Finnish Higher Education (FEIDHE)" project (see FEIDHE, 2003). This interface consists of the following main components:

- A Web page for a certificate request based on the existing key pair of the user's smartcard;
- A Web page for signing the certificate request and storing the new certificate on the user's smartcard; and
- An OpenSSL configuration file for signing certificate requests.

The Web page used for making certificate requests is created by the CGI script *makecert.cgi* (Purpose: A CGI, 2001), which for the purposes of the test WLAN was modified and simplified. The modified CGI script implements the following functionality:

- It ensures that the issuer of the smartcard certificate is trusted and
- A certificate request is generated using methods provided by the Certificate Enrollment Control (CEnroll) object, included in Windows NT/2000/XP, which can be called from Visual Basic scripts.

To create the Web page used for certificate signing and storing another CGI script, *signcert.cgi,* (Purpose: Sign, 2001) was used. After modifications this CGI script implements the following functionality:

- The certificate request, retrieved as a CGI parameter, is signed using the OpenSSL interface to the OpenSSL command provided by the Perl modules of the OpenCA project and
- The new certificate is stored on the smartcard using the Visual Basic Script methods of the CEnroll object.

After these script driven operations a FINEID smart card can be used for authentication in the WPA secured test WLAN.

The following hardware and software components are needed in order to make the smartcard working under Windows XP and accessible for the WLAN supplicant:

- Smartcard reader + driver
- PKI client software

In the test WLAN an Utimaco CardMan 2020 USB smartcard reader and the PKI client software, SetCSP Certificate Manager version 1.52 Beta1, were used.

USB Token

The Aladdin eToken is simpler to prepare for WLAN authentication purposes compared to a FINEID smartcard since private keys and certificates can be imported directly using, i.e. the Windows XP Certificate Import Wizard. In other words a X.509 certified private key can be, generated and stored in PKCS12 file format, as described in section Authentication Using Soft Tokens, and as such imported to the eToken. The certificate requirements for the eToken are the same as for the FINEID smartcard except that the maximum length of the private key is 1024 bit (FINEID cards supports 2048 bit keys).

Two software packages were installed on the Windows XP client to make Aladdin eToken usable for the operating system and WLAN supplicant:

- eToken PKI Client (RTE) Version 3.51
- eToken Utilities Version 2.10

These software packages are available for free download from Aladdin (2004).

Security Configuration Overview

The security parameters of the Radius server are configured via the freeRADIUS configuration files: *clients.conf*, *radiusd.conf* and *users*. These files are, after default freeRADIUS installation, located in the */etc/raddb/* directory. The Radius server is controlled by the *radiusd* demon. More information about the freeRADIUS configuration files can be found in FreeRadius (2003).

The Cisco AIR-AP1230B-E-K9 access point is configurable via a web interface after installation to a LAN. Note that, as shown in Table 7, the access point does

Table 7. Security configurations

RADIUS	**Clients.conf** A configuration file where accepted Access Points are specified. Here the IP address and a shared secret for the Access Point are defined **radiusd.conf** The following security features are defined in this configuration file: • Authentication type: EAP/TLS • Path to the server's private key and certificate file • Path to the file containing trusted root CA certificates **users** A configuration file in which accepted users to the test WLAN are specified in form of usernames
ACCESS POINT	**Encryption cipher:** TKIP **Authentication methods accepted:** Open authentication with EAP **Authenticated key management:** WPA **Radius server information:** IP address, port number and password for the Radius server **Use Radius server for:** EAP authentication **Broadcast Key Rotation Interval:** Once an hour
WINDOWS CLIENT	**Network Authentication:** WPA **Data Encryption:** TKIP **EAP type:** Smartcard or other certificate **When connecting:** Use my smartcard **Validate server certificate:** Yes **Trusted root CAs:** The issuer (CA certificate) of the Radius server's server certificate selected from a list of trusted root certification authorities.
LINUX CLIENT	**wpa_supplicant.conf** Configuration file where security features such as encryption algorithm and authentication types are specified. For the test environment EAP-TLS and TKIP were chosen. **xsupplicant.conf** EAP authentication parameters are modified in this configuration file such as path to the user certificate and trusted root CA file.

not perform user authentication. The authentication information is passed to the specified Radius server. The broadcast key is configured to be updated every 3600 second. The options on the Cisco AP broadcast key rotation are 10-10000000 seconds. For more detailed configuration instructions (see Cisco IOS, 2003).

The client running Windows XP is configured via the user interface provided by the Windows XP support patch for WPA. This user interface replaces the old WLAN configuration interface, "wireless network properties", after installing the WPA patch. Note that, as shown in Table 4, apart from showing its own certificate the client also is required to verify the certificate of the Radius server.

After installing the Host AP driver, the security settings of the Linux client are configurable via two configuration files, *wpa_supplicant.conf* and *xsupplicant.conf*. For starting the Linux based WPA supplicant in WPA-EAP mode the WPA supplicant and XSupplicant daemons are executed in parallel.

Security Policy Overview

The security policy for the test WLAN is quite simple. Every user needs a personal user certificate issued by a certification authority trusted by the Radius server before they can access it. Since mutual authentication is required by WPA, also the Radius server needs a server certificate issued by a certification authority trusted by the client. In the test WLAN the certification authority is the administrator of the test WLAN and this authority is the issuer of both the user- and the server certificate used for testing.

When a user connects to the WLAN, apart from the user certificate, the user also must present a username for the WLAN. If this username is not specified in the freeRADIUS configuration file *users*, and if it doesn't match the CN value of the smartcard certificate, then access to the WLAN is denied.

If a smartcard is used for authentication, a user who owns a FINEID smartcard can register to the WLAN via the web enrollment interface. With this a valid user certificate for the WLAN, issued by the administrator of the WLAN, is stored in the user's smartcard.

Test Design

In the following is given a presentation of the most interesting test results gathered from experiments made with the test WLAN. The purpose of the experiments was to find out how WPA affects connect time and bandwidth in

a WLAN compared to other security solutions, such as WEP and IPSec. Different WLAN cards were used for the tests to investigate the possible differences in the functionality between them.

Test Setup and Test Methods

The following components were used in the test environment:

- Cisco AIR-AP1230B-E-K9 wireless access point controlled by the operating system 12.2(11)JA3 connected to a 100Mbps LAN
- Radius server, freeRADIUS 0.9.0, running on Redhat Linux 9.0 on a 100Mbps LAN connection
- Notebook computer running Windows XP WPA supplicant on Windows XP Professional with a 11Mbps wireless connection
- Notebook computer running WPA supplicant , XSupplicant and FreeS/WAN 2.0.4 on Fedora Core Linux with a 11 Mbps wireless connection
- Desktop computer running FreeS/WAN 2.0.4 on Redhat Linux 9.0 acting as an IPSec Gateway

The connect time was measured from the moment of time when SSID was chosen in the wireless client until the moment of time an IP address was obtained from the DHCP server in the LAN. The time to dial passwords and WEP keys was not included in the total connect time. Each connect time value in Table 8 and Table 9 is an average of 10 measured connect time values.

Transfer rate was tested by downloading to the wireless clients a 125 Mb software package from a computer, located in the LAN. The package was downloaded and the download time measured five times in each security mode to make sure that no outside factors affected the bandwidth during the download process.

The network architecture during testing IPSec was a little bit different from the architecture presented earlier in Figure 11. Here the wireless access point was located behind a desktop computer acting as a gateway between the WLAN and the LAN. The data communication was then encrypted, using IPSec, all the way from the wireless client, through the wireless access point, to the WLAN gateway. The IPSec communication was accomplished by installing and configuring the open source Linux software FreeS/WAN 2.0.4 on both ends (Linux based client and WLAN gateway).

Table 8. Test results with Windows XP client

	Linksys WPC11 V 3.0 PCMCIA Adapter	Linksys WPC54G PCMCIA Adapter	IntelPro/Wireless LAN 2100 3B Mini PCI Adapter
UNPROTECTED			
Connect time	6,8 s	4,3 s	5,6 s
Transfer rate	601 Kb/s	754 Kb/s	631 Kb/s
WEP			
Connect time	6,9 s	4,3 s	5,0 s
Transfer rate	599 Kb/s	749 Kb/s	445 Kb/s
WPA-PSK			
Connect time	11,7 s	11,2 s	10,8 s
Transfer rate	599 Kb/s	767 Kb/s	412 Kb/s
WEP-EAP			
Connect time	14,8 s	14,6 s	11,8 s
Transfer rate	607 Kb/s	744 Kb/s	466 Kb/s
WPA-EAP			
Connect time	15,1 s	13,9 s	14,0 s
Transfer rate	604 Kb/s	740 Kb/s	407 Kb/s

Test Result Analysis

The client running Windows XP was tested in several security modes with three different WLAN cards. The results are presented in Table 8.

As visible from the table, WPA protection delays the time to connect to a WLAN with a few seconds. WPA-EAP security mode results in the longest connect time. However, compared to WEP-EAP authentication there are no significant differences in the connect time.

From these test results it can be stated that WEP- and WPA encryption significantly decreases the transfer rate with the Intel Pro/Wireless LAN card. However, no significant change on the transfer rate is discovered with any of the Linksys cards in any security mode.

The same tests were performed with the Linux client but only with the Linksys WPC 11 V3.0 WLAN card, since the HostAP driver only supports WLAN cards based on the Intersil Prism 2/2.5/3.0 chipset. Here the transfer rate with IPSec also was tested. Table 9 shows the results.

The transfer rate during downloading the 125 Mb software package to the Linux client was lower in WPA mode than in WEP mode but not crucially. IPSec caused the lowest transfer rate.

Table 9. Test results with the Linux client

	Linksys WPC11 V3.0 PCMCIA Adapter
UNPROTECTED	
Connect time	4,0 s
Transfer rate	650 Kb/s
WEP	
Connect time	4,0 s
Transfer rate	648 Kb/s
WPA-PSK	
Connect time	7,2 s
Transfer rate	639 Kb/s
WEP-EAP	
Connect time	82,9 s
Transfer rate	648 Kb/s
WPA-EAP	
Connect time	9,5 s
Transfer rate	639 Kb/s
IPSec	
Transfer rate	615 Kb/s

The differences in the connect times between different security modes were about the same as with the Windows XP client but overall the Linux client connected in less time to the AP than the Windows XP client.

By studying these test results it can be generally concluded that WPA causes negative effects on both bandwidth and connect time. Compared to static WEP, the connect time is a little bit slower with WPA, but compared to dynamic WEP the connect time is about the same. Considering bandwidth, the performance with WEP and WPA are similar while the bandwidth performance with WPA is much better than with the presently most used WLAN security solution, IPSec.

Differences in how WPA affected bandwidth using different WLAN cards were detected. WPA caused the largest loss of bandwidth with the Intel WLAN card while no loss was detected with the two other cards in the Windows XP client. However, when the Linksys WPC11 card was used in the Linux client, WPA did affect the bandwidth somehow. Concerning connect time the values are about the same with all WLAN cards.

WLAN Security Management

Online network monitoring is a cornerstone in WLAN security in order to:

- Detect intrusion;
- Detect intrusion attempts;
- Identify threats and vulnerabilities; and
- Enforce WLAN security policies.

Security Related Network Monitoring

Online airwave monitoring requires sensors in the vicinity of each access point in all BSS and ESS segments of a WLAN. For an IBSS WLAN the entire possible communication range must be covered by monitoring sensors. This can be extremely hard to achieve since the coverage of an IBSS WLAN depends on the network topology, which is dynamic and unpredictable. Monitoring sensors also must use the monitored WLAN as communication agents and they must be controlled from at least one management station.

Detection of malicious failures is essential in attack identification. The common occurrence of benign failures like transmission impairments, path breakages, and dropped packets makes it difficult to identify malicious failures especially in an IBSS WLAN with many mobile nodes (Papadrimitratos & Haas, 2002).

A monitoring system for a BSS or ESS WLANs thus consists of:

- Distributed sensors close to all access points and
- Management servers.

In an available BSS/ESS WLAN monitoring system, AirDefense, (AirDefense, 2004; Understanding, 2003) remote sensors provide 24*7 monitoring of all WLAN communication to/from access points and online monitor reporting to a management server. A management server of this monitoring system:

- Analyzes WLAN traffic in real time;
- Detects intrusion and disconnect intruders;
- Detects impending threats and protects against attacks;
- Monitors WLAN performance;
- Troubleshoots network issues;
- Offers a secure Web-based administrator interface; and
- Enforces WLAN policies like identification of policy violations for:
    - Unauthorized stations;

- Unencrypted and unauthorized traffic;
- Use of unauthorized channels; and
- Off hours traffic.

Intrusion Management

A multi-dimensional intrusion detection approach is required, since no single method can detect all similar possible intrusions into a WLAN. Intrusion detection methods for WLAN are surveyed in Wireless (Lim, 2003). These methods are either pattern recognition based like *signature recognition* and *security policy deviation* or statistical like *anomaly analysis*:

- *Signature Detection:* Known WLAN traffic patterns of earlier detected intrusions/intrusion attempts are stored in a signature database similar to the virus signature database of anti-virus software. WLAN traffic is analyzed in real time to find pattern matches in the signature database. The signature database must be updated after intrusions/intrusion attempts based on new WLAN traffic patterns. For WLANs these signatures also must include 802.11 protocol specific attacks. The intrusion detection system should be able to process data frames in the airwaves to prevent disturbances in WLAN devices from corrupted.
- *Security Policy Deviation:* The intrusion detection system generates alarm, when pre-set policy and performance thresholds are violated. Access to wireless frame data from the airwave is usually necessary.
- *Protocol Analysis:* 802.11 MAC protocols are monitored to detect deviations from standards.
- *Statistical Anomaly Detection*: Statistical anomalies are deviations from normal WLAN behavior. Normal WLAN behavior is based on earlier WLAN monitoring data.

Typical responses to intrusion events/attempt are:

- Event logging;
- Alarms to the WLAN hardware/software;
- Alerts to the WLAN administrator/intruder; and
- Disconnections of intruders from valid WLANs.

WLAN Security Related Research

General security goals in any computer network are (Papadrimitratos & Haas, 2002):

- *Availability*: It means survivability of network services despite misbehavior of network nodes and other disturbances. In a WLAN and especially in an ad hoc WLAN, DoS attacks can be launched at any network layer. An ad hoc WLAN is blocked, if the route discovery procedure is disabled.
- *Integrity*: Integrity means that data is not altered in transfer. Integrity is preserved when no information is removed, replayed, reordered or unlawfully inserted in the communication between WLAN nodes.
- *Authentication*: Authentication means that the identity of WLAN nodes and users is ensured.
- *Confidentiality*: Confidentiality ensures that certain information is never disclosed to unauthorized entities. In an ad hoc WLAN routing information can be used to locate nodes and should therefore be protected.
- *Non-repudiation*: Non-repudiation ensures that the sender of data cannot deny occurred data transfer. In a WLAN non-repudiation is useful for detection and isolation of compromised nodes.
- *Authorization*: Authorization established access rules. In a WLAN authorization defines conditions for access to information and to network resources.

Recent research of WLAN security has focused on:

- Roaming security;
- IBSS WLAN security (Secure Mobile Ad Hoc Networks); and
- Improved WLAN security management improvement.

A survey on recent wireless networking research is published in Gavrilovska, Yomo, Wijting, Popovski and Koslova (2002). Recent research on WLAN security is also covered in this survey.

Secure Roaming

Wireless roaming is defined here as a process in which a wireless station (client) can move across multiple Access Points in one domain, and across multiple Access Points in differing domains (VeriSign, 2004). A dial-up connection from a wireless station through a Wireless Internet Service Provider (WISP) to a corporate network (such as a home domain) is a typical application.

The 802.1X Framework including EAP defines secure WLAN authentication. EAP provides support for a number of authentication schemes within PPP. Such authentication schemes are smart cards, Kerberos, public key, passwords, etc. Mutual authentication based on 802.1X, EAP-TLS and certificates are used. Two relationship based and certificate based roaming models are presented (VeriSign, 2004).

As a third example, seamless roaming in a heterogeneous network environment including WLAN and GPRS/UMTS is briefly discussed (Zivkovic, Lagerberg & van Bemmel, 2004). Mobile IP has been chosen as the solution for session mobility.

Roaming within a Domain

Below is a general description of roaming between access points (Schiller, 2000):

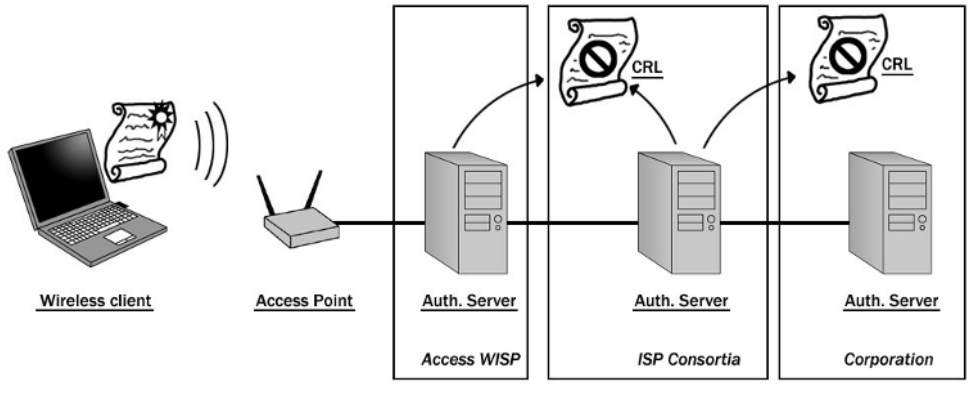
- A station starts scanning for another access point if the quality of the current link is too poor.
- The station goes into the scanning phase, for example, searching for another BSS (BASIC Service Set) or setting up a new BSS (ad-hoc) starts. Both passive and active scanning is possible. *Passive scanning* comprises listening to the medium, for example, receiving the beacon signal from another network. *Active scanning* includes sending a probe signal on each channel and collecting information from the beacon and the probe response signals.
- The access point with the highest signal strength is chosen and an association request is sent to the chosen access point.
- If the association request from the new access point to the station is successful, the station has roamed to a new BSS.
- The distribution system (DS) updates its database and informs the old access point that the station is no longer within its BSS.

Roaming across Domains

We present three models for WLAN wireless roaming (VeriSign, 2004; Zivkovic et al., 2004):

- *Relationship based roaming*: WISPs have agreements with other ISPs to allow customers to connect to one another's access points or WISPs/ISPs form a consortia acting as a clearinghouse. This model includes storing information about domains, users, routing, management, pricing and billing.
- *Certificate based roaming*: A user presenting a certificate is authenticated by the proxy access WISP. The 802.1X framework with EAP-TLS and Authentication Servers are deployed. The Certificate Revocation List (CRL) issued by the Certificate Authority (CA) is used for verification, see Figure 12. To protect end-to-end data VPN/IPSec can be deployed.
- *Mobile IP based roaming*: The WLAN supplicant authenticates to a SP's (Service Provider) RADIUS server, not to the WLAN itself. For authentication, the 802.1X/EAP-TLS protocol is used. In order to check user identity and the existence of a roaming agreement and to forward messages to the RADIUS server a RADIUS proxy is deployed. Once the authentication is verified, the certificate keys are used for data confidentiality. The supplicant obtains a local IP address via DHCP and registers itself via a Mobile IP tunnel at the home agent. Authentication to a GPRS base station involves the SIM card and the Home Location Register (HLR). The client automatically re-registers at the home agent. When the client enters a foreign WLAN the authentication is done in the same way as in the home WLAN.

Figure 12. Certificate based WLAN roaming



Secure Mobile Ad Hoc Networks

Much recent wireless network research focuses on security in mobile ad hoc networks. Basic physical security requires tamper free network nodes (Stajano & Anderson, 1999). A prerequisite for secure operation is a sufficient level of trust in network nodes (Papadrimitratos & Haas, 2002). Trust and trust relationships depend on network node behavior.

Trust Management

Introduction and use of definitions for credentials, trust levels, trust relationships and security policies are components of trust management. Trust management based on manually reconfigurable credentials and network node interaction rules (Stajano & Anderson, 1999) is possible only in small scale IBSS networks. In other IBSS networks trust management is based on cryptographic techniques such as Public Key Infrastructure (PKI). However, existing trust management solutions for distributed computer networks cannot be used in the IBSS context, because there is no network hierarchy and no central entity in an IBSS network (Papadrimitratos & Haas, 2002).

IBSS network nodes have many unpredictable and vulnerable features. PKI fault tolerance is thus a necessity in trust management based on PKI. PKI fault tolerance can be implemented with *threshold cryptography* (Cryptography, 2004). Solutions to distribute Certificate Authority (CA) functionality across multiple nodes in an IBSS network are presented in Kong, Zerfos, Luo, Lu and Zhang (2001), Yi and Kravets (2003), and Zhou and Haas (1999).

A *threshold signature* scheme is proposed in Zhou and Haas (1999). The signing key is divided into shares stored on different network nodes. A threshold signature is obtained only if a sufficiently large subset of partial signatures using signing key shares can be created. A threshold signature based on at least 2 partial signatures is called a (n,2) threshold cryptography scheme, where n is the number of signing key shares. Enhanced protection of a threshold signature set against compromised nodes is achieved by regular signing key share refreshments. In share refreshment a set of signing key shares is replaced by a new share set, which is calculated from the old share set without disclosing the signing key.

Secure Routing

The unpredictable and dynamic topology of IBSS networks is a source of routing complexity. Routing in IBSS networks is a rich research field (Giordano,

Stojmenovic & Blazevic, 2003). The IETF Routing Area Working Group "Mobile Ad-hoc Networks (manet)" proposes focusing on further development of following routing protocols (IETF Routing, 2004):

- Ad Hoc On-Demand Distance Vector (AODV)
- Optimized Link State Routing (OLSR)
- Topology Dissemination Based on Reverse-Path Forwarding (TBRPF)
- Dynamic Source Routing (DSR)

Routing protocols in computer networks with a fixed infrastructure are usually open and unprotected. A manifestation of an emerging recognition of routing security in the Internet community is the recently formed IETF Routing Area Working Group "Routing Protocol Security Requirements (rpsec)", which in December 2003 published an Internet Draft "Generic Threats to Routing Protocols" (Barbir, Murphy & Yang, 2003). In an IBSS network, which has no fixed infrastructure, secure routing protocol operation is a recognized necessity (Papadimitratos & Haas, 2002).

Secure routing in an IBSS network starts with route discovery protection, by choosing routes satisfying predefined security criteria (Yi, Naldurg & Kravets, 2001). At node initiation a route discovery defines the required minimum trust level for nodes participating in the query/reply propagation. Two security extensions of the Ad hoc On-Demand Distance Vector (AODV) routing protocol and a routing protocol designed for security in the presence of malicious network nodes, the Secure Routing Protocol (SRP), are described in Papadimitratos and Haas (2002)

Secure Data Forwarding

Although a correctly discovered route and a secure routing protocol are prerequisites for data forwarding in an IBSS network, there is still no guarantee that the trusted network nodes along a correctly discovered route will indeed relay data as expected. Also a secure and fault tolerant data-forwarding scheme like the proposed Secure Messaging Protocol (SMT) is needed (Papadimitratos & Haas, 2003).

SMT maintains for a (source, destination) pair an Active Path Set (APS) of all paths which have not been detected as failed. Messages can be dispersed into pieces, which are transmitted along different paths in an APS. Timeout of acknowledgement messages for received message pieces causes re-transmission of lost pieces along APS paths. Messages are reconstructed from pieces at

the destination. SMT relies on a unidirectional security association from source to destination and protects data communication with efficient symmetric key cryptography.

Improved WLAN Security Management

Intrusion detection is incomplete and intrusion response is insufficient in presently available tools for WLANs. A prototype for improved WLAN intrusion detection combined with active intrusion response is described in Lim and Owen (2003). Proposals for improved intrusion detection are:

- Detection of passive intruders by monitoring the IEEE 802.11b "Request to Send (RTS)" and "Clear to Send (CTS)" data frames;
- Determination of unique signatures for each kind of attack; and
- Improved detection accuracy by using attack profiling algorithms like rule based algorithms and neural network algorithms, which "learn" normal network behavior.

Proposals for better intrusion response are:

- Active intrusion responses like sending specially crafted malformed data frames to an intruder doing a DoS attack against a WLAN with flooding and
- Confusing passive attackers with broadcasts of corrupt data.

Conclusion

Present WLAN interoperability is a result of successful international LAN standardization work in the late 1990's. However, the first attempts to implement security features for the wireless radio communication media and for the connections between WLAN network units were incomplete. Wired Equivalent Privacy (WEP), introduced in 1999, included no authentication procedure and the implementation was flawed. WEP encryption can be cracked simply by recording a sufficient amount of WEP encrypted data communication even for encryption key lengths exceeding 100 bit, which is a secure key length for all standardized symmetric encryption algorithms.

The new WLAN security standard, Wi-Fi Protected Access (WPA), introduced in 2004 by the Wi-Fi Alliance is based on recent achievements in LAN standardization work of the IEEE. The proposed authentication procedure also protects network ports in wired LAN's from physical connections of unauthorized devices. Key management procedures for data communication support even per-packet keys. The security of the data communication is thus as good as the used symmetric encryption algorithms. The EAP-TLS authentication even allows user identity based authentication for user identity certificates with Enhanced Key Usage defined to support Client Authentication. This WLAN feature has been successfully validated by a test WLAN for both soft and hardware tokens.

WPA also has scalability features. Authentication based on the use of a dedicated AAA server can be used in WLANs of any size. For small WLANs WPA offers Pre-Shared Key authentication. This authentication gives the same level of security as AAA server based authentication, if the quality of the pass phrase for the Pre-Shared Key is sufficient.

The present WPA version (version 1) allows software upgrading of pre-WPA WLAN hardware to WPA. The next version, WPA2, has an improved key management protocol based on the use of the recent symmetric encryption standard AES. The data communication security level will thus be higher for WPA2 and the implementation efficiency will still outperform WPA. However, WPA2 also requires support from WLAN hardware.

References

3Com® OfficeConnect® Wireless 11a/b/g PC Card. Retrieved December 13, 2003, from *http://www.3com.com/prod/fi_FI_EMEA/detail.jsp?tab=features&sku=3CRWE154A72*

802.16a/WiMAX, The Next Step after WiFi is Here. Visant Strategies, Jan. 21.2004. Retrieved March 15, 2004, from *http://www.visantstrategies.com/pr80216.htm*

Adoba, B. & Simon, D. (1999). *PPP EAP TLS Authentication Protocol*. IETF RFC 2716. Retrieved March 15, 2004, from *http://www.faqs.org/rfcs/rfc2716.html*

Agilent Technologies. (2000) WLAN Certification. Retrieved December 13, 2003, from *http://wireless.agilent.com/WLAN/qual/wlan_full.shtml*

AirDefense Portal. Retrieved March 14, 2004, from *http://www.airdefense.net*

AirPort Extreme (2003). Retrieved December 13, 2003, from *http://www.apple.com/airport/*

AirSnort Portal. Retrieved March 3, 2004, from *http://airsnort.shmoo.com*

Aladdin Knowledge Systems – Software security, software protection, Internet security, content (2004). Retrieved March 12, 2003, from *http://www.ealaddin.com*

Barbir, A., Murphy, S. & Yang, Y. (2003). *Generic Threats to Routing Protocols*, Internet-Draft, Internet Engineering Task Force, December 17. Retrieved March 13, 2004, from *http://www.ietf.org/internet-drafts/draft-ietf-rpsec-routing-threats-04.txt*

Borisov, N., Goldberg, I. & Wagner, D. (2001). *Intercepting mobile communications: The Insecurity of 802.11*, ACM MobiCom 2001, 180-188.

Burns, J. & Hill, J. (2003). *Evolution of WLAN Security*. White Paper. Retrieved December 12, 2003, from *http://meetinghousedata.com/MDC_Evolving_Standards.pdf*

Cam-Winget, N., Moore, T., Stanley, D. & Walker, J. (2002) *802.11i overview*. Retrieved December 10, 2003, from *http://csrc.nist.gov/wireless/S10_802.11i%20Overview-jw1.pdf*

Catalyst 6500 Series Switch Cisco IOS Software Configuration Guide—Release 12.1 E (2003). Chapter 25. Configuring IEEE 802.1X Port-Based Authentication. Retrieved December 14, 2003, from *http://www.cisco.com/univercd/cc/td/doc/product/lan/cat6000/12_1e/swconfig/dot1x.pdf*

CERT® Coordination Center. Denial of Service Attacks. Retrieved February 27, 2004, from *http://www.cert.org/tech_tips/denial_of_service.html#1*.

Certified Products Listing. Wi-Fi Alliance. Retrieved December 12, 2003 from *http://www.wi-fi.org/OpenSection/Certified_Products.asp*

Cisco Aironet 1100 Serias Data Sheets. (2003). Retrieved December 12, 2003, from *http://www.cisco.com/en/US/products/hw/wireless/ps4570/products_data_sheets_list.html*

Cisco Aironet 1200 Serias Data Sheets. (2003). Retrieved December 12, 2003, from *http://www.cisco.com/en/US/products/hw/wireless/ps430/products_data_sheets_list.htm*

Cisco IOS Software Configuration Guide for Cisco Aironet Access Points. (2003). Retrieved December 14, 2003 from *http://www.cisco.com/application/pdf/en/us/guest/products/ps4570/c1069/ccmigration_09186a0080184b11.pdf*

Cryptography and Information Security Group Research Project: Threshold Cryptology. Laboratory for Computer Science, Massachusetts Institute of

Technology. Retrieved March 13, 2004 from *http://theory.lcs.mit.edu/~cis/cis-threshold.html*

Disabato, M. (2003). *Wi-Fi Protected Access: Locking down the Link*. Retrieved December 14, 2003, from *http://www.weca.net/OpenSection/pdf/Wi-Fi_ProtectedAcessWebcast_2003.pdf*

Draft Supplement to ISO/IEC 8802-11/1999(I) ANSI/IEEE Std 802.11, 1999 edition. *IEEE Std 802.11i/D3.0*, November 2002.

Ethereal Portal. Retrived March 3, 2004, from *http://www.ethereal.com*

Extensible Authentication Protocol Transport Layer Security Deployment Guide for Wireless LAN Networks. White Paper. (2002). Retrieved December 12, 2003, from *http://www.cisco.com/warp/public/cc/pd/sqsw/sq/tech/acstl_wp.htm*

FEIDHE pilots. (2003). Retrieved Decenber 15, 2003, from *http://www.csc.fi/suomi/funet/middleware/english/summary.html*

Finnish Electronic Identity Card Portal. (2003). Retrieved December 12, 2003 from *http://www.fineid.fi*

FreeRADIUS Portal (2003). Retrieved December 13, 2003, from *http://www.freeradius.org*

Funk Software Portal (2003). Retrieved December 13, 2003, from *http://www.funk.com/*

Gavrilovska, L., Yomo, H., Wijting, C., Popovski, P. & Koslova Madsen, T. (2002). *FACE Future Adaptive Communication Environment. D 1.1 – Status Report on state-of-the-art in short-range networking*, Center for PersonKommunikation, Aalborg University, Denmark. Retrieved March 15, 2004, from *http://cpk.auc.dk/FACE/documents/deliverable%201.1.pdf*

Giordano, S., Stojmenovic, I. & Blazevic, L. (2003). Position based routing algorithms for ad hoc networks: A taxonomy. In Cheng, X., Huang, X., & Du, D. Z. (Eds.), *Ad hoc wireless networking*.

Goransson. P. (2003). *WPA Will it Deliver Required Enterprise-class Security for Wireless LANs?* 30th Annual Computer Security Conference & Exhibition, Washington, D.C., November 3-5. Retrieved December 12, 2003, from *http://csiannual.com/classes/f5.pdf*

Higgins, T. (2003). *Wi-Fi Protected Access (WPA) NeedToKnow – Part II* . Retrieved December 13, 2003, from *http://www.smallnetbuilder.com/Sections-article50.php*

Hill, J. (2001). *An Analysis of the RADIUS Authentication Protocol*, InfoGard Laboratories. Retrieved March 15, 2004, from *http://www.untruth.org/~josh/security/radius/radius-auth.html*

Host AP driver for Intersil Prism2/2.5/3. (2003). Retrieved December 13, 2003, from *http://hostap.epitest.fi/*

IEEE Standards. Retrieved December 12, 2003, from *http://standards.ieee.org*

IETF Portal. Retrieved December 12, 2003, from http://www.ietf.org/

IETF Routing Area Working "Group Mobile Ad-hoc Networks (manet)" Portal. Retrieved March 13, 2004, from *http://www.ietf.org/html.charters/manet-charter.html*

Intel: VPN and WEP (2003). Retrieved March 12, 2004, from *http://www.intel.com/business/bss/infrastructure/security/vpn_wep.pdf*

IP Security Protocol (ipsec). Retrieved March 12, 2004, from *http://www.ietf.org/html.charters/ipsec-charter.html*

Kismet Portal. Retrieved March 3, 2004, from *http://www.kismetwireless.net*

Kong, J., Zerfos, P., Luo, H., Lu, S. & Zhang, L. (2001). Providing robust and ubiquitous security support for mobile ad-hoc networks. In *Proceedings of ICNP'01*.

LaRosa, J. (2003). *WPA: A Key Step Forward in Enterpriser-class Wireless LAN (WLAN) Security*. White Paper Retrieved December 12, 2003, from *http://meetinghousedata.com/MDC_WP_052603.pdf*

Lim, Y. L. & Owen, H. L. (2003). Wireless intrusion detection and response. In *Proceedings of the 2003 IEEE Workshop on Information Assurance*, United States Military Academy, West Point, NY, June 2003.

Linksys Wireless-G Notebook Adapter WPC54G. Retrieved December 13, 2003, from *http://www.linksys.com/products/product.asp?grid=33&scid=36&prid=507*

Linux FreeS/WAN Portal. (2003). Retrieved December 13, 2003, from *http://www.freeswan.org/*

McKay, R. (2002). FreeRADIUS EAP/TLS - WinXP HOWTO. Retrieved December 14, 2003, from *http://www.impossiblereflex.com/8021x/eap-tls-HOWTO.htm*

Meetinghouse Data Communications Portal (2003). Retrieved December 13, 2003, from *http://www.mtghouse.com/*

Microsoft Corporation. (2003). TLS Handshake Protocol. Retrieved December 15, 2003, from *http://msdn.microsoft.com/library/default.asp?url=/library/en-us/security/security/tls_handshake_protocol.asp*

Microsoft Knowledge Base Article – 815485 (2003). Retrieved December 13, 2003, from *http://support.microsoft.com/?kbid=815485*

Networking and Communications. Wireless Ethernet Devices. Wireless Security – 802.1X and EAP Types. Retrieved December 12, 2003, from *http://www.intel.com/support/network/wireless/seceap.htm*

Nicopolitidis P., Obaidat M. S., Papadimitriou G. I., & Pomportsis A. S. (2001). *Wireless networks*, Wiley.

NIST Ipsec Project. (2001). Retrieved December 13, 2003, from *http://csrc.nist.gov/ipsec/*

Open Source Implementation of IEEE.1X. (2003). Retrieved December 13, 2003, from *http://www.open1x.org/*

OpenSSL Portal (2003). Retrieved December 13, 2003, from *http://www.openssl.org/*

Overview Wi-Fi Protected Access. (2002). Retrieved December 12, 2003, from *http://www.wi-fi.org/OpenSection/pdf/Wi-Fi_Protected_Access_Overview.pdf*

Papadrimitratos, P. & Haas, Z. J. (2003) Secure message transmission in mobile ad hoc networks. *Elsevier Ad Hoc Networks Journal*, *1*(1).

Papadrimitratos, P. & Haas, Z. J. (2002). Securing Mobile Ad Hoc Networks. In Ilyas, M. (Ed.), *Handbook of ad hoc wireless networks*, CRC Press.

Potter, B. & Fleck. B. (2003). *802.11 Security.* USA: O'Reilly.

Purpose: A CGI script to generate a Certificate Request based on existing Smart Card Certificate. (2001). Retrieved December 15, 2003, from *http://www.tut.fi/HSTYA/xenroll/makecert_cgi.txt*

Purpose: Sign a Certificate Request and install the resulting Certificate on Smart Card. (2001). Retrieved December 15, 2003, from *http://www.tut.fi/HSTYA/xenroll/signcert_cgi.txt*

Rose RADIONET OPEN SOURCE ENVIRONMENT. (2003). Retrieved December 13, 2003, from *http://www.rosewlan.org*

Schiller, J. (2000). *Mobile communications*. London : Addison-Wesley.

Seattle Wireless. Retrieved March 3, 2004, *http://www.seattlewireless.net/index.cgi/HotNews*

Sourceforge Project wepcrack (2001). Retrieved December 12, 2003, from *http://sourceforge.net/projects/wepcrack*

Stajano, F. & Anderson, R. (1999). The resurrecting duckling: Security issues for ad hoc wireless networks. *Security Protocols, 7th International Workshop*, LNCS, Springer-Verlag.

Tanzella, F. (2003). *Wireless LAN Intrusion Detection & Management*. Technical White Paper, AirDefence, Inc. Retrieved March 14, 2004, from *http://www.airdefense.net*

The Hacker's Choice. Retrieved March 3, 2004, from *http://www.thc.org*

Understanding The Layers of Wireless LAN Security & Management. (2003). White Paper, AirDefence, Inc. Retrieved February 23, 2004 from *http://www.airdefense.net*

Vance, J. (2004). Making Sense of Evolving WLAN Standards, Part 1: Security. *E-Security Planet*, January 5, 2004. Retrieved March 15, 2004, from *http://www.esecurityplanet.com/trends/article.php/11164_3295031_2*

VeriSign, Inc., Secure Global Roaming for 802.11 WLANs. White Paper. Retrieved March 8, 2004, from *http://research.verisign.com/Papers/VeriSign-WLAN-Security.pdf*

Walker, J. 802.11 *Security Series Part III: AES-based Encapsulations of 802.11 Data*. Retrieved December 10, 2003, from *http://cedar.intel.com/media/pdf/security/80211_part3.pdf*

WarLinux. Retrieved March 3, 2004, from *http://sourceforge.net/projects/warlinux*

WAVElan SECurity using Ipsec. Retrieved December 13, 2003, from *http://www.wavesec.org/*

Wellenreiter. Retrieved March 3, 2004, from *http://www.wellenreiter.net*

WEPCrack. Retrieved March 3, 2004, from *http://wepcrack.sf.net/*

Wi-Fi Alliance Portal. Retrieved December 12, 2003, from *http://www.wi-fi.org/*

Wi-Fi Protected Access: Strong, standards-based, interoperable security for today's Wi-Fi networks. (2003). Retrieved December 13, 2003, from *http://www.wi-fi.org/OpenSection/pdf/Whitepaper_Wi-Fi_Security4-29-03.pdf*

WiFiConsulting, Inc. Web Portal. Retrieved April 29, 2004, from *http://www.wificonsulting.com/Solutions/Coffee1.htm*

Wireless LAN resources for Linux. (2003). Retrieved December 13, 2003, from *http://www.hpl.hp.com/personal/Jean_Tourrilhes/Linux/Wireless.html*

Wireless LANs: A Primer for Secure Design. White Paper, Info-Tech Research Group. Retrieved April 29,2004, from *http://www.bitpipe.com/detail/RES/1067549763_201.html*

WirelessDriver Homepage. (2002). Retrieved December 13, 2003, from *http://wirelessdriver.sourceforge.net/*

WPA Specification Documentation. Version 2.0. Link on Wi-Fi Protected Access Portal. (2003). Retrieved December 12, 2003, from *http://www.wi-fi.org/OpenSection/protected_access.asp?*

Yi, S. & Kravets, R. (2003). MOCA : Mobile Certificate Authority for Wireless Ad Hoc Networks. In *2nd Annual PKI Research Workshop (PKI03)*, April 2003.

Yi, S., Naldurg, P. & Kravets, R. (2001). *Security-aware ad-hoc routing for wireless networks.* UIUCDCS-R-2001-2241 Technical Report, Department of Computer Science, University of Illinois at Urbana-Champaign, USA.

Zhou, L. & Haas, Z.J. (1999). Securing ad hoc networks. *IEEE Network Magazine, 13*(6).

Zivkovic, M., Lagerberg, K. & van Bemmel, J. (2004). *Secure seamless roaming over heterogeneous networks*. Retrieved March 14, 2004, from *http://www.ist-albatross.org/RoamingWhitePaper.pdf*

Chapter XI

Security, Privacy, and Trust in Mobile Systems and Applications

Marco Cremonini, University of Milan, Italy

Ernesto Damiani, University of Milan, Italy

Sabrina De Capitani di Vimercati, University of Milan, Italy

Pierangela Samarati, University of Milan, Italy

Angelo Corallo, University of Lecce, Italy

Gianluca Elia, University of Lecce, Italy

Abstract

Mobile systems and applications are raising some important information security and privacy issues. This chapter discusses the need for privacy and security in mobile systems and presents technological trends which highlight that this issue is of growing concern.

Introduction

Access to general purpose Information and Communication Technology (ICT) is not equally distributed on our planet: developed countries represent about 70 percent of all Internet users while its percentage of Internet hosts has raised from 90 percent in 2000 to about 99 percent in 2002. Things change dramatically if we look at mobile and wireless technology: developing countries already represented about 40 percent of mobile connections in 2000, with a foreseen growth rate that is faster for developing countries than that for the developed one in the period 2000-2005 (mainly due to India and the People's Republic of China). This trend depends on the new perspectives mobile electronic technology applications offer, making in principle possible to do business with partners located anywhere on the globe by-passing the poor telecommunication infrastructure still common in many developing countries. On the other hand, in the developed world the set of techniques going under the name of e-Mobile is becoming more and more important in e-Business transactions. The use of smart mobile terminals will allow new kind of services and new business models, overcoming time and space limitations. The technological evolution in wireless data communications is introducing a rich landscape of new services relying on three main technologies:

- Proximity (or personal) area networks (PANs), composed by personal and wearable devices capable of automatically setting up transient communication environments (also known as *ad-hoc* networks);
- Wireless local area network technology (WLAN);
- 3rd Generation of mobile telecommunications (3G), gradually replacing General Packet Radio Service (GPRS) and the related set of technologies collectively called "2.5 Generation" (2.5G). 3G services are made available through technologies such as Wideband Code-Division Multiple Access (WCDMA), offering high data speeds.

PANs is a new technology bringing the "always connected" principle to the personal space. On the other hand, 3G systems and WLANs have coexisted since long; what is new is their interconnection, aimed at decoupling terminals and applications from the access method. While 3G is generally considered applicable mainly to fully mobile wireless devices (such as operating from a car), WLAN is more relevant to fixed and portable wireless devices (such as operating from an elevator). 3G mobile networks already provide video-capable bandwidth, global roaming for voice and data, and access to the Internet rich online content. Thanks to their increasing integration, PANs, WLANs, and 3G

networks will extend the users connectivity in a complementary and hierarchical manner; in the fullness of time, they will provide all the functionalities of a Integrated Services Multimedia Network (ISMN), enabling a whole series of new business models and applications. The fusion of these technologies will eventually result in a ultimate ubiquitous wireless system that will be operational from anywhere on the planet, including use in homes, businesses, land vehicles and even commercial aircrafts. Even today, WLAN and 3G can already promote each other encouraging WAN users to continue connections in the wider area, provided that security, roaming and mobility are fully supported.

Mobile and Wireless Security Issues

While wireless communications provide great flexibility and mobility, they often come at the expense of security. Indeed, wireless communications rely on open and public transmission media that raise further vulnerabilities in addition to the security threats found in wired networks. A number of specific open issues and even inherent dangers (some of which had been already identified and described in the early stages of wireless technology adoption [Howard, 2000]) are yet to be solved. With wireless communications, important and vital information is often placed on a mobile device that is vulnerable to theft and loss. In addition, this information is transmitted over the unprotected airwaves. Third, 3G networks are getting smaller and more numerous, causing opportunities for hackers and other abusers to increase. Currently, 2.5G security mechanisms include 40-bit encryption, but theoretical attacks against this and the authentication mechanisms have been demonstrated (van Oorschot, Menezes, & Vanstone, 1996). 3G technologies incorporate stronger cryptographic techniques, and new authentication systems. This is probably not enough, because application areas like mobile commerce require this critical information to be decrypted by a server located somewhere in the communications chain before it is encrypted again and forwarded to a new destination.

Every hop in the wireless communication chain where information is decrypted and re-encrypted represents a potential vulnerability in the overall security. Furthermore, the growing complexity of mobile terminals and the increased presence of interoperability software on them is making them vulnerable to viruses and hacking attacks. However, there is great motivation for 3G security. The boom of users demand for richer content for their mobile terminals (such as through multimedia messaging, video conferencing, voice-over-IP, m-business) is increasing the need for security solutions ensuring user and data confidentiality, quality of service (QoS), billing, and protection against intruders. The challenge for industry players now is to tackle all security issues within PAN, 3G and WLAN and create a profitable integrated wireless business comprising of

services and value. In this chapter we shall look into some of the main security issues within the whole hierarchy of 3G and WLAN systems, including network access security, network domain security, user domain security, and personal identity management.

Wireless Applications and Security Testing Methodologies

As the complexity of mobile and wireless applications increases rapidly, importance of manufacturing security test becomes more critical. Important requirements of an effective security test methodology are functional completeness and compliance with appropriate security requirements, and minimum test execution time. Activities associated with testing include the following:

- Identification of the security requirements to be satisfied;
- Identification of product security mechanisms;
- Determination of test objectives;
- Determination of test methodologies and techniques;
- Determination of expected test results;
- Conduct of the test;
- Documentation and analysis of test results;
- Feedback of test results to appropriate individuals/organizations;
- Determination of the next action to be taken (such as additional testing or corrective actions).

The *Open Source Security Testing Methodology Manual* (OSSTMM) (Herzog, 2003) came about as a need for an open, free security testing methodology in response to the numerous security testing companies who claimed to have an internal and corporate methodology for testing. The OSSTMM has become the most widely used security testing methodology. In particular, the OSSTMM provides testing methodologies for the following six security areas: information security, process security, internet technology security, communications security, wireless security, and physical security. The methodology is used by IT consultancies, financial institutions, government offices, and legal firms worldwide because it offers low-level tests for many international laws on privacy and security.

We now focus our attention on the wireless security testing section. This section includes 10 modules (such as electromagnetic radiation testing, 802.11 wireless networks testing, bluetooth network testing, and so on) that in turn include one or more tasks. Each module has an input, which is the information used in performing each task, and outputs a dataset, which can then be classified in terms of *Risk Assessment Values* (RAV). RAVs serve to quantify the results of each module, which in turn tells security testers how long information remains useful and "current". Basically, a relative risk value is assigned to systems under test, and each user is willing to accept different levels of risk. This allows end users to determine how often they want regular testing to be carried out and how much risk they are willing to support. The output of a module may then be the input for one or more sections, or in certain cases, may be the input for a previous module.

Organization of the Chapter

The chapter is structured as follows. We first present an overview of the main privacy and security issues in mobile systems. We then describe the identity management issue in 3G mobile systems and discuss the integration of different wireless technologies into ubiquitous networking. Next, we illustrate the concept of mobile identity management and present some privacy and security issues in the hotspots context. Finally, we address the privacy and security issues that may arise by introducing recovery procedures for transactions initiated by mobile users and report our concluding remarks.

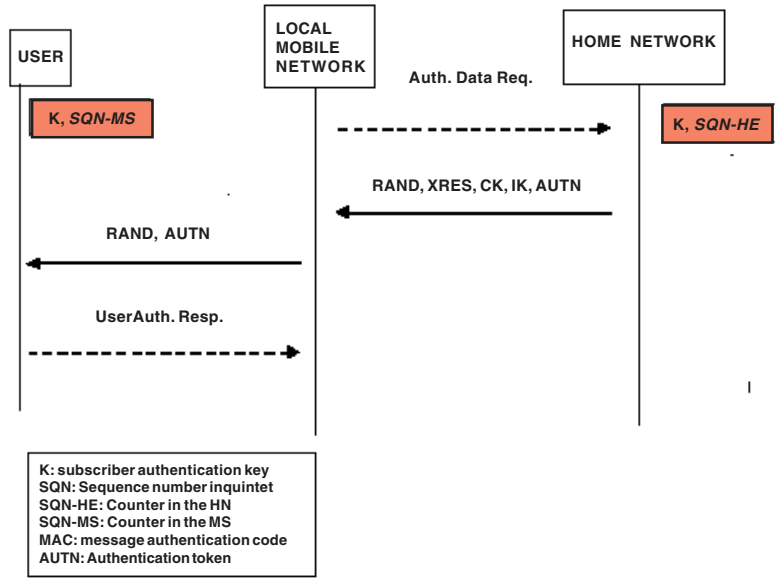
Mobile Systems Security: An Overview

Mobile systems security was conceived as a natural development of conventional POTS (*Plain Old Telephone Service*) security. Some of the objectives, therefore, were clear and well-understood: avoiding unauthorized disclosure of a user's or operator's data, repelling *denial-of-service* (DOS) attacks and preventing unauthorized access to and use of mobile service. However, as we anticipated in the previous section, a mobile communication environment presents a number of unique challenges due to the fact that mobile terminals are easily lost or stolen and to user expectations for flexibility and ease of use. In this section we shall focus on the main authentication and identity establishment techniques which are instrumental for the more complex mobile identity management solutions that will be described in the remainder of this chapter.

2G and 2.5G Mobile Authentication

First generation of analogical mobile phones relied on an electronic serial number to confirm that the terminal should be allowed access to the service (Blanchard).[1] On the other hand, GSM systems were designed with security in mind. Each subscriber to a GSM service receives a *Subscriber Identity Module* (SIM) card which contains the user's identity (see section 3) and a long-life authentication key (technically speaking, a shared secret key [van Oorschot et al., 1996]) supposed to last for the whole duration of the subscription. The SIM is a removable security module which is issued and managed by the users' home service operator (even when the user is roaming) and is independent of the terminal. SIM-based authentication does not require any user action, other than entering the familiar 4-digit *Personal Identification Number* (PIN) into the terminal. No more user awareness on security is needed than what they are already used to from their ATM cards. While certainly not unbreakable (e.g., it was subject to cloning attacks), this system was successful inasmuch it placed much of the security and authentication responsibilities with the final users holding the SIMs.[2] In GSM, after the initial access request message has been exchanged over the air back to the user's home operator, a temporary user identity is allocated which is local to the area operator where the user is located and is reassigned to another user as soon as the original requestor leaves the area. This reduces the exposure of the real user identity on the air and prevents information on a user's movements or use of a service being harvested by unauthorized eavesdroppers (e.g., for traffic flow analysis). Note that the GSM authentication mechanism is one-way only: the user sending the request cannot be completely sure that she has reached an authentic service operator. In the last few years, GSM 2G technology was upgraded to 2.5G with the introduction of the *General Packet Radio Service* (GPRS) overlaying, an IP core network on the GSM transport via two additional network elements, the Serving GPRS Support Node (SGSN) and *Gateway GPRS Support Node* (GGSN). While a complete description of GPRS technology is outside the scope of this chapter, it is worthwile to remark that enabling IP traffic via GPRS allowed 2.5G systems to take advantage of some well-known and understood authentication techniques (Smith, 2002) used on the Internet, such as certificates based on asymmetric encryption. Such authentication is performed in addition to (and independently of) GSM PIN-based authentication. Also, the GSM infrastructure already in place allows for large-scale roaming and recognition of security information. Recently, fully fledged *Public Key Infrastructure* (PKI) techniques have been enabled for mobile terminals by using enhanced SIM cards to handle the asymmetric key protocols.

Figure 1. 3G user authentication



3G Authentication and On-the-Air Confidentiality

In the design of 3G systems like UMTS, a new security architecture was specified. However, the approach that was taken was rather conservative. Indeed, the new approach maintained backward compatibility with GSM, while trying overcoming some perceived weaknesses of 2G systems. A main heritage of GSM still present in 3G systems is the *automatic integrated roaming*. 3G systems retain the basic idea of the GSM radio signaling system, that is, the concept that each user has a "home" cell and may be currently visiting another, operated by the home operator (telecom company) or by a local one. In order to find the location of its users (and bill them accordingly) the mobile network relies on distributed location registers, respectively called the *Home Location* and *Visited Location* Register (HLR/VLR). The HLR/VLR solution ensures that 3G calls can be set up with the same speed users experienced (and liked) in 2G networks. On the other hand, it preserves operator-based management of user authentication via shared authentication keys stored in SIMs. Like in 2G systems, 3G systems' users identify themselves by providing the identity stored in their

SIM and known to their home service operator, just like users accessing a computer system. 3G authentication was designed with the following requirements in mind.

- *Mutual authentication*: Both the user and the network are identified in the authentication exchange;
- *Key freshness*: Assurance that authentication information and keys are not being re-used;
- *Integrity of signaling*: Protection of service messages, for example, during the encryption algorithm negotiation;
- *Strong encryption*: Strong cryptography, obtained via a combination of key length and algorithm design, is performed inside the core network rather than at the periphery.

Figure 1 shows a 3G authentication and key agreement (AKA) mechanism involving the local and home network operators. The mechanism is based on symmetric key encryption and uses a subscriber authentication key K that is shared between the user and the home network operator. The mechanism then combines a challenge-response protocol with a sequence number-based protocol to support network authentication and to provide the user with assurance of key freshness. More precisely, the AKA mechanism works as follows:

1. Upon receiving an authentication data request, the home network operator generates a fresh sequence number (SQN) from its local counter (SQN-HE).
2. The home network operator generates a challenge RAND and prepares a quintet that includes: the challenge RAND, an expected response XRES, a cipher key CK, an integrity key IK, and an authentication token AUTN. The authentication token is obtained by combining the sequence number SQN and a message authentication code. XRES, CK, and IK are generated by applying three different key generating functions that take RAND and K as input and return, respectively, XRES, CK, and IK. The quintet is then sent to the local mobile network.
3. The local mobile network extracts from the received quintet the challenge RAND and the authentication token AUTN and send them to the user.
4. The user checks whether the authentication token AUTN can be accepted. Basically, the user verifies the integrity of the message and if the received sequence number SQN is acceptable. In case of a positive response, the user sends back to the local mobile network a response RES. The user also

computes a cipher key CK and an integrity key IK. Note that CK and IK are computed by using the challenge RAND received from the local mobile network and the shared key K. Therefore, CK and IK corresponds to the keys generate by the home network operator.

5. The local mobile network compares the received response RES with XRES and if they match, the protocol is successfully completed. The local mobile network then selects the corresponding CK and IK from the quintet.

The AKA mechanism provides mutual entity authentication and the establishment of a shared secret cipher key and integrity key between the involved parties. Indeed, after authentication took place, the established keys CK and IK are transferred to the entities that perform ciphering and integrity functions. On-the-air encryption performed at the radio interface can be used by the network operator to prevent *session hijacking*, maintaining the validity of the authentication throughout the call.[3]

Personal Identity Management in 3G Mobile Systems

In the previous section, privacy and security issues of mobile systems have been described mainly from the perspective of technological security research (access control, integrity, authentication, non repudiation, availability, and confidentiality). Recent developments in ICT-based business models reveal the necessity to approach the concept of privacy and security more broadly, embracing not only the technical aspects, but also the socioeconomic, the policy and business points of view. This approach could represent a useful attempt to create a common basis from which users' trust in mobile world can arise, opening new business opportunities, launching new services and goods (mobile payment and finance, mobile ticketing, and mobile voting), sparking new social and economic dynamics, and generating new life styles (Tsalgatidou, Veijalainen & Pitoura, 2000). The ongoing transition from monolithic and localized systems, mainly based on single technology and weakly opened to integration with heterogeneous systems, towards multi-application, multi-access, multiplayers, distributed and heterogeneous scenarios, is generating a context in which mobile applications and systems could play a strategic role. This event will occur if these kinds of scenarios will be wisely managed, taking into account both a set of internal elements and a group of context drivers that constitute important levers to enhance the users' trust in mobile systems and applications (Kagal, Finin &

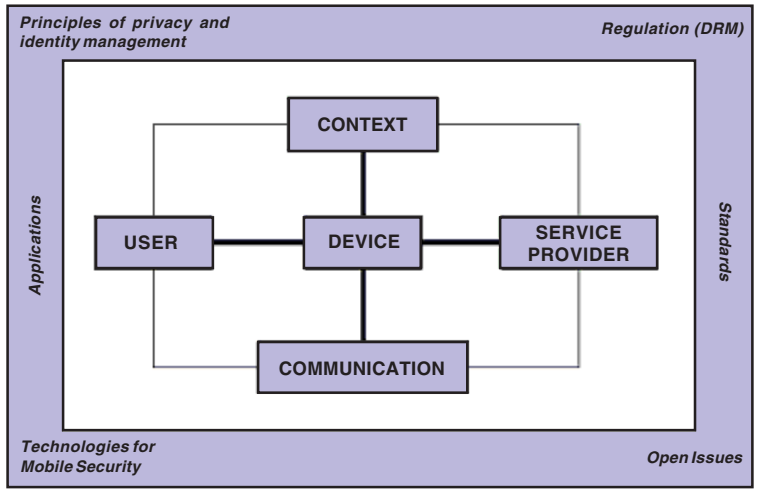
Joshi, 2001; Kagal, Parker, Chen, Joshi & Finin, 2003; and Matskin & Tveit, 2003). In other words, this means that technological potentialities, business opportunities, and joining industries complex dynamics have to be strongly internetworked with users' social dynamics, standards, policy, and regulation to create a sort of digital identity management framework where digital identity is conceived as "an electronic representation of individuals' or organizations' sensitive information" (Damiani, De Capitani di Vimercati, & Samarati, 2003). Support offered by this framework is crucial for building and maintaining trust relationships in today's globally interconnected society because it:

- Offers adequate security and availability;
- Strikes the right balance between protection of privacy and convenience;
- Allows to present different subset of the users' identity depending on the on-going and perceived application and communication context;
- Guarantees that identity, personal data, and user profile (including location based information) are safeguarded and no thefts will happen.

Starting from the late 1980s, many examples of *identity management system* (IM) have been proposed. In 1985, David Chaum considered a device that helps the user with payment transactions and upholds the user's privacy (Chaum, 1985a; Chaum, 1985b). In 1993, Roger Clark proposed the *digital individual*, that is, the individual's data shadow in the computer system which can be compared to user's identity (Clark, 1993). In 1995, John Borking published a report about the *Identity Protector* to protect the user's data (van Rossum, Gardeniers & Borking, 1995). In 1999, Martin Reichenbach proposed the *reachability manager* applied to telephone reachability (Damker, Pordesh, & Reichenbach, 1999). Starting from 2001, Jendrike, Kreutzer and Zugenmaier (2002) and tom Markotten, Jendricke and Müller (2001) proposed the concept of generic *identity management* to provide the users with usable and secure way to protect their privacy when using the Web.

Digital security and, more generally, digital identity management has been growing quickly in recent years, especially in mobile scenarios where personal communication and new computing devices will generate new security and integrity requirements for user and service information (Jendricke et al., 2002).[4] New trends include general ubiquity, new context-aware applications and services, new network and terminal technologies, flexible spectrum management and dynamic reconfiguration of terminals and networks in response to user mobility, user behavior, and capacity optimization. Most of these trends have surely an impact on the user's privacy (both in terms of access control and of published data), due to the additional user profile attributes that should be added

Figure 2. Our reference digital identity management architecture



in a mobile context (such as location, context, and terminal capability). Users are thus more and more aware of the impact of these developments on their personal privacy. Having a framework that gives a systemic view of the digital identity management represents a step to be explored to reinforce users' mobile trust in mobile systems, enhancing the penetrability level of mobile applications, and services in today society. As it is visible from Figure 2, a Digital Identity Management Framework is realized by taking into consideration both the real internal dynamics characterizing a use-case scenario, and the main external elements that may influence the architecture of an identity manager (such as regulations, standards, and so on). In particular, with respect to the internal dynamics we have identified the following five main elements.

User. The service requestor associated with a profile. According to application and communication context, a subset of personal data is extracted from the user profile to create the user's *personal identity*. The digital identity management framework should allow the user to keep her desired level of privacy depending on the situation, presenting multiple user "appearances" in different circumstances. In a mobile scenario, a portable user identity might include the following information:

- *Profile information* that consists of a number of static (date or place of birth) and dynamic attributes (technical skills and role);
- *Preferences in system usage* (browser settings) and other personal preferences that do not depend on the system (UK or US English spelling);
- *Behavioral information* that may be derived by an history of previous interactions with the system.

Service Provider. The supplier of network services and applications.

Context. The particular situation in which user interacts with the system. It includes the channel information (device and network features), the location information (cell, country, town) and time information.

Communication. It is based on well-known secure mechanisms to enable anonymity and confidentiality like *Secure Socket Layer* (SSL) (Freier, Karlton & Kocher, 1996).[5] Referring to anonymity, it is interesting to see that there are some possibilities for users to remain anonymous even in a world of SIM based authentication, since the authentication step is not repeated when roaming; rather the users hold a reusable, temporary identification provided by the local mobile network. At the network level, therefore, mobile users have no fixed device address and, in principle, are identified only by the location. Location-based addressing ensures that no information that can be traced back to a specific device is required for communication on the datalink and network level of the protocol stack.[6]

Device. The terminal that provides the physical layer services (such as a radio interface) used to communicate data and to interact with context and service providers. Moreover, the device becomes the physical place in which user profile, context and communication could be revealed and analyzed. For this reason, the terminal must be able to change the information it discloses much in the same way as the user.

Interactions among the five elements of the internal subsystem is aimed at enabling users to express and enforce their privacy and security needs, according to their specific requirements.

We are now ready to describe some of the external aspects that may influence the Digital Identity Management Framework.

Shared Principles. Mobile privacy and identity management is realized to implement the following main principles.

- *Confidentiality:* The guarantee that information is read only by the intended receiver. In turn, confidentiality can be split into three main

elements: integrity of message content, protection of location information (location-based information should be related to a specific user and device only with her consent), and support for sender/receiver anonymity. The latter element can be seen as relying on mobile terminals being capable of revealing SIM authentication data only in well-defined situations and to well-defined partners; in all other cases, users are capable to act under a pseudonym without revealing their true identity;

- *Integrity:* Transmission of information is executed by using cryptographical mechanisms (symmetric and asymmetric) to identify and detect possible manipulation of information;
- *Accountability:* Information exchange by using encryption techniques and digital signatures is necessary for security and trust;
- *Notice:* An alert service must be available to draw the user's attention to situations in which privacy and security could be affected. Notice mechanisms should be manual whenever automatic solutions could compromise user's security;
- *Data collection:* Users should be able to actively manage their own data, deciding whether and which identity is presented to device and applications (Ceravolo, 2003). Data collection must be inspired to the principle of data minimization, by which data should only be collected for a specific purpose.

Technologies for Mobile Security. As we have seen in the previous Section, technologies for 2G mobile security provide standard functions for checking the subscriber identity authenticity, for protecting the subscriber anonymity and for encrypting user and signaling data. 3G, while retaining SIM-based authentication, enhances security features organizing the issue in four domains: access, network, user and application, and adding auxiliary information on visibility and configurability. For packet data traveling over the mobile network layer, conventional security technologies apply. Two main areas can be identified:

- *Security Network Domain:* When Mobile IP is used at the network level over a mobile infrastructure, the most salient security issue is the problem of how to authenticate the registration messages that inform the server about a mobile node's current IP address, in order to avoid spoofing and IP impersonation attacks (Cheswick, Bellovin & Rubin, 2003).[7]
- *Security Transport Domain:* The well-known Secure Sockets Layer (SSL) and Transport Layer Security (TLS) protocols provide entity authentication, data confidentiality, and data authentication.

- *Trust Management:* In the previous section we saw how SIM-based authentication is the main technique for linking a terminal to a user identity. To secure this mechanism, however, specifically mobility-related threats must be addressed. As they get smaller, mobile terminals become more and more susceptible to theft. Stolen data is often regarded as being more valuable than the terminal itself. Thus, the need to protect user data and secrets is of paramount importance in a 3G mobile computing environment. Since 1999, the Trusted Computing Platform Alliance (TCPA) was created to foster industry participation in the development of an open specification for a trusted computing platform focused on two areas: ensuring privacy, and enhancing security. The TCPA provides for a platform root of trust, which uniquely identifies a particular platform, and provides various encryption capabilities, including hardware-protected storage.
- *Digital Rights Management (DRM):* DRM mobile networks rely on two crucial standards: the *Open Mobile Association* (OMA) DRM (OMA, 2003) and OMA Download (OMA, 2004). OMA DRM is the Digital Rights Management standard language for mobile phones published by the Open Mobile Alliance, while OMA Download is the application level protocol that enables reliable and secure downloading to mobile terminals of digital content whose access rights are specified using OMA DRM. OMA Download can be integrated to other channel specific services such as billing, and management of premium priced. However, OMA DRM and OMA Download are different technologies designed for independent purposes. Taken together, they enable secure downloading of digital content to mobile terminals and improve the consumer's experience of mobile content. Content protected by OMA DRM can be delivered using the OMA Download or other channel specific protocols such as the *Multimedia Message System* (MMS).

Wireless Heterogeneous Environments: Toward Ubiquitous Networking

We are now ready to discuss the integration of different wireless technologies, like 2.5 or 3G cellular networks and WiFi (IEEE 802.11b and 802.11g) (IEEE-SA Standards Board, 2000) into the more general landscape of *ubiquitous networking.* Ubiquitous networking is aimed at addressing the users' need of seamlessly roaming from one connection mode to the other without impairing their on-going operations. Accordingly, multi-mode cards (LAN-WLAN-GPRS cards) have been launched on the market and are becoming increasingly affordable. In particular, the advent of 3G is likely to make those multi-mode

cards rapidly evolve to the LAN-WLAN-3G setting thus transforming portable devices – cellular phones, laptops, and PDAs – in multi-mode devices equipped with cards that permit connections to multiple heterogeneous networks.

However, to foster effective mobility and ubiquitous computing through networks built on different wireless technologies, many fundamental issues need to be taken into consideration. An important one is the integration at link level between WiFi and GPRS/3G, which could result in a uniform network level. Realizing a uniform network layer between WiFi and GPRS/3G, in turn, may facilitate *transparent mobility*, that is, the possibility for users to automatically switch from one wireless network to another (possibly based on a different technology) without any detriment to on-going Internet transactions or application service provision. There are many high-value mobile application services that will greatly benefit from transparent mobility such as Tele-Medicine, Intelligent Transport Systems (ITS), and mobile Geographical Information Systems (mGIS). As an example, we could imagine a mobile user connected to a certain WLAN that is performing an Internet transaction or is interacting with complex application services. In the course of such a transaction, the user could be moving close to the physical limit of the WiFi Access Point range of transmission.[8] Before losing network connection, the user's device may, for example, switch to a GPRS/3G cell. Transparent mobility should permit to keep the Internet transaction alive (this could be a logical property, since physically the user connection with the first provider must be terminated and a new one established with the second). Finally, the same user that keeps moving, could enter into the range of a new WLAN and switch back to WiFi. To achieve features like the one described, the user should use several devices (laptops, PDAs, and cell phones) or a unique multimode device.

Transparent mobility is characterized by successfully migrating live TCP connections during the handoffs through different wireless technologies (WLAN→GPRS/3G handoff and GPRS/3G →WLAN handoff). To do this, it is not only sufficient a seamless internetwork handoff mechanism, but also the connectivity (as devices keep moving across environments while still minimizing any disruption to ongoing flows during switchovers) is another important aspect.

A mechanism that enables this has to exhibit a low handoff latency, incur little or no data loss (even in highly mobile environments), scale to large internetworks, adapt to different environments, and act as a conjuncture between heterogeneous environments and technologies without compromising on key issues related to security and reliability (Vidales, Patanapongpibul, Chakravorty, 2003). For all these reasons, transparent mobility is indeed one of the most challenging goals of ubiquitous computing in wireless heterogeneous environments.

Network technologies that are actively used for such systems are (Chakravorty, Vidales, Subramanian, Pratt & Crowcroft, 2003): *Mobile IPv4* (MIPv4) and

Mobile IPv6 (MIPv6). MIPv4 is the network technology traditionally used to foster seamless roaming for ubiquitous computing systems, mainly due to its compatibility with the wired IP-based network infrastructure. Nevertheless, MIPv4 limitations have forced the development of overly complex systems and protocols. MIPv6 promises to overcome some of MIPv4 limitations and improve security, although it has other disadvantages in high mobility scenarios (Chakravorty et al., 2003; Perkins & Johnson, 1996). Some current studies are then actively exploring the possibility to make use of approaches similar to those used in micro-mobility protocols that are aimed at improving the transparent roaming of mobile hosts at the subnet level of a network domain. Such protocols reduce the handoff latency and improve performance under high mobility scenario (Campbell & Gomez-Castellanos, 2001). An IETF working group, called Seamoby (Kempf, 2002), has been formed aiming to resolve complex interaction of parameters and protocols needed for seamless handoffs and context transfers between nodes in an IP access network.

Ubiquitous computing in wireless heterogeneous environments needs operational features and security requirements to be provided (Vidales et al., 2003). For instance, two interesting and fundamental open issues are the following.

- *Link-Switch Decision Rule-base*: Current schemes that regulate handoffs operate based on link layer information, such as signal strength. However, this could be insufficient to assist the handoff process in heterogeneous environments. Signal quality, overall link characteristics and robustness, link cost, as well as security considerations might be other parameters that need to be evaluated to decide the handoff. In particular, with respect to the security-related information, it has to be taken into consideration the *trust relationship* or the *reputation of the network provider* (WiFi or GPRS/3G), and the technical provisions put in place to guarantee a certain security level (message security mechanisms or *mobile identity management*);
- *Context-Awareness*: A mobile device context involves aspects such as physical context variables (device location, movement direction, velocity, and so on), application characteristics and, of course, user-based preferences. Context-awareness is necessary to take informed decisions about switching to a different network and provider. For instance, based on the exact position (available from a GPS system) and velocity available to a mobile host (speed sensors), a given proxy in the infrastructure can assist mobility by tracking and accurately predicting when a handoff should occur. This may let a user anticipate the link-switch decision before reaching the physical network connectivity limit. The user may then evaluate whether the next network provider is reliable and secure with respect to her on-going Internet transactions.

Multilateral Security

Multilateral security is an important factor to consider for wireless heterogeneous environments. Traditional security approaches assume that the whole set of actions that could be legally performed on corporate IT resources can be fully described in a security policy. Consequently, corporate security is achieved by enforcing the security policy throughout a secure and trusted entity. *Multilateral security*, instead, considers different and possibly conflicting security requirements of different parties that cannot be efficiently regulated through a static security policy (Rannenberg, 2000; Rannenberg et al., 1999). Some examples of conflicting security requirements of different parties in networks are the following.

- *Service requesters* cannot fully trust service providers (such as network operators) because they could perform unsolicited actions such as monitoring, profiling, and, in general, collecting data from the service offered to the users;
- *Service providers* might be victims of frauds or malicious service misuse caused by service requestors;
- *Network operators* could be harmed by breaches caused by network intrusions, sabotage, or other risks that could lead to network failures and downtime periods;
- *Users* of network connections might be harmed by other users.

Multilateral security copes with these competing security requirements of different parties and aims to strike a balance among them. *Open communication systems* (networked services based on the telephone or on the Internet) often exhibit these characteristics and a high degree of untrustworthiness. A possible approach for multilateral security consists in taking into consideration the security requirements of all parties involved and, at the same time, considering the parties as potential attackers.

Two technical areas are considered especially important in multilateral security (Rannenberg, 2000; Rannenberg et al., 1999):

- *Negotiation:* Negotiating security requirements is a natural way to set common security practices and foster cooperation among communicating parties. A related approach, although developed for a different technology, has been proposed by the Platform for Privacy Preferences (P3P) project

of the Word Wide Web Consortium (WWW Consortium, 1998). The P3P project, as well as multilateral security, aims to set a standard for reaching agreements with service providers about the collection and use of personal information;

- *Secure Architectures:* Security measures located on devices are not sufficient alone to satisfy all security requirements. For instance, transaction recovery after a handoff between WLAN and 3G/GPRS has associated security risks that cannot be prevented with device-based countermeasures only. Hence, architectural security measures must be put in place protecting communication functions and network operations.

A number of technical design principles that support the development of multilateral security architectures and solutions have been proposed and they can be summarized as follows.

- *Data Economy:* Data economy states that the only way to keep confidential data on which users have no control, is to avoid those data. For instance, in communication protocols only data under the control of users should be transmitted. This principle is particularly important for identification data. The strategy of data economy aims at minimizing data transmitted and at transmitting only those personal data that were explicitly authorized by the owner. This prevents security risks and reduces the cost and complexity of data protection;
- *Careful allocation:* Data that are needed to conduct Internet transactions or to obtain networked services must be carefully allocated. This means that systems must permit a strict control over both the ownership and the location of such data;
- *User ability to control:* Negotiating security requirements between communicating parties often results in trade-offs that strike a balance among conflicting requirements. Users should keep the control of the outcome of the trade-off and actively monitor how the security context evolves. This might be achieved by means of monitoring consoles, tools providing status information and the access to configuration/administration interfaces;
- *Usability of security mechanisms:* Lack of usability in security mechanisms is a well-known problem that has impaired many security solutions. Multilateral security prescribes the adoption of usable mechanisms only. This is a challenging principle since usability is a dynamic notion that may vary for different users at different stages of interest, understanding, and competence;

- *Opportunities for individual negotiation*: Negotiation can only work if there are real options and opportunities to negotiate on. This may need economic and regulatory frameworks to balance the usually different power of communicating parties and network operators.

Mobile Identity Management

Wireless heterogeneous environments present many challenges in the area of *digital identity management.* As previously mentioned, digital identities are the electronic representations of individuals' or organizations' sensitive information (Damiani et al., 2003). In everyday life experiences, the personal identity is not a unique, monolithic concept. Instead, the identity is a complex concept made up of many different attributes and each one manages her own identity according to the circumstances that ask for personal identification data. Normally, only that personal information that is needed to access a certain service are disclosed. *Digital identity management* is then defined as the ability to selectively disclose only those personal information related to the service, while preserving and enforcing privacy and security needs, such as protection from possible theft of identities for later illegal usage, requirements of anonymity, and use of pseudonyms. These security and privacy-related issues have contributed to the development of the notion of *mobile identity management* (Jendricke et al., 2002; Roussos and Patel, 2002).

In general, as described in (Jendricke et al., 2002), the strategic impact of mobile identity management can be evaluated along three major areas: increasing *operational efficiencies* without compromising security, increasing degree of *personalization of services* as well as active consumer management, and finally increasing rate of development of novel services thus increasing revenue streams. Therefore, the ultimate goal of mobile identity management is to increase trust between businesses, consumers, and trading peers, so as to enable a wider adoption and access to network services. In the context of mobile users roaming through heterogeneous environments and accessing critical corporate resources and possibly exchanging sensitive data, mobile identity management benefits represent certainly a strong incentive to its development.

Mobile identity management systems support a collection of different *interaction modes.* The simplest mode of interaction is *peer-to-peer*. In this case, identification and credential exchange is performed without the mediation of a third party (a Certification Authority) in a distributed and decentralized manner. Significantly more complicated is the support by mobile identity management systems of *nomadic* and *ad-hoc conferencing*. To do so, it is necessary to

provide mechanisms for group member identification, membership control and access to common resources, either within the context of a single organizational unit. Finally, the last operational mode we present is a fully deployed mobile identity management system in *intraorganization* or even *global scale*, which requires a global mobile identity infrastructure that should be open, fully interoperable, and distributed. However, current mobile identity management systems have to cope with the extra requirements of a heterogeneous context. Areas that are likely to require improvements are interoperability, roaming and self-configuration, as well as privacy protection and security. For instance, many mobile identity management systems are based on Public Key Infrastructure (PKI) technologies that have proved not to scale well and have shown many interoperability problems in practical contexts. These limitations may impair the development of mobile computing in wireless heterogeneous context, which in turn are heavily based on strong interoperation requirements for authentication. Modern trust management systems exhibit better characteristics that may well support mobility. Moreover, whether interactions between nomadic users and wireless network providers are carried out in a peer-to-peer style, centralized solutions (e.g. PKIs) could not be adopted: in the case of ad-hoc interactions the peers cannot resolve certificate chains without incurring in high latencies due to the indirect access to verification resources.

It has been widely recognized the relevance of requirements exposed in multilateral security for the foundation of mobile identity management. Clauss and Kohntopp have explicitly developed the SONET system for identity management on the basis of multilateral security principles (Clauss & Kohntopp, 2001). Jendricke et al. (2002) have derived the following relevant privacy principles for mobile identity management directly from multilateral security requirements.

- *Confidentiality and Integrity*: A mobile identity management system must support cryptographic techniques and key exchange protocols to achieve confidentiality and message integrity;
- *Anonymity and Pseudonymity*: There are situations where Internet transactions should not be linked to individuals. Users should have the possibility to conceal their own true identities by using pseudonyms or even by accessing services anonymously. High mobile users equipped with multimodal wireless cards should have the ability to selectively disclose personal information or use pseudonyms;
- *Availability*. Wireless connections are by nature more prone to network failures than wired systems: Handoffs between WiFi and 3G/GPRS providers might introduce new failure possibilities. On the contrary, although handoffs are critical operations, multi-mode network access might be used for alleviate disconnection problems if a backup recovery mechanism exists. For instance, a WLAN connection may experience connectiv-

ity problems and an on-going Internet transaction could unexpectedly terminate. In this case, if the 3G/GPRS mode was used as a backup to save some safe state of the user session, switching to the 3G/GPRS link, the transaction could be recovered from that safe state. Transaction recovery is indeed an extremely important area for multihop mobility in heterogeneous environments that should be integrated with mobile identity management. Security risks may arise if recovery features could be misused by attackers that impersonate other digital identities and reclaim the recovery of transactions belonging to different users;

- *Accountability*: Wireless service providers, such as hotspot providers and 3G/GPRS telco, would probably offer their network connection facilities at a market price. Billing systems, linked to digital identities or pseudonyms, are likely to be the target of subversion attacks and need to be carefully protected;
- *Security-awareness:* A basic security principle that even mobile identity management systems have to satisfy is that users must always be fully informed of security-related actions performed on their behalf by devices. In this way, an informed user must ultimately do evaluations about the security risk associated with the switch to a certain wireless network provider and decide how to negotiate security requirements. Those decisions cannot be taken automatically and transparently by the device on its own and based, for example, on PKI certificate chains or reputation mechanisms. Techniques and tools are of extreme value when effectively assist users to take an informed decision, but they may turn out to be harmful when they substitute users or even make them incapable of enforcing their own decision;
- *Data Collection*: A basic principle of identity management that should always been enforced is that users must be able to decide which personal data disclose to whom. It is not sufficient that users were fully informed of which data have been disclosed and collected. Minimizing disclosed personal data actually represents the main task of an identity management system.

Multihop Hotspots

Hotspot providers in public area represent key components of an heterogeneous wireless infrastructure for mobile users, which could be used to access WLAN services while moving. *Multihop hotspots*, in particular, are hotspots through which users could roam seamlessly. Considering heterogeneous environments,

users could hop through hotspots that are either physically contiguous, thus directly switching from one hotspot to another, or through a sequence of multimode handoffs between hotspots and GPRS/3G cells (Balachandran, Voelker, & Bahl, 2003).

With respect to security, hotspots have still significant open issues. One is *authentication* that is currently implemented with different and incompatible techniques by commercial WiFi networks. For instance, since hotspots are often under the control of different providers, users will have to repeat the authentication procedure (possibly different for each hotspot) at each hotspot location. Also, some commercial hotspot providers offer access to users through preestablished accounts, while others offer scratch-off cards containing a one-time login and password.

A uniform and shared authentication infrastructure is fundamental for effective multihop mobility since highly mobile users cannot be required to cope with different authentication schemes, mechanisms, and configurations at each handoff. Clearly, the goal of providing fast and seamless authentication, while simultaneously ensuring user accountability, raises several research problems that are today still unsolved. Examples are:

- *Ease of Access*: Single-Sign-On (SSO) features encompassing multihop hotspots are needed to support transparent mobility and reduce the latency;
- *Mechanism.* What authentication mechanisms are best suited in such an environment? Is it adequate for the network to authenticate the users through software mechanisms such as identity certificates or digital tokens or are hardware mechanisms, such as SIM cards, needed?
- *Identity Management:* The mobile identity a user presents to each network provider could change according to context-related information such as provider reputation, QoS, location or other contextual attributes. Which mechanism can permit an effective context-awareness and negotiation of identity attributes and at the same time minimize the latency of the handoff?
- *Third-Party Authenticators*: Should authentication be delegated to dedicated third parties offering such a service for the whole multihop infrastructure?

Another challenge to multihop mobility is *wireless hop security*. Traditional security mechanisms – like SSH, SSL or VPN – provide end-to-end data privacy to communicating parties. These mechanisms are not always well suited for multihop mobility because intermediaries – like network providers, gateway or

proxies – might need to access and inspect message-specific information. This could be asked for security reasons as well as for routing, accountability, or recovery.

To this end, novel approaches, for example realized in the area of Web services, have developed per-message security solutions that permit to selectively disclose information carried by messages to specific intermediaries. In this way, information carried by a network message could be targeted to different destinations. For instance, some control information could be disclosed to intermediaries for authentication, access control, or routing. Other information could be delivered to the final endpoint of the communication and then kept private during the whole multihop session.

Security and Privacy Issues in Mobile User Recovery

Recently, an important contribution to mobile computing has been published by VanderMeer, Dutta, Ramamritham & Navathe (2003). In their work, the authors address the problem of recovery Internet transactions initiated by mobile users. The issue is new and relevant since it presents many differences with respect to classical database transaction recovery. Also, it appears extremely important in the context we have considered, since there is not only the case of recovery after a network failure, but there is the peculiar situation of recovery user activity after a handoff. This aspect is an additional novel issue to the most general problem of mobile user recovery.

This issue also has significant links with security and privacy since mobile user transactions and mechanisms for recovery could become critical points of security risk and been targets of network attacks and subversion attempts. If network attacks would eventually succeed, it will be possible for an intruder, for example, to subvert the recovery mechanism and then recover transactions of other users possibly gaining their privileges. By attacking a recovery mechanism it could be possible to access transactions' state information that still could let intruders impersonate users or gathering sensitive information. Denial-of-service attacks towards the recovery systems is another threat that might severely impairing the benefits of the infrastructure for ubiquitous computing and transparent mobility.

Security researches in this area are still at the beginning since even operational features, like mobile user recovery, are in their initial stages. Despite this, the issue looks extremely important for future evolution of ubiquitous computing and

transparent mobility. In the following the characteristics of mobile users recovery are presented, according to (VanderMeer et al., 2003).

Firstly, consider a simple case study to describe the peculiarities of mobile user recovery. A user is buying an airline ticket over his wireless Internet connection. She may execute typical Web operations like: logging on to the airline site with her frequent flier number, entering travel dates and destination, selecting the preferred seat, and finally entering credit card information and receiving a confirmation of the purchase. Typically, this interaction spans multiple sites, that is, at least the travel agency and the company in charge of processing payment information. In wireless heterogeneous environments, mobile users may have switched several times on different links and modes through the life of the described Internet transaction. The flow of operations executed by the user during her ticket purchase should proceed seamlessly after each handoff.

Considering the underlying mechanisms that support transparent mobility, it has many elements in common with Internet session recovery since in both cases a transaction state has to be stored somewhere and then recovered when the new connection is established. As a consequence, requirements in terms of efficiency, performance and transparency have an increased importance.

In the scenario described above, the goal is to be able to avoid the repetition of work (computation, communication, I/O) required after a connection disruption, thus minimizing the cost for recovery. Solutions for recovery of such interactions are quite different from classical system transaction recovery. In recovering database transactions, the focus is on ensuring that the status of the underlying database system is consistent: if a transaction prematurely aborts, the transaction is rolled back and is resubmitted after the database system recovers.

In mobile computing, in addition to the classical recovery problem, we need to minimize (or to completely avoid), the user task of resubmitting the transaction again. The recovery infrastructure should permit users to efficiently and quickly restart from an appropriate point prior to disconnection. For this reason VanderMeer et al. (2003) have proposed the expression *user recovery* in place of the traditional *system recovery*.

Therefore, the first goal is to define the *user state* during her flow of operations. The proposed approach observes the sequence of operations as they occur, logs state information corresponding to each operation, stores the state information, and utilizes it to recover the user to a useful point in her interaction. User state information may span through multiple sessions and multiple service providers. For this reason, along with user recovery, the notion of *user session* was introduced to encompass all active sessions included in the on-going user transaction. A user state, after the execution of a given action, has been define as a 4-tuple composed of: the set of *cookies* valid after the given action; the

HTTP request corresponding to the given action; the site's *response* to the user HTTP request; and a *function* denoting the validity of the user state.

Based on the notion of user state, a *recovery protocol for Internet transaction* should have the ability to: *store user states* to be used in the case of connection failure; and *return a recent and valid state* to the user upon reconnection after failure.

Intuitively, the recovery protocol that has been proposed works in the following way. It logs user states for each action in *action logs* and maintains *a failover map* of various sub-transactions. This information is needed to recover after and handoff too.

According to this proposal, several important issues that may affect security arise.

- *The secure storage and access to action logs, failover maps, and recovery logic*: There could be different choices, from storing them locally to the device, to storing them at network gateways or specialized recovery hosts. Indeed, a support to recovery features from the network infrastructure is needed. This introduces security issues related to trust relationships with third parties or networked components in charge of recovery user sessions. Also, distributed authentication mechanisms should be in place because users that switch from some links (from WiFi providers), might reconnect later to different network providers with different modes (to 3G/GPRS telcos);
- *Trustworthy generation of action logs and failover maps*: The generation of state information must be secured and trustworthy. State information must be protected from tampering and disclosure since network components in charge of generating and store state information are probable points of attacks. Moreover, the usage of cookies to simulate HTTP sessions is traditional target of intrusions for gathering information that let attackers impersonate Internet users. Secure management of session state information is important to secure handoffs and transactions management;
- *Trustworthy management and usage of users action logs and failover maps:* User authentication is another important aspect for the security of user session's recovery. For instance, imagine a user that disconnect from a certain link and only after a certain time frame reconnects to a different wireless network. What if while she is disconnected someone else reclaim the recovery of that user session? How the recovery system recognizes that this is a malicious attempt? Recall that for ubiquitous computing we need to strike a balance between ease-of-use and security to foster

transparent mobility and mobile device always have to deal with the problem of power consumptions, therefore current strong authentication mechanisms might be unfeasible in this context.

Conclusion

The amount of mobile computing is expected to increase dramatically in the near future. As the user's demands increase with the offered services of mobile communication systems, the main expectation on such systems will be that they provide access to any service, anywhere, at anytime. Indeed, in today's highly connected, and highly mobile environments, the secure transmission of information is imperative for every enterprise, and will grow in significance as mobile devices, networks, and applications continue to advance. However, the promise of mobile computing technologies further increases privacy and security concerns. In this chapter we have discussed the need for privacy and security in mobile systems and have presented technological trends that highlight that this issue is of growing concern.

References

Balachandran, A., Voelker, G. M., & Bahl, P. (2003). Wireless hotspot: Current challenges and future directions. In *Proceeding of the ACM WMASH'03*.

Blanchard, C. Security for the third generation (3G) mobile system. Retrieved on December 16, 2004 from *http://www.isrc.rhul.ac.uk/useca/Other Publications/3G UMTS%20Security.pdf*

Campbell, A. T. & Gomez- Castellanos, J. (2001). IP micro-mobility protocols. *ACM Mobile Computing and Communications Review (MC2R)*, ACM SIGMOBILE.

Ceravolo, P. (2003). Managing identities via interactions between ontologies. In *Proc. of the Workshop on Metadata for Security*, Catania, Italy.

Chakravorty, R., Vidales, P., Subramanian, K., Pratt, I., & Crowcroft, J. (2003). Practical experiences with wireless integration using MobileIPv6. In *Proc. of the ACM MOBICOM 2003.*

Chaum, D. (1985a). Security without identification: Transaction systems to make big brother obsolete. *Communications of ACM, 28*(10), 1030-1044.

Chaum, D. (1985b). Showing credentials without identification. Signatures transferred between unconditionally unlinkable pseudonyms. In *Proc. of a Workshop on the Theory and Application of Cryptographic Techniques,* Linz, Austria.

Cheswick, W., Bellovin, S., & Rubin, A. (2003). *Firewalls and Internet security: Repelling the wily hacker.* Addison Wesley.

Clark, R. (1993). Computer matching and digital identity. In *Proc. of the Conference on Computers, Freedom & Privacy*, San Francisco, CA.

Clauss, S. & Kohntopp, M. (2001). Identity management and its support of multilateral security. Computer Networks, *Elsevier Science, 37*, 205-219.

Damiani, E., De Capitani di Vimercati, S., & Samarati, P. (2003). Managing multiple and dependable identities. *IEEE Internet Computing*, *7*(6), 29-37.

Damker, H., Pordesh, U., & Reichenbach, M. (1999). *Personal reachability and security management - Negotiation of multilateral security.* Chapter Technical Building Blocks, 95-112. Addison Wesley Longman.

Freier, A., Karlton, P., & Kocher, P. (1996). The SSL Protocol - Version 3.0. Retrieved on December 16, 2004 from *http://wp.netscape.com/eng/ssl3/ssl-toc.html*

Herzog, P. (2003). OSSTMM 2.1 - Open-source security testing methodology manual. ISECOM. Retrieved on December 16, 2004 from *http://www.isecom.org/osstmm/*

Howard, P. (2000). 3G security overview. In Proc. of the IIR Fraud and Security Conference. Retrieved on December 16, 2004 from *http://www.isrc.rhul.ac.uk/useca/OtherPublications/IIR-overview.pdf*

IEEE-SA Standards Board (2000). Part11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. IEEE.

Jendricke, U., Kreutzer, M., & Zugenmaier, A. (2002). Mobile identity management. In *Proceedings of the Workshop on Security in Ubiquitous Computing (UBICOMP2002).*

Kagal, L., Finin, T., & Joshi, A. (2001). Trust-based security in pervasive computing environments. *IEEE Communications, 34*(12), 154-157..

Kagal, L., Parker, J., Chen, H., Joshi, A., & Finin, T. (2003). *Security, privacy and trust in mobile computing environments.* CRC Press.

Kempf, J. (2002). Problem description: Reason for doing context transfers between nodes in an IP access network. *IETF Request for Comments*, RFC 3374.

Matskin, M. & Tveit, A. (2003). Software agents for mobile commerce services support. In Siau, K. (Ed.), *Advanced Topics in Database Research*, Volume 2. Chapter 11, 246-266. Idea Group Inc.

OMA (2003). OMA DRM requirements - version 2.0. Retrieved on December 16, 2004 from *http://member.openmobilealliance.org/ftp/Public_documents/BAC/DLDRM/2003/OMA-DRM-REQ-v2_0-20030515-C.PDF*

OMA (2004). Generic content download over the air - Approved version 1.0. Open Mobile Alliance. Retrieved on December 16, 2004 from *http://www.openmobilealliance.org/release_program/docs/CopyrightClick.asp?pck=Download&file=v1.0-20040625/OMA-Download-OTA-V1_0-20040625-A.pdf*

Perkins, C. E. & Johnson, D. B. (1996). Mobility support in IPv6. In *Proc. of the ACM MOBICOM.*

Rannenberg, K. (2000). Multilateral security - Concept and examples for balanced security. In *Proceedings of the 9th ACM New Security Paradigms Workshop.*

Rannenberg, K., Pfitzmann, A., & Muller, G. (1999). *Multilateral security in communications - Technology, infrastructure, economy*. Addison-Wesley Longman.

Roussos, G. & Patel, U. (2002). Mobile identity management. In *Proc. of the Mobile Business 2002.*

Smith, R. (2002). *Authentication: From passwords to public keys*. Addison Wesley.

tom Markotten, D., Jendricke, U., & Müller, G. (2001). Benutzbare Sicherheit { Der Identitätsmanager als universelles Sicherheitswerkzeug, chapter 7, 135-146. Springer-Verlag Berlin.

Trusted Computing Platform Alliance. Trusted computing platform alliance. Retrieved on December 16, 2004 from *http://www.trustedcomputing.org/home*

Tsalgatidou, A., Veijalainen, J., & Pitoura, E. (2000). Challenges in mobile electronic commerce. In *Proc. of the 3rd International Conference on Innovation through E-Commerce*, Manchester, UK.

van Oorschot, P., Menezes, A., & Vanstone, S. (1996). *Handbook of applied cryptography*. CRC Press.

van Rossum, H., Gardeniers, H., & Borking, J. (1995). Privacy-enhancing technologies: The path to anonymity, No. 5, Registratiekamer.

VanderMeer, D., A, D., Dutta, K., Ramamritham, K., & Navathe, S. B. (2003). Mobile user recovery in the context of Internet transactions. *IEEE Transactions on Mobile Computing, 2*(2), 132-146.

Vidales, P., Patanapongpibul, L., & Chakravorty, R. (2003). Ubiquitous networking in heterogeneous environments. In *Proceedings of the 8th IEEE Mobile Multimedia Communications (MoMuC 2003).*

W3C Consortium (1998). P3P Guiding Principles. Retrieved on December 16, 2004 from *http://www.w3.org/TR/1998/NOTE-P3P10-principles*

Endnotes

1 Such a naive system was doomed. However, before long, hackers learned to read these electronic serial numbers from the air and access unsuspecting users' accounts.

2 Long life is guaranteed by the fact that the authentication key is used to enable the user terminal and is not required by the GSM network when the user is placing or receiving calls.

3 On-the-air encryption is not mandatory in 3G networks due to concern about restrictions on the use of encryption in some countries.

4 We will talk more about mobile digital identity in Section 5.

5 These mechanisms work at the packet level and sit on top of the on-the-air encryption mechanism offered by some 3G networks.

6 Also, service discovery relies on a broadcast message on the part of the service provider. Terminals do not have to become active, and can avoid revealing their presence just for discovering services the user may not be interested in.

7 Mobile IPv4 and Mobile IPv6 solve this issue by using a protocol specific authentication extension based on a secret key shared between mobile node and home agent, and by reusing IPSec protocol to secure the binding updates, through Internet Key Exchange (IKE) protocol, respectively.

8 The available physical range might be limit by other parameters other then the pure transmission range, such as the Quality of Service (QoS).

Section VI

Turning the Threat into an Opportunity

Chapter XII

Visions for the Completion of the European Successful Migration to 3G Systems and Services: Current and Future Options for Technology Evolution, Business Opportunities, Market Development, and Regulatory Changes

Ioannis P. Chochliouros, Hellenic Tellecommunications Organization S.A. (OTE), Greece

Anastasia S. Spiliopoulou-Chochliourou, Hellenic Telecommunications Organization S.A. (OTE), Greece

Abstract

Mobile communications are fundamental to business operations, individual lifestyles and the welfare of the European economy. The proposed work aims to investigate some among the current technical, business, financial, and regulatory visions associated with the effective evolution of third

generation (3G) networks and services, in particular to fulfil the great market realities, the expectations and 3G's significant potential in building the EU Information Society. The work depicts data related to the current state of play for 3G communications in Europe, with specific emphasis given to the underlying technologies and probable standardization options (both for network and terminal equipment). Wireless mobile technologies are a major driver to realize the way forward to a knowledge-based economy, in terms of consumer demand, productivity, competitiveness and job creation. Under suitable terms, this may create enormous potential and significant investment incentives, for the full recovery of the wider ICT sector. 3G is likely to play a key role in providing widespread access to the Internet and to interactive services, thus maximizing consumer choices and providing flexibility for the market itself. Furthermore, we evaluate the related EU regulatory perspectives, to support innovation, competition, legal certainty and proportionality, the consolidation of the Single Market and the removal of technical barriers to international trade.

Introduction: The 3G Market as a Key-Component of the Information Society Sector

The development of the telecommunications market is driven forward at a great speed, to some measure by the competition and the liberalization of the telecommunications sector worldwide, and to some extent by the evolution and advances in technology (especially in major areas such as information processing and multimedia communications over the Internet).

Following liberalization in 1998, competition in European telecommunications markets has driven growth and innovation and the widespread availability of rich, advanced and diversified services to the public, to bring everyone (governments, authorities, administrations, businesses, citizens, homes and schools) into the digital age. New systems and services are under development with inputs, contributions and traditions from multiple industries--including telecommunications, broadcasting, Internet service provision, computer and software companies, media, and publishing industries.

Between 1999 and 2001 the value of telecom services in the European Union (EU) (European Commission, COM(2002) 695) rose 24 percent from € 182 billion in 1999 to € 225 billion in 2001. For the current year, the electronic communications market is expected to grow at an estimated rate of between 3.7

percent and 4.7 percent in nominal terms. This compares with a forecast rate of EU GDP (gross domestic product) growth of 0.8 percent (3 percent in nominal terms), as for the year 2003. The combined national markets of the 15 Member States were expected to be worth an estimated € 251 billion in 2003 (European Commission, COM[2003] 715).

Mobile telecommunications is experiencing rapid growth in Europe. Efficient mobile telecommunications network infrastructures, which deliver the coverage and quality demanded by customers, are essential for continued economic development. The mobile sector alone grew by 32 percent in 2000 and 21 percent in 2001 in terms of revenue, while the average mobile penetration rate in Europe is currently more than 75 percent (from 70 percent as measured in 2001). The mobile penetration rate has in some Member States almost reached saturation level (81 percent of EU citizens now have a mobile telephone, while in some Member States the penetration rate is close to 90 percent). Indeed, while the number of subscribers to mobile services continues to increase (there are currently more than 305 million mobile users and a total of some 125 million handsets sold in 2001), the rate of growth for the year 2002 was 6 percent, compared to 69 percent in 2000 and 36 percent in 2001. Such a development takes a wider dimension while supporting the option for a global Information Society without frontiers, where information is stored and communicated electronically.

The global (and the European) mobile telecommunications industry is evolving from being primarily voice telephony service providers (with extra features like short message service – SMS) to delivering mobile data and multimedia services. Improved network technologies and software in the third generation of mobile communications will improve and extend the range of services available, particularly by increasing the speed at which services will run over the underlying networks. This will enhance the usability and interactivity of the relevant applications.

However, the market is somewhat fragile following the global economic slow-down and over-investment (*especially in the backbone capacity*), combined with high levels of debt resulting from expensive acquisition strategies and the cost of the transition to third generation (Universal Mobile Telecommunications Systems-UMTS or "3G") mobile systems. The extended electronic communications sector is undergoing a multi-natured adjustment process, a fact having significant implications for its future as well as for broader economic growth in Europe.

As regards mobile licenses, the conditions in 3G network licenses relating to rollout and coverage requirements have been the subject of intense debate, in view of the difficulties generally experienced in the sector and the anticipated delays in the commercial launch of relevant services. In particular, the high level

of payments associated to 3G licenses (distributed through auctions or comparative hearings) contributed to worsen the financial situation of most of the operators involved. In addition, the availability of investment funds has been significantly reduced in many cases, especially at a time when deployment (physical and logical) of infrastructure required significant resources. This has affected operators to rebalance their finances.

The experiences gained from such processes, implicate the necessity for the full establishment of suitable and commonly coordinated policies across the EU Member States. To this aim the European Commission has recently realized series of measures and associated policies (European Commission, COM[2002] 263) to avoid probable fragmentation risks and/or diverging conditions, especially from the expected development of innovative electronic communications services and infrastructure.

Despite the difficult financial situation in the market, there are positive indicators of continued demand for services and of relevant competitive activity (European Commission, COM(2002) 655; European Commission, COM[2002] 62). In any case, market uncertainties still do not affect, drastically, the fundamental role of Information and Communication Technologies (ICT) for the deployment of the economy. The ICT sector and the Information Society still remain a valuable source of productivity gains, of competitiveness and of improvements in living standards. It also offers the potential for organizations and enterprises to make best use of their expected investment and to generate significant revenue streams.

In addition, the 3G industry is actively pursuing standardization activities to achieve adequate end-to-end interoperability, interconnection, and interworking options between services, infrastructures (including all methods of access) and terminals as well. Within the same scope, operators in the 3G-value chain recognize the value of interoperability in their service offerings with those of other service providers. Delivery of some Information Society services, *and especially the multimedia services*, will require a broadband delivery system to the end-user/consumer. 3G mobile networks offer such a platform with extensive capabilities; in addition, under suitable terms, this could be able to interoperate with other modern and enhanced platforms like, for example, digital television networks.

To achieve to the aim of the full establishment of a knowledge-based economy (European Commission, COM[2003] 65), the EU has promoted specific measures to support cooperation between the public and the private sector. These include, among others: (a) the adequate sustainability and the strengthening of various research efforts at national and EU level to ensure Europe's competitiveness (in parallel with efforts to improve aspects such as performance, capacity, quality of service, usability and cost of both infrastructures and

services); (b) well-designed standardization efforts at the EU level, in cooperation with the European Standardization Organizations (ESOs), in order to prepare modern standards able to fulfil market needs and to respond to innovative technology trends; (c) the full and timely implementation of a new regulatory framework for the electronic communications sector to underpin innovation and investment perspectives in an "open" and competitive market environment; and (d) the effective application of suitable measures (such as the eEurope-2005 Action Plan) to reorganise the e-business sector, to create incentives and to increase productivity and inclusion (especially for long term investments) via the applicability of enabling technologies.

Within such an environment of reference, 3G services can realize a fundamental role in the implementation and the progress of a competitive and dynamic information society, especially as recommended by the Action Plan *e*Europe 2005 (European Commission, COM[2002], 263).

Work is well under way to bring third generation ("3G") mobile technology, applications and services to the wider market and the first commercial launch in Europe of 3G services occurred in the first half of 2003. 3G subscriptions are currently estimated at around 375,000 in Italy, 195,000 in the United Kingdom, 12,000 in Sweden and 10,000 in Austria (European Commission, COM[2003] 715).

State of Play of the European 3G Market: Perspectives for the Future

The deployment of 3G, which finally constitutes a new wireless network and service generation, was the result of a continuous effort over the past years (European Commission, COM[2001] 141). Although operators have already started implementing such networks, 3G is not, for the moment, an effective market and commercial reality. This could be partly due to the small screen size of mobile terminals affecting the ability to access and view Web content from such devices, and due to the non-availability of alternative handsets. However, technology is likely to develop in ways to overcome this limitation, given the commercial incentives to offer Internet and video content as part of mobile data services.

The rollout of 3G services implicates a mutual interaction between different market *"players"* such as users, industry and manufacturers, network operators, service providers, software/content providers, local/regional authorities, governments, public, and/or European authorities. Because of the wide range of

the related activities, it is evident that there is a divergent ensemble of interests, dependent on various trends and distinct options.

Since the earlier stages of evolution, the industry sector looked for the rapid development of a clear regulatory framework in certain areas of major importance, such as the impact of competition rules on UMTS, the way licenses would be issued (and the conditions attached to them), interconnection and interoperability issues, future frequency allocations, and spectrum pricing (on scarce natural resources), possible health effects due to exposure to electromagnetic fields, measures to ensure the protection of privacy and personal data (and against fraudulent use), and many more.

As for the technical option, a common approach at the EU level would be essential to maximize the competitive opportunities for the European players within the global market. To this aim, it has been strongly emphasized that ETSI (European Telecommunications Standards Institute) provides the best platform for translating the notion of UMTS into key open and common standards and that technology choices should originate from the industry itself, through the formal standardization process in ETSI. Recent developments have promoted full agreement on most of the 3G key service characteristics (offering of extended coverage, higher bit rates [to support multimedia applications], better spectral efficiency and greater flexibility for the customer, both in service offering and price). In fact, 3G systems offer access to Internet services specifically tailored to meet users needs via various multimedia applications using image, video, sounds, and voice.

3G has been expected to play an important role in continuing the progress and the success of Europe's mobile communications industry in terms of technology development, competitiveness and service deployment, especially to meet particular user and societal needs, both by offering manufacturers and operators a strong market area to expand and effectively competing other parts of the world. The wider 3G sector has been considered as a strong opportunity for the potential building of the Information Society. Development has been assumed to be based on market-led criteria and to be driven by the private sector, while governments should have provided conditions to assure innovation, investment, and a broader competitive offering of mobile multimedia facilities. In addition, pricing of the relevant services would be considered as crucial to whether 3G can be finally seen as a mass-market product or a niche premium service. To this aim, it would be essential that any remaining artificial market barriers are removed and that a broader convergence aspect is promoted.

The European industry very early identified some key areas (European Commission, COM[1997] 513) where specific attention had to be given to overcome possible difficulties, before any generalized attempt for the full introduction of 3G in the EU. These were, *inter-alia*, regulatory issues mainly focused on the

granting of licenses, frequency-related issues, various standardization aspects and the definition of the underlying system to solve possible outstanding technical affairs, research and development (R&D) features, the financial context, as well as proper actions at national and/or international level. As a consequence, the existing licensing and interconnection framework was practically applicable, while rights and obligations to negotiate commercial roaming agreements with other service providers or network operators were based on common, open and internationally competitive air-interface standard. The availability and the adequate allocation of spectrum were fundamental to the UMTS launch, especially for the market expansion and exploitation. Standardization effort has been focused to permit systems to be developed, while allowing a degree of differentiation at service provision. In parallel, extended R&D efforts have been made as for the study of possible health effects related to the use of mobile handsets and to the public exposure to radio frequencies.

Overview of the Underlying Technologies for the 3G Mobile Communications

The primary aim of the third-generation mobile communications system is to achieve convergence between the mobile and the Internet environments by offering a broad range of high-quality speed and *"always-on"* multimedia services. Traditionally, the first generation in mobile networks based on cells used analogue voice communications. Then a common feature in the subsequent second generation was the fact of digital communication, both for voice and data at low speed (such as 9.6 kbit/s). Relevant examples are the well-known and quite popular GSM (global system for mobile communications) with some extra figures (such as SMS), the TDMA (time domain multiple access) solution in IS-136 and the CDMA (code division multiple access) solution in IS-95 of the Pacific/Japanese digital cellular. A basic feature of the third-generation groups of mobile services is that they allow for multimedia applications and mobile Internet at relatively high speeds (2 Mbit/s). Commonly known systems are UMTS and IMT-2000 (International Mobile Communications). In fact, five different standards are foreseen. The UMTS in Europe is expected to be implemented using UTRAN (UMTS terrestrial radio access network), which has two fundamental modes of operation (UTRA-FDD [frequency division duplex], which is compatible with IMT-DS [direct sequence] and UTRA-TDD [time division duplex], which is equal to IMT-TC [time code]). As for the technical evolution, the transition from 2G to 3G has been already accomplished.

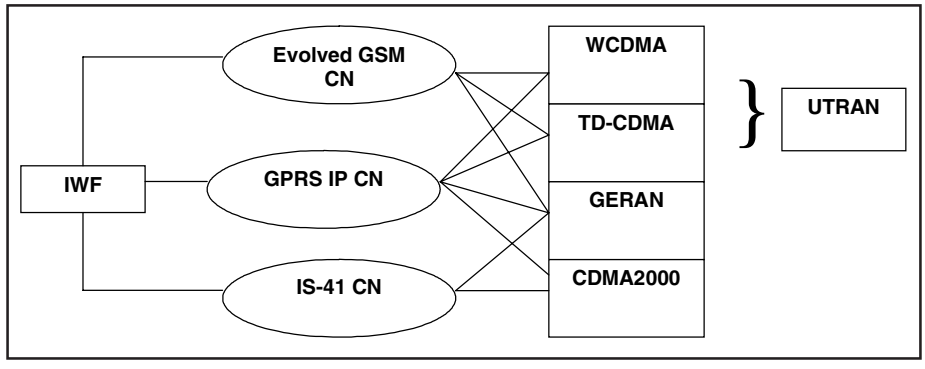
Indicative and characteristic examples of such a *"transition"* are the wireless application protocol (WAP), short-distance connectivity using Bluetooth (2.45 GHz spread spectrum link) technology, CAMEL (customized application for mobile enhanced logic), EDGE (enhanced date rates for GSM evolution, allowing data rates up to 384 kbit/s), GPRS (general packet radio services or GSM phase 2+), etc. The major enhancements to be expected from 3G mobile services are directly linked to new kinds of services, to be widely offered, such as: (a) various communications services (such as enhanced voice quality, fax, e-messaging); (b) multiple information services (such as news, weather reports, sports information, lottery results, stock exchange, etc.); (c) specific financial services (such as online banking and shopping, ticketing, e-payments); (d) enhanced business services such as virtual office, corporate Internet, etc.; (e) a variety of entertainment services (such as video and music on-demand and electronic games); (f) innovative transport-related services such as in-car real-time navigation (GPS), traffic information, etc. In particular, such an option is expected to contribute, significantly, to the development procedures of modern economies.

The evolution towards packet switching (PS) reached the realm of mobile cellular systems with the creation of packet-switched GSM or GPRS. The GPRS network uses Internet Protocol (IP) routers with statistical multiplexing. The important revolution in mobile networks, therefore, should be considered to be the implementation of GPRS in GSM core networks, whereas UMTS or 3G practically represent a straight evolution of GPRS. As for the GPRS case, a related user is constantly connected to the network (an "always-on" state), which is in fact a suitable case for Web browsing. Furthermore, this promotes the option of tariff charging per volume, instead of per connection or time duration of a connection, which facilitates new perspectives for development, especially within the innovative broadband context. The noticeable increase in data rate compared to classical GSM is also appreciated by various categories of customers, in parallel with the fact that the latter can remain reachable even while surfing the Internet (as happens with ISDN (Integrated Services Digital Network) in the local loop). With UMTS the user can enjoy all advantages of GPRS but at a higher throughput (up to 2Mbit/s).

With respect to the implementation of 3G mobile networks, the global differences relate to the preferred radio access technologies: for example, BRAN (radio access network), SRAN (satellite radio access network), GERAN (GPRS/EDGE radio access network), CDMA 2000 (multi-carrier), or UTRAN. Among the above, according to recent market overviews, most of the European operators prefer UTRAN for implementation.

A depicted 3G implementation of possible radio access technologies is given in Figure 1. The interworking functions (IWF) either can use GSM evolution as the

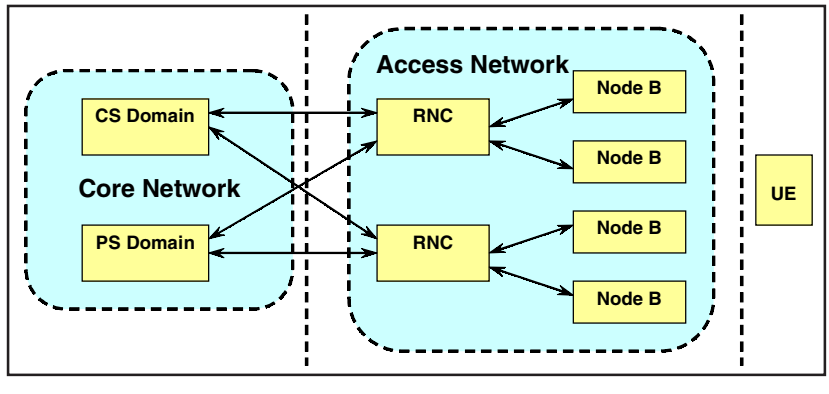
Figure 1. Implementation of 3G mobile networks via the use of different radio access technologies



core network (CN), GPRS with Internet Protocol, or in the US the ANSI-41 core network (if no GSM MAP messaging is foreseen for some providers).

In UTRAN, the user equipment consists of the mobile unit (handset) in which a new universal subscriber identity module (USIM) has to be placed. The USIM smartcard will have twice as much RAM as the SIM in GSM (64 kbyte), will be backwards compatible with SIM (the USIM will not work in GSM phones, but a GSM SIM card will operate in 3G/UMTS devices) and will allow global roaming and multiple network access. Figure 2 gives an example of the interconnectivity between the core network and user equipment. The radio network subsystems consists of radio network controllers (RNCs) connected to base stations (Node B). When compared to the GSM core networks, in UTRAN the base station controllers (BSCs) have interfacing with each other too.

Figure 2. Interconnection scheme between the core network (with either circuit-switched (SW) or packet-switched (PS) domains), the access network and the user equipment (UE)



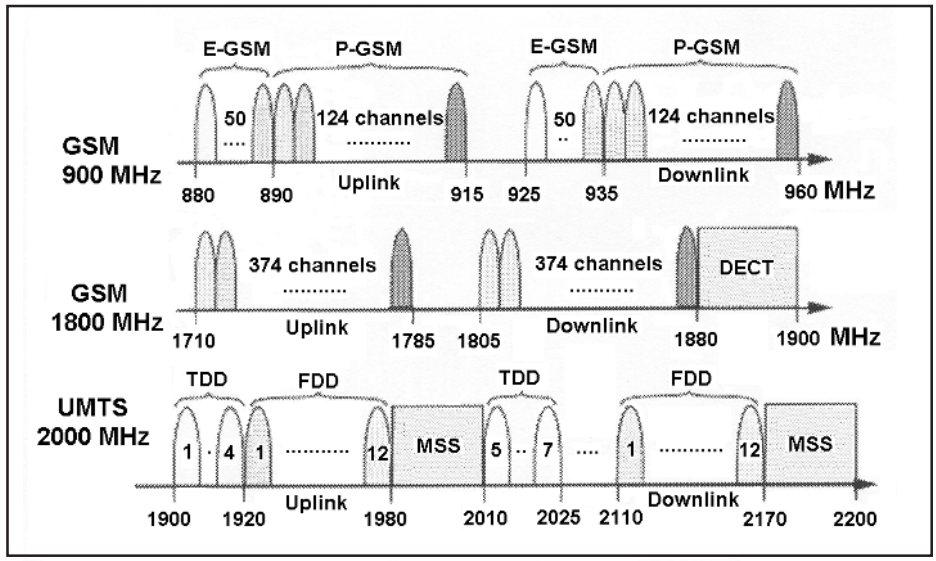
New features and technologies have been selected for the digital communication in UMTS. An important key-player is code division multiple access (CDMA) and the application of the latter uses spread spectrum techniques. CDMA has some important advantages when compared to time domain multiple access (TDMA) as it is used in the actual GSM for each allocated frequency channel. Among the possible advantages are the following: (a) flexible services (as for options referred to bit rates, quality of service, delays, packets); (b) possibility for bandwidth on demand; (c) spectrum efficiency options (also including more facilities for frequency planning); (d) robustness against interference and enhanced resistance against multipath reflections; and (e) the soft way of handover, due to base station diversity, without the usage of interrupts (as happens in GSM).

In UMTS, users in a cell occupy both the entire allocated frequency and time domain. The distinction between users is done through the use of unique codes. Codes of different users appear as noise that can be partly suppressed by the receiver at the base station; however, not all interference can be suppressed, especially if high-data-rate users are located at the edge of a cell. An important difference in network planning for UMTS compared with GSM is the fact that geographical coverage and capacity (by cells) no longer can be regarded as being independent design parameters. This is because the use of wideband code division multiple access (WCDMA) where all users in a cell share the same time and frequency resources and the same power amplifier with respect to the transmitter.

A significant change in the radio link for UMTS is the spectral assignment of frequencies and bandwidth. Figure 3 demonstrates spectral management of UMTS compared to the European GSM bands. The primary GSM (P-GSM) in the 900 MHz band has been extended first with 50 extra channels to yield the extended global system for mobile communications (E-GSM), whereas the 1800 MHz band (referred to as the "dual band" GSM in many locations worldwide), occupies the radio spectrum just below a window allocated to some DECT (Digital European Cordless Telephone) applications. (The 1900 MHz band allocated to GSM in the United States [known as the "third band"] is not shown in Figure 3). Furthermore, MSS in Figure 3 indicates mobile satellite services; TDD refers to time division duplex and FDD to frequency division duplex, with an offset of 190 MHz.

In general, the planning of a network for UMTS consists of three distinct steps: The first one is usually a static analysis considering the general outline. It defines the different classes of types of services in terms of the obtainable data rate, the unavoidable generated delay, the reachable bit error rate (BER) and quality of service (QoS) issues. These can facilitate in order to determine the local distribution of the types of services in the region to be covered (and hence some basic budget calculations). This is also the phase for the identification and the

Figure 3. Spectral management of UMTS compared to the European GSM bands



exact determination of transmitting sites (this is especially important for upgrading previous GSM sites or for the acquisition of new ones). The second step requires effective simulations in order to investigate the quality of different services at different locations, as a function of time; it also requires statistics on performance (for example, the effective data rate, information on dropouts, reachable coverage, etc.). The second stage may also include measurements for service availability and QoS from field trials. The third final phase consists of dynamic simulations to investigate the performance of a population of moving mobiles, over a certain period of time; here extremely important is the investigation of the handover parameters and of the attained network performance. From all these can be derived important physical site parameters (such as height, tilt, azimuth with respect to the site location) and some frequency parameters (such as frequency and capacity planning, handover regions, etc.).

After a transitional period as for the full development of the relevant underlying technology, the latter is now essentially available for use, as it has reached to a suitably mature level. The EU currently possesses the technological know-how and an adequate equipment manufacturing capability. Previously existing difficulties (due to the introduction of new products of considerable technological innovation) have been solved while network equipment is being deployed. As for the necessary enhanced handsets, after successful prototype trials, it is expected that they will be soon available for users and they will be able to support

multifunctional purposes (for example, terminals with a "dual-mode" capability, to support both 3G and GSM). Additionally, new features for such terminals are already available today or can be expected in the near future, such as for example, camera-equipped terminals with high resolution high contrast color screens, better energy efficiency to boost battery capacity and integration of multimedia functions (such as radio, games, recording function, micro browsers, etc.). Significant progress has been achieved, up-to-now, as for the development of various features of terminal equipment and related devices, including, for example, the bandwidth available used, access speed, screen size and display, input methods, processing power, memory, and many other capabilities. In a world of multiple vendors and service providers, interoperability can be achieved while preserving independent competitive implementations of significantly proprietary product and service offerings. A practicable implementation of Information Society services and electronic commerce, requires the presence of generic interoperable technologies across the service infrastructure; at the same time, the vast majority of solutions adopted at network servers should as much as possible be adequately supported by terminal devices and vice-versa.

Standardization Aspects: The Challenge for the Future Evolution of the Sector

The promotion of the wider standardization activities (also including all related voluntary and market-led initiatives) facilitates access to new services. Remedying or preventing market failure is an argument for generating and promoting standardization. Adherence to consistent standards can produce benefits in the form of volume, rapid adoption and ubiquity of services. In any case, as open standards meet various fundamental criteria (such as control, completeness, compliance, cost), their implementation helps realizing interoperability between existing platforms/infrastructures, so that users can be provided the ability to access (*via different methods*) many services and content. This is quite important to create critical mass and economies of scale (with resulting cost advantages and benefits to consumers) and to foster and to speed-up development. Consumers whose devices support globally interoperable open standards, are in position to use services and content, independently of the source of the service and of the user's location. Standardization activities in mobile communications are taking place at many levels for the promotion of the 3G options. Some among their fundamental features may be considered as a result of the promotion of products and services based on open, global standards, protocols and interfaces. Furthermore the application layer is bearer agnostic (GSM, GPRS, EDGE, CDMA, UMTS), while the

architectural framework and service enablers are independent of operating systems. Interoperability of applications and platforms is also a principal feature.

It should be expected that the new version of the Internet Protocol (IPv6) would overcome any addressing shortage for the full deployment of 3G services (European Commission, COM[2002] 96). The primary reason to deploy IPv6 is to improve customer communications experiences while also ameliorating civilization and living standards. It will also enable provision of additional service features, such as guaranteed quality of service and extended security for end-to-end communications and virtual private networks; furthermore, it will allow for both wireless machine-to-machine and man-to-machine interconnection. 3G systems, in combination with other existing technologies may provide the solution for the expected "exploration" of the broadband reality.

Under such circumstances, as interoperability is essential for consumer acceptance of new services, it becomes evident that standardization has a significant role to realize, in particular to provide solutions "independent" of any specific requirements which might be imposed either by the network operator or by the terminal manufacturer. This strengthens flexibility for operators to choose the technical platform to deliver wireless and mobile services. Common standards based upon market requirements are also important to define safe equipment, while overall legal and technical clarity will increase health protection and confidence of citizens and help the necessary infrastructure deployment and/or upgrading. In particular, technical specifications for safe mobile equipment have been developed as harmonized standards for mobile terminals, while there is no evidence of any related harmful health effects. A similar process is currently in progress for mobile stations. New releases of specifications also allow for full and "always-on" multimedia capability.

Market Potential, Financial Context and Business Opportunities

The continuous convergence processes, the opening of the European telecommunications sector to the competition, the fast development of Internet, the advent and the global penetration of *on-line* electronic services, have created a completely new market structure, implicating new roles for any of the participants.

As previously indicated, there are remarkable and encouraging signs of a growing market potential, especially for mobile multimedia and data (Internet-based) services, within the wider context of the ICT-related sectors. Several

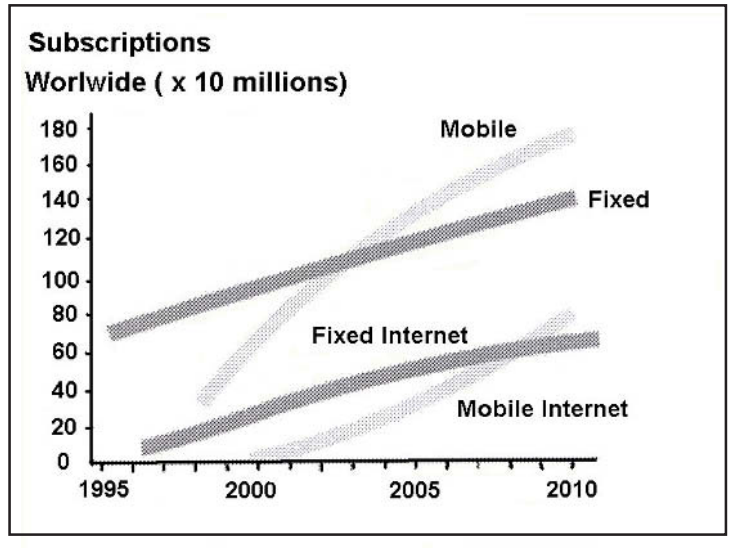
sources forecast increases in global revenues of mobile commerce. However, in some cases detailed forecasts are difficult to make: for example, although WAP has been carefully planned, has not, up-to-now completed a *"revolutionary"* penetration and market reformation; on the contrary, SMS was not as such forecasted to yield to success, but has gained an important (and unexpected) share due to young users, preferring this communication medium over voice.

In any case, the take-off of data applications, like mobile multimedia services, gaming, and so on, seems quite encouraging, especially considering that these packet switched services may lead the way towards 3G services.

Mobile operators are expected to offer a variety of services packages depending on the type of the customer targeted (combination of voice, SMS, e-mail, information and multimedia services). A special feature would be that access to the Internet (or to Internet-based services) is a key part of any service offering. Furthermore, market share should be expected to have significant impact on any relevant business cases; in a similar way, any initial mobile penetration and UMTS penetration may also have impacts on the expected financial results.

Figure 4 provides a forecast for the global expected number of mobile subscriptions. Following this forecast, the number of fixed and mobile users will be equal from 2003 onwards, whereas after that the number of mobile users will grow faster than that of fixed-line users. With respect to Internet usage, the crossing of both curves is expected to take place in 2008. All mobile operators agree that the number of mobile clients will increase at a rate much higher than is the case for fixed users.

Figure 4. Expected worldwide mobile subscriptions (Source: Siemens)



The development of 3G provides for both personal and corporate user an option to realize a full and rich interaction with the world around. As terminals and sensors become generators of data, the user is practically able to communicate with various environments anytime, anywhere, any network and any device, by exchanging information in different formats.

Furthermore, 3G enhances group communications, by making easy to communicate at all times with anyone, whatever their current communication method. Other significant applications could be related to the effective support and further evolution of personal virtual networks, personal information agents, and new families of applications as well.

Figure 5(a) demonstrates a forecast for expected long-term revenue growth to 2010, by providing estimations for worldwide revenues covering all probable services (from simple voice to customized infotainment applications). Figure 5(b) projects expected worldwide 3G revenues from data and voice (including simple voice) up to the year 2010.

Figure 5(a). Expected long-term revenue growth to the year 2010 (Source: UMTS Forum and Telecompetition Inc., August 2001)

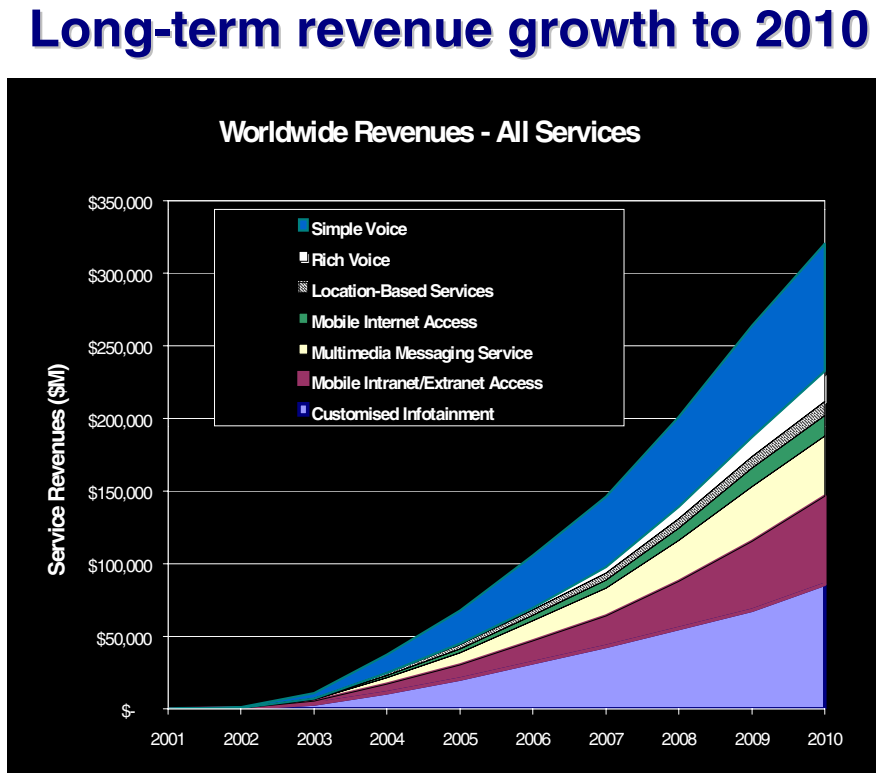
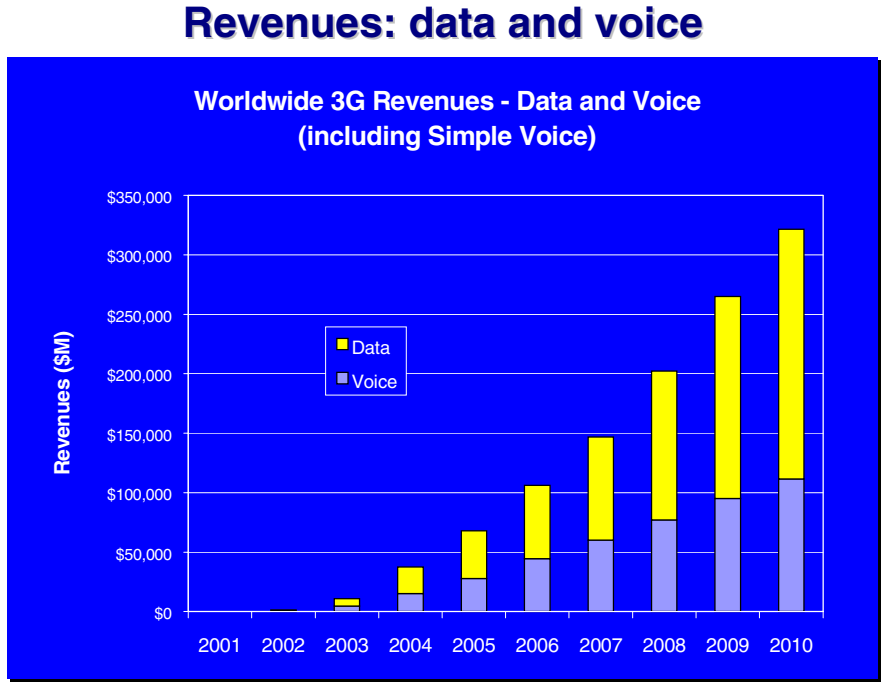


Figure 5(b). Worldwide 3G revenues: Data and voice (Source: UMTS Forum and Telecompetition Inc., August 2001)

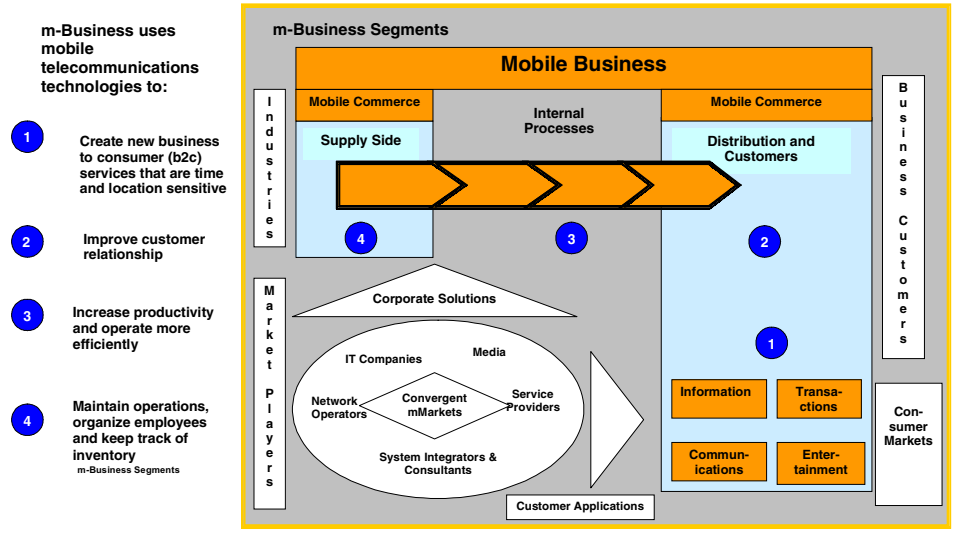


Business models may vary by service and will have a high impact on operator's retained revenue streams. Operators are expected to shift to a portfolio of modern services, within an environment of increased competition with third parties.

From the above considerations, we could summarize that some among the most important user perspectives would be listed as follows: (a) increasing mobility options to enforce mobile communications in business and private sector; (b) wider offering of a variety of services and applications (with a dynamic adaptability for some of their basic features such as, for example, wide spread of usage patterns, data rates and traffic volume), in a convenient and secure way; (c) wider business processes, especially via the penetration of e-commerce applications; and (d) offering and use of a great variety of terminals for different scenarios and various applications, with scaled capabilities.

Figure 6 demonstrates an indicative overview of the mobile business (m-business) perspective, as for possible mobile commerce applications. In fact, m-

Figure 6. Mobile business overview (Source: Detecon, Euroforum)



business supports the exchange of goods, services, information, and knowledge within and between businesses (from various and distinct market sectors, such as network operators, service providers, IT companies, and media companies) and at the interface to customers. Figure 6 also presents some distinct steps, *indicated as 1-4*, for the creation of new business to consumer services, for increasing productivity and maintaining operations. This outlines new perspectives for the global mobile market, especially if mobile services and devices become faster, smoother and cheaper.

However, 3G networks require significant and large investments. The establishment of these networks is likely to generate significant revenue streams: to operators, to service providers and to equipment manufacturers. These revenues will make an important contribution to the future prosperity of the sector but action needs to be taken to encourage appropriate rollout. 3G mobile operators are expected to develop and operate their own configuration of network infrastructure and services with a view to create distinctive commercial presence and offering. Many operators have already upgraded their GSM networks, while the launch of 2.5G services may prove to be a crucial "intermediate" step for the satisfactory uptake of the 3G, by allowing a phased market development and the extensive testing of *"3G-like"* applications (such as data-based services using GPRS (General Packet Radio Services)-enhanced GSM networks or *"i-mode"* services), offering reliable experiences. In any case, it is up to the market players to find out what the viable business models will be, to promote and to strengthen further, such options. Furthermore, planning in 3G requires market-

ing, engineering and finance functions to work closer together, in an integrated fashion.

Apart from the total charges for the granting of licenses, operators have faced comparable costs for the deployment of new infrastructure and for the marketing of new 3G services. As a consequence, operators have turned to the financial markets to finance the investments required. Both the demand for external funds and the high debt of most telecom operators have led to a down-grading of credit ratings and to substantial interest rate spreads. The readjusted financial outlook for the telecommunications sector could have implications for the development of the market since the financial burden weighs heavily, particularly on new entrants, not yet having a presence in the market and at a moment when availability of investment funds is significantly reduced. Indeed, the sector has been faced with a strong and increased pressure from the financial markets. Market players not able to control their 2G experience and customer base and to revise, *accordingly*, their strategies in order to identify ways to reduce capital expenditure, have been (*and still are*) exposed to serious financial burden before being able to obtain benefits and revenues from mobile applications. However, after a period of extreme difficulties, ICT market seems to recover its potential.

The e-economy (European Commission, COM(2001) 711) leads to significant transformations in organizational and market structures including, *for example*, the following: (a) strengthening of competition resulting from lowering of existing barriers/limitations and the creation of new paths for delivery of services; (b) new business models with cost savings, better quality and more customer-driven options; (c) new ways of trading, buying and selling, with further customization either of products or of services. Consequently, the impact of the e-economy varies substantially from sector to sector.

The success of 3G also depends on the critical development and the availability of attractive applications and content, which practically reflects market players' ability to approach and/or to (re-)form suitable market needs (e.g. by offering a variety of "service packages" depending on the type of customer targeted). It may also represent a high potential for job creation beyond the *"traditional"* mobile sector (mobile electronic commerce, booking of tickets, e-banking, transaction and financial services, micro-payments, mobile wallet services, mobile advertising, information services, and entertainment services). However, such a deployment would require suitable technical implementations, an adequate financial context, the development of new services market(s) and the adaptation of business structures to a new value chain, as well as new forms of corresponding business demands. This implicates the necessity for a very detailed approach and/or prediction of the behavior of the relevant market(s), where new business models can come up to provide innovative strategic options. In a complex *"multi-platform"* approach, it should be expected that 3G mobile communications play a key-role for future access to various Information Society

services, as 3G are able to offer data services to users on the move. Such services have to be personalized and user friendly, to assure acceptance from both individual and business users. The challenges become of greater importance for the final consumer, in particular due to the forthcoming expansion of the global broadband applications. Openness, interactivity and interoperability between the existing multimedia platforms are crucial factors to provide a guarantee for a combination of different kinds of viable applications and services, offered by various delivery systems to different end-user terminals.

Other issues such as the creation of a secure and attractive environment to inspire consumer trust and confidence (especially for *on-line* interactive services), to ensure protection against fraudulent use and protection of privacy, personal data and professional secrecy are equally important (European Parliament and European Council, Directive 58/EC, 2002). Increased protection of the networks and systems is necessary against the various types of attacks on their availability, authenticity, integrity and confidentiality. Addressing security issues is also crucial to stimulating demand for new services and applications. Consumers and businesses need a secure environment in which to conduct their communications and to transact business. Identification, authentication, encryption, electronic payment systems, protection against the disclosure of personal data are some among the most important priorities to be arranged, from different points of view. Mobile handsets with their smart card functionalities could help to significantly upgrade the current arrangements for checking authorizations to access particular services or sites.

Another very important domain is the one referred to the potential harmonization of digital rights management (DRM), including certain aspects of copyright and related rights (European Parliament and European Council, Directive 29/EC, 2001) in the Information Society sector; this may affect widespread access to systems as well as the ability of different service providers to sufficiently offer the same content available over different platforms in a pluralistic way. DRM technologies can encourage content producers to make their content available and thus entice users to take-up broadband options. Adequate DRM solutions involving all stakeholders (from content provider to the consumer) promote the development of competition in markets and support the establishment of modern business models. DRM systems and services are closely related to consumer's perception of freedom of choice in accessing Information Society services. Currently, the market is seeing a patent movement by the world's biggest electronic, software and hardware manufacturers towards securing their future with a DRM solution bundled in their offering. It is a simple concept that allows each provider to define a structure by which they will encrypt and deliver secure monetized content. The market is so at a critical point of development. DRM solutions are based on the deployment of elements which objective is to secure the content and avoid piracy. There are several elements that work together in

order to secure the media content with encryption and provide the licenses necessary for the users to access the content.

An open and interoperable environment will allow the creation and delivery of multi-service packages attractive to consumers, encouraging more players to enter the market, thus bringing greater competition with more choice of content, procuring significantly greater economic impact. Openness and flexibility are essential in order to promote innovation and creativity in customer choice, and to avoid unnecessary fragmentation, while interoperability is fundamental to respond efficiently to the demands of the markets and consumers.

In addition, a very important role is to be performed by the European governments, as they boost demand via: (a) the creation of successful business models for content producers and infrastructure operators (especially by introducing new e-government services); and (b) the creation and the distribution of secure transmission environments and by putting public information and government services *online* in a format suitable for access by mobile terminals, thus creating a variety of value-added services. Such services should help to provide citizens and companies personalized public services that meet their specific needs. Public Authorities and the private sector should aim to offer their content on different technological platforms, such as 3G. Different content types are to be put together to a wider and richer *"multimedia presentation"*. In parallel, there is a strong case for EU-level action to provide incentives for multi-lingual, European e-content (European Council, Decision 48/EC, 2001).

In any case, both the commercial offering and the consumer uptake of new services/facilities require clarity with respect to the applicable regulatory treatment of these services. The success of 3G services will also depend on the design, offering and delivery of a range of Internet-based data services for the various categories of users, adapted to the different environment of use (e.g. mobility-specific, location-based, time-dependent data). This has to be combined with the full application of proper tariff policies, based on well-defined market criteria and objective price control. It is evident that such options can have significant positive impacts on wider development activities.

Regulatory Perspectives: E-Communications as an Enabler of the European Regulation

The public sector has the responsibility to provide the most convenient terms and conditions to create a supportive regulatory environment suitable for the

promotion of 3G services, in a fully competitive environment. Furthermore, it aims to enhance legal certainty in a dynamic market and to promote options for technological neutrality, to encourage innovation and to stimulate investment in both networks and services (especially within the context of the forthcoming pan-European "broadband" perspective).

Legislative and regulatory conditions should create a favorable environment for business, attractive investments and favoring innovation and economic development, as well as safeguarding consumers' interests. Establishing suitable regulatory measures (e.g. for spectrum licensing, telecom numbering, tariffs, interconnection, roaming agreements, and roll-out of infrastructure) and increasing reliance on EU competition law are important in creating an appropriate transparent and stable environment for investment. Previous EU regulatory initiatives (European Council, Decision 128/EC, 1999) have defined the capabilities of the new 3G services and set out a suitable time frame to support their introduction via different selection procedures (auctions, comparative selections or a mixture of the two). Predictability for licensing conditions and provisions (e.g., rollout obligations, probable facility-sharing obligations, license duration, license charges and payments, etc.) allows business cases to be established in a reliable manner. As a result, any relevant changes could be considered only when circumstances have changed unpredictably, and always under proportional, transparent and non-discriminatory terms.

The new European regulatory framework (European Parliament and European Council, Directive 21/EC, 2002) will bring benefits to the users by ensuring an open and competitive environment for the delivery of services within a fully converged environment, and by increasing all possible customers' choices (in parallel with the perspective for reduction of tariffs). It encourages the provision of a wide range of high quality services at competitive prices to European citizens. In addition, one of the basic directions is to set up clear, distinct, transparent and objective rules, to improve certainty for further economic investments, to boost business activities, to offer innovation and flexibility, to create powers for intervention in order to achieve wider interoperability, to promote and to consolidate the Internal Market.

There have been some concerns expressed by the industry that the new European regulation could lead to additionally requirements on 3G service markets. However, as most of these markets are characterized as *"emerging"* markets, they will not be subject to *"ex-ante"* regulation. Regulatory intervention from National Regulatory Authorities (NRA) should be better focused only on markets where a durable lack of competition leads to abuse of dominance. Where and when necessary, such an intervention should be accompanied by a thorough analysis of the relevant market, a cost-benefit analysis of regulatory intervention and an analysis of possible less intrusive measures.

A significant challenge also raises from the fact that the relevant sector(s) could be best served by letting the market drive processes ahead, though public authorities can contribute to confidence by ensuring an *"open"* and stable regulatory environment, conducive to a competitive market serving different consumers' interests. Although most European Member States have already issued 3G licenses, there are still difficulties of the sector to become a fully viable reality. Under the scope of the current European legislation (European Parliament and European Council, Directive 20/EC, 2002), it should be expected that deployment of networks and services could be further facilitated, for example, by harmonizing the conditions and accelerating the procedures (thus avoiding probable regulatory hurdles, administrative delays and uncertainties over infrastructure sharing, secondary trading and/or restrictions on antennas placements). Although flexibility is recommended, licensing conditions are expected to remain proportional and transparent, probably within a common, more coordinated and more detailed European harmonized approach (to secure investments with long term prospects). Uneven license conditions across the EU may lead to varying investment incentives in national markets and may eventually give rise to some discrepancy with respect to the levels of mobile service developments, disadvantaging investors in new mobile services and consumers.

Major Radio Spectrum Issues and Other Issues Related to the Physical Deployment of Underlying Infrastructure

Consistency was also a key driver for the adoption of the Radio Spectrum Decision (European Parliament and European Council, Decision 676/EC, 2002), which is in fact a key-part of the new EU regulatory package. The decision establishes, for the first time, a common EU framework to ensure effective coordination for frequency allocation/assignment, and it facilitates policy making with regard to the strategic planning and harmonization of the use and availability of radio spectrum bands (given their considerable demand, especially for 3G operations). Moreover, there is a concern for the establishment of a dialogue framework with industry and national regulators on matters for secondary trading of radio spectrum and its implications, to provide more liquidity in investments. Any relevant activity is for the moment *"implicit"*, through mergers and acquisitions (subject to competition rules) or through handing back to the administration the license, which would be offered as a new one. A clear

and predictable spectrum regulation policy is important as uncertainty in this regard can act as a barrier.

The radio spectrum available in Europe for 3G today is the result of a planning process started about 10 years earlier. For any future availability of possible future requirements there will be suitable consultation and planning work. In particular, the European Commission has already mandated CEPT (European Conference of Post and Telecom Administrations) to undertake studies on the channeling of the European additional frequency bands for future 3G applications. Moreover, the EU supports high-level analysis of the relative spectrum needs for different commercial wireless communications access platforms, with the aim to enable Europe to agree on a strategy to support this important sector. In any case, it is essential for the EU to ensure that 3G systems could operate without undue technical restrictions in a spectrum band chosen by Europe itself, by protecting such systems from any probable harmful interference and within a context of a possible global harmonization of spectrum (European Commission, COM[2003] 707). In parallel, there is an effort for the EU to support and to ensure strict protection of the bands foreseen in Europe to accommodate IMT-2000 systems. The EU position is to keep all options open for additional spectrum identification in the future of 3G and for systems beyond, on the basis of actual market experience and technological progress. The migration between mobile technologies is a long-term evolutionary process, and it is preferable not to take premature decisions on a future mobile generation while 3G services are becoming commercially established, but rather to leave some time for industry and operators to develop services and applications in a stable regulatory context. This convinces potential investors to commit to particular business strategies.

A particular issue where further progress is achievable could be to provide assistance to operators for the deployment of 3G physical infrastructures. Existing difficulties are due to environmental concerns relating to the installation of new 3G masts; this may usually results to fragmented national and/or local planning (in connection with the provision and the acquisition of base stations). Recent initiatives have promoted measures for co-location and facility sharing, to accelerate local planning, in parallel with exchange of best practices between national or local authorities to find proper solutions. Facility sharing should be encouraged when technically appropriate and in accordance with license terms and competition law. Other matters originate from sensitive public opinions on health of users (European Council, "Recommendations on the Limitations", 1999) due to their exposure to electromagnetic emissions (as for both base stations (European Commission, "Report on a Short", 2001) and handsets. Work for determination of emission regulations has facilitated consensus, while various research programs have proved that there are not adverse health effects from the normal use of equipment. In any case, the lack of harmonization and of best

practices referred to the levels of emissions in the Member States, hampers both industry and the citizens, as creates confusion about what is safe or not. Thus, it would be necessary, *in the future*, to enhance applicable health and environmental policies, also by developing a common and coherent approach on site/antenna licensing. Furthermore, the full appliance of the provisions of the R&TTE Directive (European Parliament and European Council, Directive 5/EC, 1999) will support development of technical specifications for safe mobile use.

The appropriate use of the new framework for electronic communications, the harmonization in licensing conditions and reasonable radio spectrum assignment approaches across the EU, can facilitate integration in the Internal Market for communications and avoid market distortions (or uncertainty), especially in view of the future enlargement of the European Community.

Overview and Concluding Remarks

Among the priorities of the European policy (European Commission, COM[2002] 301) is to foster the use of open platforms to provide freedom of choice to citizens for access to applications and services of the Information Society, notably through 3G mobile and other platforms that technological convergence may provide in the future. This option can be considered simultaneously with the great popularity and the continued growth of the mobile penetration in the EU, *for example*, two important factors providing a significant basis for the acceptance (under "suitable" terms) and the success of 3G's *"approval"* by the consumers. However, although most of the EU Member States have already issued a significant number of 3G licenses, the introduction of the new mobile communications generation is still limited, in most of the existing markets.

The difficult global financial situation for most of the operators has obliged them to reassess strategies, to explore new revenue sources and to reduce investment expenditure, to what is absolutely necessary. However, the financial environment is still challenging, especially due to the fact that there are strong indications that market is recovering its potential, after a period of extremely pessimistic predictions not materialized.

Moreover, the total sector has succeeded to overpass a certain number of previously existing barriers and/or other limitations, and has reached to a significant evolutionary progress. In particular, existing technologies have been developed to offer innovative solutions, while research efforts still provide major input for further exploitation. 3G technologies seem to be more stable today than in the near past. Indeed, standardization works carried out in the context of 3GPP

(Third Generation Partnership Project of ETSI) have been continued to be successful, with the release of new sets of specifications which gradually support transition for the 3G development, based on internet protocols. Significant progress has been also achieved for the handset technologies, most of which are almost ready to be commercialized very soon. In the global 3G environment, industry is seeking to achieve interoperability of services via open infrastructure, to assure adequate access by all citizens and businesses.

Market has also responded, by itself, to the expected emerging potential for the promotion of the Information Society applications, especially within current EU initiatives to achieve widespread access to new services and similar facilities. The challenge from the expansion of broadband opportunities has reformed the total sector to become a real commercial reality, especially in terms of consumer demand, productivity and competitiveness. Thus, the rollout of 3G networks and services is an evolving process, structuring an important and effective aim for many actors participating in ICT market(s).

Experience from previous (and sometimes time-extended) activities has driven to the establishment of a new and more stable EU regulatory framework, able to support innovation perspectives and investment aspects. Confidence has been already guaranteed in the sector as for the majority of the relevant regulatory issues, while further progress is expected to take place. This will provide assurance and certainty for the development of healthy competition and will contribute to assurance for the benefits of all parties involved. The new framework will also contribute to avoid probable market distortions and to accelerate both infrastructure and services' deployment. In particular, within the proposed framework, market is expected to forward processes, though governments, public authorities and the European Commission can contribute to assure stability (without intervene in the financial environment). Coordination across Europe and further coordination of policies might be necessary to overcome present challenges (such as, *for example*, efforts to improve public awareness on safe use of equipment together with speeding up procedures for the acquisition of base station sites).

The long-term goal is to fasten and to secure 3G as one of the fundamental service platforms in an *"always-on"* and innovative Information Society environment.

Acknowledgments

The major author of the present work, Dr. Ioannis P. Chochliouros, would like to express his profound gratitude to Mrs. Anastasia S. Spiliopoulou-Chochliourou

(Lawyer, LL.M., Member of Athens Bar Association and OTE's Lawyer-Partner) as a co-author of the paper, especially for her valuable contribution for the full completion of the exposed work.

References

European Commission. *Communication on Strategy and Policy Orientations with Regard to the Further Development of Mobile and Wireless Communications (UMTS),* COM(1997) 513 final, 15.10.1997.

European Commission. *Communication on The Introduction of Third Generation Mobile Communications in the European Union: State of Play and the Way Forward,* COM(2001) 141 final, 20.03.2001.

European Commission (2001). *Report on a Short Term Mission of Base Station Exposure within the Context of COST 244bis*, European COST 244bis Programme, Belgium.

European Commission. *Communication on The Impact of the e-Economy on European Enterprises: Economic Analysis and Policy Implications,* COM(2001) 711 final, 29.11.2001.

European Commission. *Communication on eEurope Benchmarking Report,* COM(2002) 62 final, 05.02.2002.

European Commission. *Communication on Next Generation Internet - Priorities for Action in Migrating to the New Internet Protocol IPv6,* COM(2002) 96 final, 21.02.2002.

European Commission. *Communication on eEurope 2005: An Information Society for All, An Action Plan,* COM(2002) 263 final, 28.05.2002.

European Commission. *Communication on Towards the Full Roll-Out of Third Generation Mobile Communications,* COM(2002) 301 final, 11.06.2002.

European Commission. *Communication on eEurope 2005: Benchmarking Indicators,* COM(2002) 655 final, 21.11.2002.

European Commission. *Communication on the Eighth Report from the Commission on the Implementation of the Telecommunications Regulatory Package,* COM(2002) 695 final, 03.12.2002.

European Commission. *Communication on The Road to the Knowledge Economy,* COM(2003) 65 final, 11.02.2003.

European Commission. *Communication on Results of the World Radiocommunication*, COM(2003) 707, 19.11.2003.

European Commission. *Communication on European Electronic Communications Regulation and Markets 2003 - Report on the Implementation of the EU Electronic Communications Regulatory Package,* COM(2003) 715 final, 19.11.2003 [SEC(2003) 1342].

European Council. *Recommendation on the Limitations of Exposure of the General Public to Electromagnetic Fields (0 Hz to 300 GHz) (1999/519/EC)* OJ L199, 30.07.1999, pp. 59-70.

European Council. *Decision No2001/48/EC adopting a Multi-Annual Community Programme to Stimulate the Development and Use of European Digital Content on the Global Networks and to Promote the Linguistic Diversity of the Information Society* OJ L14, 18.01.2001, pp.32-40.

European Parliament and European Council. *Decision No128/1999/EC on the Co-ordinated Introduction of a Third-Generation Mobile and Wireless Communications System (UMTS) in the Community* OJ L17, 22.01.1999, pp.01-07.

European Parliament and European Council. *Directive 1999/5/EC on Radio Equipment and Telecommunications Terminal Equipment and the Mutual Recognition of their Conformity ("R&TTE Directive")* OJ L091, 07.04.1999, pp. 10-28.

European Parliament and European Council. *Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society* OJ L167, 22.06.2001, pp. 10-19.

European Parliament and European Council. *Decision No 676/2002/EC on a Regulatory Framework for Radio Spectrum Policy in the European Community ("Radio Spectrum Decision")* OJ L108, 24.04.2002, pp. 1-6.

European Parliament and European Council. *Directive 2002/20/EC on the Authorisation of Electronic Communications Networks and Services ("Authorization Directive")* OJ L108, 24.04.2002, pp. 21-32.

European Parliament and European Council. *Directive 2002/21/EC on a common regulatory framework for electronic communications networks and services (*"*Framework Directive*") OJ L108, 24.04.2002, pp. 33-50.

European Parliament and European Council. *Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the communications sector ("Data protection Directive" or "e-Communications Privacy Directive")* OJ L201, 31.07.2002, pp.37-47.

Appendix

List of Abbreviations

136HS	136 High Speed
1xRTT	First phase of CDMA2000 (next phase is 3xRTT)
2G/2.5G	2nd Generation (Cellular System), General term for digital wireless networks using IP
3G	3rd Generation (Cellular System)
3GPP/3GPP2	3G Project Partnership (3GPP – UMTS/WCDMA and 3GPP – CDMA2000)
4G RAC	Radio Access Controller of 4G network
5GPP	5 GHz Partnership Project
5GSG	5 GHz Service Group
802.11	Original IEEE standard for “Wireless Ethernet ” at 2.4Ghz (FHSS or DSSS)
802.11a	IEEE standard for “Wireless Ethernet ”at 5.2GHz (OFDM)
802.11b	IEEE standard for High Rate “Wireless Ethernet ” at 2.4Ghz (DSSS)
802.11g	IEEE (proposed) standard combining 802.11a and 802.11b
AA	Authenticator Address
AAA	Authentication, Authorisation, and Accounting

AAAH	Home AAA
AAAL	Local AAA
AAL	ATM Adaptation Layer
ACELP	Algebraic Code Excited Linear Predictive
AES	Advanced Encryption Standard
AF	Assured Forwarding
AMPS	Advanced Mobile Phone System
Anonce	Authenticator Nonce
AP	Access Point – the end network equipment providing network access to the end users.
ARIB	Association for Radio Industry and Business
ARPU	Average Revenue Per User
AS	Authentication Server
ASCII	American Standard Code for Information Interchange
ATM	Asynchronous Transfer Mode
AuC	Authentication Controller
B-ISDN	Broadband – Integrated Services Data Network
BPSK	Binary Phase Shift Keying
BRAIN	Broadband Radio Access for IP based Networks (IST-1999-10050)
BRAN	Broadband Radio Access Networks
BSC	Base Station Controller
BSS	Basic Service Set
BSSID	Basic Service Set Identifier
BTS	Base Transceiver Station
CA	Certificate Authority
CBC	Cipher-Block Chaining
CBC-MAC	CBC Message Authentication Code
CCK	Complementary Code keying
CCM Mode	Counter Mode with CBC-MAC
CCMP	CCM Protocol
CDMA	Code Division Multiple Access
CDMA2000	Code Division Multiple Access 2000
CDPCISCO	Discovery Protocol

CDPD	Cellular Digital Packet Data
CGI	Common Gateway Interface
CHAP	Challenge Handshake Authentication Protocol
CL	Convergence Layer
CN	Core Network
COPS	Common Open Policy Service
CRC	Cyclic Redundancy Check
CRL	Certificate Revocation List
CS	Circuit Switched
CSMA	Carrier Sense Multiple Access
CTR	Counter Mode
DAMPS	Digital Advanced Mobile Phone System
DBPSK	Differential Binary Phase Shift Keying
DCCH	Digital Control Channel
DCS	Dynamic Channel Selection
DECT	Digital Enhanced Cordless Telephone
DES	Data Encryption Standard
DFS	Dynamic Frequency Selection
DHCP	Dynamic Host Configuration Protocol
DLC	Data Link Control
DLEN	MPDU Data Length
DoS	Denial of Dervice
DPSK	Differential Phase Shift Keying
DQPSK	Differential Quadrature Phase Shift Keying
DRNC	Drift Radio Network Controller
DSCP	Differentiated Services Code Point
DSSS	Direct Sequence Spread Spectrum
DTCH	Digital Traffic Channel
DWDM	Dense Wavelength Division Multiplexing
EAP	Extensible Authentication Protocol
EAPOL	EAP Over LAN
EAP-TLS	EAP Transport Layer Security
EAP-TTLS	EAP Tunneled Transport Layer Security

ECC	Elliptic Curve Cryptography
(E)DCF	(Enhanced) Distributed Coordination Function (IEEE802.11e)
EDGE	Enhanced Data Rates for GSM Evolution
EF	Expedited Forwarding
EFR	Enhanced Full Rate
EK	Encryption Key
EPOC	New operating system for mobile devices (Symbian)
ESN	Enhanced Security Network (WEP2 (=TKIP) and AES devel opment)
ESS	Extended Service Set
ETSI	European Telecommunications Standards Institute
FA	Foreign Agent
FDD	Frequency Division Duplexing
FHSS	Frequency Hopping Spread Spectrum
FINEID	Finnish Electronic Identity
GERAN	GSM/EDGE Radio Access Network
GGSN	Gateway GPRS Support Node
GMK	Group Master Key
GMSK	Gaussian Minimum Shift Keying
Gnonce	Group Nonce
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications
GTK	Group Transient Key
GTP	GPRS Tunnelling Protocol
GTP-U	GPRS Tunnelling Protocol-User plane HA Home Agent
HCF	Hybrid Coordination Function (IEEE 802.11e)
HDML	Handheld Devices Markup Language (now called WML)
HiperLAN	High Performance Wireless LAN - Versions 1 and 2 (European Standard)
HiperLAN/2	High Performance Radio Local Area Network 2 (ETSI/BRAN standard at 5.2GHz using OFDM)
HIRAN	HiperLAN /2 Radio Access Network
HLR	Home Location Register

HR/DSSS	High Rate Direct Sequence Spread Spectrum
HSCSD	High-Speed Circuit-Switched Data
HSS	Home Subscriber Server
i-mode	See PHS
IAPP	Inter-Access Point Protocol (transfers client's state between APs)
IAS	Internet Authentication Server (Microsoft's Radius)
IBSS	Independent Basic Service Set
ICV	Integrity Check Value
ID	IDentifier
IEEE	Institute of Electrical and Electronic Engineers
IETF	Internet Engineering Task Force
IKE	Internet Key Exchange
IIS	Internet Information Server
IMEI	International Mobile Equipment Identity
IMSI	International Mobile subscriber Identity
IMT-2000	International Mobile Telecommunications for year 2000 and beyond
IP	Internet Protocol
IPPM	IP Performance Metrics
IPSec	IP Security Protocol
IrDA	Infra-red Data Association
ISDN	Integrated Services Data Network
ISM	Industrial, Scientific and Medical
ISP	Internet Service Provider – see also WISP
ITU	International Telecommunications Union
IV	Initialization Vector
IWU	InterWorking Unit
J2ME	Java 2 Micro Edition (Java enabled browsers in the handset)
L2CAP	Logical Link Control and Adaptation Protocol (Bluetooth)
LAN	Local Area Network
LEAP	Lightweight Extensible Authentication Protocol
LED	Light Emitting Diode
LMP	Link Manager Protocol (Bluetooth)

MAC	Medium Access Control protocol
MAN	Metropolitan Area Network
MAP	(GSM) Mobile Application Protocol
MCC	Mobile Country Code
MD4	Message Digest 4
MD5	Message Digest 5
MGW	Media Gateway
MH	Mobile Host
MIC	Message Integrity Check
MIP	Mobile IP
MK	EAPol-Key MIC Key
MM	Mobility Management
MMAC	Multimedia Mobile Access Communication Systems (Japan)
MNC	Mobile Network Code
MPDUs	MAC Protocol Data Units
MPLS	Multi-Protocol Label Switching
MS	Mobile Station
MSC	Mobile Switching Centre
MS-CHAPv2	Cryptanalysis of Microsoft's PPTP Authentication Extensions
MSDU	MAC Service Data Unit
MT	Mobile Terminal
NAI	Network Access Identifier
NAS	Non Access Stratum
NIC	Network Interface Card
NIST	National Institute of Standards and Technologies
NMT	Nordic Mobile Telephone
OFDM	Orthogonal Frequency Division Multiplexing
OSI	Open System Interconnect
PBCC	Packet Binary Convolution Coding (Texas Instruments)
PBCCH	Packet Broadcast Control Channel
PCCCH	Packet Common Control Channel
PCF	Physical Control Field
PCMCIA	Personal Computer Memory Card International Association

PCN	Personal Communication Network (Equivalent to GSM 1800 MHz)
PCS	Personal Communications Service
PDA	Personal Digital Assistant
PDC	Personal Digital Cellular
PDCH	Packet Data Channel
PDH	Plesiochronous Digital Hierarchy
PDP	Packet Data Protocol
PDTCH	Packet Data Traffic Channel
PDU	Packet Data Unit
PEAP	Protected EAP
PHB	Per Hop Behaviour
PHY	Physical Layer
PHS	Personal Handyphone System (i-Mode/DoCoMo)
PKI	Public Key Infrastructure
PLCP	Physical Layer Convergence Protocol
PLMN	Public Land Mobile Network
PLP	Packet Loss Probability
PMK	Pairwise Master Key
PN	Packet Number
PPP	Point to Point Protocol
PPTP	Point to Point Tunneling Protocol
PRF	Pseudo Random Function
PRNG	Pseudo Random Number Generator
PS	Packet Switched
PSK	Pre-Shared Key
PSTN	Public Switched Telephone Network, the local telephone operator
PTK	Pairwise Transient Key
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAC	Routing Area Code
RADIUS	Remote Authentication Dial-in User Service

RAN	Radio Access Network
RANAP	Radio Access Network Application Port
RC4	A variable key-size stream cipher with byte oriented operations
RC-LED	Resonant Cavity - Light Emitting Diode
RF	Radio Frequency
RFCxxx	Request for Comments xxx
RLC	Radio Link Control
RNAS	RAN Access Server
RNC	Radio Network Controller, 3G switch controlling the radio part.
RNSAP	Radio Network Subsystem Application Port
RNTI	Radio Network Temporary Identifier
ROSE	Radionet Open Source Environment
RrK	Rapid reKeying (IEEE Submission, August 2001)
RSN	Robust Security Network
RSN IE	RSN Information Element
RSVP	Resource reSerVation Protocol
SAC	Service Area Code
SACCH	Slow Associated Control Channel
SAI	Service Area Identifier
SAP	Service Access Point
SCAM	Supplemental Channel Assignment Message
SCRM	Supplemental Channel Request Message
SDH	Synchronous Digital Hierarchy
SDP	Service Discovery Protocol (Bluetooth)
SDU	Service Data Unit
SGSN	Serving GPRS Support Node
SIM	Subscriber Identification Module – see also (U)SIM
SIP	Session Initiation Protocol
SIR	Signal-to-Interference Ratio
SLA	Service Level Agreement
SLP	Service Locator Protocol
SM	Session Management
SMG	Special Mobile Group

SMS	Simple Messaging Service (~ 100 characters)
Snonce	Supplicant Nonce
SONET	Synchronous Optical NETwork (network with SDH over fiber optical link)
SRNC	Serving RNC
SRP	Secure Remote Password
SSCS	Service Specific Convergence Sublayer
SSH	Secure Shell
SSID	Service Set Identifier
STA	Wireless Station – Any 802.11 device other than an AP.
SWAP	Shared Wireless Access Protocol
TA	Transmitter Address
TACS	Total Access Communications System
TDMA	Time Division Multiple Access
TEID	Tunnel Endpoint Identifier
TIA	Telecommunications Industry Association
TK	Temporal Key
TKIP	Temporal Key Integrity Protocol
TLS	Transport Layer Security
TMSI	Temporary Mobile Subscriber Identity
TOS	Type of Service
TPC	Transmit Power Control
TSC	TKIP Sequence Counter
TTA	Telecommunications Technology Association
TTAK	TKIP mixed Transmit Address and Key
UNII	Unlicensed National Information Infrastructure
UMTS	Universal Mobile Telecommunications System
URA	UTRAN Registration Area
USB	Universal Serial Bus
USF	Uplink Status Flag
(U)SIM	User Subscriber Identification Module – see also SIM
USSD	Unstructured Supplementary Services Data (<182 characters)
UTRAN	Universal Terrestrial Radio Access Network

UWCC	Universal Wireless Communications Consortium
VLR	Visitor Location Register
VoIP	Voice over IP
VPN	Virtual Private Network
VSELP	Vector Sum Excitation Linear Predictive
WAE	Wireless Application Environment
WAMN	Wide Area Mobile Network
WAN	Wide Area Network
WAP	Wireless Application Protocol
WAP-NG	WAP Next Generation (WAP 2.0) based upon XHTML
WASP	Wireless Access Service Provider
WCDMA	Wideband Code Division Multiple Access
WCMP	Wireless Control Message Protocol
WDP	Wireless Datagram Protocol
WECA	Wireless Ethernet Compatibility Alliance
WEP	Wired Equivalent Privacy
WEP IV	WEP Initialization Vector
Wi-Fi	Wireless Fidelity
WIM	Wireless Identity Module
WISP	Wireless Internet Service Provider
WLAN	Wireless Local Area Network
WML	Wireless Markup Language
WPA	Wi-Fi Protected Access Version 1
WPA2	Wi-Fi Protected Access Version 2
WPA IE	WPA Information Element
WPAN	Wireless Personal Area Network
WPKI	Wireless Public Key Infrastructure
WRAN	Wireless Radio Area Network
WRR	Weighted round robin (scheduler)
WSG	Wireless Second Generation
WSL	Wireless Session Layer
WSN	Wireless Support Node
WSP	Wireless Session Protocol

WTA	Wireless Telephony Application
WTAI	Wireless Telephony Application Interface
WTL	Wireless Transport Layer
WTLS	Wireless Transport Layer Security
WTP	Wireless (Transaction, Transport) Protocol
WWAN	Wireless Wide Area Network
X.509	ITU standard specifying contents of a digital certificate
XHTML	Extensible Hypertext Markup Language
XML	Extensible Markup Language

Glossary

1G: First Generation systems, which are analog and were designed for voice communications.

2G: Second-generation systems, which are digital and capable of providing voice/dat/fax transfer as well as a range of other value-added services, including SMS.

2.5G: Evolving second-generation systems that are intermediary options before the introduction of multimedia cellular.

3G: Third-Generation systems, which enable multimedia and are standardized under 3GPP.

3GPP: Third Generation Partnership Protocol.

4G: Fourth-generation systems. In early 2001, Alcatel, Ericsson, Motorola, Nokia and Siemens founded the Wireless World Research Forum (WWRF), whose "vision of the wireless world" was identified as that of 4G systems.

4-Way Handshake: The final procedure in the authentication protocol defined by the IEEE 802.1X standard.

802.1p: IEEE standard that defines techniques for supporting dynamic group multicast filtering and traffic prioritization in 802 LANs. The latter is accomplished by the use of a 3-bit user priority field contained in the 802.1q VLAN tag.

802.1q: IEEE standard that defines the method for supporting virtual LANs (VLANs) across 802 LANs, It works by inserting a 4- or 10-byte VLAN tag in each frame as it traverses one or more VLAN-capable bridges (switches).

AAA: Authentication, Authorisation and Accounting: An AAA server performs these functions, processing requests using a AAA protocol such as RADIUS.

AAAH: Home AAA: Logical function within the loose coupling architecture that provides AAA functions to support subscribers who have a permanent relationship with that network.

AAAL: Local AAA: Logical function within the loose coupling architecture that enforces the AAA policy within the local HiperLAN /2 network.

Access Point: A device through which computer clients connect to a WLAN.

AES: Advanced Encryption System: An encryption method based on the Rijndael algorithm that will be the basis for future wireless encryption standards.

Agent advertisement: The procedure by which a mobility agent becomes known to the mobile node. An agent advertisement message is constructed by attaching a special extension to a router advertisement message.

Agent discovery: The process by which a mobile node can obtain the IP address of a home agent or foreign agent, depending upon whether the mobile node is home or away from home. Agent discovery occurs when a mobile node receives an agent advertisement, either as a result of periodic broadcast or in response to a solicitation.

AP: Access Point: A wireless station that also provides services such as association and distribution of frames to other station or a network.

AP: Access Point: Interface between the radio network part and the wired network part of a HiperLAN /2 network, offering wireless connectivity to MTs.

AS: Authentication Server: A network component that performs authentication. A RADIUS server is an example of an AS.

Association: A set of mutually agreed security parameters for protected communication.

Authentication: User identification or client computer identification for legitimate access to computing resources.

Authentication Protocol: Rules and procedures for authentication.

Authentication Server: A computer program dedicated to authentication of users and computer clients.

Authenticator: 802.1X term for an entity that facilitates authentication. Access points act as authenticators.

Beacon: A radio transmitter or the signal emitted by it when the emission is used as a directional guide, such as a homing beacon or a localizer beacon.

Binding: See Mobility binding.

Binding update: The message that supplies a new binding to an entity that needs to know the new care-of address for a mobile node. The binding update contains the mobile node's home address, new care-of address, and a new registration lifetime.

Bridge: A device connecting two or more network segments within the same logical local area network.

Broadcast Traffic: The same data packet is sent to all hosts in a network.

BSS: Basic Service Set: A set of 802.11 stations that communicate with each other.

BSSID: Basic Service Set Identifier: A unique identifier for a particular BSS. In infrastructure mode, the MAC address of an access point.

CA: Certificate Authority: An authority that issues and manages digital security credentials such as public-key certificates.

Certificate: A record, cryptographically signed by a CA, with a public key and identity information about the key owner.

CN: Correspondent Node: A peer with which a mobile node is communicating. A correspondent node may be either mobile or stationary.

COA: Care-of Address: An IP address which identifies the mobile node's current point of attachment to the Internet, when the mobile node is not attached to the home network. The protocol can use two different types of care-of address: a foreign agent care-of address is an address of a foreign agent with which the mobile node is registered; a colocated care-of address is an externally obtained local address which the mobile node has associated with one of its own network interfaces.

COPS: Common Open Policy Service: A protocol used for policy control and provisioning.

Credentials: Identity proof information.

CRL: Certificate Revocation List: A list of client certificates that were revoked by the authority before they expired.

Cryptographic key: A symmetric key or a public key or a private key.

Cryptographic Protocol: Rules and procedures for applying cryptographic algorithms.

Cryptographic Signature: A hash of a digital document or of a digital message encrypted with the private key of the signer.

DER format: A format for storing certificates in files.

Digital Certificate: Usually same as certificate.

Digital Signature: Usually same as cryptographic signature.

EAP: Extensible Authentication Protocol: A general protocol for authentication which support multiple authentication mechanisms.

EAPOL-Key Packet: Created and transmitted by the Authenticator in order to provide media specific key information in WPA key management.

Eavesdropping: Unauthorized read access to transmitted information.

Encapsulation: The process of incorporating an original IP packet (less any preceding fields such as a MAC header) inside another IP packet, making the fields within the original IP header temporarily lose their effect.

Encryption: The process of obscuring information to prevent it from being read by unauthorized parties.

Encryption Algorithm: A set of operations implementing encryption.

FA: Foreign Agent: A router on the foreign network that assists a locally reachable mobile node on that network. The foreign agent assists the mobile node in receiving datagrams delivered to the care-of address.

Firewall: A network traffic filter.

Foreign Network: The network to which the mobile node is attached when it is not attached to its home network. The mobile node's care-of address is local to the foreign network and is reachable from the rest of the Internet.

FreeBSD: An open source UNIX based operating system.

FreeS/WAN: Open source VPN software for Linux computers.

Gateway: A device connecting a network to an internet.

Group Key: A symmetric key for protected communication between an AP and all client computers authenticated by this AP.

HA: Home Agent: A router on a mobile node's home network that tunnels datagrams for delivery to the mobile node when it is away from home.

Handover: To maintain a path between an MT and a correspondent node when the MT moves between cells of the same radio technology or between different radio technologies with a minimum of involvement from the user.

Hardware Token: A certified private key stored in a separate computer chip.

Hash: A fixed size, unique, cryptographic bit pattern derived from a document or from a message.

Header: The control information in a data packet before the payload.

HiperLAN/2 Network: Consists of a number of Access Points with continuous radio coverage and of associated mobile terminals.

HLR: Home Location Register: Centralized entity containing subscription data that is required for user authentication and encryption in a 2nd generation GSM network on a per user basis.

Home Address: A static IP address on the home network that is assigned for an extended period of time to a mobile node. The home address may be permanently assigned to the mobile node, or may be dynamically assigned for the duration of the mobile node's session.

Home Network: The network associated with the mobile node's home address. IP routing mechanisms will deliver datagrams destined to a mobile node's Home Address to the mobile node's Home Network.

HSS: Home Subscriber Server: A centralized entity containing subscription data that is required for user authentication and encryption in a 3rd generation mobile network (UMTS) on a per user basis.

IAPP: Inter-Access Point Protocol: A protocol for communicating information between access points in order to support roaming.

IEEE: Institute of Electrical and Electronics Engineers: A non-profit, technical professional association for electrical and electronics engineering.

IEEE 802.11i: A forthcoming standard for OSI layer 1 and 2 WLAN security

IEEE 802.1X: A standard for authentication of computers clients connected to a Local Area Network port or to a WLAN access point.

Internet: A network consisting of interconnected networks. An internet is any network, which has the same structure as Internet.

IETF: Internet Engineering Task Force: The principal body engaged in the development of new Internet standard specifications.

Key Hierarchy: Rules for deriving symmetric session keys from a symmetric master key.

Key Management: Rules and procedures for creation, distribution, storage, and use of cryptographic keys.

LEAP: Lightweight EAP: A Cisco vendor-specific authentication method that provides mutual authentication and dynamic WEP key generation.

Linux: An open source UNIX based operating system.

Macrocell: cell with coverage radius of many Km.

Microcell: cell with coverage radius of up to many hundreds of meters.

Master Key: The symmetric key used to generate symmetric session keys.

Multicast Traffic: The same data packet is sent a subset of all hosts in a network.

Minimal Encapsulation: An encapsulation technique which uses fewer header bytes for tunneling packets to the care-of address than the default IP-within-IP method uses.

MN: Mobile Node: A computing device that may change its point of attachment from one network or sub-network to another through the Internet. The mobile node is assigned a fixed home address on a home network, which correspondent nodes may use to address their packets to, regardless of its current point of attachment.

Mobility: Ability of an MT to be used in different network environments, within a single and in different administrative domains, with minimum user intervention.

Mobility Agent: A node (typically a router) that offers support services to mobile nodes. A mobility agent is either a home agent or a foreign agent.

Mobility between administrative domains: Ability for a MT to function in a serving network different from the originating network mobility between network environments: refers to the ability of an MT to be used in different network environments, such as home, corporate and public roaming: mobility between administrative domains.

Mobility Binding: The triplet of numbers that associates a mobile node's home address with its care-of address and registration lifetime.

MSC server: Switching Center for Circuit Switched traffic in 3G networks.

MT: Mobile Terminal: End system equipment providing the interface to people.

Nonce: A unique random value that is never reused.

OpenBSD: An open source UNIX based operating system.

Open Source Software: Software, which is available also as source code without a commercial software license.

OpenSSL: Open source software implementing the Secure Sockets Layer protocol.

Open1x Project: Open Source Implementation of the IEEE 802.1X standard.

Pairwise: Two entities associated with each other.

Pairwise Key: A symmetric session key used by two entities associated with each other.

Pass phrase: A sentence used as a password.

Payload: The information content of a data packet in data communication .

PEM Format: A format for storing certificates in files.

Per-Packet Key: Every data packet is encrypted with a different symmetric key.

Picocell: cell with coverage radius of many meters.

PKCS12: A format for storing a Public Key Cryptography key pair in a file.

PKI: Public Key Infrastructure: A configuration of systems and components required to manage and administer a public key environment.

Pre-Shared Key: A pre-installed symmetric key.

Private Key: The key in a Public Key Cryptography key pair known exclusively by the key pair owner.

Probe Response: An action taken or an object used to learn something about the state of a network.

Public Key: The publicly known key in a Public Key Cryptography key pair.

Public Key Cryptography: Two different but interrelated keys are used for encryption and decryption. One key is public and the other key is private. The key interrelation is easily created but too complex to crack.

Radio AP: BTS's, Node B's, etc.

RADIUS: Remote Authentication Dial-in User Service: A protocol used to perform authentication, authorization and accounting (AAA).

RC4: A variable key-size stream cipher with byte oriented operations. A registered trademark of RSA.

Registration: The process by which the mobile node informs the home agent of its current care-of address.

Registration Lifetime: How long the mobility agents may use a mobility binding.

Replay Attacks: A security violation whereby a malicious third entity attempts to imitate a transaction recorded during a previous and valid transaction between two protocol entities. Both protocol entities have to be aware that the subsequent identical traffic streams may no longer be valid. Since the previous transaction was valid, the algorithms for detecting replay attacks need to incorporate data that can never be reproduced in any correct subsequent transaction.

Route Optimisation: A process that enables the delivery of packets directly to the care-of address from a correspondent node without unnecessarily detouring through the home network.

Secret Key: The same as a symmetric key.

Security Association: A collection of one or more security contexts, between a pair of nodes, which may be applied to Mobile IP protocol messages exchanged between them.

Security Context: A security context indicates an authentication algorithm, a secret (a shared key), and a style of replay protection in use.

Security Protocol: Rules and procedures to be applied in protected data communication.

SIM Card: The SIM (Subscriber Identification Module) card is a smart card that identifies a user to the network and contains a microprocessor chip, which stores unique information about an account, including phone numbers and security numbers (PIN and key). There are two different sizes of SIM cards used for GSM phones, one is the same size as a credit card and the other is about the size of a stamp. Both cards contain the same electrical circuits; only the plastic surrounding it different. Functions on the SIM card includes: memory space to save up to 100 names and phone numbers, in addition to as many as 15 SMS, short text messages.

SLP: Service Locator Protocol: A protocol which provides a method to discover and select network services.

Smart card: A computer chip embedded in a plastic card.

Soft Token: Same as software token.

Software Token: A certified private key stored in a file.

SPI: Security Parameter Index: An index identifying a security context between a pair of nodes among the contexts available in the Security Association.

SRP: Secure Remote Password: A cryptographically strong authentication mechanism suitable for negotiating secure connections and performing a secure key exchange using a user-supplied password.

SSID: Service Set Identifier: An arbitrary string naming an access point or set of access points for purposes of identifying the WLAN to clients.

STA: Wireless Station: Any 802.11 device other than an AP.

Supplicant: 802.1X term for an entity that is being authenticated. Often a synonym for client, workstation, or user.

Symbian: Company which develops operating system for hand-held devices. Most wireless manufacturers have adopted this OS.

Symmetric Key: The same key is used for encryption and decryption.

TLS: Transport Layer Security: A protocol designed to provide privacy and data integrity between two communicating applications. Specifically, EAP-TLS provides protected ciphersuite negotiation, mutual authentication, and key management. transactions are eventually handled by the AAAH, possibly via one or more intermediaries.

Tunnel: The path followed by a datagram while it is encapsulated. An encapsulated datagram is routed to a knowledgeable de-encapsulating agent, who de-encapsulates the datagram and delivers it to its ultimate destination.

Unicast Traffic: A data packet is sent to only one receiver.

USB Adapter: Provides connectivity through a computer's USB port.

Virtual Private Network: Software implementing IPSec.

Visited Network: A network other than a mobile node's home network, to which the mobile node is currently connected.

Visitor List: A list of mobile nodes visiting a foreign agent.

Visual Basic: A programming environment from Microsoft for graphical Windows applications.

VPN: Virtual Private Network: A method of using encryption and tunneling to securely connect users over a public network.

WEP: Wired Equivalent Privacy: An 802.11 privacy service – encrypts data over wireless medium.

Wi-Fi Alliance: The Wi-Fi Alliance is a nonprofit international association formed in 1999 to certify interoperability of wireless Local Area Network products based on IEEE 802.11 specification.

WLAN: Wireless Local Area Network: A network that provides the features of traditional LAN technologies such as Ethernet and Token Ring using wireless technology.

X.509: International Telecommunications Union standard specifying contents of digital certificate.

About the Authors

Margherita Pagani is a principal researcher at the I-LAB Centre for Research on the Digital Economy of Bocconi University (Italy) and she teaches courses at the Department of Management of Bocconi University. She is an associate editor of the *Journal of Information Science and Technology* and serves as an Ad Hoc reviewer to the Editorial Advisory Board of the *International Journal of Cases on Electronic Commerce (IJCEC)*. She is the editor of the *Encyclopedia of Multimedia Technology and Networking* and she is the author of three books on digital interactive television and full Internet mobility. She has published her research works in such journals as the *Journal of Interactive Marketing*, the *Journal of Media Management*, *Economy & Management* and many book chapters, and conference papers in the research area of interactive digital television, full Internet mobility, content management, and Digital Rights Management. She was a visiting scholar at the Sloan School of Management — MIT (Massachusetts Institute of Technology) and a short visiting professor at the School of Business of the University of Redlands (California). She worked with RAI Radiotelevisione Italiana and as an associated member of the Permanent Forum of Communications (work group "Digital Terrestrial") for the Ministry of Communications (Italy).

* * *

Ioannis P. Chochliouros is a telecommunications electrical engineer who graduated from the Polytechnic School of the Aristotle University of Thessaloniki (Greece). He also holds an MSc and Ph.D from the University Pierre et Marie Curie, Paris VI, France. He worked as a research and teaching assistant at the University Paris VI, in cooperation with other European countries. His practical

experience as an engineer has been mainly in telecommunications, as well as in various construction projects in Greece and the wider Balkan area. Since 1997 he has been working at the Competition Department and then as an engineer-consultant of the chief technical officer of OTE (Hellenic Telecommunications S.A.), for regulatory and technical matters. He has been strongly involved in major OTE's national and international business activities, as a specialist for technical and regulatory affairs. He currently works as the head of the Department for Technical Regulations of the Division for Standardisation & Technical Regulations of OTE, within the context of the General Directorate for OTE Group Technology. He has been involved in different national, European, and international projects and activities, while he has published numerous scientific and business papers and reports, especially for technical, business and regulatory options arising from innovative e-Infrastructures and e-Services. He has also participated to many international conferences, workshops, forums, and other events, in most of which as an invited speaker—official OTE's representative.

Angelo Corallo was involved until 2000 mainly in the field of physics working on data modelling data, acquisition, and data analysis processes. In a first period he was involved in the high energy physic field, working on a MACRO project, and on a neutrino oscillations issue, developed at the Gran Sasso laboratory by an international collaboration. In the second period he worked on the ROSETTA project, which focused on the design of a space mission devoted to analyse the Wirtanen comet tail. This project was developed by a European collaboration funded by ESA. Starting in 2000, he was involved in higher education and research activities of eBMS-ISUFI, and since 2004 he has been a researcher at the Department of Innovation Engineering at the University of Lecce. His research field concerns knowledge management and e-business adoption from SMEs, both from the technological and organisational aspects.

Marco Cremonini is assistant professor at the Department of Information Technology of the University of Milan (Italy). He earned a Ph.D in electronic and information technology engineering at the University of Bologna. He has been a research assistant at the Institute for Security Technology Studies (ISTS) of Dartmouth College, USA. Among his interests, there are Web-based secure systems and applications, secure protocols, and network security.

Ernesto Damiani holds a degree in electronic engineering from Universita‘ di Pavia and a Ph.D degree in computer science from Universita‘ di Milano. He is currently a professor at the Department of Information Technology of the

University of Milan (Italy). His research interests include distributed and object oriented systems, semi-structured information processing, and soft computing. He is the vice-chairman of the ACM Special Interest Group on Applied Computing (SIGAPP). He is the author, together with I. Sethi and R. Khosla, of the book *Multimedia MultiAgent Systems* (Kluwer, 2001).

Sabrina De Capitani di Vimercati is an associate professor at the Department of Information Technology of the University of Milan (Italy). She received her Laurea and Ph.D degrees both in computer science from the University of Milan (1996 and 2001, respectively). Her research interests are in the area of information security, databases, and information systems. She has been an international fellow in the Computer Science Laboratory at SRI, CA (USA). She is co-recipient of the ACM-PODS'99 Best Newcomer Paper Award. For more information, visit: *http://www.dti.unimi.it/~decapita*.

Claus Dietze graduated in technical computer science at the University of Co-operative Education in Mannheim, Germany. He has dealt with smart cards since August 1996, starting with the definition and product management of smart card solutions for payment systems at Giesecke & Devrient. Since May 1999, he has worked as a project manager for smart cards used in mobile communications systems and defined the first commercially used TETRA SIM as well as managed the introduction of multiple mobile services on SIM and USIM cards. From March 2002 until December 2003, he was responsible for standardisation bodies responsible for the specification of second and third generation smart cards at the European Telecommunications Standards Institute (ETSI) (France).

Gianluca Elia has mainly focused on innovative technologies development for the touristic territorial system, attends the first edition of eBMS Master. Since 2000, he has been involved in higher education and research activities of eBMS-ISUFI. From 2002-2003, he was a teacher at the Engineer Faculty of University of Lecce. Since 2004, he has been a researcher at Department of Innovation Engineering, Faculty of Engineering, University of Lecce (Italy). Currently, his research has concerned the analysis of innovative model and solutions for managing knowledge resources in e-Learning platforms strictly aligned with Knowledge Management platforms, creating innovative solutions for business innovation. He is also involved in coordination activities of working-teams regarding some research projects funded by national and European institutions.

Daniel Escartin Daniel is a technical engineering student in information technology at Escuela Universitaria Politécnica de Teruel (Universidad de Zaragoza), Teruel (Spain). Since September 2003, he has been working on his final thesis on WLAN security, as an exchange student at Arcada Polytechnic, Espoo, Finland.

Elizabeth Fife is a principal researcher at the Center for Telecom Management at the University of Southern California (USA), where she is the editor of the Telecom Outlook Report and a lecturer in the field of Business and Technical Communications in the Marshall School of Business and the School of Engineering. She received her Ph.D in international relations from USC, and her areas of research include European telecommunications, wireless, e-commerce, SMEs, and R&D innovation. She has seven years research experience in the telecommunications field.

Bardo Fraunholz is a senior lecturer in project management, enterprise modelling, and business communication systems. He has a masters degree in business, specialising in information systems/accounting and has done post graduate work in legal studies, specialising in IT, Media, and Corporate Law, from London. Bardo spent several years in the information communication technologies sector as co-editor/board member of a publisher specializing in IT and telecommunication magazines. Currently, he is actively involved in managing/consulting with a number of projects dealing with trade and information systems. His main research interests are information systems projects, IT, and law and mobile applications.

Fiona Fui-Hoon Nah is an associate professor of management information systems at the University of Nebraska-Lincoln (USA). She received her Ph.D in management information systems from the University of British Columbia. She is an associate editor of *Journal of Electronic Commerce Research*. She also serves on the editorial boards of six other major journals. She has published her research works in such journals as *Communications of the ACM*, *Journal of the Association for Information Systems*, *Communications of the Association for Information Systems*, *Information Resources Management Journal*, *Journal of Electronic Commerce Research*, *International Journal of Electronic Business*, and *Journal of Computer Information Systems*. Her research interests include ubiquitous and mobile e-commerce, human-computer interaction, computer-supported collaborative work, and theory building in information systems research.

Holtjona Galanxhi-Janaqi is a second year doctoral student of management information systems at the University of Nebraska-Lincoln (USA). She has co-authored a paper on Ubiquitous Commerce that is forthcoming in *Industrial Management and Data Systems*. Her research interests include ubiquitous and mobile e-commerce, silent commerce, human-computer interaction, and the strategic role of IS in organizations.

Kaj J. Grahn, Dr. Tech., is presently senior lecturer in telecommunications at the Department of IT and Electronics of Arcada Polytechnic, Espoo (Finland). He is also program manager of the Electrical Engineering Programme.

Athanassios C. Iossifides received his Diploma in Electrical and Computer Engineering and his Ph.D from the Department of Electrical and Computer Engineering of Aristotle University of Thessaloniki (AUTh) (1994 and 2000, respectively). He participated with others in the research projects "ATTACH" (1996-1998) and COST 252 (1997) in the area of telecommunications. He served as an associate lecturer in the areas of informatics and telecommunications for over three years for AUTh and Technological Institute of Thessaloniki. In 1999, he joined COSMOTE Mobile Telecommunications S.A. (Greece), serving now as the manager of the "Core and Access Management Group" of Northern Greece Maintenance Division. He has published over 15 papers in international journals and conferences. His research interests lie in the area of mobile communications, CDMA, modulation and FEC techniques, transmission media and technology, and optical fibers.

Jürgen Jung is a lecturer in enterprise modelling, software development and mobile applications. He has a masters degree in computer science specialising in information systems. Jürgen led and has been involved in several projects regarding the development of mobile applications. He is currently involved in an e-commerce-project regarding the introduction of e-business-related processes in small and medium-sized enterprises. His main research interests are business process modelling, mobile applications, and object-oriented software-development.

Jonny Karlsson is a BSc (Eng) student in Information Technology at Arcada Polytechnic, Espoo (Finland). He has since May 2002 been working at Arcada Polytechnic as a research trainee and research assistant in VPN technology and WLAN security research.

Chi Ming Leung received his BE in electrical and electronic engineering from the University of Hong Kong (1992), and his MSc (DIC) in Communication and Signal Processing Department at Imperial College, University of London (1998). He then joined the Department of Electronic Engineering, Queen Mary College, and was awarded a Ph.D in 2004, for research into the measurement of packet networks. He is currently a researcher in the Department of Electronic Engineering at Queen Mary Unviersity of London (UK); his research interests are in the area of quality-of-service theory, algorithms, measurement, and architectures.

Spiros Louvros graduated from Physics Department of University of Crete, Hellas (1993). He received financial assistance from the Public Benefit Foundation ALEXANDROS S. ONASSIS for graduate studies in the UK, where he received a master of science in telecommunications from the University of Cranfield, UK (1994). In 2004, he received a Ph.D in telecommunications from the Physics Department of University of Patras, Hellas. He joined SIEMENS telecommunications in 1996 for six months as a microwave engineer and the same year he joined Vodafon-Hellas as a switching engineer. In 1998 he joined his present employer, COSMOTE, as a BSS engineer. He is involved as an external researcher in the wireless lab, Electrical Engineering Department, University of Patras. His research interests involve communication engineering, handover algorithms, queueing theory, optical communications, next generation mobile networks, and lately quantum computing.

Mikko Martikainen is a BSc (Eng) student in Media Engineering and Information Technology at Arcada Polytechnic, Espoo (Finland). In the summer of 2003, he worked as a research trainee at Arcada Polytechnic researching security in WLAN environments.

Phillip Olla is an independent telecommunication consultant in the UK. Over the past 10 years, Phillip has worked on a wide variety of pioneering and innovative European telecoms projects in conjunction with mobile network operators and mobile service providers including MMO2, Hutcinson 3G, T-Mobile and IBM Global Services. He received his Ph.D from the Department of Information Systems and Computing at Brunel University (UK). His research papers have been published in a wide range of international journals including *Information Technology and People, Telecommunication Policy*, and *Internet Research*; he is a member of the editorial board for the *Industrial Management & Data Systems Journal.*

Francis Pereira is a principal researcher at the Center for Telecom Management. He also is an assistant professor of clinical information and operations management at the University of Southern California (USA). Francis received his Ph.D in political economy and public policy from the University of Southern California and teaches courses in e-commerce, economics and statistics. His areas of research include trade and financial flows in the Association of South-East Asian Nations. Current research focuses on e-commerce applications, particularly in the SME market. He has seven years research experience in the telecommunications field.

James B. Pick is a professor in the School of Business at the University of Redlands (USA), former department chair of management and business and former chair of the Business School Faculty Assembly. He holds a BA from Northwestern University and a Ph.D from the University of California, Irvine. He is the author of eight books and 110 articles, book chapters, and conference papers in the research areas of management of information systems, geographic information systems, urban studies, and population. He has received faculty distinguished teaching and research awards from University of Redlands, Senior Fulbright scholar award for Mexico, and the Thunderbird Award from the Business Association for Latin American Studies.

Göran Pulkkis, Dr. Tech, is presently senior lecturer in computer science and engineering at the Department of IT and Electronics at Arcada Polytechnic, Espoo (Finland).

G. Keith Roberts is academic associate dean and assistant professor in the School of Business at the University of Redlands (USA). He holds a BBA and JD from the University of Oklahoma, an MSIT from the University of Redlands, and an LLM from the National Law Center, George Washington University. He teaches and has authored papers in the research areas of law (including intellectual property), ethics, telecommunications policy, and information technology.

Pierangela Samarati is a professor at the Department of Information Technology of the University of Milan (Italy). Her main research interests are in data and application security, information system security, access control policies, models and systems, and information protection in general. She has been a computer scientist in the Computer Science Laboratory at SRI, CA. She has been a visiting researcher at the Computer Science Department of Stanford University, CA, and at the ISSE Department of George Mason University. She is co-author of

the book, *Database Security* (Addison-Wesley, 1995). She is co-recipient of the ACM-PODS'99 Best Newcomer Paper Award. For more information, visit: *http://www.dti.unimi.it/~samarati*.

John Schormans graduated in 1984, having been a sponsored undergraduate with GEC Marconi. In 1987 GEC sponsored him to join the Department of Electronic Engineering at Queen Mary, University of London, gaining a Ph.D (1990). He returned as a lecturer in 1994, researching packet based networks. He has been a Principal Investigator for EPSRC and the recipient of a BT Short Term Fellowship, and is the special editor of the IEE Special Edition of Proceedings Communications "Superhighways Technology and Broadband VPNs". Promoted to senior lecturer in 2000, he is co-author of *Introduction to ATM Design and Performance* (Wiley 1996), and *Introduction to IP and ATM Design and Performance* (Wiley 2000).

Anastasia Spiliopoulou-Chochliourou is a lawyer, LLM, member of the Athens Bar Association. Since 1992, she has gained extended experience as a lawyer, while she has been involved in various affairs. Her LLM's post-graduated degree, from the Athens University Law School, has been taken place with specific emphasis given to the investigation of the multiple regulatory aspects related to the Internet (infrastructure, services, software, content). During the latest years, she had a major participation in matters related to telecommunications and broadcasting policy, in Greece and abroad, within the framework of the Information Society. She has been involved in current legal, research and business activities, as a specialist for e-commerce and e-business, electronic signatures, e-contracts and e-procurement, e-security and other modern Information Society applications. She has published numerous scientific papers, with specific emphasis given on regulatory, business, commercial, social, and technical aspects. She currently works as an OTE's (Hellenic Telecommunications S.A., Greece) and serves as a lawyer-partner for the Division of Procurement and Services' Contracts of the Legal Department of OTE Group of Companies.

Chandana Unnithan is an associate lecturer in business information systems and project management. She has a master's degree by research in business computing and an MBA. She has spent 14 years in the information communications technology sector highlights being with IBM GSA and TATAs of India. She is actively consulting and focusing on research relating to mobile applications in project management, implications of mobile technology for global IT/IS projects in project-driven organisations. She has a special interest in comparative studies relating to mobile communications diffusion and growth of mobile technologies.

Index