telektronikk 3/4.05

Future Mobile Phones

Contents

Telektronikk
Volume 101 No. 3/4 – 2005
ISSN 0085-7130

Editor:
Per Hjalmar Lehne
(+47) 916 94 909
per-hjalmar.lehne@telenor.com

Editorial assistant:
Gunhild Luke
(+47) 415 14 125
gunhild.luke@telenor.com

Editorial office:
Telenor R&D
NO-1331 Fornebu
Norway
(+47) 810 77 000
telektronikk@telenor.com
www.telektronikk.com

Graphic design:
Design Consult AS (Odd Andersen), Oslo

Layout and illustrations:
Gunhild Luke and Åse Aardal,
Telenor R&D

Prepress and printing:
Rolf Ottesen Grafisk Produksjon, Oslo

Circulation:
3,600

Networks on networks

Connecting entities through networks – in technological, societal and personal terms – enables telecommunication. Networks occur on different levels, form parts of larger networks, and exist in numerous varieties. The artist Odd Andersen visualises the networks on networks by drawing interconnected lines with different widths. Curved connections disturb the order and show that networks are not regular but are adapted to the communication needs.

Per H. Lehne, Editor in Chief

Future Mobile Phones

Special

Guest editorial – Future mobile phones

DO VAN THANH

Do Van Thanh is Senior Research Scientist at Telenor R&D

In technology, things are quite often turned upside down and become the opposite of what they were originally meant to be. Hardware components of computers are becoming "softer" and more replaceable than software components. Telephones go wireless while televisions get "wired" through cable. The mobile terminal, which was originally meant as a peripheral device, has become very central in the mobile telecommunication world. Indeed, from a rather heavy and unpractical mobile station, the mobile terminal has evolved to becoming a handy device that has found its way to the heart of more and more people. The mobile phone is getting more advanced every day in terms of functionality and offers; in addition to telephony, valuable services like messaging, contact, calendar, mobile commerce, gaming, etc. are available The mobile phone is incorporating more and more gadgets like digital camera, MP3 player, memory stick, remote control, etc. It is becoming more visible to the user and offering greater user experience with higher resolution screen, better sound quality, easier navigation facility, etc. It is no longer considered as a communication device, but more as a jewel that one is expressing oneself with.

The battle in the mobile telecommunication world is nowadays very much more over the mobile phone than on the mobile network. The competition is no longer between telecom manufacturers like Nokia, SonyEricsson, Motorola, Samsung, etc., but the telecom operators are also starting to show interest in influencing both the functionality and appearance of the mobile phones. Some major telecom operators such as Vodafone, Orange, T-mobile, etc. view customisation of mobile phones as a decisive differentiation since the mobile phones are what the customers see and want to pay for. Another motivation for joining the battle of mobile terminals is the dissatisfaction on the fragmentation of the hardware components and the incompatibility of the operating system and software platforms. These constitute one of the major hindrances to the promotion and acceptance of advanced data services. Indeed, the mobile operators tend to be in favour of the evolution of the mobile terminal from a black box to an open and standard computer where software applications can be installed or removed on demand.

Another factor that favours the "opening" of the mobile phone is the advance in micro technology that has enabled the production of high-integration mobile phone chipsets which are smaller in size and consume less power. These chipsets contribute to the standardisation of the mobile phone's hardware. Parallel to this is the emergence of open and "de-facto" standard operating systems based on Linux for mobile phones like Monte Vista, Trolltech, BlueCat, etc. Even native Java operating systems like Savaje have appeared. This forces the established mobile phone operating systems like Symbian, Windows Smartphone, PalmOS to provide open and standardised APIs. The fundaments for open and computer-oriented phones are hence laid, but it may not be desirable that mobile phones become pure computers which are vulnerable to virus attacks. Indeed, the minimum requirement is that no matter what happens, a phone should, at least, function as a phone, i.e. be able to make and receive calls. This calls for a new architecture where security is seriously taken into account.

This is one of the activities of the OMTP (Open Mobile Terminal Platform), an operator initiative to open and standardise mobile phones. Another major goal of the OMTP is the operator's customisation across heterogeneous mobile phones, i.e. to define a uniform user experience across different mobile phones. But, the mobile operators are not the only players that want to have something to say on the mobile phones. The first and most obvious player is the user who wants to personalise his mobile phone according to his preferences. Next is the third party service provider that delivers its services on the mobile phone and may want to customise to differentiate with other service providers. Last, but definitely not least, is the enterprise that is the employer of the user, the payer of the mobile subscription and most often the owner of the mobile terminal. The enterprise may or may not want to have company logo or colour on the mobile display. However, they will probably want to have the right to approve the applications running on the mobile phone in the same way that they select and grant the installation of the clients running on the employee's laptop that communicate remotely with the servers on the company's intranet and behind the firewalls.

Will the mobile phones continue to be distributed in the market in the same way as today with most of the functionalities and features decided by the mobile phone manufacturer? Will the user have the possibility to buy separately the hardware and software components and tailor their own mobile phone as in the

ISSN 0085-7130 © Telenor ASA 2006

case of the PC? Will the operator or the enterprise do the assemblage and customisation of the mobile phones? Actually, nobody can predict what will happen in the future. The most reasonable solution may be a "multi ownership" scheme where each player owns one part of the mobile phone. But for this to happen, the future mobile must have a new and sound architecture that provides autonomy, adequate security and privacy to each owner at the same time as the scarce resources of the mobile phone are shared in an optimal way. There are surely many technology challenges but the major obstacles seem to be political and economical making the future scenery in the mobile phone arena quite unclear.

The objective of this issue of *Telektronikk* is to shed light on the state-of the art of the mobile phone but also its history and future. The issue will provide views on technologies, user's requirements, social consequences and economic aspects of the mobile phone. This *Telektronikk* issue on future mobile phones opens with a recapitulation of the mobile phone's evolution from its birth to the present day with a prospect to the future. Since the history of the mobile terminal cannot be understood separately the history of mobile telephony is organised to be the second overview paper.

The next three papers focus on the human and social aspects of the mobile phones. The first one talks about mobile phones and fashion while the second investigates the interest in future net based services. The third one studies how people use their mobile multimedia devices.

The mobile phone arena is currently led by the telecom vendors and their views have great influence in the evolution of the mobile phones. The next three papers reflect the vision of the vendors. The paper from Microsoft expresses the importance of deploying services across heterogeneous devices. The Qualcomm's paper focuses on the flexible selection and installation of applications on the mobile phone. The Philip's paper attempts to predict the gadgets that can be added to the mobile phones.

The paper about OMTP – Open Mobile Terminal Platform – summarizes international activities that for the first time ever focus solely on the mobile phone. As the mobile phone evolves to be more than a telephone, more interesting data services are emerging. While one paper describes how emails can be made available on the mobile phone, the other explains how the user can fetch music from his home PC to his mobile phone and send home the photos from his mobile phone. The usage and role of RFID on mobile phones are also covered in one paper.

This issue concludes with a view on the future with three articles. The first one presents a future device management system that is capable of providing dynamic and complex management operations on heterogeneous mobile phones. Another predicts that mobile personality will replace personal mobility that allows the mobile user to develop her own online personality in terms of personal preference, usage and service profiles over time, as well as the offered services to acquire a unique proactive behaviour. The last paper presents a futuristic mobile phone, "the virtual device", which consists of all the autonomous devices around the user as one big "Virtual Device" having multiple input and output units and providing a coherent and surround interface to the user.

Finally, I would like to thank Richard Seyler Ling, Anders Spilling, Birgitte Yttri and Marianne Jensen for their precious support, advice and contributions to the elaborations of this *Telektronikk* issue.

I hope that you will find this issue both entertaining and informative no matter whether you want to jump into the mobile phone arena or not.

Have an enjoyable read.

Prof. Dr. Do Van Thanh obtained his MSc in Electronic and Computer Sciences from the Norwegian University of Science and Technology (NTNU) and his PhD in Informatics from the University of Oslo. In 1991 he joined Ericsson R&D Department in Oslo after seven years of R&D at Norsk Data, a minicomputer manufacturer in Oslo. In 2000 he joined Telenor R&D and is now in charge of two Eureka Celtic projects, ADPO and Fidelity, which focus on personalization of mobile services and Identity Management. He also holds a professorship at the Department of Telematics at NTNU in Trondheim. He is the author of over 100 publications in international conferences and journals and the inventor of 14 patents and a dozen of pending ones.

email: thanh-van.do@telenor.com

The mobile phone: Its evolution from a communication device to a universal companion

DO VAN THANH AND IVAR JØRSTAD

Do Van Thanh is Senior Research Scientist at Telenor R&D

Ivar Jørstad is a PhD student at Norwegian University of Science and Technology

Nowadays, most people know what a mobile phone is but not so many know all the functions this little device has and how many metamorphoses it has encountered before becoming what it is today. This paper attempts to provide a comprehensive description of the mobile phone and a presentation of its evolution in terms of architecture, functionality, bandwidth, wireless access technologies, etc. from a simple communication device to becoming a universal companion of the user. The paper also gives an overview of development trends of the mobile phone.

1 Introduction

At the time (80s) when fixed telephones became a commodity present in every household and lost its preponderant economic position in the telecommunication, the mobile phone emerged as the main source of revenues for telecom equipment manufacturers and mobile telecommunication rose as the biggest success ever in the telecommunication history. Indeed, most of the telecom manufacturers were about to close their production of terminals and concentrate on the network equipment since the business was not lucrative anymore. The general belief was that the competition should be on the network side rather than on the terminal side. One company did have a visionary view on the role and potential of the mobile phones and put considerably resources and efforts on the construction of the mobile handsets to make them the most used electronic device in the world. Today, everybody knows that this company is Nokia, which has grown from a modest telecom vendor to be a telecom giant. This is just to depict the important role of this little handy device, which is the mobile phone. A mobile phone is now a very familiar device to most people but not everyone really knows about the advanced functions and capabilities incorporated in the mobile phone. Not many people are aware of all the metamorphoses the mobile phone has undergone before becoming what it is today. However, the changes will not stop there but will continue at an even faster pace. Indeed, nobody can predict what the mobile phone will look like in the future. It could be the combination of everything one can think of from a watch, a camera, a camcorder, and an MP3 player to document storage, a universal security token or a remote control. It may become "invisible" with all the components taken apart and discretely mounted into the user's clothes.

In this paper, we are going to re-discover the mobile phone from the day of its invention. Its functions and capabilities will be revisited and elucidated. At the end of the paper, we will attempt to depict a picture of the future mobile phone.

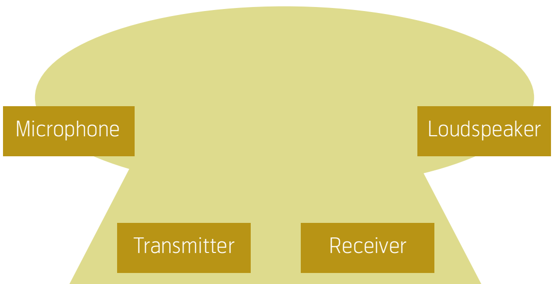


Figure 1 A Plain Old Telephone

2 The basic functionality of the mobile phone

2.1 The mobile phone is a terminal

The mobile phone is quite often referred to as a *mobile terminal* in telecommunication terminology. A terminal is an apparatus that terminates or is at the terminating endpoint of the telecommunication network. A terminal is a peripheral device which is rather simple and completely dependent on the network, i.e. it cannot function when disconnected from the network. Typically, a PSTN terminal, also called Plain-Old Telephone (POT) is terminating the Public Switched Telephone Network. It is a rather simple device that enables the user to make or receive phone calls and consists basically of four components:

- A receiver to accept signals
- A transmitter to send signals
- A microphone to capture the voice of the user
- A loudspeaker to generate sounds to the user.

To operate, the PSTN terminal is alimented by an electric current generated by the local exchange on the local loop, i.e. the circuit between the local exchange and the terminal. When disconnected from the line, the PSTN terminal is completely inoperative.

ISSN 0085-7130 © Telenor ASA 2005

Photos courtesy of Thomas Farley, www.privateline.com

The phone does not have any identity on itself but the identification is based on the line between the local exchange and the household. Any phone hooked up to the line will be viewed in the same way by the local exchange and will operate without any difference.

2.2 The mobile phone is a station

The mobile phone is a terminal and has similar capabilities to making and receiving calls as the fixed phone. However, contrary to the fixed phone, the mobile phone is an advanced and quite complex device. In the specifications of the Groupe Special Mobile (GSM), which was later renamed Global System for Mobile communication, the mobile terminal is called *mobile station* that denotes a bunch of equipment necessary for transmitting and receiving while moving. Indeed at the time, it was anticipated that the mobile terminal would be quite big and heavy. It is *transportable* but not very *portable*. According to the specifications [1], a GSM Mobile Station can be:

- A vehicle mounted station;
- A transportable station;
- A handheld station;
- A vehicle mounted/transportable station;
- A vehicle mounted/handheld station.

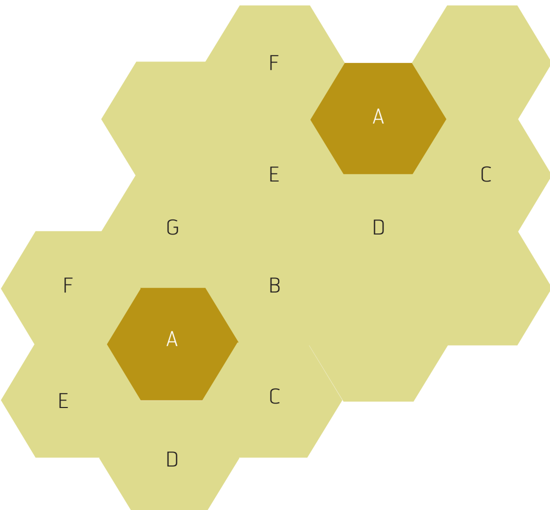


Figure 2 Frequency reuse in cellular networks

A handheld station was only one of the five alternatives. It was also considered to be the least realistic since it was difficult to incorporate so much functionality in a small and portable device.

2.3 The mobile phone is a cellular phone

Quite obviously, a mobile phone must be equipped with radio transmission and reception capabilities. However, a mobile phone does not communicate directly with another mobile phone as in the case of the walkie-talkie. It communicates via a cellular network. A mobile phone is also called cellular phone because it is connected to a cellular network.

A cellular network is a radio network made up of a number of radio cells (or just cells), each served by a fixed transceiver, normally known as a base station. The reason for having cells is to extend both the coverage area and the number of simultaneous calls.

According to the laws of physics, a radio frequency can be used at a location by one and only one radio transmission. Since the usable frequency spectrum is limited, the number of simultaneous transmissions is also limited.

A genius solution to the problem invented at the end of the 1940s was to divide the area into small cells and to reuse the same frequencies in these cells. It is worth noting that although a cell is shown in most books as a hexagonal, a cell is amorphous and there is quite often overlapping between neighbour cells. Due to signal leakage that can cause interference and deterioration of the signal quality, the same frequencies cannot be used in neighbour cells. In Figure 2, the cells denoted with different letters get allocated different frequencies. In order to avoid interference, cells of same type are placed away from each other.

The usage of multiple cells means that when the mobile phone moves from one cell to another, it will communicate with different base stations and change channels. If there is an ongoing call then it should not

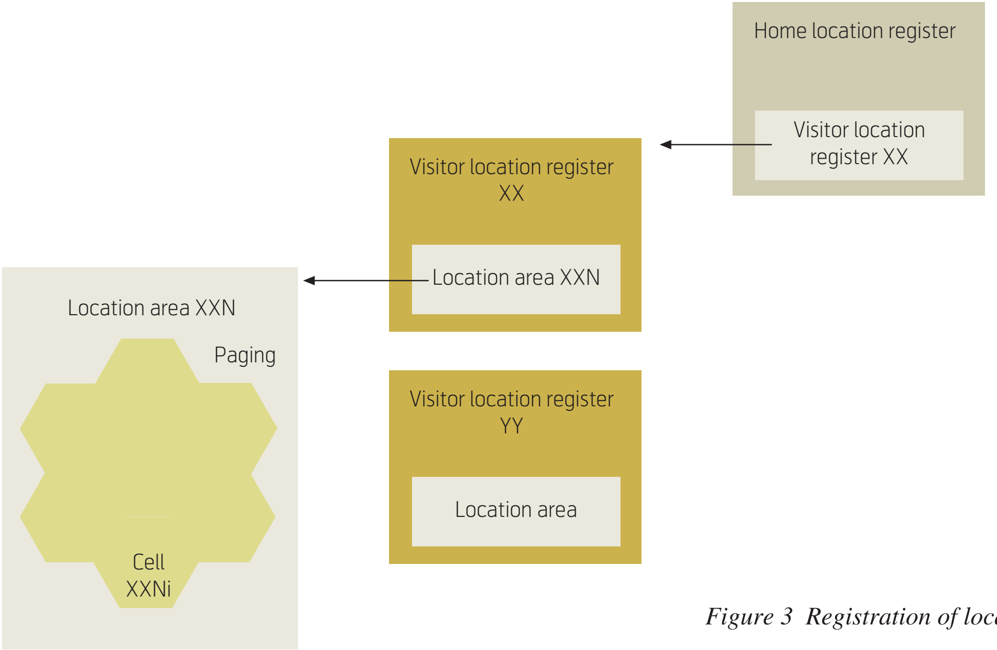


Figure 3 Registration of location in GSM

be interrupted. To prevent interrupt, it is necessary to carry out *handover*, also called *handoff*.

Handover requires naturally coordination between the mobile phone and the base stations. But first of all, the mobile phone must be "intelligent" enough to know which base station it is currently served by and which will take over. In GSM (Global system for mobile communications)[7], the mobile phone will detect that it is about to leave one cell and enter another. It will ask to get a channel reserved with the new base station while still maintaining the current channel. The signal from the former base station continues to deteriorate and when it falls below a threshold, the mobile phone will order the switching to the new channel and terminate the old channel. This is called *hard handover*.

In American system IS-95 and WCDMA the same frequency handovers, i.e. the mobile phone changes cell but uses the same frequency, both channels will actually be in use at the same time. This is called a *soft handover*. Handover is carried out not only when the mobile phone moves to another cell but also in load balancing or when signal deterioration occurs to avoid interrupt.

In addition to the wireless transmission and reception capabilities, the mobile phone must also be equipped with advanced functions to perform the *management of the radio link*.

2.4 The mobile phone is an intelligent device

As a mobile phone moves and changes cells or points of attachment to the network, it must have an identity of its own, independent of the network access point identity for the network to use to communicate with it. In GSM [7], every mobile phone has a unique identity called International Mobile Equipment Identity (IMEI) and stolen mobile phones can be internationally barred based on this identity.

Since the mobile phone is constantly changing position, when there is a call addressed to it, the mobile network must locate and deliver the call to the mobile phone. It can do this by broadcasting over the network a message "Mobile phone ID XX, please report yourself". When receiving the message, the addressed mobile phone ID XX will respond and a call can be set up. This procedure is called *Paging*. The paging may be time consuming when the mobile network is geographically large. Flooding the network with paging message is also a waste of bandwidth. The *registration of location* is hence introduced.

Two stages of registration are used in GSM to reduce the updating frequency. The Home Location Register (HLR) registers which Visitor Location Register (VLR) the mobile phone is currently visiting. The VLR registers which Location Area the mobile phone is located in. A Location Area (LA) or Routing Area (RA) in GPRS (General Packet Radio Service) [9] groups a number of cells and constitutes the smallest registration unit. An LA is used instead of a cell to reduce the number of location updates since the coverage area of an LA is larger than a cell. In addition, if the mobile phone is moving in high speed the registration on cell basis may contain obsolete information. Within an LA, paging is used to pinpoint which cell the mobile phone is located in at the current time. So in addition to the radio link management function,

ISSN 0085-7130 © Telenor ASA 2005

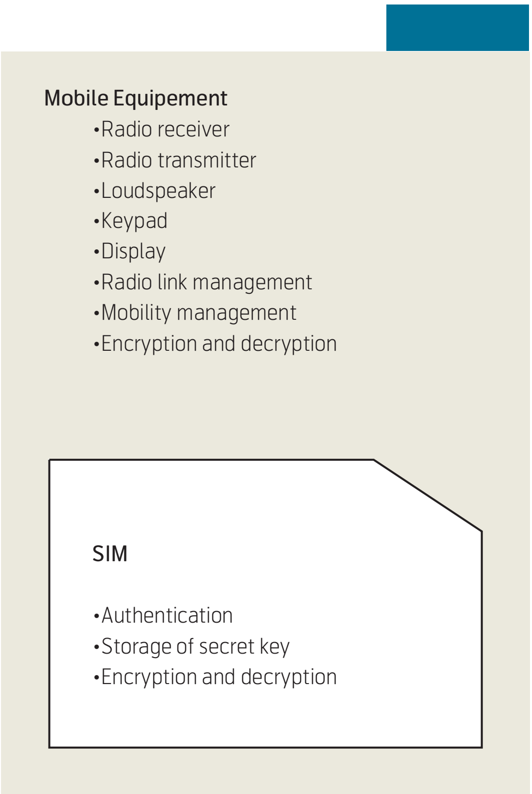


Figure 4 The mobile phone with its components and functions

the mobile phone needs to be equipped with the *mobility management function*. This again proves that wireless does not necessarily mean mobile.

2.5 The mobile phone is a composite device

As a mobile and wireless device the mobile phone is more exposed to fraud and abuse. It is hence necessary to have adequate protection. Strong symmetric authentication is used to authenticate the mobile phone and in 3G systems, network authentication is also carried out by the mobile phone to avoid communicating and delivering confidential information to the unauthorised network. Encryption of the communication channel is used to preserve the confidentiality and integrity. Most mobile phones are also equipped with locking and PIN code mechanisms for protection against unauthorised use of the phones.

In NMT (Nordic Mobile Telephone) and some American systems, the identity of the subscriber is assimilated to the identity of the mobile phone. With GSM the identity of the subscriber, the International Mobile Subscriber Identity (IMSI) is separated from the terminal identity, International Mobile Equipment Identity (IMEI). In order to provide personal mobility, the identity of the subscriber is moved from the mobile phone to a special component, called the Subscriber Identity Module (SIM). The SIM is actually a smart card having both processor and memory. It is a tamper resistant device which is used to store the secret key allocated to the subscriber and to perform the authentication, encryption and decryption functions.

The mobile phone is hence a composite device consisting of the mobile phone itself, also called Mobile Equipment, and the SIM.

Description of the SIM

Smart cards have had a long history for telecommunication network operators with the pre-paid phone card development. However, the most extensive usage started with the introduction of the SIM card, a specialised form of smart card in GSM.

A SIM card contains the subscriber identity module. It is normally removable from the mobile phone and comes in two sizes. Each SIM card has a unique identification number called IMSI (international mobile subscriber identity). However, SIM cards continue to evolve and host new applications with the increasing demand for convergent mobile data (Internet) and voice telephony. New applications exceeding the original purpose of user identification include on-line banking, m- or e-commerce, bill payment and trading, and further personalised usage of value-added services.

Authentication normally takes place when the MS is turned on and with each incoming call and outgoing call. A verification that the 'Ki' (secret key) stored in the authentication center (AuC) matches the 'Ki' stored in the SIM card of the MS completes this process.

The AuC is a protected database that holds a copy of the secret key stored in each subscriber's SIM card, which is used for authentication and encryption over the radio channel. The AuC provides additional security against fraud. It is normally located close to each HLR (Home Location Register) within a GSM core network.

The user must enter a PIN (Personal Identification Number) code on the handset in order to activate the SIM before this automatic procedure can start.

Two physical types of SIM are specified by ETSI/SMG and 3GPP. These are the "ID-1 SIM" and the "Plug-in SIM".

The corresponding Telecommunication Standards (GSM 11.11 "Specification of the Subscriber Identity Module – Mobile Equipment (SIM – ME) interface" or 3GPP TS 11.11 [2]) define the interface between the Subscriber Identity Module (SIM) and the Mobile Equipment (ME) for use during the network operation phase of GSM as well as those aspects of the internal organisation of the SIM which are related to the network operation phase. This is to ensure interoperability between a SIM and an ME independently of the respective manufacturers and operators.

The specification TS 11.11 defines
- The requirements for the physical characteristics of the SIM, the electrical signals and the transmission protocols;
- The model which shall be used as a basis for the design of the logical structure of the SIM;
- The security features;
- The interface functions;
- The commands;
- The contents of the files required for the GSM application;
- The application protocol.

Unless otherwise stated, references to GSM also apply to DCS 1800 (GSM at 1800 MHz) and PCS 1900 (GSM at 1900 MHz as usually operated in USA).

The physical characteristics of both types of SIM shall be in accordance with ISO 7816-1,2 ("Identification cards – Integrated circuit(s) cards with contacts, Part 1: Physical characteristics", Part 2: "Dimensions and locations of the contacts") unless otherwise specified. Format and layout of the ID-1 SIM shall be in accordance with ISO 7816-1,2, whereas Annexes A.1 and A.2 of ISO 7816-1 do not apply to the Plug-in SIM which has a width of 25 mm, a height of 15 mm, a thickness the same as an ID-1 SIM and a feature for orientation.

SIM Application toolkit

GSM 11.11 defines that the ME communicates to the SIM using a protocol specified in ISO/IEC 7816-3 ("Identification cards – Integrated circuit(s) cards with contacts, Part 3: Electronic signals and transmission protocols"). The ME is always the *"master"* and initiates commands to the SIM, and therefore there is no mechanism for the SIM to initiate a communication with the ME. This limits the possibility of introducing new SIM features requiring the support of the ME, as the ME needs to know in advance what actions it should take.

A new interface was thus defined which allows the SIM to act as master and the ME as a slave. The ME can transfer control to an application on the SIM device, and the SIM can issue a variety of commands through this mechanism, which cover text display on screen, operation of a dialogue between the SIM and the user, requiring either a single character or a more detailed user response, playing a tone on the ME, polling the 'idle' activity of status information exchange between ME and SIM, the exchange of local information between ME and SIM, a SIM initialisation and restart of a card session by resetting the SIM, provision of a list of items the user is expected to choose one from, sending a short message or an SS to the network, setting up a call, and incorporation of new features into the ME's menu structure.

SIM toolkit features can be classified into the following categories:

- Control of the man-machine interface;
- Communication services;
- Menu management and application control;
- Accessory management;
- Miscellaneous.

Standard GSM 11.14 "Specification of the SIM application toolkit for the Subscriber Identity Module – Mobile Equipment (SIM – ME) interface" or 3GPP TS 11.14 [3] define the interface between SIM and ME and mandatory ME procedure specifically for the "SIM Application Toolkit".

SIM Application Toolkit (SAT) is a set of commands and procedures for use during the network operation phase of GSM, in addition to those defined in GSM 11.11.

Specifying the interface is to ensure interoperability between a SIM and an ME independently of the respective manufacturers and operators.

GSM 11.14 defines the commands, the application protocol, and the requirements on SIM and ME for each of the SAT mechanisms like

- Profile Download;
- Proactive SIM;
- Data download to SIM;
- Menu selection;
- Call control by SIM;
- MO (Mobile Originated) Short Message control by SIM;
- Event download;
- Security;
- Multiple card.

SIM applications

In general, SIM services require a strong level of security (authentication and/or encryption) like

- Payment services;
- Banking services;
- Network access (WLAN-SIM);
- Any other information like sports updates, news etc. that require subscriptions;
- Typically operator controlled services; provided access to through an "operator button" on the handset.

SIM is originally an operator-controlled service environment in the mobile phone. SIM suppliers have custom built the SIM applications based on requirements from operators. However, lately the SIM has been opened up for application providers by adding open development environments based on Java Card and Dynamic SIM Toolkit [4]. The SIM has been transformed into a powerful platform to host multiple applications and enabled complete application lifecycle management. SIM applications can now be provisioned by using an Over-The-Air (OTA) approach combined with Java technologies. Although the operators still control which applications reside on the SIM, they can now focus on picking applications from a vast list of services and easily provision these to all their customers.

Figure 5 A candy bar mobile phone

Figure 6 A clamshell phone

Figure 7 A flip phone

Figure 8 A slide phone

USIM-Card

The document 3GPP TS 21.111: "USIM and IC Card Requirements" defines the requirements of the USIM (Universal Subscriber Identity Module [5]) and the IC card for 3G (UICC: Universal Integrated Circuit Card) [6]. These are derived from the general service and security requirements for 3G systems. The USIM is a 3G application on an IC card. It inter-operates with a 3G terminal and provides access to 3G services.

3 The mobile phone form factors

Mobile phones come in several different form factors, i.e. physical styles. While manufacturers are continually coming up with new types of designs, there are several common categories used on this site to describe form factors:

- *Bar:* (AKA candy-bar or block) This mobile phone type is the most basic style. The entire phone is one solid monolith, with no moving parts aside from the buttons and possibly antenna. To prevent keys from being pressed accidentally while the phone is in a pocket, etc., a "key guard" feature is usually provided, requiring a special key combination to "unlock" the keys.

- *Clamshell:* (AKA folder) This mobile phone type consists of two halves, connected by a hinge. The phone folds up when not in use. The top half usually contains the speaker and the display or battery, with the bottom half containing the remaining components.

  Most clamshell phones have a feature called Active Flip, which means that calls can be answered and ended by simply opening and closing the phone.

- *Flip:* This mobile phone type is a cross between the Bar and Clamshell types. Most of the components of the phone are in one part, but a thin "flip" part covers the keypad and/or display when not in use. The flip may be all-plastic, or it may contain one or two minor components such as a speaker or secondary keys. Most flip phones also feature Active Flip, as described above.

- *Slide:* This mobile phone type is usually designed similarly to a clamshell, with a large main display and speaker in one half, and the keypad and battery in the other half. But the two halves slide open instead of using a hinge. Slide designs allow the main display to be seen when closed, and are generally easier to open and close one-handed.

4 Evolution of the mobile phone

4.1 Evolution of the SIM

With the technology progress, the smart card has evolved from a single application card hosting a unique application to becoming a multi-application one. Multi-application smart cards are cards that can be used for more than one application at the same time. They may be seen as multiple cards jointly put into one. Multi-application smart cards have shown that they can deliver highly secure transactions and enforce true protection between applications held within the card itself. Their powerful encryption and digital signature capabilities are ideal for the emerging new technology sectors. Smart cards have built-in tamper resistant qualities, and PKI (public key infrastructure) incorporating digital signatures functionality embedded in their chip. These fundamental features contribute to creating a totally secure environment for applications owned by different parties.

As a smart card, the SIM card can evolve from a device containing the Subscriber Identity and the cryptographic functions necessary for the authentication to being a universal tamper resistant device that can host other applications such as:

- *Identification:* This application contains user personal information such as name, date of birth, home address, telephone number, etc. It may also contain photos of the user or fingerprints. It has the capability to display these contents upon eligible request.

- *Authentication and Authorisation:* This application makes use of the cryptographic functions of the smart card to offer authentication and authorisation of the user for other applications. These applications can be located on the SIM card, the mobile phone or even in other devices such as a laptop. There exist currently several authentication applications such as SIM authentication, One-time password authentication.

- *Personalisation:* This application enables the user personalisation of services and applications, i.e. to allow the user to customise his services regarding both functionality and presentation. Basically, it exploits the safe storage capability of smart cards to save the user 's preferences and settings for other applications and offers interfaces to retrieve, modify and store them.

- *Financial transaction:* This application allows carrying out financial transactions, e.g. transferring money from one account to another. Thanks to the strong authentication available on the smart card, this application enables the user to log on to finan-

Mobile Commerce

Briefly, Mobile Commerce can be defined as commerce done using mobile devices, especially mobile phones. As in traditional commerce, mobile commerce consists of the following phases:

- goods examination
- selection
- payment
- delivery.

An example of a total solution for mobile commerce is Telenor's SmartPay. The commerce phases, i.e. goods examination, selection, payment and delivery, can be carried out separately to increase flexibility. The user can search information and goods independently of payment. They may use any channel for goods searching, e.g. WAP, SMS, Web, fax, phone calls, etc. The purchase is carried out on the mobile phone using either SIM Application Toolkit or SMS depending on the security level that the user has chosen.

As shown in Figure 9 the purchase process is initiated by the shop after the user has selected the goods. The purchase transaction involves the shop, the Payment server connected to the financial institutions (bank, credit card companies) and the secured access control to the mobile phone.

The SIM card is a multi-application smartcard, which besides the Subscriber Identity contains the SIM Application Toolkit and the PKI (Public Key Infrastructure) necessary for the authentication and authorisation of the user and digital signing.

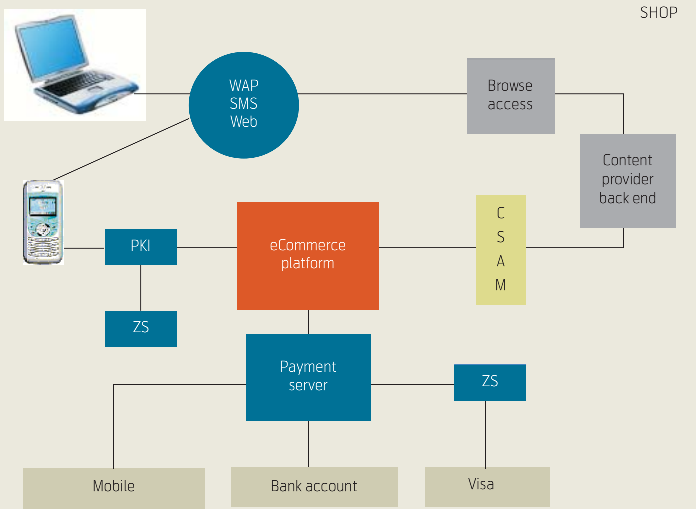


Figure 9 The SmartPay System architecture

To use SmartPay the user has to activate the PKI application on the SIM card and download a SIM toolkit application over the air (OTA) by sending an SMS to a number such as 2500. The registration of the user data, e.g. bank account numbers can be entered on the phone, via the Internet for customers with net bank accounts, or at the bank or post office.

ISSN 0085-7130 © Telenor ASA 2005

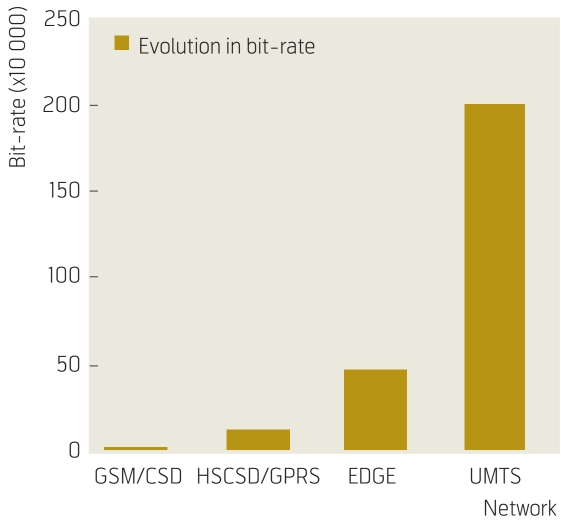


Figure 10 Evolution in transfer rate

cial institutions to perform financial transactions using a digital signature.

- *Payment:* This application, thanks to the digital signing functions on the smart card, allows the user to carry out payment functions towards banks or credit card companies. It may offer both micro payment (e.g. less than 100 Euros) or macro payment (e.g. over 100 Euros).

- *Loyalty:* This application can be owned and administered by a third party. It registers the bonus points that the user has accumulated through his service usage. It may be equipped with the ability to display the balance of points to the user.

- *Health care:* This application holds the health information of the user. As a tamper resistant device the smart card is a secure storage for confidential personal information. This is only possible thanks to the tamper resistant characteristics of the smart card. Since the mobile user is carrying the mobile phone with him most of the time, his health information will always be available when he needs them.

- *M-commerce:* This application enables performing M-commerce and it will be described more thoroughly in the box below.

4.2 Evolution in bandwidth and bit rate

For service platforms on mobile phones to be of great value, the mobile phones must allow services to access adequate levels of resources both on the device, e.g. in terms of processing power and storage, but also in terms of network bandwidth.

When mobile data services were introduced through Circuit Switched Data (CSD) (i.e. for dial-up services), the maximum theoretical network throughput achievable by a handset was 14,400 b/s. This was due to the underlying mechanisms for sharing the radio spectrum (FDMA and TDMA in GSM) and the algorithms for encoding/decoding information on the wireless bearer.

High-Speed Circuit Switched Data (HSCSD) increased the theoretical maximum throughput for each user to 115.2 kb/s by combining up to 8 timeslots for a user (where each timeslot provides 14,400 b/s).

Enhanced Data Rates for GSM Evolution (EDGE) is an update to the air-interface in GSM networks, and by using other schemas for modulation onto the wireless carrier, the efficiency of the radio spectrum is increased. This can offer up to 69.2 kb/s per timeslot, for a total theoretical rate of 473.6 kb/s. The actual rate depends among other things on the error-correction schema employed.

UMTS brings along promises for further increased throughput rates. For UMTS, a completely new radio interface is introduced, and the theoretical maximum throughput rate at high mobility is 384 kb/s, the maximum rate in stationary situations is 2 Mb/s.

4.2.1 Multiple Air Interfaces

In parallell to the development of the mobile telecommunication standards, other short range wireless standards are also steadily being improved. The future mobile terminals will include not only one wireless interface, but at least three; cellular network interface (e.g. UMTS), medium range wireless interface (e.g. WLAN, Bluetooth) and short range wireless interface (e.g. RFID). Another option is to include medium range low rate interfaces like the ones specified by ZigBee.

4.2.2 Bluephone

The Bluephone project by British Telecom (BT) allows a normal cellular phone to connect, place and receive calls through a Bluetooth basestation when inside a house or in the office. Since the project was announced, the WiFi[1)] standards have evolved, and BT is now aiming at using WiFi instead of Bluetooth in upcoming versions of Bluephone. However, Bluephone does not provide Voice over IP service, so it is not part of the ongoing convergence. Calls are still transferred over the GSM network.

1) *WiFi (sometimes written Wi-fi, Wi-Fi, Wifi, wifi; and an acronym for "wireless fidelity") is a trademark for sets of product compatibility standards for wireless local area networks (WLANs). WiFi is quite often used interchangeably with WLAN.*

ISSN 0085-7130 © Telenor ASA 2005

It is highly likely that many of the future cellular phones will include a WiFi interface, and due to the increased processing and memory capabilities of these phones it is likely that Voice-over-IP will become a common way of placing voice calls from cellular phones when within range of WiFi base stations.

4.2.3 WLAN and UMA phone

With the rising popularity of WLAN and the increasing number of WLAN hotspots, it is not surprising to see the appearance of mobile phones with WLAN. The biggest hindrance of WLAN until now has been the high power consumption which is due to the fact that with WLAN scanning is done continuously over the whole channel. This issue is partly solved by the introduction of Ultra low power WLAN chip that has higher integration, less component and most importantly can have sleep operation mode.

The WLAN equipped with a SIP client becomes hence a wireless IP phone that can operate within the coverage of an access point or a hotspot. Figure 11 shows a wireless IP phone, WIP 5000 from Unidata, a Korean vendor. Soon the IMS phones, i.e. phones equipped with the 3GPP IP multimedia system API, will appear and offer VoIP, push-to-talk, instant messaging services over the cellular network.

Philips went a step further to promote UMA (Unlicensed Mobile Access) into mobile phones. UMA Technology provides an alternative access to GSM and GPRS core network services via IP-based broadband connections. In order to deliver a seamless user experience, the specifications define a new network element and associated protocols that provide for the secure transport of GSM/GPRS signaling and user traffic over IP. The caller can automatically switch from GSM to IP-based networks and back again without interruptions. UMA also helps to make better use of the scarce GSM spectrum: as soon as a phone is in reach of an UMA access point, it will leave the GSM spectrum and continue its call over WiFi, leaving the GSM spectrum for those callers that really need it.

Figure 11 A wireless IP phone

4.3 Evolution in functionality and services

There has been a continuous evolution in the functionality and services provided by cellular phones and networks since the introduction of Nordic Mobile Telephony (NMT). Initially, the goal was to provide mobile telephony with roaming opportunities across the borders of the Nordic countries.

Voice and supplementary services	Personal SMS messaging	Premium SMS services	Standalone Java services	Distributed Java services	?
1993	1995	1998	2002	2005	2010

Figure 12 Timeline displaying development of the major service enablers

A timeline approximately illustrating the development and emergence of new services and service enablers for mobile phones in GSM is shown in Figure 12.

With the introduction of GSM, supplementary services were added. Examples of these services are call-forwarding, call-barring, caller-id, voice-mail etc. The Short Message Service (SMS), as well as the parallell development of the Internet, helped spawn the work with support for more advanced data services on cellular phones.

The first generation of data services on cellular phones (disregarding services used through dial-up Internet with CSD) were based on SMS, and this category of services is still the most successful.

The second generation of data services were introduced with the Wireless Application Protocol (WAP). WAP was branded the Internet/WWW for mobile, handheld devices. However, and probably due to its overexposure as the Internet-equivalent for mobile, WAP has not been as great a success as SMS. There can be several reasons for this:

- From a developer and service provider viewpoint: In the beginning, Web-pages had to be completely re-designed to fit WAP and with the form-factor of WAP-enabled devices the technology was not in place to ease this process;
- From a user viewpoint: Browsing complex Internet pages is too cumbersome on small devices. Devices have just recently got displays that can render visually good pages;
- A lot of services provided by WAP are more easily provided and consumed by SMS; e.g. simple look-up and directory services (e.g. a public phone directory); easier to use and easier to develop.

The third generation of data services has in the last years been emerging, and these are enabled by Java 2 Micro Edition (J2ME). This generation of services are mostly realised by standalone software on the

ISSN 0085-7130 © Telenor ASA 2005

handsets, e.g. games, calendars, Personal Information Managers (PIM) and similar. However, as the specifications of the J2ME platforms are maturing, support for more advanced functionality is being added to the devices. For example, some new cellular phones allow J2ME applications to take control over camera functionality, Bluetooth functionality and the file system on the device. In addition, some devices now include support for higher-layer communication protocols like SOAP and SIP. The former will propel the development of services realised by distributed systems, as opposed to the early services that were realised by standalone applications. As the 3rd Generation (3G) mobile telecommunication system has been launched, the feasibility of providing highly distributed data services has been further increased.

In addition to the three generations of data services described above, cellular phones increasingly include advanced functionality like mp3-players, video playback and productive applications like Office applications.

4.4 Evolution from black box to open device

As a terminal, the mobile phone is originally a black box, i.e. its functionality is defined at fabrication time and only the manufacturer fully knows about its internal structure. As a telephone, the mobile phone should be able to support telephony service. All other services and application are considered as extra offered by the terminal manufacturers. A black box device is usually tested carefully and is both simple and reliable.

The biggest drawback is the lack of flexibility compared to a computer where the user can purchase separately the hardware and software components. One can argue whether the mobile phone needs to be an open computer, but there is obviously a demand for functionality upgrading and customisation after fabrication. Indeed, although still being in minority, the smartphones, i.e. mobile phones with an operating system and supporting the installation of applications by the operator or the user, are increasing rapidly in number. Let us have a brief overview on operating systems of smartphones.

4.4.1 Mobile phone operating systems

- Symbian

Symbian OS is an Operating system designed for wireless information devices such as communicators, smartphones and handheld computers. It is developed, licensed and supported by Symbian Ltd. (www.symbian.com). Newer versions (6.0 and later) are called Symbian OS or Symbian platform, but EPOC is still a widely used name.

The first public release was made by Psion in April 1997, when it emerged in the Psion Series 5. Psion Software was de-merged from the Psion Group and became Symbian Ltd, formed as a joint initiative between Ericsson, Nokia, Motorola and Psion. Lately, Nokia holds the majority of the shares and this makes Symbian less attractive to other manufacturers.

- BREW

BREW (Binary Runtime Environment for Wireless) is a wireless-application-development platform for phones that use the Qualcomm-developed code-division multiple access (CDMA) cellular technology. In the future, BREW may run on PDAs and also work on other popular cellular technologies, including time-division multiple access (TDMA) and the global system for mobile communication (GSM).

Developers can create games and other applications that work on any BREW-enabled phone, regardless of the device's operating system.

On the phone itself, BREW is a thin client (about 150 K) that sits between a software application and the Application Specific Integrated Circuit (ASIC) level software. Thus, developers can write to BREW without knowing or caring about the device's chipset or air interface. The second major component of BREW is the BREW Distribution System (BDS). The BDS is a wireless data services delivery and billing environment. Leveraging an open and extensible technology, the BDS enables network operators to rapidly deploy wireless data services to subscribers network-wide via mobile devices.

- Microsoft Windows Mobile

Microsoft has developed its own operating system named Windows CE for the mobile devices, which has a look and feel of traditional windows GUI (Graphical User Interface). The usage of Windows CE covers a broad range of devices like handheld PCs, high-speed data acquisition devices, Internet access devices, web terminals, industrial automation, retail and point of sale devices.

While both Pocket PC 2000/2002 (Phone Edition) and SmartPhone 2002 are based on Windows CE 3.0, which is a modular, embedded real time operating system for mobile 32-bit devices, they address Personal Digital Assistants (PDAs) and smart phones, respectively. However, the latest version released in April 2005, called Windows Mobile 5.0, supports both smart phones and PDAs.

- SavaJe OS

SavaJe OS developed by SavaJe Technologies, is a richly-featured, high performance, Java-based operating system for mobile multimedia handsets, delivered with a complete set of easy-to-use applications, and a fully-customizable user experience layer. It uses J2ME CDC (Connected Device Configuration) instead of CLDC (Connection Limited Device Configuration). SavaJe OS is available pre-ported to a number of common wireless chipsets. The first reference implementations are available on TI OMAP and Intel processor families, but any ARM9-based CPU can also be supported.

- Palm OS

The Palm operating system is a proprietary operating system from Palm Inc (in fact a subsidiary of 3Com Corp). The Palm Inc provided the Palm OS for its handhelds and licensed it to other manufacturers. Recent models of Palm and compatible devices include a colour display and an interface for accessories, which range from a fold-in keyboard to a mobile phone. Palm devices are functionally simple devices that concentrate on the user interface and experience. The available memory on them is usually severely limited, so there is a strong trend of minimalism in Palm applications. Palm Computing provides a software development kit, but it requires a separate C compiler or an integrated development environment, such as Metrowerk's CodeWarrior. This section will introduce the main features of PalmOS, the platform availability, the specific platform capabilities and the existing development environments.

- Embedded Linux

Embedded Linux is a feature-rich subset of Linux, stripped down to include only those elements of Linux that are useful on an embedded system. For example, embedded systems typically use FLASH memory in place of a hard drive to store the file system. Embedded Linux systems transparently support the Linux file system. Linux applications do not know that FLASH memory is in use and your products have the added benefits of intelligent FLASH memory management.

Embedded Linux versions are also mostly stripped of the overhead of a graphical interface, embedded Linux boots in just a few seconds and fits into a few megabytes of FLASH memory. In addition, Linux is known to be stable and reliable, and it is based on open-source. This has had the result that several different vendors have developed their own embedded operating system based on Linux. As an attempt to standardize an embedded Linux platform, the Embedded Linux Consortium (ELC) released a specification of such a platform in December 2002.

a. MontaVista

MontaVista is a developer of embedded Linux. Until now the most used version is called the Linux Consumer Electronics Edition (CEE) 3.0[2]. It is the first Linux operating system and cross-development environment specifically designed for consumer electronics applications such as mobile phones, digital televisions, set-top boxes and automotive telematics. CEE 3.0 is based on the 2.4.20 Linux kernel. The most recent embedded Linux version from MontaVista is called MontaVista Linux for Mobile Devices (MOBILINUX). This version is based on the Linux 2.6 kernel, and it includes X Windows support through Tiny X (an Xfree86 implementation), graphical widgets through GTK (Gimp Toolkit) and two different WindowManagers.

b. Trolltech Qt/Embedded

Trolltech provides the Qt/Embedded which is an application development framework targeted at embedded devices. It is currently used to develop numerous types of products and devices on the market, ranging from consumer electronics (mobile phones, web pads and set top boxes) to industrial controls (medical imaging devices, kiosks, mobile information systems and others).

Qtopia Phone Edition builds on Qt/Embedded and is an application platform and user interface specifically targeted at PDAs and mobile phones. This means that existing Qt/Embedded applications easily can be plugged into Qtopia-based devices with minimal effort. This has resulted in the availability of an unlimited number of applications for Qtopia-based devices.

c. OpenWave

OpenWave is the provider of different content delivery solutions for mobile phones. The solutions can be delivered integrated in what is called the OpenWave Phone Suite Version 7, or they can be delivered as standalone components (instant messaging client, mobile browser, SMS/EMS/MMS messaging clients,

---

2) *http://www.mvista.com/cee/index.html*

mobile e-mail etc.). All these products are compatible with any third party Java Virtual Machine (JVM), which means that it can be integrated with Embedded Linux distributions like MontaVista.

d. LynuxWorks BlueCat

LynuxWorks provides an embedded Linux distribution called BlueCat, which is based on the Linux kernel version 2.6. This distribution also supports Qtopia and Qt/Embedded from Trolltech.

4.4.2 Distributed Computing middleware

To pave the way for innovative applications on the mobile terminal, portability must be ensured. Indeed, both the cost and the time-to-market will be considerably reduced if applications are written once and run on every phone. Seen from the user, it is also desirable to have access to the same applications on different mobile phones.

Figure 13a shows the original architecture of a computer with three layers: Hardware, Operating System and Applications. The computer is a general-purpose machine in the sense that its functionality is determined by the applications installed on it. The applications running on a computer are bound to this computer in two ways:

- The source code is bound to the operating system through the usage of the operating system API.
- The executable code is bound to the hardware, especially the processor and the display.

An application (executable code) is hence not portable, i.e. it cannot be moved and run in another computer having different hardware and operating system. For a computer with different hardware a recompilation of the application is required. For a computer with different operating system, code modification is necessary before recompilation. For example, an application written in C/C++ on Symbian v6.0 may need some modifications to run on Symbian v7.0. Recompilation is required to move and run a C++ application from one phone to another.

In order to make the applications portable, a middleware is introduced between the applications and the operating system. Distributed Computing middleware such as CORBA removes the application binding to the operating system but recompilation is still needed. By Interpretation middleware in Figure 13b we refer to middleware that also offers interpretation and conversion of the intermediate code to machine code at execution. At compilation time the program code is compiled onto intermediate code and not onto machine code. An example of such interpretation middleware is the JAVA Virtual Machine (JVM). In fact, the application code is executing on a virtual machine and is not aware of the hardware machine underneath. It is worth noting that the API offered by such an Integration Middleware must be open and standardised to ensure portability. The telecom and security APIs must also be standardised.

It is desirable to have Interpretation Middleware in the mobile phone in order to ensure Application Portability. It is also crucial that APIs for telecom functionalities are provided for this Interpretation Middleware layer.

4.4.3 XML Web Services support

In Figure 13c an additional layer called Web Service Framework has been added on top of the Interpretation layer. This layer will enable the execution of XML Web Services.

A Web service is application logic that is made available on the Internet using the following standards: XML for data description, SOAP for message wrapping, WSDL for service description and UDDI for service discovery.

This layer is particularly important for mobile devices that can only host thin XML Web Service clients due to the limitations in processing, storage and battery life. The thin clients can collaborate with thicker XML Web Services on the network to offer valuable services to the user. Furthermore, a Web Service run on a Web server can be accessed by various clients installed in different mobile devices as shown in Figure 14. The Web services exposed legacy applications that can be implemented in proprietary technologies.

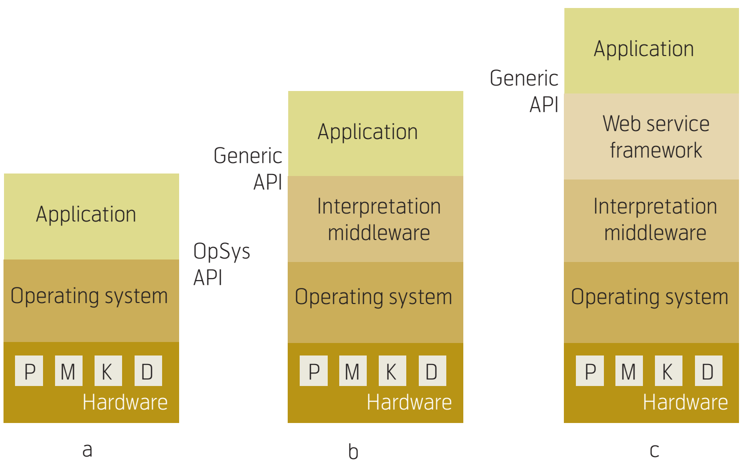


Figure 13 Evolution in the computer architecture

It is desirable that XML Web Service Framework is supported in the mobile phone.

4.4.4 Web services layer

As mentioned earlier, data services for mobile phones will increasingly be realised by distributed systems. In the early days of J2ME, proprietary protocols had to be devised on top of HTTP when building distributed services. This is no longer necessary to create this type of services. The introduction of XML Web Services support for mobile phones, first by kSOAP and then more recently by standardization through the Java Community Process (JCP) as Java Specification Request (JSR) 172, allows developers to more easily implement communication between the client side software and server side software. JSR172 defines a subset of the standard XML Web Services protocols. Some functionality, like support for binary attachments, is not supported, but the standard simplifies developing distributed systems where a mobile terminal should act like a client. Much like CORBA, it is possible to generate client stubs (as well as skeletons for the server side) based on a description of the service interfaces. These interfaces are specified using the Web Services Description Language (WSDL), and should include all information necessary to interact with a Web Service. Although JSR172 is an optional package to the J2ME platform, some recent mobile phones come fully equipped with this library from the manufacturer (e.g. Nokia N91). For testing it is also possible to use the wireless toolkit provided by Sun which includes a reference implementation of the JSR172 specification.

Server side components can be developed with standard Java or .NET tools and deployed in standard application servers like the Jakarta Tomcat with Axis for handling the XML Web Service layer.

The major benefits of XML Web Services for handheld devices are:

- *Standardized and open interfaces.* This can allow a multitude of developers to develop their own clients that communicate with the same server side components. Services are exposed using the Web Services Description Language, which is a complete description of the service interface in XML.

- *Ease of implementation.* Tools exist that ease the development efforts of the client-server communication mechanisms. Stubs can easily be generated for clients, and for a developer, remote method calls can look exactly like local calls (similar to RMI and CORBA). This allows the developer to focus on higher-layer tasks, like the value-added service that should be provided to the user whilst leaving protocol details to the system.

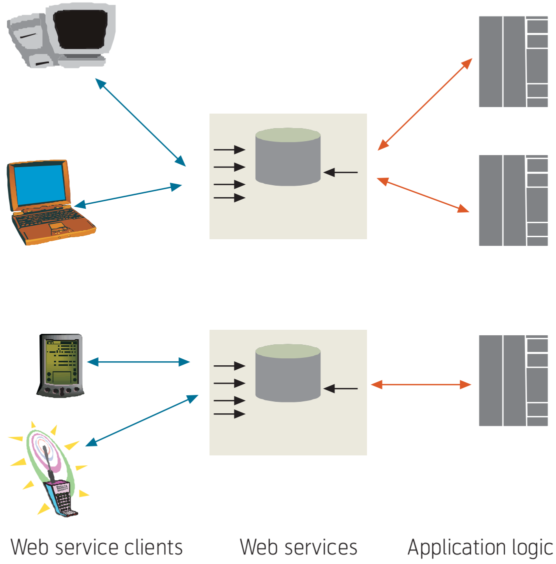


Figure 14 Ubiquitous access to applications via heterogeneous devices

The major drawbacks of XML Web Services for handheld devices are:

- *Overhead in the communication.* Using XML for encoding data yields some overhead in the communication, which is problematic with today's limited capacity on the wireless bearer.

- *Overhead in the processing.* Parsing and processing XML also yields some overhead, which can be crucial for limited capacity handheld devices.

- *Currently no support for attachments.* The mobile equivalent to XML Web Services, i.e. JSR172, does not currently support attachments. This can be a limitation for certain services requiring transport of binary data. The solution is to encode the binary data with the base64-algorithm before transferring it.

4.4.5 Operator's customisation

The current situation where the mobile phone's functions and presentation are decided by the manufacturers is not very satisfactory to the operators, who want to be able to differentiate in their service offering at the same time as the number of service platforms is low. There are currently so many different mobile terminal platforms which are completely different and they pose big problems to the wireless operators.

Further, operators wish to be able to offer customisation of handset in order to differentiate in their ser-

Figure 15 Telenor Nordic Mobile customisation with Entry

vice offering to reduce customer churn and to improve loyalty. Figure 15 shows an example of operator's customisation.

The Open Mobile Terminal Platform, OMTP Ltd. ("OMTP") has been established to gather the needs of the mobile operator community for mobile terminal platforms and to define and agree on those requirements necessary to deliver openly available standardised application interfaces to provide customers with an improved user experience across different devices. This will also enable individual operators and manufacturers to more easily customise and differentiate their products and services.

The founding members of OMTP are mmO2, NTT DoCoMo, Orange, SMART Communications, Telefónica Móviles, TIM (Telecom Italia Mobile), T-Mobile and Vodafone. The goal of the Open Mobile Terminal Platform (OMTP) is to achieve the openness of the mobile terminal in the same way as the computer experienced in the 70s, where the computer is a black box that the customer has to buy as a whole.

OMTP Open Mobile Terminal Platform

4.4.6 User's personalisation

Parallel with the operator's demand for customisation, there is a user demand for personalisation of their handset. Indeed, the mobile phone is no more only a communication terminal bit but a personal accessory that the owner uses to express himself. The personalisation of a handset can include the following adaptation:

- Ringtones
- Logos
- Wallpapers
- Celebrity Voice-mail Messages
- Handset facias
- Personalised Location Based Services (e.g. targeted navigational and advertising services)
- Themes that include icons, wallpapers, backgrounds, colors, etc.

There are also websites like MyFoneThemes.com that offer a service for converting any image or photo of the user (examples: own photo, a loved one's photo, own artwork, a favorite celebrity's picture) into a phone theme for his mobile phone. They produce personalised phone themes on demand according to the submitted photos and specifications.

At first glance, the user's personalisation may appear to be in conflict with the operator's customisation but it does not have to be. In fact, a balanced combination of operator's customisation and user's personalisation will be the best for both parties. An example of such combination is the FT (Financial Times) personalised mobile phone [10], which allows the user to select "My industry News", "My Share Quotes", etc. from the Financial Times content source.

Figure 16 Example of mobile phone's theme

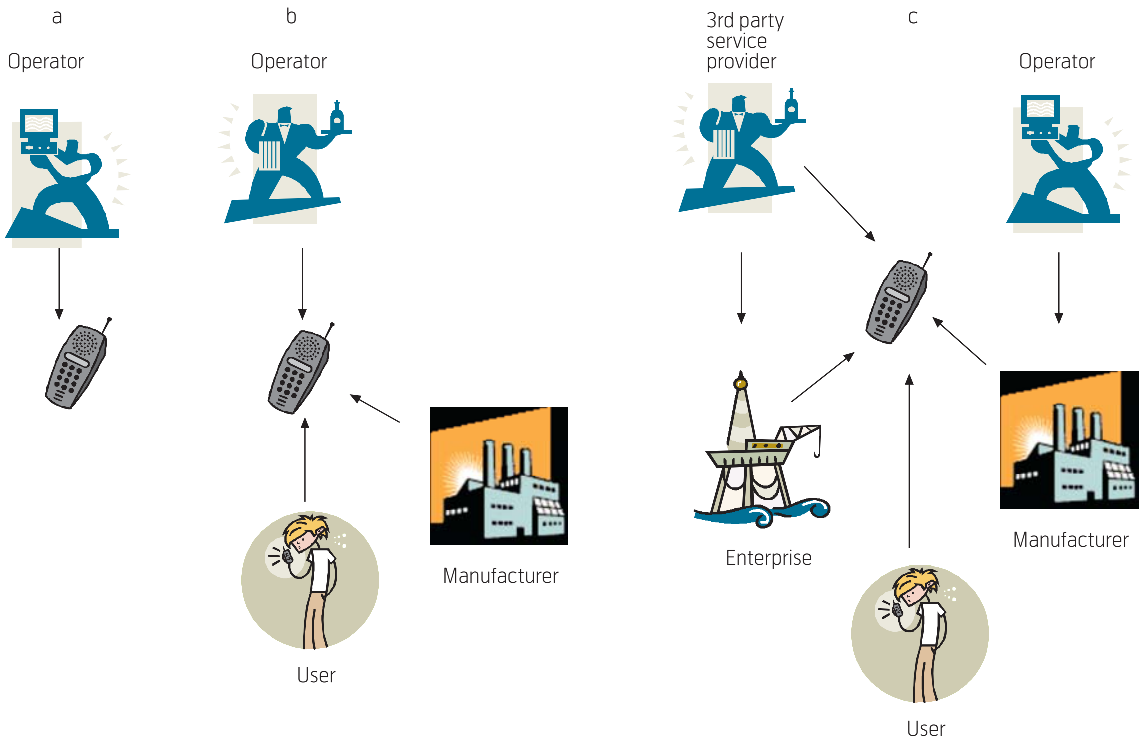


Figure 17 The ownership models

4.5 Evolution in ownership model

In the early NMT system and also now in the American and Japanese mobile telecommunication systems, the subscription includes the mobile phone which is the property of the operator lent to the subscriber during the subscription time. The operator in this case (see Figure 17a) has the total control of the mobile phone and can, in principle, decide both the functionality and the presentation of the mobile phone.

In the GSM system, with the introduction of the SIM card, the ownership of the mobile phone is transferred to the user. The manufacturers then had the freedom to decide and compete on the functionality, form factor, look and feel of mobile phones.

The evolution does not stop there, because the mobile phone is not a stand-alone device but rather a communicating one. First, there is the emergence of the operator's services located on the operator's domain and accessed through a thin client, which could be a browser, on the mobile phone. In order to function properly, these services might require proper configuration and customisation from the operator. This leads to the operator's claim to control the mobile phone.

Second, there are more and more services that are offered by third party service providers such as banks, credit card companies, online game, webshops, etc. that may or may not have an agreement or alliance with the mobile operator. To ensure that their services work well, they are also interested in getting some control of the mobile phone.

Last but not least, the enterprise, i.e. employer of the user that pays for both the mobile phone and the mobile subscription, naturally wants to be able to control the functionality and presentation of the mobile phone. Indeed, when the mobile phone is getting advanced and allows the users to access emails, documents, information, etc, that are located within the enterprise's intranet, it is crucial that the enterprise has the right to control.

One way to solve this conflict situation is to design a mobile phone with five discrete execution environ-

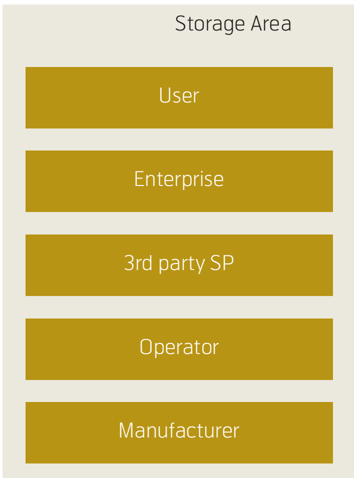


Figure 18 Separate storage area for each player

ISSN 0085-7130 © Telenor ASA 2005

ments for the five parties or at least, five separate storage areas (EEPROM, Flash memory or memory card) as shown in Figure 18.

This solution does not, however, prescribe any hierarchy telling what actor should have higher priority than the others. In fact, such a hierarchy should be changed according to the different situations as follows:

- The mobile phone is purchased and totally paid for by the user: It is then reasonable that the user should have the highest right to decide whatever he wants. However, it might be necessary for fundamental functions such as making and receiving calls, sending and receiving SMS, etc. to be locked such that the mobile phone can still operate as a phone even when being attacked.

- The mobile phone is purchased by the user but subsidized by the operator: In this case the operator will have the right to customise the mobile phone as it wants. However, a balanced combination between operator's customisation and user's personalisation will be the best solution.

- The mobile phone is purchased and totally paid for by the enterprise: In this case, the enterprise might want to have branding on the wallpaper, logos, etc. and to install clients for communication with enterprise applications. Here again, some level of user personalisation can be a good thing to have.

Figure 19 Mobile phone with camera

5 Mobile phone trends

With the advance in microelectronics, the mobile phone had the opportunity to evolve from a communication device, i.e. a mobile terminal to be a mobile computing device, e.g. a handheld computer or a Personal Digital Assistant (PDA). In fact, the smartphones have invaded and conquered the market of PDAs. The evolution of mobile phones will not stop there and will continue to expand and threaten other devices that we are going to discuss successively.

5.1 Mobile phone as low end camera

Mobile phones with digital cameras appear in the market both in large numbers and in varied quality. Quite affordable mobile phones of the middle segment use CMOS matrixes and the quality of photos is quite good under conditions of a sunny day, yet indoors a noise appears as photosensitivity of the matrixes is slender. Siemens S65 is an example of such photos.

Currently, telecom manufacturers strive to produce matrixes of a higher quality and position these products as a partial substitute for digital cameras. One can dispute whether the quality of matrixes is good, but production of hybrid devices is in full swing. The main task for such devices is just providing an acceptable quality of photos. It will not be long, probably beginning of the autumn 2006, before the quality of cameras will be comparable to current low-end digital cameras. This will be attractive to many users since they will not need a separate digital camera and can still take pictures of acceptable quality.

The main improvements of cameras mounted in mobile phones are as follows:

- Introduction of the *macromode* that allows a digital camera to photograph close-up pictures of small objects like flowers, insects, coins, etc. Macro mode potentiality varies from camera to camera but almost all digital cameras offer this exclusive feature to enable the users to widen up their imagination and their weirdest fantasies in creating art.

- Appearance of *autofocus,*

- Offering *resolution of 2 megapixels* or higher,

- Unifying the camera interface to make them as similar as possible to the digital camera.

---

3) *Flash memory (sometimes called "flash RAM") is a type of constantly-powered nonvolatile memory that can be erased and reprogrammed in units of memory called blocks. It is a variation of electrically erasable programmable read-only memory (EEPROM) which, unlike flash memory, is erased and rewritten at the byte level, which is slower than flash memory updating.*

ISSN 0085-7130 © Telenor ASA 2005

Figure 20 A memory stick

5.2 Mobile phone as portable storage

Mobile phones contain more and more memory. On the advanced ones, there is from 4 Mbytes up to 80 Mbytes Flash memory[3]. In addition, memory cards can also be installed in the mobile phone to considerably increase the storage capacity up to Gigabyte level. There are many card types such as MS Duo promoted by Sony Ericsson, RS-MMC by Nokia, MMC/SD by Japanese manufacturers. The mobile phone will then be an *ideal portable storage device*.

The mobile phone will of course not threaten the USB flash drive also called USB drive, flash drive, keychain drive, or disk-on-key since, with their quite affordable prices they have become consumables like diskettes that are used in documents and data transfer between people. However, a mobile phone equipped with a USB cable to connect to a computer will become the user's private and portable information container that always stays with him.

5.3 Mobile phone as MP3 player

Equipped with Flash memory, the mobile phone can easily be turned into an MP3 player. Quite soon, the quality of the players in the mobile phones will be comparable to the standalone MP3 player. At this early phase the prices are still high but they will soon be competitive and the mobile phone will really become a threat to the MP3 player. The typical example of mobile phone in this class is the Sony Ericsson W800 that is promoted with the trademark of Walkman which is famous in the market of portable audio equipment.

5.4 Universal security token

The mobile phone consisting of the phone itself and the SIM card constitutes both a smard card and a smart card reader that can be exploited to provide security services like authentication, encryption/decryption and digital signing to all the user's applications. These applications can also be located on other computers or devices that communicate with the mobile phone via infrared or Bluetooth.

The SIM card on the mobile phone will perform the security function and return the result to the mobile phone that in turn passes it to the application using either infrared or Bluetooth.

Figure 21 A mobile phone as MP3 player

For authentication, the SIM card can offer three authentication methods:

- Using the GSM network authentication
  - A3/A8 algorithms and radius server

- Using the Short Message Service (SMS) security 3GPP TS23.048

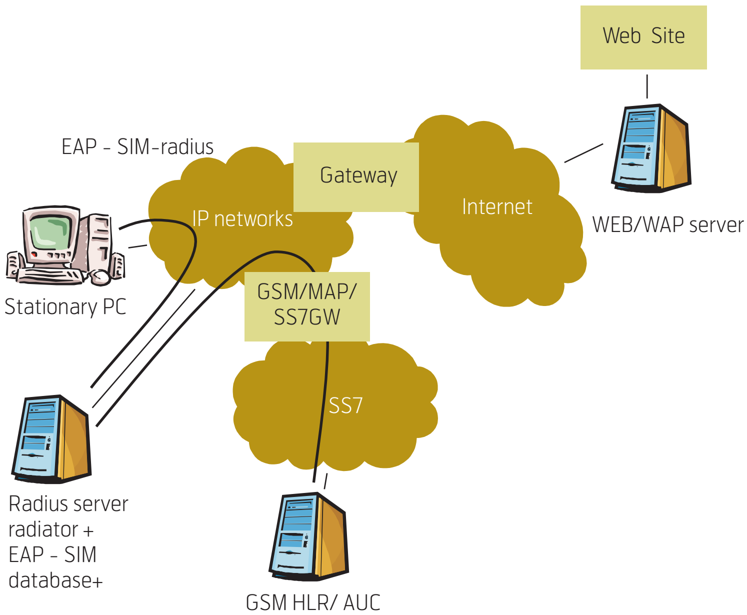


Figure 22 Using SIM in application authentication

ISSN 0085-7130 © Telenor ASA 2005

Figure 23 A remote control for mobile phone

  - Encryption and Signature/Checksum/Redundancy keys
  - Over The Air (OTA) server

- Using SIM Toolkit One-time password applets
  - Dedicated symmetric key and HOTP applet
  - HOTP authentication server

5.5 Universal remote control

A mobile phone may act as remote control for almost everything from TV, stereo music system, door, electricity switch, etc. since it is equipped with short-range wireless communication capabilities such as infrared, Bluetooth, etc. and a keypad for issuing command or eventually entering the required pin code.

There are currently numerous remote control applications that can be downloaded and installed in the mobile phone. An example of such applications is for example the Psiloc that can be used in a range of terminals like Sony Ericsson P900.

5.6 Payment device

As mentioned earlier, the mobile phone can be a quite valuable payment device allowing both micro- and macro payment thanks to the strong security functions offered by the SIM card. However, it is worth noting that the mobile phone is used to pay only for goods purchase electronically and not for merchandise in the shop. With the emergence of Near Field Communication (NFC) technology that equips the mobile phone with the possibility to interact with its surroundings by simply "touching" them, the mobile phone can become a payment device for the mobile user. The SIM card can store credit cards, tickets, ID cards and e-money that can be transferred by the mobile phone using NFC to the point-of-sale (POS) terminals.

6 Conclusion

Although originally intended to be a modest communication terminal, the mobile phone has unexpectedly evolved to become an inseparable companion of every people from businessmen, housewives to youngsters or retirees. It has shown precious and irreplaceable qualities like few other devices. It is useful and enables the user to keep in touch with whomever they want. It is loyal and follows its master wherever they go. It is small, light and easy to use at the same time as it is quite robust and endurable. It is also very personal and allows its master to express himself as jewellery. The mobile phones will continue to increase in numbers. More people will have mobile phones and each individual will have more mobile phones. The mobile phone will have more functions and play more roles such as personal data storage, Personal Information Manager (PIM), MP3 player, camera, portable storage, etc. The evolution of the mobile phone will continue in terms of both functionality and form and probably also at a quicker pace. It is quite difficult if not totally impossible to predict what the mobile phone will look like in the future, but it will surely come in numerous variants from the attiring one to the invisible phone that is entirely integrated in the surroundings.

References

1 ETSI. *Mobile Station Conformity Specifications, Version 3.4.0*. Sophia Antipolis, European Telecommunications Standards Institute, March 1991. ETSI Technical Specification GSM 11.10.

2 ETSI. *Digital cellular telecommunications system (Phase 2+) (GSM); Specification of the Subscriber Identity Module – Mobile Equipment (SIM – ME) interface*. Sophia Antipolis, European Telecommunications Standards Institute, 1996. ETSI Technical Specification GSM 11.11.

3 ETSI. *Digital cellular telecommunications system (Phase 2+) (GSM); Specification of the SIM application toolkit for the Subscriber Identity Module – Mobile Equipment (SIM – ME) interface*. Sophia Antipolis, European Telecommunications Standards Institute, 1996. ETSI Technical Specification GSM 11.14.

4 Gemplus. *Applications – The Benefits for wireless applications providers, operators, and users. A Gemplus paper*. Gemplus, October 2004. online: http://www.3gamericas.org/pdfs/gemplus_sim_applications_oct2004.pdf

5 ETSI. *Universal Mobile Telecommunications System (UMTS); USIM and IC Card Requirements*. Sophia Antipolis, European Telecommunications Standards Institute, 1999. 3G TS 21.111 version 3.0.1, Release 1999.

6 ETSI. *Universal Mobile Telecommunications System (UMTS); UICC-Terminal Interface; Physical and Logical Characteristics*. Sophia Antipolis, European Telecommunications Standards Institute, 1999. 3G TS 31.101 version 3.0.0 Release 1999.

7 ETSI. *Universal Mobile Telecommunications System (UMTS); Characteristics of the USIM Application*. Sophia Antipolis, European Telecommunications Standards Institute, 1999. 3G TS 31.102 version 3.0.0 Release 1999.

8 Mouly, M, Pautet, M-B. *The GSM System for Mobile Communications*. France, Palaiseau, 1992. (ISBN 2-9507180-0-7)

9 ETSI. *TS 44.064 Mobile Station – Serving GPRS Support Node (MS-SGSN) Logical Link Control (LLC) Layer Specification*. Sophia Antipolis, European Telecommunications Standards Institute, 2002.

10 FT Mobile (Financial Times). *Personalise your FT mobile phone*. 22 Nov 2005 [online] – URL: http://globalelements.ft.com/Mobile/Specials/personalise.html

For a presentation of Do Van Thanh, please turn to page 2.

Ivar Jørstad (29) received his CandScient degree in informatics (communication systems) from the Department of Informatics, University of Oslo, in December 2002. He is currently pursuing a PhD degree at the Department of Telematics, Norwegian University of Science and Technology, Trondheim (NTNU), where the work is focused around the requirements of future service platforms for mobile services with particular focus on personalisation. His major fields of interest include service architectures and platforms for mobile services and applications, mobile distributed computing, Service-Oriented Architectures and Web Services, mobile terminal platforms and the personalisation of mobile services. Visit http://ongx.org for more information.

email: ivar@ongx.org

Mobile telephone history

TOM FARLEY

Tom Farley is a freelance telecom writer

This article describes how mobile telephones, for decades a near dormant technology, became the dynamic and perhaps most important communication tool of our lives. Commercial mobile telephony began in 1946. The cellular radio concept was published in 1947. But only since 1995 have mobiles become low cost, rich in features, and used world wide. We first examine mobile telephony's early and bulky beginnings. Next, the long journey to analog cellular. Finally, full digital working, exemplified by GSM and now CDMA, providing services and features that make the mobile indispensable and ubiquitous. We'll see how early mobile telephony battled the same problems of today: government regulation, scarce spectrum, and hardware limitations. How Scandinavian, Japanese, and United States groups independently crafted their own radio-telephone solutions. At 58, the relatively recent, spectacular success of today's mobile telephone could hardly be guessed by its age. But its history reveals why this technology took so long to mature. And the present shows us that it was worth the wait.

Introduction

Public mobile telephone history begins in the 1940s after World War II. Although primitive mobile telephones existed before the War, these were specially converted two way radios used by government or industry, with calls patched manually into the landline telephone network. Many New York City fireboats and tugboats had such radiotelephones in the 1930s. These were private services. For this article, though, a mobile telephone is a wireless device which connects to the public switched telephone network and is offered to the general public by a common carrier or public utility. Further, mobile history is not just a study of the telephone, the handset itself, but a look at the wireless system it is connected to.

After World War II badly neglected civilian communication needs could finally be addressed. Many cities lay in ruin; their infrastructures need years of reconstruction. Post, Telephone and Telegraph administrations, the PTTs, and private telephone companies concentrated on providing landline telephones and services first, but some mobile radio research and development still went on. Americans lead this low priority movement for three reasons. The United States was physically intact after the war, Bell Telephone Laboratories had a large group of radio engineers and scientists to use, and the Motorola corporation had grown significantly during World War II. Consumer demand, research facilities, and manufacturing capability all existed for US mobile telephony. But was that enough? And what kind of mobile system would be created?

Already in 1924, Bell Labs tested mobile radio telephony (from http://www.bell-labs.com/history/75/gallery.html)

On July 28, 1945 a cellular radio or small zone system was first described in print. The head of the United State's Federal Communications Commission, the FCC, outlined a two way radio service in the 460 MHz band to the *Saturday Evening Post*. Commissioner J.K. Jett had just been briefed by AT&T personnel. They had speculated about American wireless communications after World War II. Deceptively titled "Phone Me by Air", Jett's *Post* interview didn't suggest connecting mobile radios to the landline telephone system. But he did describe frequency reuse within a small area, the main element of cellular radio. Millions of users, he said, could use the same channels across the country. Low powered transmitters using high band radio frequencies would keep signals in nearby cities from interfering with each other. Despite Jett's initial enthusiasm, the FCC never allocated the spectrum needed for this service. Still, radio engineers were thinking of cellular, even if they couldn't build such a scheme just yet.

A year after that landmark article, the first American commercial mobile radio-telephone service began. On June 17, 1946 in Saint Louis, Missouri, AT&T and one of its regional telephone companies, South-

 ISSN 0085-7130

western Bell, began operating MTS, or Mobile Telephone Service.[1] Motorola built the radios and the Bell System installed them. MTS was modeled after conventional dispatch radio. A centrally located antenna transmitted to mobiles moving across a wide area. The mobiles, all of them car based radio-telephones, transmitted to several receivers situated around the city. The traffic from the receivers and to the transmitter were connected by an operator at a central telephone office. MTS used six channels in the 150 MHz band with 60 kHz wide channel spacing. Unexpected interference between channels soon forced the Bell System to use only three channels. Waiting lists developed immediately in every one of the twenty five cities MTS was introduced.

Cellular telephone systems first discussed

In December, 1947 Bell Laboratories' D.H. Ring, with help from W.R. Young, articulated a true cellular radio system for mobile telephony in an internal company memorandum.[2] Young said later that all the cellular radio elements were known: a network of small geographical areas called cells, a base station transmitter in each, cell traffic controlled by a central switch, frequencies reused by different cells and so on. He stated from1947 Bell teams "had faith that the means for administering and connecting to many small cells would evolve by the time they were needed."[3] But more mobile telephones were always needed. Then, in 1947, and for decades after. Better technology would help, but more spectrum, more channels, were essential to developing a high capacity mobile telephone service.

Conventional mobile telephony

In 1947 the Bell System asked the FCC for more frequencies. The Commission allocated a few more channels in 1949, but they also did something unexpected. They gave half of those frequency allocations to other companies wanting to sell mobile telephone service. These firms were called Radio Common Carriers or RCCs. The FCC thus created wireless competition for the Bell System while allowing capacity to increase only slightly. These small businessmen, however, advanced early mobile telephony further and faster than AT&T. As proof of their competitiveness, the RCCs serviced 80,000 mobile units by 1978, twice as many as AT&T. This growth began with an excellent start, the introduction of automatic dialing in 1948.

On March 1, 1948 the first fully automatic radiotelephone service began operating in Richmond, Indiana, eliminating the operator to place most calls.[4] AT&T by comparison didn't provide automatic dialing until 1964. Most systems, though, RCCs included, still operated manually until the late1960s. While these small, independent wireless companies could provide service to a few dozen customers at a time, they did not have the money or the resources to research, design, and then build a high capacity mobile telephone system.

On July 1, 1948 the Bell System unveiled the transistor, a joint invention of Bell Laboratories scientists William Shockley, John Bardeen, and Walter Brattain. It would revolutionize every aspect of the telephone industry and all of communications. Fragile and bulky vacuum tubes would eventually be replaced by transistors. Compact, low cost, rugged radios could now be speculated about. Vacuum tubes, though, would dominate the radio and telephone industry for another twenty years.

Outside of the United States mobile telephony developments came slowly. Most governments or PTTs did not allow the public radiotelephones. There were exceptions. In 1949 the Dutch National radiotelephone network inaugurated the world's first nationwide public radiotelephone system. And in 1951 the Swedish Telecommunications Administration's Sture Lauhrén and Ragnar Berglund designed a novel automatic mobile telephone system called the MTA. This scheme began with a Stockholm trial and soon encompassed the city and its surrounding area.
A similar system was soon set up in Gothenburg, although both networks did not become fully operational until 1956. As with all car mounted radio telephones, the equipment was huge and required much power. The transmitter and receiver were mounted in the boot or trunk, while the dial and handset went inside the cab. A car's headlights dimmed while a

---

1) *Peterson, A C, Jr.* Vehicle Radiotelephony Becomes a Bell System Practice. *Bell Laboratories Record, 137, April, 1947.*

2) *Roessner, D et al.* The Role of NSF's Support of Engineering in Enabling Technological Innovation: Phase II, Chapter 4: The Cell Phone. Final report to the National Science Foundation. *Arlington, Virginia: SRI International, 89, 1998. citing Ring, D H, "Mobile Telephony – Wide Area Coverage," Bell Laboratories Technical Memorandum, December 11, 1947. Online: http://www.sri.com/policy/stp/techin2/chp4.html*

3) *Young, W R. Advanced Mobile Phone Service: Introduction, Background, and Objectives.* Bell System Technical Journal*, 7 January, 1979.*

4) *McDonald, R. Dial Direct: Automatic Radiotelephone System.* IRE Transactions on Vehicle Communications*, 80, July, 1958.*

customer transmitted. On the other side of the planet, an electronics giant was gaining life.

In 1952 Japan regained its independence, seven years after World War II ended. Nippon Telephone and Telegraph became privatized, its research division strengthened, and various government sponsored laboratories escalated radio and telephone studies. Although private radiotelephones were not allowed, consumer demand for commercial radio and television broadcasting sets would come about quickly, and the Japanese soon looked to making this equipment for export. Quality control pioneer Edwards Deming had been lecturing Japanese industry leaders since 1950. He stressed quality first, something American manufacturers were not receptive to. But the Japanese took Deming's advice quite seriously. Japanese cameras, cars, and electronics became so good over the next thirty years that other countries were forced to rethink and often retool entire industries.

In 1953 the Bell System's Kenneth Bullington wrote "Frequency Economy in Mobile Radio Bands."[5)] This dull sounding paper appeared in the *Bell System Technical Journal*, circulated around the world. For perhaps the first time in a publicly distributed paper, the 21 page article hinted at, although obliquely, cellular radio principles. Three years later the Bell System began providing a manual radio-telephone service at 450 MHz, a new frequency band assigned to relieve overcrowding on their lower frequency band. This system also filled to capacity wherever it was introduced.

In July, 1958 Jack Kilby invented the integrated circuit at Texas Instruments in Dallas, Texas. A toothpick size piece of germanium contained his complete electrical circuit or IC. It used no soldered connections and consequently was reliable and stable. He also showed how resistors, capacitors, diodes, and transistors could co-exist on a single block of semiconductor and that they could all be made of this same material. As Texas Instruments itself puts it, "The roots of almost every electronic device we take for granted today can be traced back to Dallas more than 40 years ago."

In 1958 the innovative Richmond Radiotelephone Company improved their automatic dialing system. They added new features to it, including direct mobile to mobile communications. Other independent telephone companies and Radio Common Carriers made similar, incremental advances to mobile telephony throughout the 1950s and 1960s. In this same year the Bell System petitioned the FCC to grant 75 MHz of spectrum to radio-telephones in the 800 MHz band. Despite the Bell System's forward thinking proposal, the FCC ignored their request for ten years.

During the late 1950s little cellular radio research and development was accomplished. Without enough spectrum to make it economically feasible, a high capacity cellular system could not be built in the United States. Still, two important papers by Bell System employees were published in 1960. They appeared in the *Institute of Radio Engineers Transactions on Vehicle Communications*. The articles discussed handoffs, that process of transferring a call from one cell to the next, with different frequencies used in adjacent cells.[6)] This was the first time the entire cellular system concept was outlined in print to a worldwide readership.

In 1961 Ericsson subsidiary Svenska Radio Aktiebolaget, or SRA, reorganized to concentrate on building radio systems. This forerunner of Ericsson Radio Systems was already selling paging and land mobile or dispatch radio equipment throughout Europe. SRA would go on to become a central part of Ericsson, helping create their wireless consumer business.

In 1964 the Bell System introduced Improved Mobile Telephone Service or IMTS, a replacement to their badly aging Mobile Telephone System.[7)] With IMTS people didn't have to press a button to talk. Conversations went back and forth just like a regular telephone. IMTS finally permitted direct dialing, automatic channel selection, and reduced bandwidth from between 25 and 30 kHz. Some regional telephone companies like Pacific Bell, owned by AT&T, took nearly twenty years to replace their old MTS systems. Again, although demand was great, there were not enough channels to accommodate more users.

Other countries in the mid 1960s were also replacing their first mobile telephone systems. The Swedish Telecommunication Administration began replacing their MTA system with MTB. Ragnar Berglund developed this new system and, thanks to the transistor, made possible smaller phones which required less power and were therefore less expensive. MTB was

5) *Bullington, K. Frequency Economy in Mobile Radio Bands.* Bell System Technical Journal*, 32 (42) et. seq. January,1953.*

6) *Lewis, W D. Coordinated Broadband Mobile Telephone System.* IRE Transactions*, 43, May, 1960; and Schulte, H J Jr. and W A Cornell. Multi-area Mobile Telephone System.* IRE Transactions*, 49, May, 1960.*

7) *Douglas, V A. The MJ Mobile Radio Telephone System.* Bell Laboratories Record*, 383, December, 1964.*

available to the public from 1965. Like MTA, the MTB soon ran out of capacity with 660 customers served.[8)9)]

In 1967 Nokia was formed by consolidating two companies: the Finnish Rubber Works and the Finnish Cable Works. Nokia expanded Finnish Cable Works electronics division to include semi-conductor research. These early 1970s studies helped Nokia develop digital landline telephone switches. Also helping the Finns was a free market for telecom equipment, an open economic climate which promoted creativity and competitiveness. Unlike most European countries, Finland's PT&T was not required to buy equipment from a Finnish company. And other telephone companies existed in the country, any of whom could decide on their own which supplier they would buy from. Nokia's later cellular development was greatly enhanced by this free market background and their early research.

In 1967 Televerket, now Telenor, began operating a public mobile telephone system known as the OLT. It was a manual system using the 160 MHz band. It, too, ran out of capacity soon after introduction. A few years later an additional system was introduced in the 450 MHz band in southern Norway.

By the late 1960s it is certain that every major telecommunications company and manufacturer knew about the cellular radio idea. In 1967, for example, NT&T may have begun research for a nationwide cellular system at 900 MHz for Japan.[10)] But how to make it work technically and economically? There was no way to evolve the existing radiotelephone infrastructure to cellular. New base station radio equipment and new customer mobiles were needed. Instead of a single, central antenna site with one fairly simple transceiver, several to dozens of cell sites would be required, each needing its own transceiver, all of them interconnected to each other, with a network switch to manage the traffic, and software to make everything work. The cost would be enormous.

The Federal Communications Commission in the United States in 1968 reconsidered the Bell System's ten year old request for more frequencies. They made a tentative decision in 1970 to grant them, asked AT&T to comment, and received the system's technical response in December, 1971. The Bell System submitted a frequency-reuse cellular radio scheme. Their proposal was based on the patent Amos E. Joel, Jr. and Bell Telephone Laboratories filed on December 21, 1970 for a mobile communication system. Six long years passed before the FCC allowed AT&T to start a trial.

Besides bureaucratic sloth, this delay was also caused by lawsuits and objections from radio common carriers, independent telephone companies, and their suppliers. All three groups feared the Bell System would dominate cellular radio if private companies weren't allowed to compete equally. They wanted the FCC to design open market rules, and they fought constantly in court and in administrative hearings to make sure they had equal access. And although its rollout was delayed, the Bell System was already working with cellular radio, in a small but ingenious way.

The first commercial cellular radio system

In January, 1969 the Bell System made commercial cellular radio operational for the first time by employing frequency reuse in a small zone system. Using public payphones. Passengers on what was called the *Metroliner* train service running between New York City and Washington, DC found they could make telephone calls while moving at more than 160 kilometers per hour. Six channels in the 450 MHz band were used again and again in nine zones along the 225 mile route. A computerized control center in Philadelphia, Pennsylvania, managed the system. Thus, the first cell phone was a payphone! As Paul described it in the *Bell Laboratories' Record* article on the project, "[T]he system is unique. It is the first practical integrated system to use the radio-zone concept within the Bell System in order to achieve optimum use of a limited number of radio-frequency channels."[11)12)]

Around 1969 the first all transistor mobile telephones appeared from a large manufacturer. The tube era for radio telephones was ending. Motorola's 'Mark 12' was an IMTS telephone designed to work in the 450 Mhz band. This transistor rig was still big and bulky and mounted in a vehicle. The first commercial

---

8) *Online: http://www.tekniskamuseet.se/mobilen/engelska/1960_70.shtml*

9) *Olle Gerdes, citing* Dædalus 1991, The Yearbook of the National Museum of Science and Technology. *Stockholm.*

10) *Ikegami, F. Mobile Radio Communications in Japan.* IEEE Transactions On Communications*, 744, Vol. Com-20 No. 4, August, 1972.*

11) *Paul, C.E. Telephones Aboard the 'Metroliner'.* Bell Laboratories Record*, 77, March, 1969*

12) *For many more details on the* Metroliner *or "High Speed Train Project", please see http://www.privateline.com/PCS/metroliner.htm*

ISSN 0085-7130 © Telenor ASA 2006

In the early 1970s Bell System tested the cellular concept, which had already been used in a commercial system since 1969. (Photo supplied by John Winward)

portable radiotelephones in the United States also appeared at this time. In 1969 or 1970 SCM Melabs, owned by Smith Corona, produced an attaché phone, a complete MTS telephone built into a briefcase. Almost immediately Canyon Communications and Livermore Data came out with their attaché phones. These were all MTS even though IMTS had been introduced in 1964. Only small firms made these units. Harris, Motorola, and GE never did. All these phones were essentially made by hand.[13]

In November, 1971 Intel introduced the first commercial microprocessor, the 4004, a miniature computer on a silicon chip. The original contained 2,300 transistors and did 60,000 operations a second. Today's microprocessors can contain 5.5 million transistors, performing hundreds of millions of calculations each second. Intel's 4004 was designed originally for a desktop calculator, but microprocessors were soon improved on and eventually put into all kinds of electronics, including telephone switches and cell phones. That integration could have come sooner for one telecom group.

During the late 1960s and early 1970s the Nordic Mobile Telephone group was planning a Scandinavian wide mobile telephone network. Their 1970 report concluded that the microelectronics needed to build an analog cellular network would not be available until 1980. The group decided therefore that instead of using new technology, they'd design a conventional, manual mobile telephone system. It started in Örebro, Sweden in 1971. It required 400 operators to serve just 19,800 subscribers. MTD shut down in 1987, eclipsed, of course, by an automated cellular radio system made possible by microprocessor technology.[14]

On October 17, 1973, Motorola filed a patent for its own cellular radio system.[15] Although Motorola had supplied the Bell System with radiotelephones for decades, AT&T was now considered a threat, not a friend. Motorola's main business was dispatch radio systems for taxi companies, utility fleets, police departments, and so on. If cellular was successful then dispatch customers might move in whole or in part to the new service. So Motorola needed a cellular offering to compete with AT&T. A rivalry developed between the two companies to field working equipment. In 1973, after completing Motorola's first prototype cellular telephone and its base station, Dr. Martin Cooper called his competitors at Bell Labs. Ferranti says "Cooper couldn't resist demonstrating in a very practical manner who had won."[16] What Cooper's team invented was the first handheld cell phone. But not the cell phone itself. That had already been done on the *Metroliner* train. Motorola's successful field work caused the American magazine *Popular Science* in July, 1973 to picture the portable phone on their cover. The accompanying article said that with FCC approval New York city could have a Motorola cellular system operating by 1976. No approval came.

On May 1, 1974 the FCC approved an additional 115 megahertz of spectrum for future mobile telephone use. Cellular loomed ahead, although no one knew when FCC approval would permit its commercial rollout. American business radio and radio-telephone manufacturers begin planning for the future. The demand was certainly there. In 1976 only 545 customers in New York City had Bell System mobiles, with 3,700 customers on the waiting list. In the United States overall, 44,000 Bell subscribers had AT&T mobiles but 20,000 people were on five to

[13] *Geoff Fors. Personal correspondence.*

[14] *Online: http://www.tekniskamuseet.se/mobilen/engelska/1970_80.shtml*

[15] *US Patent Number 3,906,166, granted September16, 1975.*

[16] *Ferranti, M.* Father of Cell Phone Eyes a Revolution. *IDG News Service, New York Bureau, 14 (31), October 12, 1999.*

ISSN 0085-7130 © Telenor ASA 2006

ten year waiting lists. Demand always existed but licensed spectrum to accommodate them did not. Until now.

In1975 the FCC let the Bell System begin a trial. It wasn't until March, 1977, though, that the FCC approved AT&T's request to actually operate their cellular system. A new wireless industry was developing in America and the FCC sought to control every aspect. They'd decide the number of wireless carriers in each market, the companies allowed to operate, standards for the equipment, frequency assignments, channel spacing, and on and on.[17] Suffering less bureaucratic trouble, Japanese and Scandinavian manufacturers diligently worked on trialing first commercial analog cellular systems. The NMT group ran a satisfactory trial in Stockholm in late 1977 through early 1978. Nippon Telephone and Telegraph probably started field tests in Tokyo as early as 1975.[18]

NTT produced the first cellular systems for Japan, using all Japanese equipment. The Japanese also contributed important studies to cellular research. Y. Okumura's 1968 "Field Strength and its Variability in VHF and UHF Land Mobile Service," is an often cited, pioneering work. But Japan's greatest contribution to cellular radio was quality control. American industry and those who emulated its practices, in the final analysis, favored quantity over quality. The Japanese insisted on both.

In the mid to late 1970s, Japan's goal to produce electronic goods without defects forced manufacturers around the globe to ask themselves if they could compete. Self-examination was a wrenching but necessary process that for many companies would go on for years. Before completing the turn to better quality shipping dates would be missed, production quotas lost, profits reduced. It was all very necessary; assembly line production of mobiles by the millions could not have happened with the one at a time techniques of producing conventional mobile telephones.

In January, 1978 Andy Affrunti Sr. warned Motorola management that the biggest threat to their company was quality competition from the Japanese. He asked his bosses, "Do we have a quality organizational structure that could meet this Japanese competition and achieve zero defects?" As if to highlight the issue, the next week Affruniti found factory workers beating on warped metal housings with a board and mallet to make them true, and, to make a deadline, radios deliberately shipped with a missing part. Motorola immediately began institutional changes toward quality control.[19]

Analog cellular systems begin

In May, 1978 The Bahrain Telephone Company (Batelco) began operating the first commercial cellular telephone system. The simple two cell scheme had 250 subscribers, operated on 20 channels in the 400 Mhz band, and used all Matsushita (Panasonic) equipment.[20] Cable and Wireless, now Global Crossing, installed the equipment for Batelco.

In July, 1978 Advanced Mobile Phone Service or AMPS began operating near two American cities. The first area was around AT&T Labs in Newark, New Jersey, and the second place was near Chicago, Illinois. Ten cells covering 21,000 square miles made up the Chicago system. Oki Electric provided the mobile terminals. This equipment test started with 90 Bell System employees acting as customers. After six months, on December 20, 1978, a market trial began with paying subscribers who leased the car mounted telephones. This was called the service test. The system used the newly allocated 800 MHz band.[21] Although the Bell System bought an additional 1,000 mobile phones from Oki for the lease phase, it placed orders from Motorola and E.F. Johnson for the remainder of the 2,100 radios.[22] This early network, using large scale integrated circuits throughout, a dedicated computer and switching system, custom made mobile telephones and antennas, proved a large cellular system could work.

In 1979 INMARSAT was born, an international group fostering and coordinating satellite telephony. Originally developed for ships at sea, INMARSAT's charter later extended to telephone calls made on land

---

17) *Online: http://www.fcc.gov/Bureaus/OGC/Reports/cellr.txt*

18) *Ito, Sadao and Yasushi Matsuzaka. 800 MHz Band Land Mobile Telephone System – Overall View.* IEEE Transactions on Vehicular Technology*, 205, Volume VT-27, No. 4, November 1978, as reprinted from Nippon Telegraph and Telephone's* The Review of the Electrical Communication Laboratories*, Vol. 25, nos 11–12, November-December, 1977. (English and Japanese)*

19) *Affrunti, Andy.* A Personal Journey: 50 Years at Motorola*. 132–133. Motorola University Press, Rolling Meadows, Illinois, 1994.*

20) *Gibson, Stephen W.* Cellular Mobile Radiotelephones*. Englewood Cliffs, Prentice Hall, 141, 1987. See also online: http://www.privateline.com/PCS/Bahrain.htm*

21) *Blecher, F. Advanced Mobile Phone Service.* IEEE Transactions on Vehicle Communications*, Vol. VT-29, No. 2, May, 1980.*

22) *Fewer busy signals for mobile phones.* Business Week*, Industrial Edition, Number 2546: 60B, August 7, 1978.*

1984. The first portable units were really big and heavy. Called transportables or luggables, few were as glamorous as this one made by Spectrum Cellular Corporation

and from aircraft. MARISAT or Marine Satellite was the first mobile communications satellite service, beginning in 1976. Both satellite groups sought to make more dependable radio-telephone traffic which had previously gone over High Frequency or short-wave radio links. Shipboard satellite customers first talked with an international operator who then manually patched their call into the landline telephone system. Echo and reverberation problems were common in those days, an operator might need 6 to 9 call setups for 1 call.[23)] Let's return now to terrestrial radio-telephony.

Worldwide commercial cellular deployment blossomed in the late 1970s and then continued into the early 1980s. An 88 cell system in the challenging cityscape of Tokyo began in December, 1979, using Matsushita and NEC equipment. The first North American commercial system began in August, 1981 in Mexico City. It was a one cell system. The world's first Nordic Mobile Telephone network started on September 1, 1981 in Saudi Arabia. It used 20 cells and operated at 450 MHz. The next month, starting on October 1, 1981, and opening in stages until March, 1982, Sweden, Norway, Denmark, and Finland, began operating a Scandinavian wide NMT network. It also operated at 450 MHz, and used three Ericsson switches. The first multi-national cellular system, the NMT450 had 600 cells and offered roaming, an important first. As the Scandinavians operated the most advanced cellular system in the world, rollout of cellular radio in America was stopped again by government bureaucracy.

New regulations and AT&T's impending breakup caused American cellular to be delayed once again. The Federal Communication Commission in 1981 required the Bell System regional operating companies, such as Bell Atlantic, to have competition in every cellular market. The FCC thought this would provide better service and keep rates low. In reality prices between the wireline and non-wireline carriers were always about the same, and service no better between the two. Rules governing this state imposed duopoly were many: Applications to operate in each city were required and a lengthy licensing award process needed to be followed.

On March 25, 1980, Richard Anderson, general manager for Hewlett-Packard's Data Division, shocked American chip producers by saying that his company would henceforth buy most of its chips from Japan. After inspecting 300,000 standard memory chips, what we now call RAM, HP discovered the American chips had a failure rate six times greater than the worst Japanese manufacturer. American firms were not alone in needing to retool. Ericsson admits it took years for them to compete in producing mobile phones. Let's skip ahead five years to make this point.

In 1987 Panasonic took over an Ericsson plant in Kumla, Sweden, 120 miles west of Stockholm to produce a handset for the Nordic Mobile Telephone network. Meurling and Jeans explain: "Panasonic brought in altogether new standards of quality. They sent their inspection engineers over, who took out their little magnifying glasses and studied, say displays. And when they saw some dust, they asked that the unit should be dismantled and that dust-free elements should be used instead. Einar Dahlin, one of the original small development team in Lund, had to reach a specific agreement on how many specks of dust were permitted."[24)] Let's go back now to the early 1980s, when telecom changed forever.

On August 24, 1982, after seven years of wrangling with the American federal Justice Department, American Telephone and Telegraph was split apart, succumbing to government pressure from without and a carefully thought up plan from within. The Bell System, serving 80 % of the American population, and custodian of Bell Laboratories, was broken apart. Complete divestiture took place on January 1, 1984. After the breakup new companies, products, and services appeared immediately in all fields of American telecom, as a fresh, competitive spirit swept the coun-

23) *Online: http://www.privateline.com/Snyder/TSPS_history_recollections.htm*

24) *Meurling, John and Richard Jeans.* The Ugly Duckling: Mobile phones from Ericsson. *Stockholm, Ericsson Radio Systems AB, 46, 1997.*

try. The AT&T divestiture caused nations around the world to reconsider their state owned and operated telephone companies, with a view toward fostering competition in their own countries.

European analog systems

Europe saw cellular service introduced in 1981, when the Nordic Mobile Telephone System or NMT450 began operating in Denmark, Sweden, Finland, and Norway in the 450 MHz range. It was the first multinational cellular system. In 1985 Great Britain started using the Total Access Communications System or TACS at 900 MHz. Later, the West German C-Netz, the French Radiocom 2000, and the Italian RTMI/RTMS helped make up Europe's nine incompatible analog radio telephone systems. All services used analog for sending voice, signaling was done with a variety of tones and data bursts. Handoffs were based on measuring signal strength except C-Netz, which measured the round trip delay. Early C-Netz phones, most made by Nokia, also used magnetic stripe cards to access a customer's information, a predecessor to the SIM cards of GSM/PCS phones. All of these mobiles were car phones.

The Motorola Dyna-TAC from 1983 was the first handheld cellular phone (from: http://www.motorola.com/mediacenter/graphics/)

On October 12, 1983 the regional Bell operating company Ameritech began the first United States commercial cellular service in Chicago, Illinois. This was AMPS, or Advanced Mobile Phone Service. United States cellular developed from this AT&T model, along with Motorola's system known as Dyna-TAC, first introduced commercially in Baltimore and Washington DC. AMPS or Dyna-Tac, often both, were soon installed and operating within three years in each of the ninety largest markets in America.[25)]

Cellular's popularity in the United States was unexpectedly strong. Estimates say there were 340,213 customers in 1985; 681,825 by 1986, and 1,300,855 by 1987.[26)] Conventional mobile telephones by comparison served less than 100,000 subscribers before cellular began. This 100 % growth each year attracted overseas equipment makers. Ericsson supplied switches and eventually base station equipment, while companies like Nokia sold handsets. AMPS systems were sold throughout the world. One country was especially interested in the technology, not just to use but also to develop as an industry.

In March, 1984 the government KMT or Korea Mobile Telecommunications Company was formed. On May 1, 1984 KMT began AMPS service in South Korea. They had some experience with mobile telephony; a Motorola IMTS system had been operating in Korea since the late 1960s. But cellular was new and something the Koreans thought they could participate in. They started with manufacturing. In 1984 Nokia and Tandy formed Tandy Mobira Corporation in Korea. The Finns wanted to sell AMPS phones in America. The Tandy corporation had electronics stores across the United States which could distribute those phones. By 1992, 824,000 handsets had been sold under the Tandy label and 885,000 under the Nokia brand.[27)] South Korea thus entered the mobile telephone business, taking the first step toward becoming a leader in cellular radio.

Analog cellular was also booming in Europe by the mid-1980s. The main problem was that systems worked well by themselves but they wouldn't work together. A German customer, for example, couldn't operate their mobile in Italy. Planning began during the early 1980s to create a single European wide digital mobile service with advanced features and easy roaming. While North American groups concentrated on building out their robust but increasingly fraud plagued and featureless analog network, Europe planned for a digital future.

Why didn't America build a fully digital system earlier? The United States suffered no variety of incompatible technologies as in Europe. Only AMPS or an AMPS compatible system existed in America. Roam-

---

25) *Gibson, Stephen W.* Cellular Mobile Radiotelephones. *Englewood Cliffs, Prentice Hall, 19–22, 1987.*

26) *Online: http://ctia.org/research_statistics/index.cfm/AID/10030*

27) *Haikio, M. Nokia: The Inside Story. Prentice Hall, London, 160, 2001.*

ISSN 0085-7130 © Telenor ASA 2006

The size and weight of cellular telephones shrank considerably between 1990 and 2000

ing agreements between operators and a common networking standard, IS-41, allowed customers to make calls from whatever city or state they were in. Little desire existed to design an all digital system when the present one was popular and working well. To keep the current phones working (and producing money for their carriers) any new system would have to accommodate them. Chances lessened for an all digital future with each analog phone sold.

The Rise of GSM

Europeans saw things differently. No existing telephone system could accommodate their different cellular systems. They decided instead to create a new technology in a new radio band. Cellular radio but fully digital, the new service would incorporate the best thinking of the time. No backward compatibility with existing systems. They patterned their new wireless standard after landline requirements for ISDN, hoping to make a wireless counterpart to it. The new service was called GSM.

GSM first stood for Groupe Speciale Mobile, after the study group that created the standard. It's now known as Global System for Mobile Communications, although the "C" isn't included in the abbreviation. In 1982 twenty-six European national phone companies began developing GSM. This Conference of European Postal and Telecommunications Administrations or CEPT, planned a uniform, European wide cellular system around 900 MHz. A rare triumph of European unity, GSM achievements became "one of the most convincing demonstrations of what cooperation throughout European industry can achieve on the global market." Planning began in earnest and continued for several years.

By the late 1980s the American wireless industry began searching for a higher capacity system. In September, 1988 the Cellular Telecommunication Industry Association published a set of User Performance Requirements, urging a new digital technology be built with 10 times the capacity of existing analog schemes. Two choices quickly emerged, one digital, one analog, but neither came close to the capacity goal.

In 1990, cellular telephones were typically like this "Bag phone" from Motorola

In December 1988 Japan's Ministry of Posts and Telecommunications ended NTT's monopoly on mobile phone service. Although technically adept, NTT was also monolithic and bureaucratic; it developed a good cellular system but charged too much to use it. Growth was slow. They also required customers to lease phones, not to buy them. After 1989 competition and new networks increased cellular sales. But not until 1994, when telecom was completely deregulated, did cellular prosper. In the late 1980s Japan was also studying the next generation of cellular. Their first generation systems were modeled after AMPS but it was unclear if their second systems would be analog or digital.

In 1989 The European Telecommunication Standards Institute or ETSI took responsibility for further developing GSM. In 1990 the first recommendations were published. The specifications were published in 1991. The United States cellular industry knew time based systems would work well but wanted a digital system of their own, a dual mode technology that could keep existing analog phones working.

In January, 1989 the Telecommunication Industry Association (TIA) selected a time based or TDMA approach to North American digital cellular radio. The Cellular Telecommunication Industry Association (CTIA) also endorsed the TIA's pick, although it did not contain the 10 time capacity gain it asked for the year before. The CTIA hoped that eventually capacity gains would increase. The TIA next wrote a standard for this new digital system, soon to be called IS-54. It was unofficially called D-AMPS or Digital AMPS. After publishing the standard manufacturers would know how to build for the system. Few suspected the technology to get the most gain was already being developed.

On November 3, 1989 in San Diego, California, Qualcomm successfully demonstrated a prototype CDMA cellular system to a group of 250 network operators and suppliers from around the world. Three months later they repeated this demonstration in New York City. Code Division Multiple Access had come to mobile telephony. It appeared too late to be considered as the digital choice for new North American cellular networks. Over the next few years, however, it would come into the American market and show the wireless industry that CDMA, in one form or another, would eventually replace time division systems.

North America goes digital: IS-54

In March, 1990 the North American cellular network formally adopted a digital standard: IS-54. It worked with existing AMPS systems. This choice won over Motorola's Narrowband AMPS or NAMPS, an analog scheme that increased capacity by reducing channel size. IS-54 by comparison increased capacity by digital means: sampling, digitizing, and then multiplexing conversations, using a technique called TDMA or time division multiple access. It tripled call capacity. GSM also uses time division.

An operator had great flexibility with IS-54. It could convert any of its analog voice channels to digital. Customers got digital service where available and analog where it wasn't. Existing customers weren't left without service; they simply couldn't access IS-54's new features. CANTEL started IS-54 in Canada in 1992. Many other AMPS countries also adopted TDMA as a digital choice, like Japan in 1994 with their Personal Digital Cellular or PDC system.

Commercial GSM networks started operating in mid-1991 in Europe. On July 1, 1991 Finland's Radiolinja launched the first commercial GSM network. Radiolinja was the wireless consortium of privately owned regional telephone companies. Nokia provided the equipment. The all digital GSM increased capacity three times over analog. Every mobile contained or accessed encryption to prevent eavesdropping, authentication to prevent fraud, short messaging services or SMS, and a SIM card to easily add accounts to a handset. GSM would go on to be installed around the world and become the most popular cellular radio service. In February 2004 it was announced that GSM had one billion customers.

In the summer of 1991 Pacific Telephone, a former regional Bell System telephone company, decided to invest in Qualcomm. This was an unusual and controversial decision for a regulated telephone company. Pacific Bell's Los Angeles wireless customers were growing by 200 % a year. A CDMA solution seemed the only way to handle that growth. But Qualcomm didn't have the equipment yet to build a network; they needed money to finance production and for research and development. Besides funding them, PacTel advised Qualcomm throughout the standards making process. They also gave them vital consulting and contacts in Korea, where the government was deciding on what digital system should replace their analog cellular network. The Koreans were eager to use a scheme they played a role in. They were reluctant to buy more equipment from Japan, Scandinavia, or the United States. Manufacturing chips and handsets for Qualcomm and sharing in their research and development efforts would strengthen Korea's wireless industry.[28)]

In July 1992 Nippon Telephone and Telegraph created a wireless division called NTT DoCoMo, officially known as NTT Mobile Communications Network, Inc. It took over NTT's mobile operations and customers. And as noted before, in April 1994 the Japanese market became completely deregulated. Japanese cellular took off.

By 1993 American cellular was again running out of capacity, despite a wide movement to IS-54 or D-AMPS. Subscribers grew from one and a half million customers in 1988 to more than thirteen million subscribers in 1993. Demand now existed for other technologies, like GSM, and spread spectrum, to handle the growing number of customers. Qualcomm continued working to get their CDMA system approved as another American interim standard. If sanctioned, manufacturers and carriers would have confidence to build for and use Qualcomm's system. GSM specifications were already published and their technology was continuing to spread around the globe. But GSM hadn't come to America. Yet.

In July 1993 the Telecommunication Industry Association approved Qualcomm's CDMA scheme as an alternative digital standard for the United States. It was called IS-95 and it was a two mode system. As with D-AMPS, IS-95 defaulted to the analog AMPS protocol where the primary signal, in this case CDMA, was not present. A mobile could thus work throughout most of North America where there was cellular coverage, even in places where IS-95 hadn't been installed yet. Qualcomm's system traded greater capacity for complexity in the network and in the mobile. Also known as narrowband CDMA, each channel's bandwidth is 1.25 MHz. IS-95A later gained the trade name cdmaOne.

---

28) *Mock, D.* The Qualcomm Equation*. New York, Amacon, 82, 2005.*

29) *Meurling, J. and R. Jeans. The Mobile Phone Book: The Invention of The Mobile Phone Industry. London, Communications Week International, Ericsson Radio Systems, 78–179, 1994.*

The Nokia 9000 Communicator was introduced in 1996. It was the first mobile phone and handheld computer, a PDA (from http://www.nokia.com)

In August, 1993 the carrier Nextel Communications began operating a new, proprietary wireless network in Los Angeles. They used Motorola phones which combined a dispatch radio with a cellular telephone. Even though Nextel established a nationwide network, their iDEN technology proved unpopular within the wireless industry. iDEN's chief legacy is the push to talk button (PTT), something emulated on many of today's mobiles.

As mentioned before, Japan in 1994 began operating their own digital standard called PDC in the 800 MHz and 1.5 GHz frequency bands. Ericsson, Motorola, AT&T and Japanese suppliers all furnished different equipment for PDC to different wireless carriers. Modeled after IS-54, PDC was a D-AMPS system, it accommodated existing analog customers. Based on TDMA, carriers hoped to eventually replace their three analog cellular systems with digital working and thereby increase capacity.[29)]

A new cellular band and systems in America

In the mid-1990s more wireless channels and carriers were allowed in America. The FCC auctioned off new blocks of frequencies at 1900 MHz starting on December 5, 1994 and ending on January 14, 1997. A new, lucrative market opened for GSM and CDMA. Several carriers were licensed in each metropolitan area. CDMA, TDMA, and GSM proponents spread out across the United States, urging license holders to use their systems.

GSM vendors quickly tailored a system for the American 1900 MHz band. In November, 1995 American Personal Communications, eventually an affiliate of Sprint Spectrum, launched the first commercial GSM service in the US. This network operated in the Washington-Baltimore area. After just six months there were 15 more GSM 1900 networks in the United States. In perhaps a hint of things to come, Sprint PCS in 2000 replaced APC's GSM network with a CDMA system.

IS-136 started shortly after these new spectrum blocks were opened. This was the successor or evolution of IS-54. It again used TDMA and offered a number of new services. AT&T Wireless was its chief proponent. It is still used in America and other countries but its use is declining. In the places it remains it is slowly being cleared out for GSM systems.

On July 1, 1995 the NTT Personal Communications Network Group and DDI Pocket Telephone Group introduced the Personal Handyphone System or PHS to Japan. Also operating at 1900 MHz, sometimes referred to as 1.9 GHz, PHS is an extremely clever system, allowing the same phone used at home to also roam across a city. A cordless phone acting like a mobile.

In September, 1995, Hong Kong's Hutchison Telecom turned on the world's first commercial CDMA/IS-95 system. A year later in San Diego, California, the operator NextWave PCS launched the first American IS-95 system on August 16. The next ten years might well be called the Triumph of CDMA.

The mid-1990s: Fundamental change

On August 15, 1996, Nokia introduced the Communicator, a GSM mobile phone and handheld computer. It had a QWERTY keyboard and built in word processing and calendar programs. Besides sending and receiving faxes, the 9000 could check e-mail and access the internet in a limited way. But its effectiveness was limited since cellular networks were optimized for voice, not data.

To be a telephone an instrument must convey speech. By the mid-1990s, however, delivering quality speech was assured with every cellular radio scheme. Voice, with adjustments, was as good as it needed to be. With the speech requirement settled, data became the first interest of system designers. Voice remained

In the mid-1990s, the Blackberry, essentially a two-way pager only capable of sending and receiving e-mail or SMS, forced us to rethink what a cellular phone was (http://www.blackberry.com) the essential service for the large majority of mobile phones, but developing better and faster data networks over cellular radio became the priority.

To best conduct voice cellular had always used circuit switching, just as the landline telephone network did. But data isn't efficiently conducted by circuit switching. An example is the GSM service called High Speed Circuit Switched Data or HSCSD. It needs four GSM channels to achieve, in theory, speeds between 28.8 kbits and 43.2 kbits a second. Actual speeds are lower. A fundamental change was needed, therefore, from circuit switching to packet switching. And the kind of packet switching needed was obvious from the start.

The internet became commercial in the mid-1990s with the advent of graphical browsers like Mosaic and then Netscape. Internet user growth rivaled cellular telephony between 1995 and 2000. The internet runs on the aptly titled Internet Protocol or IP, a packet switching technique cellular data network operators quickly chose to adopt. Today's General Packet Radio Service (GPRS), its improvement, EDGE, and short range wireless networks like Bluetooth all employ IP. All 3G systems use IP as all of us head toward "an all IP world."

By the mid 1990s the mobile became as small as practically possible. The keypad and display limited any more reduction in size. Cell phone circuitry started getting built into laptops and PDAs and instruments like the Blackberry, forcing us to rethink what a cellular telephone was. Is an SMS only device a mobile telephone or a two way pager? Handsets evolve to provide a variety of services, mostly non-voice, such as ring tones, image capturing, text messaging, gaming, and so on. While cell phone services seem limited only by the imagination, the systems they run over become fewer.

GSM and CDMA systems would continue to be installed around the world but by 2005 no new cellular radio scheme would emerge. Flarion's technology was tested extensively by the American carrier Nextel but the system was not adopted. The lone exception was China. To keep its market closed they choose a hybrid technology called TD-SCDMA, a cross between TDMA and CDMA. The history of cellular telephones from the mid-1990s, therefore, is mostly a chronicle of improvements to existing systems.

In Europe, the idea of a 3rd generation mobile system called UMTS or Universal Mobile Telecommunications System, was developed in the early 1990s through several European Union funded research projects. In 1991, ETSI established a new group, SMG5, to be responsible for standardizing the system. From 1999, the standardization of UMTS has been done by the 3rd Generation Partnership Project – 3GPP. UMTS is a wideband CDMA standard. A 5 MHz channel spacing is used with data rates up to 2 Mb/s. ETSI and 3GPP provided more than just a European response to Qualcomm's narrowband CDMA technology. While acknowledging that future capacity gains could only be achieved by using CDMA, a step by step migration plan to WCDMA for GSM, PDC, TDMA and IS-95 operators was provided. This evolution plan was carefully planned to use most of GSM's core components.

On December 1, 2001 Telenor Mobil trialed a UMTS system in Oslo. Commercial UMTS systems followed, with the technology now installed in different parts of the world. Rollout of UMTS tends to be slow and expensive, since the change from time division to code division requires more than software updates. Hardware changes are needed, especially at the cell site. One can't, for example, reuse existing antennas without severe performance problems. The radio spectrum is an inherently fragile, vexing medium, of course, and operators are struggling to bring data rates close to those promised. While the UMTS Forum assures us that 384 kbps is a minimum for UMTS, and only then in "high mobility situations", 300 kbps may be the working, upper limit for this technology.

In November, 1998 the greatest mobile telephone disaster began when the Iridium project was launched. Using 66 satellites, and costing almost 5 billion US dollars, the service went bankrupt after only 16 months. The lead design firm and largest investor was Motorola. Hoping to make satellite phone service a mass market item, planning for the system began before cellular became widespread and reduced demand. Iridium gathered only 10,000 customers before it folded. Due to the high cost of handsets and services, and an inability to work indoors, satellite telephone service remains a niche market to this day.

In October 2000 Sharp produced the first integrated camera phone. It supplied them to the Japanese Operator J-Phone. The J-SH04 mobile phone let users take, send, and receive images by email.[30] (The Nokia 9110 Communicator in 1998 was the first mobile to enable image transfers but the device relied on a camera supplied by each user.) At the end of

[30] *Online: http://sharp-world.com/corporate/info/his/h_company/2000/*

ISSN 0085-7130 © Telenor ASA 2006

2004 it was estimated that 75 % of the mobiles sold in Japan were camera phones.

The CDG or CDMA Development Group promotes narrowband CDMA. They are the equivalent to the wideband CDMA oriented UMTS Forum. During the late 1990s and early 2000s, the CDG outlined coming improvements to IS-95. They gave these system changes, unfortunately, names which look and seem alike. They even changed the name of IS-95. cdmaOne is now the marketing term for IS-95A, the original CDMA scheme. cdmaOne includes IS-95B which is little implemented. We can look at these evolutions by the dates they debuted.

CDMA2000 1X was first launched by SK Telecom in Korea in October, 2000. Building on an existing IS-95 network, CDMA2000 1X, doubles the voice capacity of cdmaOne networks. It delivers packet data speeds of, supposedly, 307 kbps in mobile environments. But it's doubtful this rate is maintained while the mobile is at speed or while conducting handoffs from one cell to another.

In May, 2002 SK Telecom again made another first, introducing CDMA2000 1xEV-DO service in May, 2002.[31] This is a high speed data only service and an odd one at that. It's actually a CDMA/TDMA hybrid, and uses various modulation techniques, depending on the data rate.

On August 27, 2003, Nokia announced it completed a call using CDMA2000 1xEV-DV, and that they achieved a peak data rate of 3.09 Mbps. In a San Diego, California laboratory. CDMA2000 1xEV-DV combines data and voice, something UMTS does already. The CDG claims speeds up to 3.09 Mbps. Perhaps. Both DO and DV are backward compatible with CDMA2000 1X and cdmaOne.

In April 2004 Cingular became the first carrier in North America to offer UMTS. They now cover six markets in the United States. Acceptance is slow due to limited coverage, bulky handsets, and the high cost of service. UMTS and CDMA upgrades are very expensive for the carriers. Operators around the world are now spending billions for networks that won't pay for themselves for quite some time. The potential demand for service is certainly there, as cell phone subscriber levels attest.

In January, 2005 industry analysts Deloitte & Touche predicted mobile phone users will top 2 billion by the end of 2005. They say mobiles currently number over 1.5 billion. Many countries have over 100 % penetration, as people have second phones or multiple SIM cards, one for business, another for personal use. As throughout its history, regulatory, technical, and competitive problems remain for mobile telephony. But the desire for people to communicate, and for business to cater to that need, insures an imaginative and successful future for the mobile. What will the future look like? I'll leave that for the other authors in this issue to answer.

Tom Farley is a freelance telecom writer living in West Sacramento, California. Since 1995 he has produced the website privateline.com, an educational resource devoted to the telephone. Formerly publisher of the magazine private line, *Farley writes mostly on wireless and telephone history. Edwin Grosvenor, biographer of Alexander Graham Bell and the great grandson of same, has called Tom a telephone historian. Tom welcomes your comments and corrections.*

email: tom@privateline.com

[31] *Personal correspondence, Hanjoo Kim of the IITA (Korean Institute of Information Technology Assessment) July 13, 2004.*

Mobile phones and fashion in post-modernity

LEOPOLDINA FORTUNATI

Leopoldina Fortunati is Professor of Sociology at the University of Udine, Italy

The aim of this article is to explore the relation between mobile phones and fashion and how this has developed in the course of the last decade. It is important to reconstruct this path not only on the historical but also on the sociological plane, as it will enable us to ask certain questions and also reflect on modernity. After a careful reconstruction of the works that have appeared on this topic, we will dedicate the second part of this article to discussing the role and meaning of fashion and the mobile phone. We will deal with the relation that the mobile phone has at present with clothing and appearance (Floch, 1995; Tambini, 1997) and above all the human body (Fortunati, Katz, Riccini, 2003), by examining how this device is at the center of a vast spectrum of social processes, much vaster than any other technology. It is a fact that people tend to always carry a mobile phone on them that makes the difference in respect to other technologies. In the second section, we shall see how in the interaction between the mobile phone and fashion it is fashion that is the stronger element. It is fashion that domesticates technology, and especially the mobile phone, and not vice versa. In the third part of the article we will analyze the basic reasons for buying and using mobiles in order to see if these correspond to those of any other object of fashion. In the fourth section we will analyze the most important theories on fashion and try to see if the advent of the mobile phone as an object of fashion can be interpreted as being part of them or if it is necessary to produce a specific theory to explain its widespread use. Finally, in the fifth and last part of the article, the relation between mobile phones and fashion is seen as a lens through which to explore the concept of fashion itself, and the motivations that lie at the base of fashion as reflecting certain processes of modernity. This exploration will be guided by the idea that the mobile phone is the ICT that best represents and incarnates post-modernity (Lipovesky, 1987), an idea that will be verified here (Fortunati, 2005c). In the course of the article I will support the analysis with results that have emerged in two Italian research projects that explored how the new generations perceive and experience the encroachment of fashion on the world of the mobile phone.[1)]

A short review of the literature on the mobile phone and fashion

In 1996 in a research project on the social representation of telecommunications there emerged a significant association between fashion and the mobile phone (Fortunati, Manganelli, in press). In that same year fashion, mobile technologies and the body began to come under scrutiny, ending up in the book "Mass Moda" written by Calefato (1996), which looked into the co-penetration of the language of clothing and that of telecommunications.

The following year, 1997, the paper "Wearing Technologies" was presented at the Lyon Study meeting "Confluences: Fashioning Intercultural Perspectives", analyzing the statute of the mobile phone as part of physical appearance or the "look" of individuals. This paper, which was later published in the French journal *Réseaux* (Fortunati, 1998), is where scientific attention to this issue really starts. "Wearing Technologies", which was received with great interest in Europe, started from this significant association between fashion and the mobile phone and began to

1) *They are research projects at a quantitative level conducted in Italy by my students and myself on the mobile and fashion, which overall involved more than a thousand respondents. The first was administered in the spring of 2003 in north-eastern Italy in two secondary schools in two little towns, a classical lycée and a scientific one (Cianchi, D'Alessi, Fortunati, Manganelli, 2003). The data collected consist of 716 questionnaires with open and closed questions, which all the young people, between the ages of 14 and 21, answered (questionnaires valid: 714). The second was administered in two rural villages, again in north-eastern Italy, in an elementary school of Pasiano, in the secondary schools of Pasiano and Fiume and with a group of adolescents attending the high-schools of the provinces of Pordenone and Udine or already working (Danelon, 2005). The data collected consist this time of 371 questionnaires, again with open and closed questions, which all the young people of these schools answered. In this research were involved 153 children between the ages of 8 and 11, 174 boys and girls between the age of 12 and 14, and 40 adolescents from 15 years of age upwards. The data were analysed by means of descriptive techniques (frequency analysis) and the construction of contingency tables created by intersecting observed variables with socio-demographic variables such as gender, age, education, activity, status, geographical area, size of home town. Inferential analysis was conducted by means of* χ^2 *test, and when the statistics of* χ^2 *were significant, the analysis was developed by means of an examination of bi-varied log-linear models. The purpose was to discover the associations between variable modalities that are at the root of the significance of the relation shown by the* χ^2*. In some cases, t-test and models of one-way variance analysis were employed.*

 ISSN 0085-7130

look into the mobile's relation to the body and so to the development of the body's artificialization which this implied. In the same article it was prophesied that the mobile phone persisting on the surface of the body would necessarily fall under the dominion of fashion and as an object of fashion it would be obliged to match the individual's "look" (Floch, 1995). If this was the prophecy as to what was going to happen, the present was speaking in different terms. The first mobiles had caused such a break in social norms that they had triggered off strong hostility of the have-nots, who had ended up by immediately conferring on users a reputation for vulgarity.

In 1999 the entry "Technological clothing" in the Italian Encyclopedia "The Universe of the Body" appeared (Fortunati, 1999). In this entry, the mobile phone was analyzed as a cultural object that formed part of the shift from the artificial – make-up, tattoos, piercing, fleshing, etc. (Duflos-Priot, 1987; Borel, 1992) – to artificiality *in* the body – prostheses, therapeutic machines, etc. (Verni, 2002; Chiapponi, 2003). The mobile phone was seen as the most widespread of the worn technologies and the one that had opened the way to other electronic clothing technologies: luminous or sound accessories or details, both of a play-aesthetic nature – flashing brooches and bracelets, shoes that lit up – and of a sporting nature, such as step or speed gauges, and watches with multiple measuring parameters. At the same time, the mobile had, according to this approach, created a symbiosis between technology and clothing, which was inspired by poor art and the punk movement, and which had come up with various "technological" garments, such as for instance the Etro jacket with an attachment for charging a mobile phone, as well as the use of technological fabrics, such as metallic organdie, silk latex, and reflecting fabrics (Tortora, Eubank, 1998). The techno and robot style (Polhemus, 1994: 23) then led, still under the influence of the mobile phone and other ICT technologies, to the cyberpunk style, which was a mixture of space, technological and cyber styles, created with bits and pieces of industrial residue and high-tech materials, such as electronic circuits, holograms, and metal insertions.

In the workshop "Perpetual Contact: Mobile Communication, Private Talk, Public Performance" (Katz & Aakhus, 2002) at Rutgers University, this debate was taken further. The paper *Italy. Stereotypes: true or false* illustrated the growing role of the mobile phone as jewelry and/or a fashion accessory (Fortunati, 2002). In 2001 there was a study meeting in Milan entitled "The Human Body and Technologies, Communication and Fashion", which contained a more detailed discussion of the relation between communicative technologies and fashion (Fortunati, Katz, Riccini, 2002; Am. Tr. 2003). These themes were especially brought out in Rich Ling's contribution (2003) on the fashion and vulgarity dynamics in the use of the mobile phone by Norwegian adolescents, and Patrizia Calefato's contribution on the concept of "wearing communication" (2003). Other research on the social representations of ICTs and the human body, where fashion was used as a cue word, signaled the persistence of the fashion–mobile association (Contarello & Fortunati, in press). In the same year (2002), James Katz and Satomi Sugiyama presented – at the ICA conference in Seoul, Korea and at the conference entitled "Front stage/ Back stage. Mobile Communication and the Renegotiation of the Social Sphere," in Grimstad, Norway – two papers on a research project conducted on a group of students in the US and Japan on mobile phones seen as statements of fashion (Katz & Sugiyama, 2005).

In 2003 I began a research project with two students on fashion and mobile phones in two Italian high schools. I presented the first data from this research at the international conference held in Erfurt (7–9 October 2003) "Mobile Communication – Current Trends in Research". These data are set out in the paper *Der menschliche Körper. Mode und Mobiltelefone* [On the Human Body, Fashion and Mobile Phones] (Fortunati, 2005b). In this paper the analysis of the relation between mobile and fashion has been taken forward considerably. After a critical review of the most important theories on fashion (Balzac, 1830; Sombart, 1913; Simmel, 1919), I proposed to assume as an approach the co-construction of fashion and society. The mobile phone was now already seen as forming part of clothing, which is at the center of physical appearance, of "look". Following Flugel (1930), I also analyzed how the mobile phone forms part of the problematic of "transport of essential objects", which remains one of the most serious problems of fashion. Within this conceptual framework, I showed how the diffusion of the mobile phone might be analyzed also in terms of an overturning of the usual logic of fashion. And lastly I analyzed the three main tendencies that the mobile phone as a fashion object is developing: 1) becoming more and more an object of fashion; 2) disappearing in the various items of clothing; and 3) becoming a soft machine.

In the same year the paper "Fashion and Mobile Phones between Criteria of Purchase and Practices of Use" was presented in Milan at the conference organized by the Catholic University – Italian Association of Sociology "Fashion: Work of the Future" (13–14 November), where we discussed further data from this research conducted in Italy in two high schools (Cianchi, D'Alessi, Fortunati, Manganelli, 2003).

In short, from the work produced in the last decade the mobile phone seems to have gradually acquired the status of an object of fashion, considered as an accessory, and increasingly forming an integral part of an individual's look. That is, it is a part of that absolutely unnatural territory that encloses the material nature of the body and the cultural modernity of fashion (Troy, 2003). Not only is possessing a mobile phone a fashion, but it also appears as an object of fashion, ever more subject to the stylemes of fashion (Calefato, 1996; Fortunati, 1998; Fortunati, Katz, Riccini, 2003; Fortunati, 2005b). However, while for fixed technologies the problem of making them aesthetically attractive has been generally resolved by a good designer – for the mobile phone, which is an individual and mobile technology, the situation is more complex (Grenon, 1998). It is the very specific complexity of the mobile phone that makes its identity in respect to the fashion system still incomplete. The integration of the mobile phone in the look of an individual remains vague (it is often brought about by changing covers), because it is its original integration in the world of fashion that is bumpy and rough. Both because of mobile manufacturers' incapacity to understand the system of fashion and because of the indifference to the mobile phone on the part of fashion houses. Even its status as an accessory is uncertain. First of all because the mobile phone has not yet found its precise spatial collocation inside clothing (it can be kept in a pocket, in a bag, on a belt, in an inner pocket, in a rucksack or round the neck). Second, there is still an uncertainty as to whether it is an accessory and/or an ornament. To conclude, the mobile phone, while being a machine which has developed furthest from the form of a black box and while it has come closest to the world of fashion, is still an obscure object of scientific analysis. Our impression is that the re-semanticization of the mobile phone from technological device to fashion object is a process that has only just begun, while it has already brought about modifications in the accessories that contain it, such as bags and rucksacks.

What is fashion? What is the mobile phone?

Discussing the meaning of the phenomenon of fashion will enable us to better understand the meaning of the mobile phone as well. To say that it is a technology of information and communication means to stay at the level of pure description, as Donner underlines (2005). For a more complete analysis of the mobile phone it is necessary to understand its social significance. Elsewhere we have claimed that the mobile phone is a work instrument of reproduction (Fortunati, forthcoming). Here we wish to analyze recent developments of another aspect of its significance: its being an "object of fashion", placed upon the human body.

As an all-important social phenomenon, fashion is very complex, because it is made up of many elements and processes (Flugel, 1930). On the social plane fashion is born as a filter between the natural state of the body and its "civilization" (Amsterdamski, 1980) and socialization. The naked body is an undifferentiated body, which speaks just the basic language of gender, generation and ethnic group, while the social body is a clothed body, that is, a body that can give other information: from society of appurtenance to social class, from geographical provenance to profession, from the kind of hobby or sport practiced to personality and state of mind, etc. (Solomon, 1985; Kaiser, 1997). Bodies that speak abundantly facilitate social organization, in that they are more reassuring. They allow members of society greater mutual control and if this information then follows precise norms and ritualizations and keeps to socially shared rules, these bodies become even more reassuring. The clothed body is a historicised and social body, in that it has introjected the various normative regulations through which repression and regulation of sexuality have been elaborated and imposed (Foucault, 1976-84). The body clothed with a mobile phone is a body that articulates its distance from nature even more and which is set even more precisely in a particular social and communicative network. It is a body that is even more reassuring because it is a body that offers a large amount of direct and indirect information as to who is using it, and it is a political body, which shows a right to citizenship in respect to communication and can cope with the rules and communicative and informational requirements of post-modern society. It is also a body that is reassured, as when needed it can use the mobile phone as a security blanket (Wearing, 2002).

Fashion is a system of signs that is negotiated collectively and individually on the terrain of self-presentation. In addition, it gives a great contribution to self-presentation as by expressing itself over a wide expanse of body it offers one of the most important communicative codes. Fashion is one of the premises that enable us to form part of society, to be recognized and accepted. It is the terrain that is the necessary condition for very rapid mechanisms of recognition and social acceptance, and hence socialization, to take place. There cannot be coexistence and social cohesion without fashion. It is obvious, writes Simmel (1919: 34), that people who dress in the same way behave in a relatively homogenous way. Fashion, furthermore, is a social practice that also takes on a political significance, as it not only bears witness to, but can support processes of social democratization or strengthen class division and generational segregation. A powerful aid to fashion in self-presentation is the mobile phone, which supplies a very precious possibility of ritualization, given that the cigarette has lost much of its fascination in that sense (in the light of the damage that it has definitely done to the health of those who smoke actively and passively) (Stewart, 2003; Fortunati, 2005, in press).

On the structural plane, fashion is one of the communicative systems that pass through the portal of the "human body". We use this reductive but effective metaphor launched by Barry Wellman (2001) to signal how the body becomes a unit of localization of a complex network of communicative systems that go from the physical body itself (height, weight and so on), to verbal and non-verbal language (facial expressions, body relaxation, etc.), from paralinguistics, gestures and movement (kinaesthetics) to the spatial position and posture of the body (prozemics), from information to consumption, from advertising to technology, etc. (Fortunati, 2005a). Moreover, given that the body is the vehicle of consumption and use of the fashion system, it is further enriched by the vestimentary structure. Obviously, these communicative systems that come to form part of the portal/human body have in some way to converge with the general logic that supports it and integrates inside it. In this framework, fashion acquires a specific importance within the human body because this general logic can only pass through the language of fashion, since clothing – as we said – is the communicative system that is most extended over the human body. At a structural level, the mobile phone is a crucial part of the language of fashion: even if it occupies a minimal part of the space of clothing, it is a nomad on the body, which means that for example in the movement that goes from the pocket to the ear it crosses a long distance over the body and its clothing. Apart from this dynamicity of movement and direction, the mobile phone has put fashion's system of meanings into crisis. For example it has definitely disturbed the inanimate order of objects on the body, and produced important effects of disturbance in self-presentation (people who seem off their heads talking to themselves in the street). It has furthermore overturned the order of the intimate by stripping words of their privacy. We may say that on the one hand, fashion has made the mobile phone one of its objects but on the other, the mobile phone has created a great deal of disturbance in the world of fashion, obliging it for better or worse to come to terms with technology.
Its specific statute as fashion object is based on its mediation between strength of style and power of technology.

The motivations for fashion and the mobile phone

If the mobile phone is a fashion object, are the basic motivations behind its purchase and use the same as for any other fashion object, or are they more because it is, above all, a technology? To see how far the motivations that are generally attributed to fashion are valid also for the mobile phone, we shall discuss them starting from the results of the two research projects mentioned above.

According to numerous scholars, the motivations for fashion are variously magical, functional, protective, aesthetic, to do with play, discipline, identity, expressive, connected with modesty and the softening of social conflicts (Flugel, 1930; Veblen, 1899; König, 1971; Marchetti, 2004; Segre, 2005). To examine these, it is posited that fashion has a magical function, which was very strong in archaic societies and today is more subdued (Squicciarino, 1986). A trace of this magical aspect remains, for instance, in the perception of the mobile phone as a totem or talisman, which is to be found in the first mobile users (Fortunati, 1995a). Looking at survey material we asked children the criteria they would use when selecting a mobile phone. The material shows that "magic" is apparent, even if only in a few cases among our adolescents (2.3 %) and children (4.6 %). The functional meaning is also moderately important in clothing, but not dominant, as is mistakenly believed. Similarly, functionality is a minor element of the mobile phone, even if, being a young technology, it has only been under the dominion of fashion for the last few years. Fewer than half the children and adolescents (43.1 % and 48.2 %) give functionality as the reason that they would choose a particular model of mobile. Significantly, the ornamental/aesthetic dimension is more important for fashion objects (Rudofsky, 1972; De Carolis, 2003).

Aesthetics is an element which weighs much more than functionality when we look at a dress. This is the reason for which many elements of fashion (as the high heels, for example) have a perverse logic. They are elegant but they oblige the body to unnatural postures.

Specifically for the mobile phone, the importance of this dimension is given by 44.5 % of the children and by more than half of the adolescents (58.6 %). Thus we can see that in order of relevance, style has already overtaken functionality and "handiness" as a criterion of purchase for the mobile phone. Continuing the analysis of the dimensions of fashion, while for fashion objects the play dimension is generally considered quite a relevant element, for the mobile phone it is unexpectedly not very important. In fact only 28 % of the children and 8.5 % of the adolescents mention it.

Another strategic dimension of fashion is connected to modesty, in that as the various social regulations have imposed the repression and regulation of sexuality, the construction and imposition of modesty have made it increasingly more necessary, at least in western countries, to cover the genital organs by means of clothing. This has obviously led to a very special dynamic in covering and uncovering, of the clothed and the naked, which has gone to form part of social negotiations (Guindon, 1998). The actions of covering and uncovering, which are connected with a well known children's game (Fortunati, 1995b), crops up again in the mobile phone at a dialogic level, and supports Lurie (1981), who drew a parallel between the structure of language and that of clothing. With the mobile phone at the verbal, and not the visual, level the fact of saying and not saying, or half-saying, ends up by challenging modesty and discreetness. It summons an exhibition of the intimate sphere at an immaterial level. But the mobile phone has managed to strip words of their discreetness. In a visual parallel in the fashion world, in the 90s there was the appearance of the nude look and the confusion of over and under in clothing. These operations re-semantizised underwear (the under) as items of clothing (clothes under, bra and panties on show, etc.), and created the premises also for the public exhibition of private communication.

A further dimension that is generally attributed to fashion is its being a powerful vehicle in the construction of identity (Davis, 1992). In the past, fashion has assumed the strategic function of continually redesigning the structure of social identities. This is opposed to the continual devaluation of objects produced by fashion that are destined to rapidly becoming passé. In post-modernity however, this role of fashion as a producer of identity has been greatly modified. It is no accident that we speak in various places of the insufficiency of identity that distinguishes the contemporary individual. On the other hand, the drift of individual identity had in a certain sense already been foreseen before: by Hegel (1807), who called it an "empty tautology", and then Marx, who observed that individuals were forced to leave their singularity underdeveloped, because of the powerful processes of homogeneization, uniformity and standardization to which they were subjected, and finally by Baudrillard (1999: 58), who has recently defined identity as actually a "make of existence without qualities." Not by chance, adolescents, in the words of Lobet-Maris (2003), tend to use mobiles for creating specific tribes. Resorting to the commercial make of a phone as an instrument of definition is also confirmed also by our research. We find that 58 % of the children and as many as 66.2 % of the adolescents retain that the make of mobile is a very relevant element.

Social control must also be added to these dimensions of fashion. The mobile phone has as a part of its destiny to integrate with the possibilities of identification, surveillance and security. As we found in our analysis, 39 % of the children and 39.2 % of the adolescents declare that parents use the mobile phone to attempt to control them, while only17.4 % of the former and 13.5 % of the latter state that this control is effective. Fashion, the philosophers of jurisprudence teach us, has always been a "hypothetical norm." It is a norm with a compulsory force that is rather contained in respect for instance to categorical norms, in establishing "which formalities are necessary to carry out a certain act or task" (Bobbio, 1980: 888). But it is not because of this that it is less powerful. In concrete terms, fashion clothes the processes of indifferentiation of individuals and imposes on everyone the same social discipline and standardized self-presentation. A citizen who follows fashion is a citizen who recognizes the need for a social pact in its many expressions and forms. What is more, the power of control and discipline of social action on the part of fashion has been all the stronger the less rigid it has been as a norm. This was so at least until the 70s, when the power of fashion as "a hypothetical norm", able to guarantee a certain continuity of the social order by disciplining the masses, and inside them strongly disciplining some sections rather than others began to weaken considerably. Up to then, fashion had been an ideal form of control and discipline of the public sphere in democratic societies, both at the social and political levels. The mobile phone, in a certain sense, took its place in this role, carrying out a new kind of mass control, exercised this time at the level of communicative production. The millions of words that every day are uttered over mobile phones, the vast numbers of text messages and the growing number of MMSs, may be seen also as a quite sophisticated form of social control. Aren't those individuals, so intent on continually communicating among themselves – hindered by the restrictions of the spatialization of the word both oral and written in the mobile phone – perhaps involved in a terrible devaluation of the word itself and its meanings? Aren't they for this reason perhaps more easily controllable at a social level, distracted as they are by the communicative congestion that they create and with which they have to endure? Finally, don't these individuals transpose to their intense communicative activity the necessities of social action and the concrete management of class tensions and conflicts, gender, generation, ethnic group, and so on?

Instead, what at the level of research has so far been focused less on, and which is however also important, is the fact that fashion imposes above all a basic and widespread discipline on the body, movement and attitude to space (Flugel, 1930). Clothing influences the gait of the individual (high heels and tight skirts, for instance, certainly do not make for fast walking), the tempo and rhythm of gestures (narrow shoulders, for instance, slow down the movements of the arms), in other words it disciplines all the movement of the body. As a consequence, it conditions the psychological attitude of individuals in relation to the space that surrounds them and their capacity to act while they are in movement. The mobile telephone obviously forms part of this picture. It imposes its own specific discipline of the body, movement and attitude in relation to space (Nyiri, 2005) while also suggesting the necessity to extend the study of ergonomy to fashion as well. While in fact this discipline has developed greatly in the individual–machine interaction, it has not developed so far in the individual–clothing interaction.

Lastly, among the various dimensions of fashion we must not forget its role of social pacification. Fashion makes it possible to work off our impatience against social uniformity, to sublimate social tensions and to calm and elaborate class conflicts. What is there more suitable than fashion to guarantee a certain "singularity" (diversification) in a situation of basic "conformity"? Fashion enables us to set up a small-scale defense against serialization, which continually deprives us of our identity (Calefato, 1996). Certainly what fashion offers us is a slight margin of maneuver in constructing a personalized self-presentation, even if it does resort to serial products. The margin of maneuver is provided by the *ars combinatoria*; that is, how we compose and put together the various items of clothing, or how we individualize them. Also the mobile phone, as a fashion object, is subject to a certain degree of individualization: 40.1 % of the adolescents interviewed individualize it with a ring tone, 27.8 % with a cover, 15.2 % with inscriptions and 37.3% with a logo. Regarding the last three items we find that girls do this more often than boys. Children are even more taken with individualization: almost all of them declare that they do it in one way or another; 51 % with a ring tone, 39 % with a cover, 36 % with a logo, and 35 % with inscriptions and adhesives.

Ever since the 19th century, hardships and social conflicts have been exported on the terrain of fashion (Borrelli, 2000), entrusting to it the strategic function of managing them. This function of social mediation on the part of fashion developed further in post-Fordist society, when fashion became one of the most powerful mechanisms in its role as a social safety valve. It is indeed capable of appropriating cultural products elaborated by those social movements that invent new differences and languages and to redesign them and re-propose them within its collections and proposals. Fashion is the social system that, more than any other, is able to absorb and neutralize rebellions, contestations and movements, dismantling them from within their symbolic meanings, because it has made change and continual renewal its mission. The anti-fashion of today often becomes the fashion of tomorrow. Self-presentation and fashion are the terrains of continuous redefinition of power relations between the classes. It was not by chance that in the last decade, after the advent of the mobile phone, there was a de-emphasis on the public display of wealth. This has had the result that these classes find themselves, in a certain sense, pointlessly wealthy. That is, they possess a wealth that they cannot show off but rather they are forced to shelter themselves in separate zones, the famous VIP reserves.

In conclusion, this analysis has shown how the basic reasons for purchase and use of mobiles are more similar to those of a fashion object than to those of a purely technological device. This explains why the mobile phone is able to camouflage itself perfectly in the fashion system, in which it presents itself as the technological element, and why its identity as a fashion object prevails over its being a technological object. Its functionality is a necessary premise, but not sufficient: if beauty and style are not there too, even "handiness" moves to second place. This fact is destined to grow in importance also in the future, and become a central element in its diffusion and use.

Beyond trickle-down and trickle-up: or the theory of inversion

In this section I shall analyze the most important theories on fashion, trying to understand if the advent of the mobile phone as a fashion object can be interpreted within them, or if it is necessary to produce a specific theory to explain its diffusion. I shall also analyze three important points of the present dynamic of fashion – its relations with gender, the advent of casual clothes and the shift from clothing to skin – to try to see what can be the implications for a fashion strategy of the mobile phone.

From a sociological point of view, where is the phenomenon of fashion situated in the social structure? We might answer by remembering that for Balzac (1830) "clothing is the expression of society", while for Simmel (1919) fashion is able to make society: in fact clothes, as we said, mark gait, tempo and the rhythm of gestures; that is, they discipline movement.

We may maintain that if we combine the considerations of Balzac and Simmel, we arrive at the most pertinent way of describing the dynamics of the phenomenon, which is that there is a co-construction of fashion and society (Fortunati, 2005b). At times, or on certain occasions, it can be one or other that swings the needle of the balance in one direction or another, but substantially there is a dynamic between them that is explicated in their mutual influence and conditioning. This perspective of co-construction in fashion and society however, if it is correct as a theoretical perspective, must be developed by considerations of various order and degree. Especially in a moment in which traditional, "historical" interpretations of the phenomenon are flagging.

Let us remember Simmel's two most important ideas on the social mechanisms at the basis of fashion. The first idea is that which singles out fashion as the result of two opposing tendencies: on the one hand, the aspiration to fusion with one's own group (I dress in the same way as others) and on the other, individual distinction in respect to others (I dress differently from everybody else to express my individuality). From our research projects on the mobile phone it emerges that in the choice of this or that model of mobile the strongest motivation is "being different from others" in 33.7 % of cases for the adolescents and in 36 % for the children, while the motivation "to be like others" is weaker (22.8 % for the adolescents and 25 % for the children).[2] This means that Simmel's theory is able to account for only little more than half of the responses. But this result is not surprising because the social movements from the 60s onwards have undermined from inside this mechanism, and broken the rigidity of the separation between the classes in self-presentation. The figures given above suggest that if today there remains a certain fusion, it is rather in social indistinctness, where what varies is a determined individual or group expression that, if anything, is revitalized in post-modern society. Also Simmel's other idea (shared by Flugel, 1930 and Bourdieu, 1979) on the dynamics of fashion that filters from the higher classes to the lower, the famous trickle-down, does not work any longer (Segre, 2005). Again from the 60s, the lower and middle classes have stopped recognizing in the higher classes the capacity to construct taste and style for the whole of society. Because of this they have stopped imitating them. Rather, they have started to create their own originality and to express their own style, to the point of taking in hand the production of new fashion ideas completely. Today in fact we speak of trickle-up, by which we mean the complete overturning of this dynamics. Ted Polhemus (1994) has coined the term "street fashion", and highlighted the new creative role assumed by the lower classes as regards fashion. Even the public space and the media space are dominated by the middle-lower classes, seeing that the system looks to large numbers in that they amplify the possibility of profit.

Inside this evolution of fashion, the mobile phone has come to represent a specific dynamic that we propose to call "theory of inversion." In fact it has been a typical case of trickle-down, as its adoption and public use started from the affluent classes, but was stigmatized as vulgar. The higher classes, who had shown with clothes that they were no longer the driving force of style, in the precocious adoption of the mobile phone were considered actually the bearers of vulgarity, as illustrated by Ling in a study on Norwegian adolescents (2003). Only after it spread among the masses did the mobile phone come to be considered a trendy and fashionable object (Fortunati, 2005b). The advent of the mobile phone and its rate of diffusion therefore was a sort of explosive short-circuit which marked a phase of trickle-down characterized however by the stigmatization of vulgarity, of demodé behavior, with a phase of trickle-up, in which only the massification of the spread of the mobile phone marked the possibility of grafting it on to fashion. It was this accident that was one of the reasons that in my opinion contributed greatly to pushing the higher classes towards a form of social self-segregation.

Let us pass now to analyze three topics that may have very important implications for a fashion strategy as regards the mobile phone: gender, casual clothes, and the new role of the skin. Let us start from gender. It must be said first of all that there has been an epochal overturn in the interest of men and women as regards fashion. Women, who historically were most involved in fashion, have taken their distance from it, while men have progressively come closer. This change of direction by women took place on the wave of the mass refusal of domestic work connected with creation, production and upkeep of clothing, which ended up by causing tertiarization (Fortunati, 1994).

---

[2] *It must however be added that in the case of adolescents there are also 16.1 % of them who respond "because I like it", while one fifth of respondents give answers that are lost because they are all different. So, on the whole, around one half of adolescents choose their model of mobile conditioned by their peer group, whereas a little short of one fifth choose a mobile centering their motivations on themselves. In the case of children, there are 14 % of answers that are lost because they are all different, but there are also 25 % of children who don't know which answer to give. It is also significant that it is girls more than boys who lead in this centering the choice of mobile on themselves and their own tastes rather than on others.*

This change of direction regards the mobile phone very closely, seeing that it is the intersection between two opposing tendencies: the masculinization of fashion and the femininization of technology, and that therefore it is one of the places where the social construction of masculinity and femininity is modified. It must however be specified that the changes registered in relation to the mobile phone are still somewhat contradictory. On the one hand, the research on adolescents that we have used to present some figures is the first research in which girls tend in Italy to have more mobiles than boys (96.5 % vs 93.4 %). On the other, these changes are slow: in fact, when they choose a mobile, it is today the girl adolescents as opposed to their fellow male adolescents who are significantly more interested in the aesthetic appearance of the mobile phone (54.5 % vs 40.6 %), in wanting a mobile phone model because it is attractive (41.1 % vs. 32.6 %) or considering the appearance of the mobile phone as important (62.6 % vs. 52.4 %). The difference however, even if it is significant in respect to male adolescents, is not huge. Obviously a fashion strategy as regards the mobile phone must absolutely keep account of these processes and their continuous internal dynamics, otherwise they risk being impoverished and will re-propose old stereotypes.

Another very important change in the field of fashion has been the diffusion at a mass level of casual clothes, which is made up of items of clothing that communicate at a very low level. They provide an advantage to those who wear them, because they respect their privacy and support their resistance to speaking of themselves. The meaning of the sign shifts from the information plane to the expressive or chance one, all to the detriment of those who have to decodify them. It is as if there had been set up a new social pact, in the sense that we renounce knowing about others in exchange for our own and others' freedom to not say anything about our/themselves. We accept having very little information about those who walk by our side. Fashion, from a means of regulating a social relationship, thus becomes a kind of interlocution with the inanimate, is a relationship that is set up between oneself and the fashion object, the item of clothing. There is therefore a stronger tendency to de-communicate and therefore to the underdevelopment of sociability in favor of a strong theatricalization of the body. The mobile phone has rooted itself perfectly in this picture, expanding these tendencies at the verbal level (Katz, 2003): beside an overabundance of purely phatic or instrumental calls, there are short calls, which are therefore less communicative, and SMS and MMS messages, which are also "poor" communicatively (Fortunati, 2002). The mobile phone welds perfectly with the casual style at the vestimentary level because it represents the casual style at a communicative level.

Lastly, there has been in the past a tendency in fashion to transfer the meaning of symbols and meanings from the skin to the garment. Today the direction has begun to go from the garment to the skin (Borel, 1992). It is on the skin that the motives and designs of fabrics go; it is again on the skin that jewels and various ornaments penetrate. Today it is the body that makes fashion and which is fashion. Clothes are no longer packaged, but it is bodies that are "touched up", if not re-done, according to the kind of body that is in fashion at the time. The body is exhibited and becomes a mass theatrical costume (Fortunati, 2003). The social construction of the body from being a mediatic, publicity and political strategy with its myths and models becomes a mass practice, continuing with a large number of techniques and technologies: from plastic surgery to hormonal therapies, etc. In this construction of the body the mobile phone has for now limited itself to staying on the surface. However, some are experimenting with the use of subcutaneous microchips to put the body on line. While jewelry is increasingly used for piercing the body and tattooing is incorporating the patterns of fabric by being imprinted permanently, the mobile phone, however, remains on the surface of the body.

Fashion, modernity and the mobile phone

In this last section I wish to verify the hypothesis that the mobile phone is the technology that best represents post-modernity. Post-modernity – we can say – is based among other things on the hypertrophy of two social phenomena: fashion and information. The mobile phone, as we shall see below, actually includes them and carries them both. Let us begin from Simmel's famous hypothesis (1900: 41), according to which fashion is a result of modernity. "The lack of something defined in the core of the soul," he writes, "leads us to search for temporary satisfaction in ever newer stimuli, emotions, external activities; and we are thus enveloped in that confusion, instability, permanent anxiety which manifests itself sometimes in the bustle of the big towns, sometimes as an urge to get away from it all, sometimes as the uncontrolled fury of competition, sometimes as that typically modern phenomenon which is inconstancy in the field of tastes, styles, conventions and personal relations."
I agree with Simmel that fashion is a dimension of modernity, but in my opinion it is in the post-modern world in particular that it finds its greatest triumph. Post-modernity takes the importance of fashion further, making it the motor that drives the dynamics of life style (Fortunati, 2003) and bringing more and

more spheres under its dominion. Fashion, which until a few decades ago was all about clothing and furnishing, has extended its tentacles today over the whole of society, over the whole social body (Veblen, 1899). The same attitude of "inconstancy in the field of tastes, styles and conventions" of which Simmel speaks, has extended transversally to many other spheres of society, including that of scientific discourse. Here, too, there are themes or topics that in a given period are in fashion and then go out of fashion and are no longer given any attention. So fashion has taken control of unsuspected sectors and imposes its logic everywhere. The continuous need/desire of change imposed by fashion requires of course a higher level of work. Think, for example, of a certain modality of fashion (ephemeral) in the scientific debate. The very speed in the change of topics, methods, or authors cited etc. pushes scholars to work more intensively in order to be able to catch up. So, the result is that there is more productivity in the quantity of the work done, but perhaps the results have content that is not as seriously developed.

But the unarrestable advance of fashion in society has always obviously had its hard core in clothing. However, fashion has not limited itself to governing the clothing sector, but has modified the very concept of consumption of clothing and relative practices (König, 1971). The purchaser/consumer has been completely reinvented, in that from a certain point he/she has no longer been considered a bearer of needs or desires, but as a possible purchaser, for whom it was necessary to create needs surreptitiously and continually (Baudrillard, 1970). He/she has therefore been trained to a nervous and transitory consumption of objects destined to being discarded well before they are really worn out. The subjection of the consumer takes place however also by intervening at the level of supply, in the sense that up to then the consumer has been placed in a situation in which he/she cannot but be subjected to fashion. For the consumer it becomes difficult if not impossible to purchase items of clothing that are not fashionable. Even the so-called "classic" item, which by definition is that destined to pass over the continual changes of fashion, because it refers to the basic linear structures of a certain historical period, is even so subject to corrections and variations that bring it up to date.
In this context, the personal taste of an individual is completely neutralized: if he or she loves yellow and in a given season black and fuchsia are in fashion, rarely will he/she manage to dress in his or her favorite color. He or she will have to choose among the colors that are fashionable at the time. The same is true also for the typology of forms: if a woman loves drainpipe trousers in a period in which wide trousers are in, very likely she will not be able to find a pair of trousers to her taste. The same can be said for the mobile phone: even if someone wanted to buy a primitive mobile (as were the first Tacs) they can no longer do so, because they no longer exist on the market. Individuals under the regime of fashion are obliged to modulate and adapt their personal tastes to the supply that the industrial clothing sector proposes.

In the re-invention of the consumer, fashion acts as a pathfinder also in relation to sectors with a stronger industrial pedigree. Let us take the example of household appliances. While immediately after the war they were made to have certain duration, today they are conceived and proposed as technological objects of half the duration or reduced to one third, for the precise reason that purchasers/users are used to having a relation of use that is much shorter with objects. In general, the ever shorter life cycle of goods reflects a relation with progressively more oblique values of use. It is as if objects presented a core of usability of short duration, surrounded by structures and peripheral parts that are made superfluous by a programmed limited life. The tendency of the post-modern individual to hurriedly consume the core of usability of objects seems to indicate a sort of vampire-like relationship with them. In the relationship with objects consumers behave as if they were sucking the vital lymph and then throwing away the carcass. The populations of industrialized countries manage to use 80 % of world resources, only because they have this sort of relation with use values. General consumption, organized in this way also thanks to fashion, inevitably leads to the exploitation, increasingly short, of the core of objects. It is interesting to observe that faced by the evanescent duration of objects, our average life span is becoming longer and longer. In general, the lower durability of the exchange values that are bought, deriving from a lowering of the quality of the materials used to produce them, is acceptable to the eyes of a consumer accustomed to the seductive handling of novelty on the part of fashion. So fashion grafts and imposes on to the sphere of the everyday a different philosophy of consumption and a relation of a perverse kind with objects. It finds in this a powerful ally in consumers themselves, whose psychological center is increasingly fragile, with the result that they produce changing attitudes and behaviors. This spasmodic tendency for a "use without consumption" and thus to an irrational use involving enormous waste, also ends up by limiting affective and symbolic investment in objects. Far from being dear travelling companions to look after with love and attention in the course of their existence, objects come to play a role of sacrificial victims hurriedly sacrificed on the altar of continual change imposed by fashion.

Consumption of the mobile phone is reflected perfectly in this picture (Gaglio, 2004). We consume and rapidly get rid of our mobiles to pass to a more attractive model or with new functions and services. The advance of fashion in the world of mobile phone manufacture as a material object, by means of the implementation both of an ever more sophisticated design (an adjective that not accidentally fashion shares with technology) and new software, functions, services and contents, actually has the underlying aim of progressively shortening the mobile's life span. The elaboration and implementation of the Siemens's Xelibri collection, for instance, had the declared aim of reducing the average life span of the mobile phone from 18 months to 12 months. Just to give an idea of how often the mobile phone is replaced in practice, in the already cited research we find that 29.9 % of the children have already had more than one mobile in their short lives. The other important objective of the advance of fashion and the ever-richer technological supply is to lead to pluri-purchasing. Among the adolescents interviewed 9 % declare they have had more than one mobile, while among the children 6 % declare they have more than one.

There is another more general discourse that regards fashion and as a consequence also the mobile phone as a fashion object, and it is that relative to fashion's function of topicalization. In post-modernity, fashion, together with information, manages the topical, giving a rhythm to social life and marking, also exteriorly, the passing of time, with the introduction of continuous changes. That is, fashion and information give individuals the illusion of following the movement of the world. As life is movement, the continual change produced by fashion and information gives post-modern individuals the sensation of being alive. It makes them feel connected to life, especially in a period in which on the one hand, the increase of the inorganic in the human body is causing a lowering of its vitality, and on the other hand, the re-semanticization of the categories of space and time, conducted by the doubling of the real on the part of the media and technologies of communication and information, confounds perception of what is a direct relationship and what is virtual (Virilio, 1993; Maldonado, 1998; Fortunati, in press). Fashion and information, in other words, are the two priestesses that officiate at the altar of the continual conquest of historicity, which Hegel (1807) called *Zeitgeist*.

Fashion gives rhythm to the constant updating of self-presentation, made compulsory by the need to show both the implementation of the continual adaptation to changes in age and the march of life, and the capacity of rapidly connecting with social changes. For its part, information imposes the cadenced and in a certain sense obligatory consumption of a particular form of knowledge: news. Those who see fashion and information as two secondary elements of social life are mistaken. On the contrary, they are bulwarks inside the forces that build – as said before – social cohesion. Social cohesion begins in fact from the recognition of the other as an individual "conforming" to certain dictates, above all, to those of clothing launched by fashion. In a recent research project on social representations of the human body and of technologies of information and communication (Contarello, Fortunati, in press), from an exercise of free association it emerged that fashion, apart from being perceived as closely connected with the human body, is experienced by the 280 respondents (a convenience sample composed of university students of various faculties) on the whole more as constriction and conformity than as freedom and difference.[3)]

Social cohesion presupposes, at the same time, knowledge of the topical as of that flow of information that becomes a minimal, and "obligatory", updating, of the "contemporary history" of the world. Fashion and information become strategic territories inside which social cohesion is played out, because they create a new that is old, an originality that is conformism, a difference that is sameness. They are the ideal spheres for creating artifacts that give the illusion of a possible identity. In this context, the mobile phone is the object of fashion that is also able to produce and deliver information, newscasts, breaking news, and photograph important events. The object that at a maximum level incarnates the fusion between fashion and information is the mobile phone, which couples to elegance of design technological capacities that make rapidity of information possible. But where exactly does this need of the post-modern individuals to "actualise" themselves spring from? Certainly from this a-spirituality that leads to a decentering of the ego, as underlined by Simmel, for which reason, alienated to ourselves and our lives, we have a continual and obsessive need to have the world narrated to us, the context in which we live. How central and fundamental to the social politics of post-modernity fashion and information are, emerges clearly from the fact that to allow both to carry out their function of marking the progress of time, society is willing to waste a large amount of energy and resources. Nietzsche (1967), who well understood the

3) *And to be precise it is lived in terms of organisational structure (stylists, catwalks, models, trends), or consumption (store, shopping, élite and money), positive values (imagination, beauty and elegance), but even more negative (appearance, uselessness, stupidity, conformism, anorexia).*

great amount of resources that individually and socially we are prepared to use in order to actualize ourselves, invited everyone to become unactualized; that is, to learn to live in a meta-historical condition.

The capacity for actualization that fashion shares with information is especially important in this historical moment, marked by the development of the information society. Actualization is a social process and a crucial point of reference, because it is the sole element that is able to continually reconstruct a specific shared temporality. Any site, for instance, can show that it is up to date, by giving the day's news, and its date shows that it is live. In the same way as anyone can show that they can live their times by simply wearing a fashionable garment. Fashion and information are strategic elements in the social management of the spatialization of time. They work as a foil to the virtual world that is developing with fixed and mobile networks. In this context, it is the topical that gives the virtual its connection with reality, because it gives it proof of its existence (Fortunati, 2005d). In other words, the need to topicalize is born from the dematerialization of the real, thanks to the creation of the double of the world (Baudrillard, 1999), as well as the geometric increase in immaterial work and immateriality. If we take any good, the percentage of immaterial labor needed to produce them is ever higher. The cost of their manufacture is small because most of the overall cost is connected with design, distribution, communication, packing, advertising, and so on. The same process occurs also as regards individuals. In the reproduction of individuals, the quota of immaterial work needed to reproduce them is higher and higher: psychological work, communicative, educational, affective, emotive, sexual, instructional, and, importantly also "packaging". Fashion may be considered from a certain point of view as the "packaging" of the individual, that is, as the way in which the individual presents him/herself to potential "purchasers", and the mobile phone as the most important technological component in this packaging.

Conclusions

Following on what has been said, we may conclude that the reasons at the basis of purchasing and using the mobile phone among young people are more similar to those of a fashion object than those of a pure device. Fashion has the better of the mobile phone, in the sense that it appropriates it as an object, but above all it makes it work in the framework that it constructs. Or better, the mobile phone is the technology that has expressed the highest mimetic capacity as regards the system of fashion. This extraordinary capacity of the mobile phone of adaptation derives from the symbiosis with fashion in which it has had to live on the surface of the human body. To understand the spread of the mobile phone traditional theories as to fashion are worth little, so we have proposed the "theory of inversion".

Lastly, the mobile phone, as a technological symbol of prestige, the place of evaporation of meanings, object of fashion and source of information and entertainment, has emerged as the emblem of post-modernity. In 1911 Sombart (p.141) underlined how the special "spirit" of a technology corresponds to general principles on which the technology is based. Of the various technologies, the mobile telephone is definitely the device that allows the individual to better qualify his/her self-presentation as a hybrid element, fusional, a revealer of unknown places (where are you?), which has catalyzed technology, fashion and information. As such, it is the technology that most contributes to forming that atmosphere determining the fundamental orientations of our age. In other words, to use an expression of Hegel's, it represents the spirit of the age, and more precisely, the spirit of post-modernity.

References

Amsterdamski, S. 1980. Naturale/artificiale, [Natural/Artificial] entry, *Enciclopedia Einaudi*, vol. 9. Torino, Einaudi.

Balzac, H. 1830. *Traité de la vie élégante* (It. tr. *Trattato della vita elegante*. Milano, Longanesi, 1982).

Baudrillard, J. 1970. *La société de la consommation*. [The Consumption Society] Paris, Denöel.

Baudrillard, J. 1999. *L'échange impossible*. [The impossible exchange] Paris, Éd. Galilée.

Bobbio, N. 1980. *Norma*, voce, Enciclopedia, vol. IX, Einaudi, Torino, 876–907.

Borel, F. 1992. *Le vêtement incarné : les metamorphoses du corps*. [The embodied dress : the metamorphoses of the human body] Paris, Calmann-Lévi.

Borrelli, D. 2000. *Il filo dei discorsi. Teoria e storia sociale del telefono*. [The Thread of discourses. Theory and social history of the telephone] Roma, Sossella.

Bourdieu P. 1979. *La distinction: critique sociale du jugement*. [Distinction: Social critique of judgement] Paris, Editions du Minuit.

Calefato, P. 1996. *Mass Moda. Linguaggio e immaginario del corpo rivestito*. [Language and imaginary of the dressed body] Genova, Costa & Nolan.

Calefato, P. 2003. Wearing Communication: Home, Travel, Space. In: Fortunati, L, Katz, J, Riccini, R (eds.) 2003. *Mediating the Human Body: Technology, Communication and Fashion*. Mahwah, Erlbaun, 163–168.

Chiapponi, M. 2003. Health Care Technologies: The Contribution of Industrial Design. In: Fortunati, L, Katz, J, Riccini, R (eds.) 2003. *Mediating the Human Body: Technology, Communication and Fashion*. Mahwah, Erlbaun, 187–194.

Cianchi, A, D'Alessi, F, Fortunati, L, Manganelli, A. 2003. *Moda e Cellulare tra criteri d'acquisto e pratiche d'uso* [Fashion and mobile phone: criteria of purchase and practices of use]. Paper presented at the conference "Cultura: lavoro del futuro" [Culture: work of the future], Milano, 13–14 November.

Contarello, A, Fortunati, L. In press. ICTs and The Human Body. A Social Representation Approach. In: Law, P, Fortunati, L, Yang, S (eds.) *New Technologies in Global Societies*. Singapore, World Scientific Publisher.

Danelon, E. 2005. *Il cellulare: solo un mezzo di comunicazione?* [The mobile phone: just a means of communication?] Tesi di laurea: relatore, prof. Leopoldina Fortunati.

Davis, F. 1992. *Fashion, Culture and Identity*. Chicago, University of Chicago Press.

De Carolis, P. 2003. Cravatte, jeans, boxer: le insidie (nascoste) degli abiti [Ties, jeans, boxer shorts: the dangers (hidden) of clothes]. In: *Il Corriere della Sera*, 7 agosto 2003.

Donner, J. 2005. *Research Approaches to Mobile Use in the Developing World: A Review of the Literature*. Paper presented at the "International Conference on Mobile Communication and Asian Modernities", City University of Hong Kong, 7–9 June.

Duflos-Priot, M-T. 1987. Le maquillage, séduction protocolaire et artifice normalisé. In: *Communication*, 46, «Parure, pudeur, étiquette», 245–254.

Floch, J-M. 1995. L'intramontabile look di Coco Chanel [The Eternal look of Coco Chanel]. In: Ceriani, G, Grandi, R (eds.) *Moda: regole e rappresentazioni. Il cambiamento, il sistema, la comunicazione*. Milano, Angeli, 154–170.

Flügel, J C. 1930. *The psychology of clothes*. London, International Universities Press. [Trad, it., 1974, *La psicologia dell'abbigliamento*, Milano, Franco Angeli].

Fortunati, L. 1994. *Fashion and Invisible Work*. Report presented at the session "Aspects of fashion" of XIIIth World Congress of Sociology, entitled Contested boundaries and shifting solidarities – ISA, Bielefeld, 18–23 luglio.

Fortunati, L (ed.) 1995. *Gli italiani al telefono* [Italians on the phone]. Milano, Angeli.

Fortunati, L. 1995. *I mostri nell'immmaginario* [Monsters in the Imaginary]. Milano, Angeli.

Fortunati, L. 1998. Revêtir des technologies. In: *Réseaux*, 90, juillet-aout 1998, 85–91.

Fortunati, L. 1999. Technological Clothes, Entry of the *Encyclopedia "L'Universo del Corpo"*, vol. II. Rome, Istituto dell'Enciclopedia Italiana, 11–12.

Fortunati, L. 2002. Italy. Stereotypes: true or false. In: Katz, J, Aakhus, M (eds.) *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*. Cambridge, Cambridge University Press, 42–62.

Fortunati, L, Contarello, A. 2002. Social Representation of the Mobile: An Italian Study. In: Shin Dong Kim (ed.) *The Social and Cultural Impact/Meaning of Mobile Communication*. Chunchon, Hallym University, 87–100.

Fortunati, L, Katz, J, Riccini, R (eds.) 2003. *Mediating the Human Body: Technology, Communication and Fashion*. Mahwah, Erlbaun.

Fortunati, L. 2003. La moda e il viaggio nel 'mondo globale [Fashion and travel in the global world]. In: Valli, B, Barzini, B, Calefato, P (eds.) *Discipline della moda. L'etica dell'apparenza*. Napoli, Liguori, 85–100.

Fortunati, L. 2005a. Is Body-to-Body Still the Prototype. In: *The Information Society*, 21 (1), 53–62.

Fortunati, L. 2005b. Der menschliche Körper. Mode und Mobiltelefone [On the Human Body, Fashion and Mobile Phones]. In: Höflich, J R, Gebhaedt, J (eds.) *Mobile Kommunikation. Perspektiven und Forschungsfelder*. Frankfurt am Main, Peter Lang, 223–248.

Fortunati, L. 2005c. *The mobile phone, the principle of fusion and the constructing of the double of the world*. Paper presented at the workshop "Mobile Communication & the Network Society", Rutgers University, New Brunswick, NJ, USA, May 21.

Fortunati, L. 2005d. The mediatization of the net and the internetization of the mass media. In: *Gazette. The International Journal for Communication Studies*, 67 (1), 27–44.

Fortunati, L. 2005. Mobile phone and the presentation of self. In: Ling, R, Pedersen, P (eds.) *Mobile communication and the re-negotiation of the social sphere*. London, Springer.

Fortunati, L, Manganelli, A. In press. *The Social Representation of Telecommunications, Personal and Ubiquitous Computing*.

Fortunati, L. Forthcoming. *User Design and the Democratization of the Mobile Phone*. Paper presented at the international conference "Command Lines: The Emergence of Governance in Global Cyberspace", Milwaukee, April 29–30, 2005.

Foucault, M. 1976–1984. *Histoire de la sexualité* [History of sexuality], 3 volls. Paris, Gallimard.

Gaglio, G. 2004. Introduction to the special issue «Nouvelles technologies et Consommations», of the electronic journal *Consommations et sociétés*, 4. Available at www.consommations-sociétés.net.

Grenon, N. 1998. Il design degli italiani: dal rito della caffettiera al mito del cellulare [Italian design: from the ritual of the coffee-pot to the myth of the mobile phone]. In: Calabrese, O (ed.) *Il modello italiano: le forme della creatività* [The Italian model: the forms of creativity], Milano, Skira, 129–150.

Guindon, A. 1998. *L'habillé et le nu. Pour une éthique du vêtir et du dénuder*. Ottawa, Cerf.

Hegel, G W F. 1807. It. tr. *Fenomenologia dello spirito*. Milano, Rusconi, 1995.

Kaiser, S B. 1997. *The Social Psychology of Clothing: Symbolic Appearances in Context* (2nd ed. revised). New York, Fairchild.

Katz, J, Aakhus, M. 2002. *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*. Cambridge, Cambridge University Press.

Katz, J, Sugiyama, S. 2002. *Fashion and Mobile Phones*. Paper presented at the International Communication Association Annual Meeting, Seoul, Korea, July 21.

Katz, J. 2003. A Nation of Ghosts? Choreography of Mobile Communication in Public Spaces. In: Nyiri, K (ed.) *Mobile Democracy. Essays on Society, Self and Politics*. Wien, Passagen Verlag, 21–32.

Katz, J, Sugiyama, S. 2005. Mobile Phone as Fashion Statements: The Co-creation of Mobile Communication's Public Meaning. In: Ling, R, Pedersen, P (eds.) *Mobile communications: Re-negotiation of the social sphere*. London, Springer.

König, R. 1971. *Macht und Reiz der Mode*. Düsseldorf und Wien, Econ Verlag.

Ling, R. 2003. Fashion and Vulgarity in the Adoption of the Mobile Telephone Among Teens in Norway. In: *Mediating the Human Body*, 131–143.

Lipovesky, G. 1987. *L'Empire de l'ephémère : la mode et son destin dans les sociétés modernes*. Paris, Gallimard.

Lobet-Maris, C. 2003. Mobile Phone Tribes: Youth and Social Identity. In: *Mediating the Human Body*, cit., 87–92.

Lurie, A. 1981. *The Language of Clothing*. Henry HoltPublishing.

Maldonado, T. 1998. *Reale e virtuale*. [Real and virtual] Milano, Feltrinelli.

Marchetti, M C. 2004. *Moda e Società* [Fashion and society] Rome, Meltemi.

Nietzsche, F. 1967. *Richard Wagner a Bayreuth e frammenti postumi*, vol. IV, I delle Opere. Milano, Adelphi.

Nyiri, K. 2005. *A Sense of Place. The Global and the Local in Mobile Communication*. Wien, Passagen Verlag.

Polhemus T. 1994. *Street Style*. London, Thames & Hudson.

Rudofsky, B. 1972. *The unfashionable human body*. New York, Doubleday.

Segre, S. 1999. *Mode in Italy*. Milano, Guerrini.

Simmel, G. 1900. Die Herrschaft der Technik. In: Philosophie des Geldes. Leipzig, Dunker und Humblot, 520–535 (It. tr. Il dominio della tecnica. In: T.

Maldonado, T (ed.) *Tecnologia e cultura. Il dibattito tedesco fra Bismark e Weimar*. Milano, Feltrinelli, 1991, 37–46. [Technology and Culture. The German Debate from Bismark to Weimar]

Simmel. G. 1919. *Philosophische Kultur* (Tr.it. La moda e altri saggi di cultura filosofica. Milano, Longanesi).

Solomon, M R. (ed.) 1985. *The Psychology of Fashion*. Lexington, MA, Lexington Books.

Sombart, W. 1911. Technick un Kultur. In: "Archiv für Sozialwissenschaft und Sozialpolitik", 33, 305–347 (tr.it. 1991, Tecnica e cultura. In: Maldonado, T (ed.) *Tecnica e cultura. Il dibattito tedesco fra Bismark e Weimar*. Milano, Feltrinelli, 137–170).

Sombart, W. 1913. *Luxus und Kapitalismus*. München, Duncker & Humblot Verlag.

Squicciarino, N. 1986. *Il vestito parla*. [The dress speaks] Roma, Armando.

Stewart J. 2003. *Mobile phones: Cigarettes of the 21st Century*. http://www.rcss.ed.ac.uk/people/jamess.html

Tambini, M. 1997. *Il look del secolo* [The look of the century]. Milano, Mondadori.

Tortora, P, Eubank, K. 1998. *Survey of Historical Costume*. New York, Fairchild Books.

Troy, N J. 2003. *Couture Culture. A Study in Modern Art and Fashion*. Cambridge, MA, MIT Press.

Veblen, T. 1899. *The Theory of the Leisure Class. An economic study in the evolution of institutions*. New York, Macmillan/Macmillan.

Verni, G. 2002. Le protesi di arto inferiore e la loro evoluzione. [Prostheses of the lower limbs and their evolution] In: Fortunati, L, Katz, J, Riccini, R (eds.) 2002. *Corpo futuro. Il corpo umano tra tecnologie, comunicazione e moda*. Milano, Angeli, 284–294.

Virilio, P. 1993. *L'art du moteur*. Paris, Éd. Galilée.

Wearing, S. 2002. *Mobiles a confidence booster for young women*. Available at: http://www.uts.edu.au/new/releases/2002/July/24.html

Wellman, B. Physical place and cyberspace: The rise of personalized networking. *International Journal of Urban and Regional Research*, 23 (1), S3–43.

Leopoldina Fortunati is Professor of Sociology of Communication and Sociology of Cultural Processes at the University of Udine. She is the author of many books and articles and associate editor of the journal The Information Society, *she is on the advisory board of the journal* New Media and Society, *and serves as referee for the journal* Communication, Information, Society *and* Journal for the Theory of Social Behaviour. *She represents Italy in the COST Technical Committee for Social Sciences and Humanities. She was part of the European research project SIGIS – "Strategies of Inclusion: Gender and the Information Society" and of several COST Actions. She is the co-chair of the International Association* "The Society for the Social Study of Mobile Communication" *(SSSMC). Her works have been published in nine languages: Chinese, English, French, German, Italian, Japanese, Korean, Russian, Spanish.*

email: fortunati.deluca@tin.it

Interest in future net-based services for a sample of Norwegian interviewees

RICH LING

Rich Ling is a sociologist at Telenor R&D

This paper examines interest in advanced mobile services among Norwegians. It is based on a survey of 1000 randomly selected Norwegians that was carried out in June of 2004. The survey examined Norwegians' interest in services such as mobile instant messaging, access to time and place relevant information, use of the device to identify the user in various situations, use of the mobile phone as a type of electronic wallet, and the possibility to remotely control equipment in the home. The analysis shows that there is a split between those services that facilitate social interaction and those that are more oriented around facilitating person-machine interaction. The analysis here indicates that there is interest in the more person-machine oriented services.

Introduction and method for the survey

In this article I will describe the findings from a survey of a random sample of Norwegians with regard to their expressed interest in network-based mobile services. Net-based services are those wherein a user profile of some type is stored in the net and is available to the user in different situations. This analysis is based on a sample of 1000 randomly selected Norwegians[1)] over the age of 12. The survey was carried out via the telephone in June of 2004.

The material examined here constituted a single battery of eight questions that focused on future services that are generally net-based. In addition to this battery of questions the questionnaire covered 1) access and payment, 2) use of voice, SMS and MMS, 3) use of WAP services, 4) characteristics of the handset, 5) use of some PC-based services and the respondent's socio-demographic status. The analysis of the material points in the direction of a difference between the more socially focused services and those that are more based on interaction with third-party information such as data retrieval.

The clustering was developed through the use of factor analysis as will be described below. The specific items in the "future net-based services" battery, arranged respectively into what the analysis indicates are more "social" and "data" based clusters include:

- The possibility to chat with the mobile telephone in the same way that you chat with the PC with, for example MSN or Yahoo chat;

- Have a list of friends where I can see if my nearest friends and family are available;

- Be able to use your mobile like a walkie-talkie (push a button and talk) to talk with several family/ friends at the same time;

- The mobile telephone automatically knows where you are and can give you information that is relevant for that time and place, for example traffic information;

- Use the mobile telephone to identify yourself (for example as a key to the house or a ticket at the movies or a concert);

- Use the mobile telephone like you use your bank card; that is to pay for goods and services by sending a PIN code with your telephone;

- The possibility to remotely control electronic equipment in the home, for example heating, lighting and things like that;

- The possibility to get information from your home PC regardless of where you are.

As will be developed below, the first cluster is the more "social" of the two groupings and the second is more "data" oriented. These clusterings arise from an analysis of how the respondents reacted to the material. There is perhaps some curious juxtaposition, most pointedly the "time/place" item in the social cluster. While this is perhaps an odd placement, it is nonetheless grounded in the empirical results from the material.

Findings

This section of the deliverable will first look rather carefully at the specific items in the battery of future net-based services. This will be followed by a so-

1) *Approximately 6870 persons were contacted in order to complete 1000 interviews.*

ISSN 0085-7130

called factor analysis wherein the battery of items will be clustered into groups of items that covary with each other.[2)] Following this there is a discussion of a regression analysis wherein other items in the questionnaire were examined in the context of the "future net-based services" battery of items. This analysis shows how, in particular, interest in handset features, use of existing services such as SMS and gender play into the respondents' interest in future net-based services.

The findings here can be criticized in that they are "hypothetical." Basically people were asked to comment on services with which they had little direct experience. This means that they were often using only their imagination in order to understand how, and to what degree such services might fit into their lives. Experience has shown that there is often variance with actual adoption when compared to such hypothetical assessments. There are many cases wherein services that receive low initial estimations actually become highly integrated in social life (SMS) while others that have received high estimations never really take off (video telephony?). In all likelihood there are complex social network and psychological gyrations that need to be worked out before a service is accepted or not. Thus, it is difficult to really "hang our hat" on any of these estimations.

This said, the findings here do point to people's understanding of various services. They provide a reading of how people understand the future of mobile communication and the types of things that are potentially of interest. Thus, while the analysis digs deep into the details of the data, we always need to remember that the broader view is perhaps the most appropriate here, namely that there is a distinction between "social" oriented and "data" oriented services and that these two realms have captured our minds in distinct ways.

General interest in the services

Figure 1 shows the level of interest in the various future net-based services. As the reader can see from the figure, Instant Messaging (IM) is the least interesting and the items describing remote control of the home, time/place sensitive information and access to information from the home PC are the most popular

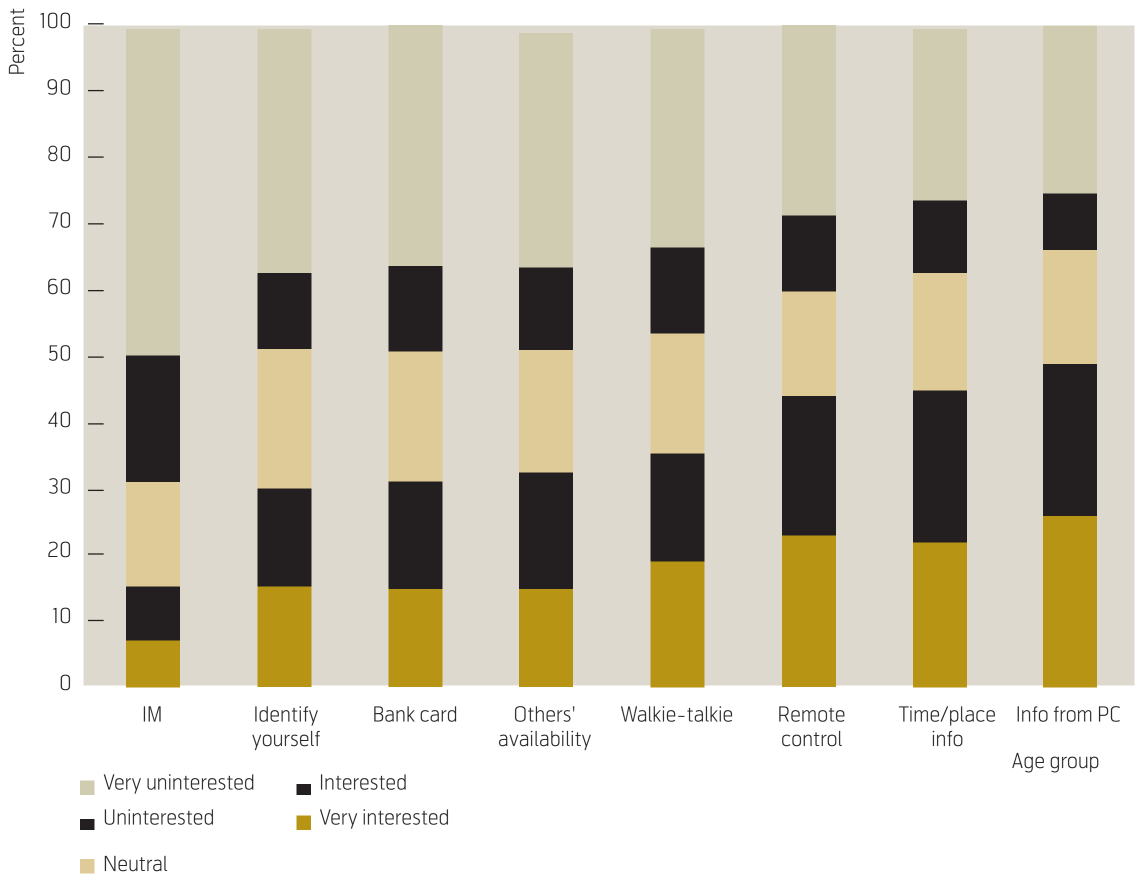


Figure 1 Interest in future net-based services: Norway 2004

---

2) *In this case, the best solution is a single factor solution. However, a slightly revised threshold allowed us to tease out two factors or clusters in the material, the aforementioned social and data clusters or factors.*

ISSN 0085-7130 © Telenor ASA 2005

new services. In general the data here indicates that those in the extreme categories drive the results. About 15 % of the respondents are neutral in all the items. In addition, the "uninterested" category contains about 10 – 12 % of the cases. The exception here is those who were "uninterested" in IM (19 %) and those who were "uninterested" in the ability to retrieve information from a home PC via the mobile telephone (8 %). When looking at the "interested" group, there are roughly three groups of items. The first is the IM item (8 %). The second (about 15 % of the respondents) included the items describing using the mobile to "identify yourself," using it as a bank card, the ability to check on others' availability and using the mobile phone as a walkie-talkie. Finally, there were three items wherein slightly more than 20 % of the respondents were "interested". These were using the mobile to remotely control items in the home, to get time/ place sensitive information and to get information from the home PC.

This means that there were relatively many respondents who were in the extreme categories. Indeed, the extreme categories accounted for about half of all the respondents. This leads to the conclusion that people consider these items in either an extremely positive or a negative light. The respondents were the most critical towards IM where there were six respondents who were "very uninterested" for each one who was "very interested." There was a group of four items where the ratio was about two "very uninterested" to one "very interested". This group included 1) using the mobile to "identify yourself", 2) use as a bankcard, 3) the ability to check on others' availability, and 4) to use it as a walkie-talkie. Finally there was the group of three items where there was roughly a one-to-one relationship. This group included the questions describing 1) remote control of things in the home, 2) to get time/place sensitive information, and 3) to get information from the home PC.

Evaluation of the individual items

I will now look through the individual items and examine how they have been distributed across age groups and gender. For the sake of simplicity, the "Interested" and "Very interested" items have been collapsed into a single variable. In general, as will be discussed at several points below, there are obvious age based differences. The nearly universal rule here is that younger persons were more positive than the older groups. In addition it is often the case that males were more interested in an item than were females.

Instant messaging

It is IM that received the worst reception among the respondents (see Figure 2). While this is not necessarily a net-based service it is a service that may possibly follow up today's SMS as a way of mediating text-based messages.

The specific item in the questionnaire was "The possibility to chat with the mobile telephone in the same way that you chat with the PC via, for example, MSN or Yahoo chat." For every person who was very interested in this type of service, there were six who were very uninterested. It is teens and to some degree teen boys who are most interested in this type of service. In spite of a seemingly large gap between genders in this age group, the difference is not statistically significant.[3] Indeed there are no real gendered differences in the responses to this item when considering all age groups. There is, however, a significant difference in interest when comparing the age groups, the teens and young adults being significantly more interested in this type of service than the older groups.[4]

This is probability the result of several issues. First of all, IM has only a marginal foothold in the PC world in Norway. Only 28 % of males and 17 % of females use IM on a daily basis in Norway. In step with the general use of IM, the users are clustered among the teens where about 55 % of the males and 40 % of the females in the 13 – 24 year age group use this form of interaction on a daily basis.[5] Perhaps more to the point, IM is not often seen as a part of the mobile world where texting is associated with SMS. Finally, there are asymmetries associated with the use of PC and mobile-based IM when considering text entry. In

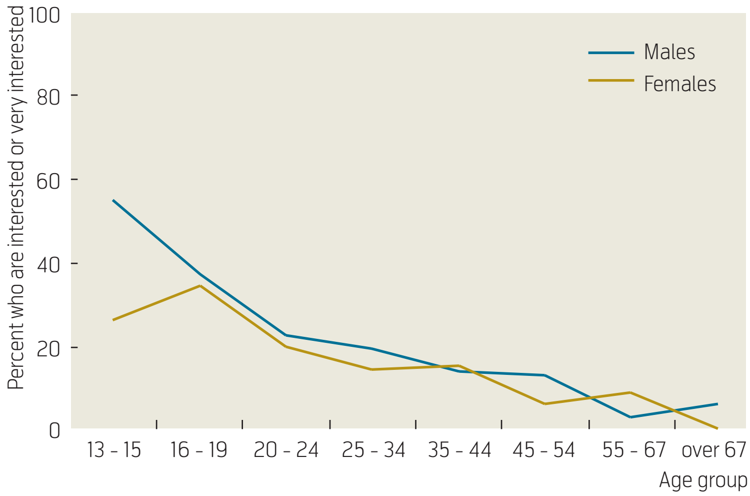


Figure 2 IM via the mobile by age and gender, Norway June 2004

3) *f (1 51) = 1.806, sig. = 0.185*

4) *chi² (28) = 148.9, sig. < 0.001*

5) *These statistics come from the same database, as do the general results reported here.*

the best case, text entry on the mobile is much slower than via a keyboard. Thus, interaction between a PC and a mobile based user would mean that the PC user perhaps expected responses at a higher tempo than the mobile user is able to produce.

All of this is not to say that IM will not become a part of the mobile world. Indeed, the tariff structure, the more open format (longer messages) and the potential to include more persons in a "conversation" mean that this type of text messaging may well gain popularity.

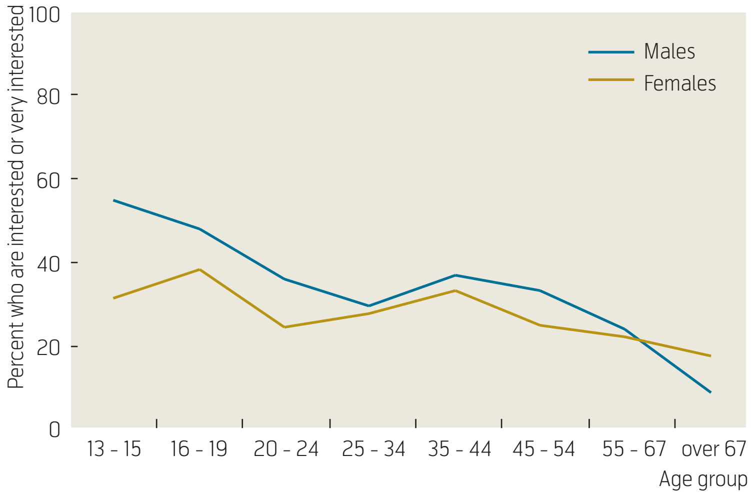


Figure 3 Identify yourself via the mobile by age and gender, Norway June 2004

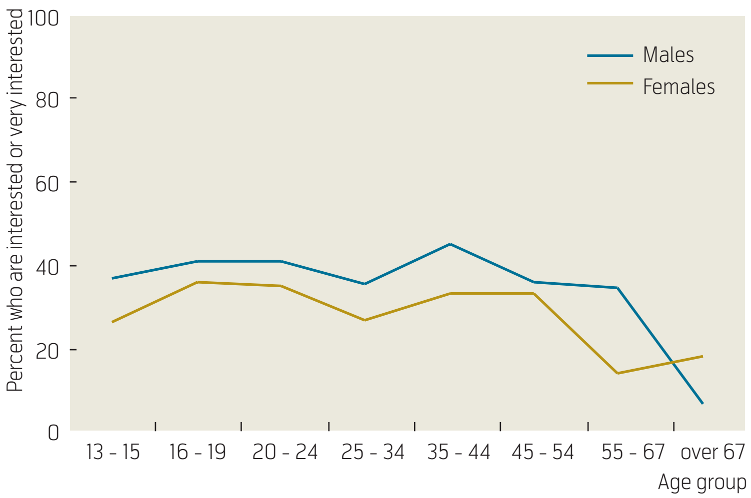


Figure 4 Bank card via the mobile by age and gender, Norway June 2004

Using the mobile telephone to identify yourself

Now the analysis moves to the "middle" group of items. These include the responses to the following four items in ascending order of interest:

- Use the mobile telephone to identify yourself (for example as a key to the house or a ticket at the movies or a concert);
- Use the mobile telephone like you use your bank card that is to pay for goods and services;
- Have a list of friends where I can see if my nearest friends and family are available;
- Be able to use your mobile like a walkie-talkie (push a button and talk) to talk with several family/friends at the same time.

The first two are more data oriented and the last two seem to fit better into a "social" constellation. That is, this type of service does not necessarily imply interaction with other persons via the mobile phone, but is an interaction with net-based functions. In the case of using the mobile to identify yourself in various situations, men were generally, but not significantly more interested in this type of service (see Figure 3). The material also shows that there were significant age based differences.[6)] Interestingly while there are clear age based differences for men by age group, the same cannot be said as strongly for women. Thus, the attitude toward this is largely stable across age groups while with men, the younger respondents were significantly more positive than the older ones.[7)]

The mobile phone as a bank card

The material shown in Figure 4 is that which most directly deals with monetary issues. While the use of the mobile telephone as a type of key to the home or ticket to a concert/film has some of the same sensitivity, it is this question on the use of the mobile telephone as a bank card that is most directly focused on monetary issues.

Along with the item on remotely controlling the home, the use of a mobile telephone as a credit card is not as popular among teens as it is among middle-aged respondents. The analysis of the bank card item shows that there are significant age[8)] and gender

---

6) *chi^2 (28) = 76.75, sig. < 0.001*

7) *chi^2 for men across age groups was (35) = 74.751, sig. < 0.001 while the same statistic for women across age groups was chi^2 (35) = 49.139, sig. < 0.057.*

8) *chi^2 (28) = 84.895, sig. < 0.001*

ISSN 0085-7130 © Telenor ASA 2005

based differences.[9] Interestingly it is women who account for the age based differences in the results of this item. Indeed, it is the only item in which the differences across the age groups are significant for one gender while they are insignificant for the other. In spite of the lower estimation of the item by women, the item does not show significant age based differences for men across the age groups.[10] By contrast, there are significant age based differences across age groups for the women.[11] Thus, it is the skepticism of the older women (and partially the younger teens) that accounts for the general age based differences on this item.

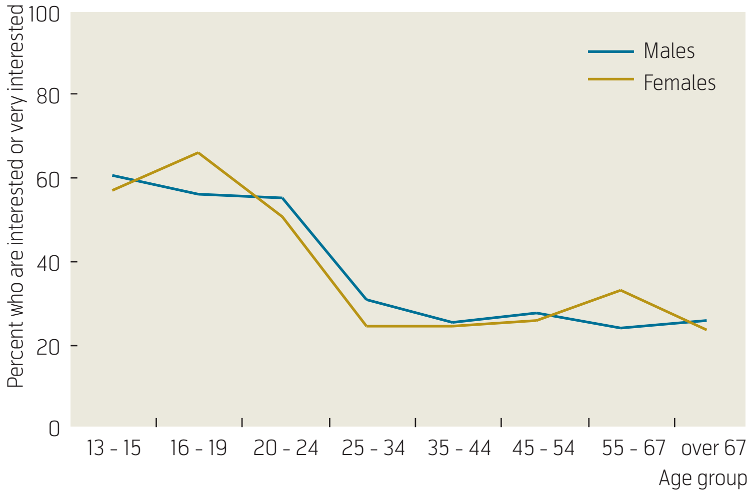


Figure 5 Others' availability via the mobile by age and gender, Norway June 2004

A partial reason that this, as well as the previous item on the use of the mobile phone as a bank card score low in terms of respondent interest is that they are items that deal (at least in part) with monetary and security issues.

Determining others' availability

Moving on to two more socially oriented items – the ability to determine the availability of friends and family and the walkie-talkie function. These two items showed the more usual pattern of being popular among teens and less popular among the older respondents. Looking first at the item on having the ability to determine the availability of family and friends (see Figure 5), there is a clear age difference in this item age.[12] There is also a marginal gender difference.[13] It is interesting to consider this item in connection with the "IM" item described above. There was generally a negative attitude towards that item. However, when extracting only the "presence" aspect of IM and placing it into a mobile context, there was an increase in interest.

The message here seems to be that there is an interest among teens and young adults in having some insight into the availability of their friends/family. However, the additional PC-based function of interactive messaging is less interesting. It may well be that the insight into others' availability is seen as an indication of whether one should initiate mobile-based interaction (voice or SMS). However, the additional functions associated with PC-based IM were judged to be less interesting.

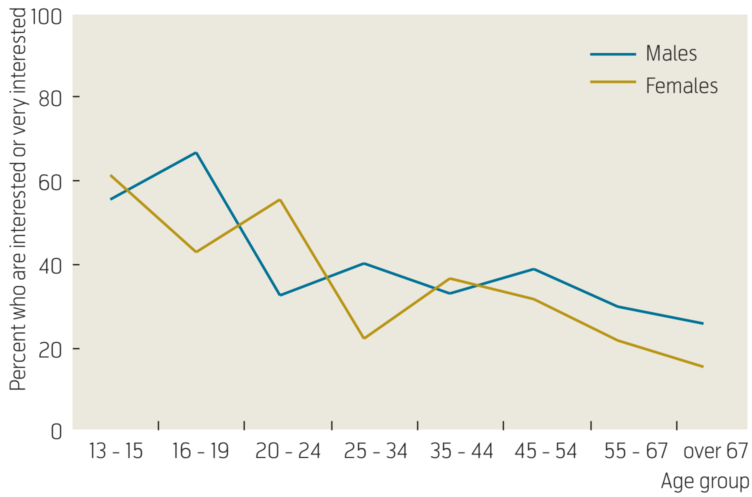


Figure 6 Walkie-talkie via the mobile by age and gender, Norway June 2004

Walkie-talkie (push to talk) via the mobile phone

Another socially directed service is that of using the mobile telephone as a type of walkie-talkie (see Figure 6). This item is a description of the function that is somewhat popular in the US, namely push-to-talk. It is an interesting service in that it is not a duplex form of communication. The participants must speak sequentially and indicate turn taking more explicitly than in traditional telephony. There is not the ability to interrupt another person in the middle of their utterances. In this respect it is perhaps more akin to radio interaction than to telephony. It enjoys some

---

9) *chi^2 (4) = 11.831, sig. = 0.019*
10) *chi^2 (35) = 45.988, sig. = 0.101*
11) *chi^2 (25) = 72.651, sig. < 0.001*
12) *chi^2 (35) =126.458 , sig. < 0.001*
13) *chi^2 (5) = 10.004, sig. = 0.075*

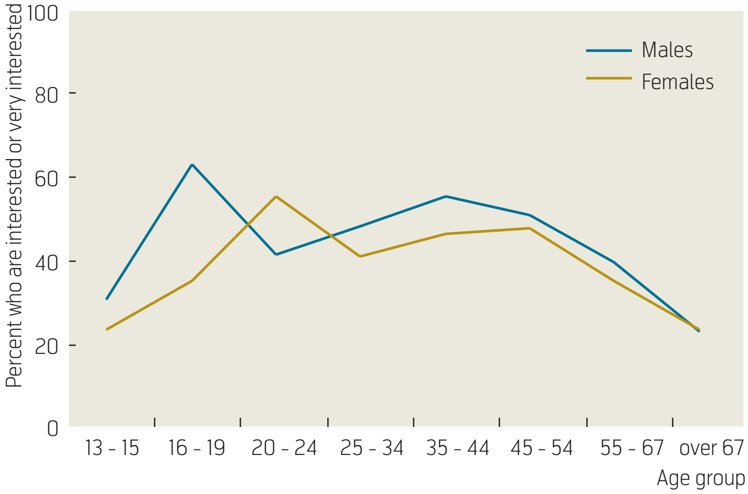


Figure 7 Remotely control home via the mobile by age and gender, Norway June 2004

popularity in the US and it also is a rather "light weight" service in that it is IP-based and does not have the same latency requirements as does traditional telephony.

The analysis shows that younger persons are significantly more interested in this type of service than are older respondents.[14] Men are also marginally more positive to this type of service than are women,[15] particularly in the 16 – 19 year old group[16] and in the 25 – 34 year age group.[17] The data here is somewhat unstable, but it shows the standard youth orientation that one also sees in the IM, friends' availability, and the time/place information items.

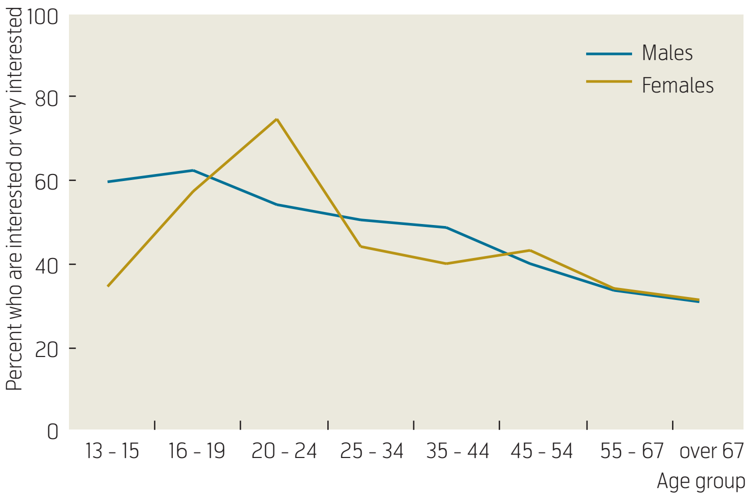


Figure 8 Time/place info. via the mobile by age and gender, Norway June 2004

Remotely control the home via the mobile phone

The final three items were those that received the greatest level of interest among the respondents. They included:

- The possibility to remotely control electronic equipment in the home, for example heating, lighting and things like that;
- The mobile telephone automatically knows where you are and can give you information that is relevant for that time and place, for example traffic information;
- The possibility to get information from your home PC regardless of where you are.

The first of these was a typical "middle aged" item, namely remotely controlling various items in the home via the mobile telephone (see Figure 7) and was most popular among those in the middle of the age spectrum. The youngest and the oldest respondents were the most skeptical to this type of service. Aside from the somewhat erratic data from the 16 – 19 year old males,[18] it was the middle-aged people who reported the most interest. Indeed there are significant age-based differences in the data.[19] The data also shows that there are loosely significant gender-based differences.[20]

Time and place relevant information

The next item was that describing the use of the mobile telephone in order to retrieve time- and place relevant information. There was also a slight tendency for the younger users to prefer this type of service.[21] In general there are significant age-based dif-

14) *chi^2 (35) = 101.204, sig. < 0.001*

15) *chi^2 (5) = 10.518, sig. = 0.062*

16) *chi^2 (4) = 17.616, sig. < 0.001*

17) *chi^2 (5) = 12.142, sig. = 0.033*

18) *There was a significant difference between genders in this group (chi^2 (4) = 8.843, sig. = 0.065). This is the only age group in which there was a gendered difference.*

19) *chi^2 (35) = 107.977, sig. < 0.001*

20) *chi^2 (5) = 10.193, sig. = 0.070*

21) *This is particularly true if one overlooks the findings from the youngest females.*

ISSN 0085-7130 © Telenor ASA 2005

ferences for the whole population here.[22] Somewhat interestingly, this fits generally into the more socially oriented items (see Figure 8). That is, as we will discuss below, responses to this item covaried in a loose way with the items on IM, others' availability, and the push-to-talk function.

Retrieving data from a home PC

The final individual item is that covering the ability to retrieve information from your home PC via the mobile telephone regardless of where you are (Figure 9). This is one of the more "data" oriented items in the battery. As with many of the other items there is a significant age-based difference in the material.[23] While the data shown in Figure 9 shows that in some cases males are more inclined for this type of service and in other age groups it is the females, when all age groups are considered together there are no gendered differences. Further, the differences by gender shown in the figure, while seemingly dramatic, are not statistically significant.

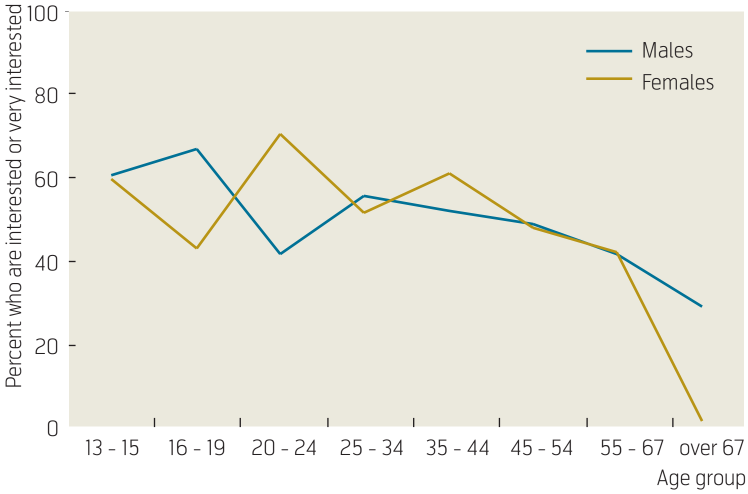


Figure 9 Info. from home PC via the mobile by age and gender, Norway June 2004

Factor analysis of the material

A way to generally summarize the interest in the items described above is to collect them into coherent groups and then to look at levels of interest in a collective manner. Factor analysis is one approach to this. Factor analysis in effect draws together items in a battery of questions that covary with each other. In many cases, there are a number of clusters. In the case of the items discussed here, there is basically one cluster (called a factor). As will become clear below, it is possible to tease out more factors from the material by adjusting the criteria for sorting the material. In the case of this material, the general analysis revealed one factor while the adjusted analysis provided the insight into the difference between the social and the data related items.

The advantage of using factor analysis (as opposed to simply summing up the factors) is that first; it alerts one to the various items that fall into different categories. Second, it provides a weight for each item indicating the degree to which that item contributes to the total factor. Thus factor analysis allows one to construct a variable that is available for further analysis. The researcher can, for example, look into the socio-demographic, psychological or behavioral dimensions that might covary with the particular factor.

As noted, the eight items describing future net-based mobile services all fell into the same factor. The material in Figure 10 shows how this collective variable is distributed in terms of age groups and gender. As with many of the other items that constitute the variable, this complex variable has significant age[24] and gender[25] based differences. That is, the material shows that younger persons and males were more

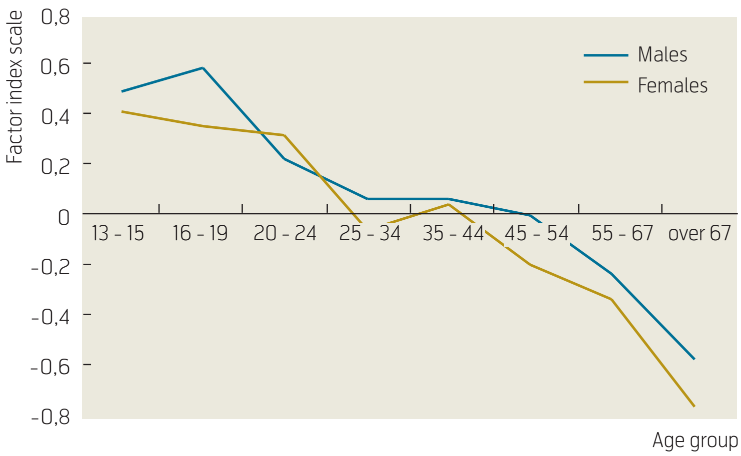


Figure 10 Interest in future net-based services (1 factor solution) by age and gender, Norway 2004

---

22) *chi^2 (35) = , sig. < 0.001. The analysis shows that there is a gendered difference for those in the 45 – 54 year age group. The reader needs to remember that the material in the chart shows only the people who were interested. The material on those who were not interested indicates that it is women in the 45 – 54 age group who are particularly uninterested in this type of technology chi^2 (4) = 13.00, sig. < 0.011.*

23) *chi^2 (35) = 142.356, sig. < 0.001*

24) *f (7 807) = 15.754, sig. < 0.001*

25) *f (1 814) = 6.357, sig. = 0.012*

ISSN 0085-7130

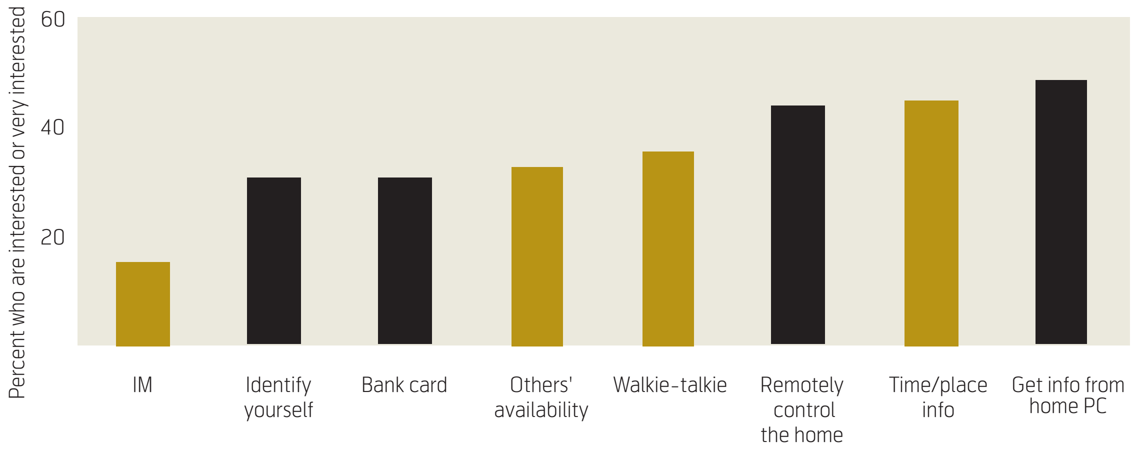


Figure 11 Interest in future net-based services (with two-factor solution marking), Norway 2004

interested in the services described in the material than were women and older people. When looking at the specific age groups there is only one case in which the results approach a significant gender difference; that being the 45 – 54 year age group.[26)]

A theme that has gone through the discussion is the suggestion that the individual items, if pressed to a certain degree, fall into two factors (see Figure 11).[27)] In this two factor solution the items describing 1) IM, 2) others' availability, 3) the walkie-talkie function, and 4) the time/place information fall into one factor while the other factor includes using the mobile 1) to identify oneself (i.e. as a key or a ticket), 2) as a bank card, 3) to remotely control the home, and 4) to remotely retrieve information from the PC.

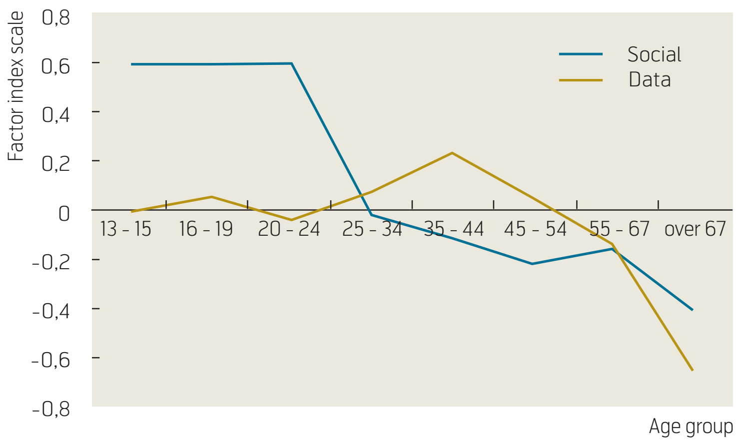


Figure 12 Interest in future net-based services (2 factor solution) by age and gender, Norway 2004

An examination of these two factors shows that the first factor generally describes social interaction.[28)] With the exception of the item describing interesting time/place information, the items describe forms of interpersonal interaction. They describe services that either help the individual judge the accessibility of others that they know, or they describe services for the actual mediation of information between individuals.[29)]

By way of contrast, the second item focuses on various types of interaction with inanimate objects in the network. These include payment, the management of accessibility (which in the case of ticketing can also include payment issues), and the remote control of either a PC or different items in the home.

It is perhaps not surprising that these two factors showed different distributions in terms of age grouping (see Figure 12). The social dimension was particularly popular among the younger respondents. There was nearly uniformly high interest among those who were under 25 years of age. For the age groups over 25 there was a declining interest in the social services described here. The oldest group had the least interest.

The data related items had a different profile. In a similar way it was the oldest persons who were the least interested. However, the youngest respondents were basically neutral to these items. The greatest interest was among those in the 35 – 44 year age group. The persons in the age groups 25 – 34 and 45 – 54 were also somewhat more interested in these

---

26) *f (1 142) = 3.358, sig. = 0.069*

27) *The two-factor solution first arises if one reduces the eigenvalue to 0.89 as opposed to the traditional eigenvalue of 1.*

28) *This factor is the best at clarifying the data. It has an eigenvalue of 4.193 and clarifies 52.409 % of the variance in the data while the second factor has an eigenvalue of 0.89 and clarifies 11.147 % of the variance.*

29) *Interestingly if one does a factor analysis by age group, the social and the data factors fall out with eigenvalues of greater than 1 in the case of the youngest groups.*

ISSN 0085-7130 © Telenor ASA 2005

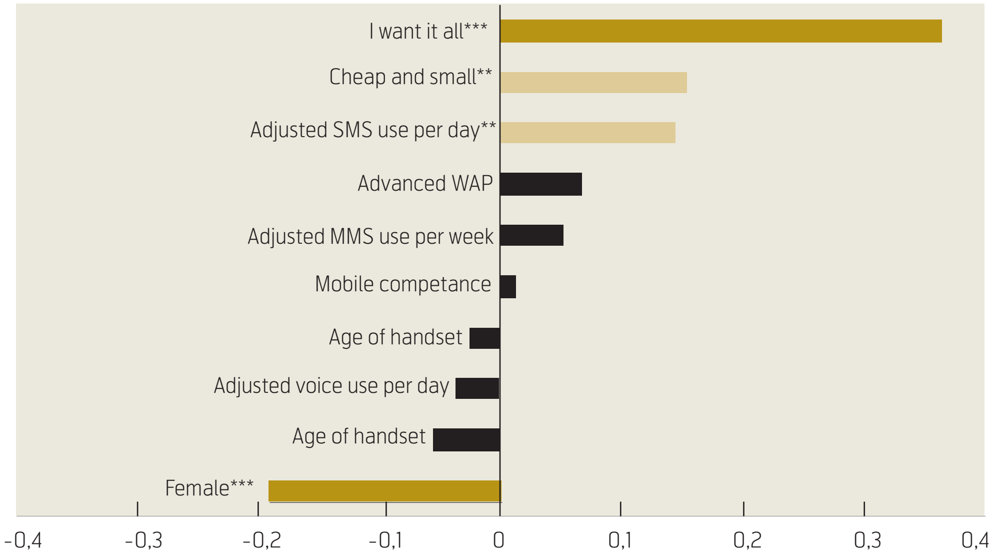


Figure 13 Regression analysis

items than the normal user. There were no major gender-based differences in the interest in these two factors.

Regression analysis

The final analysis here is a regression analysis using other items in the questionnaire to try and clarify the interactions between them and the future services factor that has been described above.[30] The regression analysis results in a model of interactions between the "explanatory" variables and the target variable which in this case is the "interest in future net-based services" variable described above. It is perhaps tempting to say that the "explanatory" variables cause changes in the target variable. That is, however, to overstate the case. The best that one can do with these data is to assert that there is an interaction between them and not that there is causality. The results of this analysis are shown in Figure 13.

The results in the figure show those elements that have either a direct or an inverse relationship to the future net-based services factor. Those going to the right have a direct relationship to the interest in future net-based services (the more one increases the variable, the more the other increases) and those lying to the left have an inverse relationship. It needs to be noted that there is not the assertion of causality here. The longer the line is, the greater the interaction. In addition, there is a collection of items in the middle of the chart that make little contribution to the general model.

The material from the regression analysis indicates that those respondents who were quite interested in the future net-based type services were, perhaps not surprisingly, interested in advanced features on their terminals. These features include many "business" features such as for example GPRS, the ability to synchronize with a PC, "the newest on the market", and the ability to surf on the Internet via the mobile telephone. They also have more typically teen features such as color screens, MMS, "flip" covers etc. It is this variable which is the strongest and along with the gendering variable, the most significant of the variables in the regression model.

The second strongest variable in the regression analysis is one of the two variables describing the gendered interest in the items. The analysis shows that women were adverse to the future net-based like services when considered as a whole. In other words, some men were positive to the services described while others were less positive. This is also of interest here.

There were two other variables that made a significant contribution to the regression model. These were interest in small and inexpensive terminals (the size dimension is perhaps the most important issue here) and the number of SMS messages that a person reported sending per day. Both of these items made a positive, if somewhat weaker contribution to the model. It is interesting to consider the SMS use. In Norway, this form of interaction is tightly connected to teens and young adults. Thus, the fact that it makes a contribution to the model here seems to verify the material that is shown in Figure 10, that it is generally

[30] *Adjusted $r^2 = 0.245$*

ISSN 0085-7130 © Telenor ASA 2005

younger persons who express strong interest in the future net-based services. The caveat here is that there are some “middle aged” services (specifically the bank card and the remotely controlling items in the home).

Discussion

The material here has outlined a tendency towards a split in the types of net-based services that are outlined. On the one hand there seems to be a clustering around what can be called social oriented services, i.e. services that facilitate direct interaction between individuals. In addition, the data points to a clustering of interest around more “data” related services. That is, there is a divide between interpersonal services and person-machine services.

Currently, mobile telephony is tilted dramatically towards the social type of services. From the perspective of the operators, the preponderant source of income comes from mediation of interpersonal messages, predominantly in the form of voice and SMS. Indeed when looking at daily use, there are more people who use SMS than use traditional landline telephony. Only a small portion of income comes from data related services. In a slightly different way, the same can be said of the Internet, namely that the service that is used most often is e-mail. Indeed Oldyzko traces the history of mediated communication from letters and newspapers sent via the post through the telegraph to the modern Internet and mobile communication and finds that there has traditionally been a greater willingness to pay for interpersonal communications than there has been to pay for information from third parties such as news, entertainment etc. (2000). Thus, the findings from this analysis, however vaguely, point to a traditional divide between mediation of social interaction and mediation of what can be called third-party information. These results can be seen through the filter of other mediation. When considering them from this perspective it seems that the obvious direction with which to develop services is to better support various forms of social interaction.

It is somewhat ironic, however, that respondents expressed the greatest levels of interest in data related services (remote control and interacting with a remote PC). The respondents gave a relatively positive rating to these two items. Nonetheless, the factor analysis indicated that there is a far greater consensus around the social variables, particularly when a two-factor solution is used.

Part of this paradox can perhaps be resolved by considering the additional baggage that the various services might be carrying in the minds of the respondents. The items covering remote control of the home, retrieving information from a PC and time/place information are all relatively abstract items. The respondents would not have had any real experience with these items and thus they are left free to muse as to the interesting potentials of their eventual development. By contrast, the other items in the battery are often weighted down with the respondent’s prejudices and worries as to their eventual misuse. The items on monetary exchange (the bank card and the ticket/ID items) can be interesting from a technical perspective but in order to be accepted by the public they will need to prove themselves reliable and tamper-resistant since they play in an area of life where the public is cautious. The items remaining items that describe social interactions (IM, others’ availability and the walkie-talkie) also operate in an area in which the respondents have a relatively well-developed understanding. These items play into the way in which people develop and maintain their social networks. People have a sense of how a walkie-talkie connection to their friends or family might (or might not) be of use. They can imagine how useful or perhaps how intrusive it might be. It may be that the two “monetary” items and the social items – aside from the somewhat oddly sorted “time/place info.” item – all operate within the respondents’ range of experience. This means that they can have a sense of how these might operate in real situations. By contrast, those items in which there is the greatest interest hang more loosely in the imagination of the respondents.

Looking into the seemingly misplaced item on time/place information this may illustrate another division in mobile inter-personal communication, namely the difference between instrumental and expressive interactions. On the one hand, many interactions are expressive (a love letter, a get-well call etc). At the same time there are also many instrumental interactions (e.g. “where are you?” or “when are you coming?”). The item on time/place information clearly fits into the notion of instrumental communication. Indeed, much of mobile communication is concerned with this theme (Ling and Yttri 2002). It is not surprising, then, that an item describing a service in which the individual interacts with a machine is sorted into the same factor as those that describe more direct social interaction.

It is also interesting to examine the attitudes of the respondents in terms of the other information in the questionnaire. In general, it is the younger persons who are the most interested in the new possibilities. A further nuance of this is that the younger persons are perhaps more interested in the social potentials of the technology where middle aged persons often see the possibilities associated with remotely dealing with their homes, monetary issues and the transmission of data from one place to another.

The people who were most interested in advanced services were also the most interested in advanced handsets. They were active SMS users and women were somewhat uniformly not in this group. Thus one is presented with the image of a portion of the male respondents, who were also relatively active SMS users, who were fascinated by the technology.

Putting this into a somewhat critical light, this may be a technical fascination with mediation technologies such as the mobile telephone that does not fully extend the use into the area of social networking, i.e. that is using the telephone in social interaction. A classic theme in the adoption and use of mobile communication technologies and services is that males are quite often the early adopters based, perhaps, on an interest in the technology. They seem to have a willingness to put up with the early frustrations of configuring the systems, and dealing with the eccentricities of the technology. Various examples show that they work through the set-up issues and perhaps map out the potential areas of inter-personal use. Their fascination with the technology in itself facilitates the eventual diffusion of the innovation. In these analyses, females seem to arrive on the scene somewhat later and to approach the technology from a slightly different perspective. Their contribution seems to be more in the direction of making the technology into a social networking tool. Where the males seem to be carried by a fascination with the technology and its possibilities, the females seem to be carried by using the technology in order to interact with friends and family.

This is the experience of early mobile telephony. In the case of teens in Norway, it is well documented that males were the first to adopt in the late 1990s but by about 2001 female teens were significantly more likely to have a mobile phone than were same-aged males (Ling 2004). Teen boys are significantly more likely to have more than one mobile telephone (read: fascination with the technology) while teen girls use the device to write significantly longer SMS messages (read: social networking) (Ling forthcoming). The same tendencies were seen in the system developed in the "Youngster" trial (Ling and Sollund 2002). This was a location-based service that allowed one to see the approximate location of friends as they moved about the small town of Grimstad. In that case the teen boys seemed to adopt the use of the service, but the teen girls basically ignored the service during the short trial period. MMS has seen the same contours in its adoption (Ling 2004).

This is not to say, however, that just any service will be adopted. There has to be a sense the service is accessible, easy to understand and useful. If either the innovators or in turn those who employ it for social networking feel that the service is not accessible or useful, there is a good chance that the service, regardless of how well developed or technically advanced, will die on the vine. This is the major barrier confronting these services. Thus, it is essential that the developers not isolate themselves in the laboratories, but confront the real situations in which eventual use will take place.

References

Ling, R. The adoption, use and social consequences of mobile communication. *Telektronikk*, 100 (3), 69–81, 2004.

Ling, R. The socio-linguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In: Ling, R, Pedersen, P (eds.). *Mobile Communications: Renegotiation of the social sphere*. London, Springer, 2005.

Ling, R, Sollund A. *Deliverable 13: Planning, operation and evaluation of the field trials*. EU IST program, 2002.

Ling, R, Yttri, B. Hyper-coordination via mobile phones in Norway. In: Katz, J E, Aakhus, M (eds.). *Perpetual contact: Mobile communication, private talk, public performance*. Cambridge, Cambridge University Press, 2002.

Oldyzko, A. *The history of communications and its implications for the Internet*. AT&T Labs, 2000.

Rich Ling is a sociologist at Telenor R&D. He received his Ph.D. in sociology from the University of Colorado, Boulder in his native US. Upon completion of his doctorate, he taught at the University of Wyoming in Laramie before coming to Norway on a Marshall Foundation grant. Since that time he has worked at the Resource Study Group and has been a partner in the consulting firm Ressurskonsult. Since starting work at Telenor he has been active researching issues associated with new ICT and society. He has led projects in Norway and participated in projects at the European level. Ling has published numerous articles.

email: richard-seyler.ling@telenor.com

 ISSN 0085-7130 © Telenor ASA 2005

Mobile multimedia and users: On the domestication of mobile multimedia

ILPO KOSKINEN AND ESKO KURVINEN

Ilpo Koskinen is Professor of Industrial Design at the University of Art and Design, Helsinki

Esko Kurvinen is an Industrial Designer at the Helsinki Institute for Information Technology

With camera phones rapidly becoming commonplace, mobile multimedia is spreading fast in the market. This paper focuses on how people use their mobile multimedia devices. Based on a series of studies conducted in Helsinki, this paper reviews existing arguments about mobile multimedia, and then explains why the Helsinki studies have been informed by an interaction-oriented argument based on ethnomethodology. Data and methods for the paper come from two major studies called *Mobile Image* and *Radiolinja*. The main body of the paper provides three illustrations of the framework. First, it studies how people build their mobile multimedia messages. Secondly, it shows what kinds of methods of response recipients use. Third, it looks at how people learn new methods of action from each other in the course of exchanging messages. Discussion compares the Helsinki studies to other research perspectives.

Introduction

Mobile multimedia is a new, but quickly expanding addition to mobile telephony. Phones with multimedia capabilities allow users to compile and send messages including text, still images, sound, and, increasingly, video clips. Message 1 provides an example of a typical multimedia message. A vacationing friend sent it to the first author in Helsinki from Greece. It consists of a snapshot photograph and a written greeting. Although users have been slow to adopt multimedia technology, it is spreading fast in handsets. This paper asks what this technology adds to mobile telephony. More specifically, it looks for ways in which people adopt multimedia. What drives its "domestication"? The focus of the paper is primarily on MMS, but messages from older applications are commented on as well.

The metaphor of "domestication" has its origins in British media studies of the 1980s and 1990s. Domestication denotes the phases of a product life cycle which are not controlled by industry and (here) mobile carriers. In the most significant articles in this area, Roger Silverstone and his colleagues [1, 2] argue that this process is directed not just by economic, but also by moral, affective and aesthetic considerations. As a result of these considerations, technologies find their place in people's spatial environments, practices and communication modes. This paper explores how mobile multimedia is domesticated in mobile telephony by reporting the perspective and key observations of design research conducted in Helsinki between 1999 and 2003. The reason for focusing on human action instead of technology is the conviction that although technology changes rapidly, user needs and practices do not.

Message 1. Postcard from Greece 02-09-2004 11:47:46

Greetings from Greece. Imagine ... An old man lives at the top of a windmill. Downstairs he has a beach café and sells small snacks. It's really relaxing to be out here. Remember to work hard while I'm vacationing ... Best wishes, Junior, Mari and kids

What drives domestication?

In addition to several descriptive accounts [3], four theoretical perspectives have been proposed to explain drivers for the use of mobile multimedia. The first and simplest argument relates mobile multimedia to user motivations. In brief, what people see as worth capturing and sharing is explained by functional needs and affective reactions – ranging from aesthetic experience to missing loved ones. The best example of this line of argumentation is in Kindberg et al. [4], who classified multimedia messages sent by the British and Americans into individual, social, affective, and functional categories. Out of the 349 messages they analyzed, 82 % were affective. 58 % of the affective messages were social in nature – targeted towards enriching shared experience, or communicating with absent friends – and 42 % individual, intended for personal reflection. Out of 63 functional messages, 53 % were for social purposes; they were sent to help other people do some task, 47 % supported tasks not involving sharing – that is, they helped the senders of messages in their own tasks.

From a more sociological perspective, Scifo [5] has studied the domestication of camera phones in Milan. Unlike mobile phones, mobile multimedia roots experience back in users' physical and social surroundings. Multimedia messaging is almost exclusively linked to one's network of strong relationships, where it primarily nourishes sentimental bonds. With

ISSN 0085-7130 © Telenor ASA 2005

MMS, people give others access to places, individual and social situations, and emotions. People share images of familiar objects and people, private life (like objects, relatives), and social networks. Mobile multimedia also provides an extension of one's experience and memory – with it, one's affective world is reportable in a visual and shareable form.

While Scifo focuses on the meanings of pictures and, by extension, camera phones, Ling and Julsrud [6] focus on the "genres" of multimedia messaging. The argument is that existing media – and other – genres provide people formats for compiling multimedia messages. For example, Ling and Julsrud studied mobile salespersons of a soft drink company, real estate agents and carpenters for six months in Norway, finding that the main genres were

- documentation of work related objects
- visualization of details and project status
- snapshots (developing camaraderie)
- postcards and greetings
- chain messages (standardized messages, usually downloaded from Web sites).

Soft drink sales persons used MMS most, followed by carpenters. These groups used it for all the purposes mentioned above. Real estate sales persons used the technology only for postcards, greetings, and snapshots. Carpenters also used MMS for "clarification": taking and sending pictures of problems at work to get advice from colleagues.

The fourth perspective builds on the idea that people interact with each other through mobile multimedia. To this end, they use "age-old practices" familiar to anyone from other realms of life: greetings, postcards, questions, riddles, family portraits, and so forth. Building on the initial idea of Licoppe and Heurtin [7], Taylor, Harper and Berg have generalized this argument first to SMS and then to MMS [8]. Their example of an age-old practice is gift-giving. People share things to delight recipients, who are obliged to return the gift and show gratitude, or face sanctions if they do not reciprocate.

These perspectives explain the use of mobile multimedia on a variety of levels, ranging from affective and cognitive processes through small group dynamics to macro-level cultural processes. However, each perspective has shortcomings. For example, the attempt to classify motivations does not account for messages which serve a number of purposes at once; as people routinely do several things in any individual message, classifying messages using categories created by researchers is risky. The cultural argument works somewhat better in this respect. For example, the genres approach situates the locus of action in those cultural structures that people use to construe lines of action. It is not necessary to classify messages into "affective" and "functional" motives. Still, neither the genres framework, nor the gift-giving analogy are sufficient to account for several types of interactional patterns in messages.

Domestication and ethnomethodology

The argument of this paper is based on the notion that mobile multimedia is essentially a technology for communication and, accordingly, interaction. The argument builds on classical ethnomethodology [9], a perspective followed in the series of studies in Helsinki [10, 11, 12, 13, 14, 15, 16]. According to this perspective, communication is situated action [17]. Any message is interpreted by a recipient, who may follow the literal meaning of the message, but need not do so. The recipient's interpretation becomes evident in his response, which can be a word, turn in talk, a piece of text, or a message that contains a digital image. This, in turn, prompts a new response. Thus, each action is shaped by the context in which the previous action is typically the most prominent element. At the same time, any action renews the context, and prepares it for the subsequent action [18]. Because of this situated quality of interaction, we do not normally know what is going to happen after a few turns. Instead, we have to act here and now.

When composing multimedia messages, people often utilize commonsense cultural resources, including those used in conversation, in professional practice, and in media. Thus, despite the fact that the medium is new, messaging as such is largely based on familiar methods. There are good reasons for such conservatism. When incorporated into the new medium, these methods provide intelligibility, coherence, and predictability for communication. The main difference between multimedia messaging and traditional postcards is transmission speed: the mobile card in Message 1 came from Greece to Helsinki in minutes, not in days as a traditional postcard.

However, mobile multimedia is more than just a process of using "age-old practices" in a new technological domain. From an ethnomethodological perspective, domestication is also a collective process of discovering new methods for capturing and sharing. In this process, conventions in media presentation are not followed blindly. Pre-existing media formats or types of content are not just replicated, but reconstituted to be more useful for the participants, who are thus able to display themselves as knowledgeable,

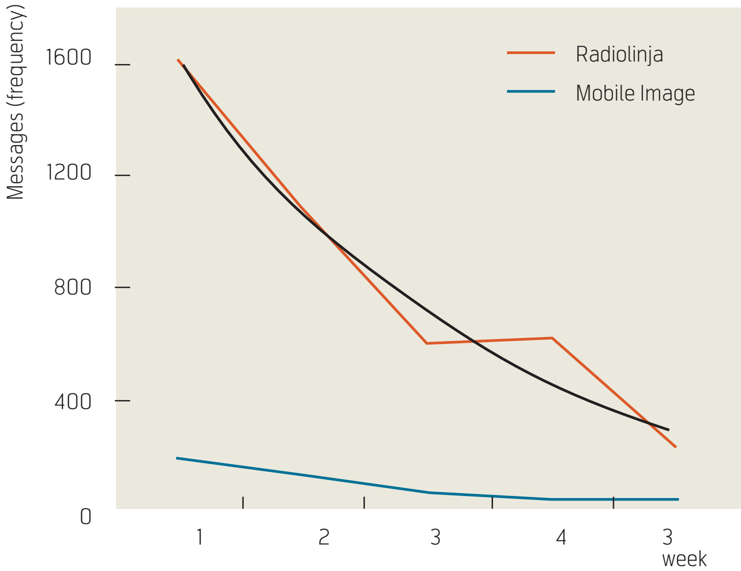


Figure 1 The Trajectory of Use in Mobile Image and Radiolinja (first 5 weeks)[1)]

considerate, entertaining, witty, and so forth. For example, the members of one research group realized that they could also use images on their computer to produce trick effects [10]. This invention took place first without conscious reflection, but gradually led to a joking culture in which members of the group constantly tried to outdo each other. Still, its origins were modest. One member experimented with an automatic signature, but instead of writing his name, address and phone number, he downloaded a picture of the actor Liam Neeson from the Internet and used it as a visual signature. Once the idea of image manipulation was invented, messaging got increasingly more inventive; it became the group's second nature in a few weeks. For this group, there was no going back to the original meaning of the images they had used.

Data and methods

Data for this paper comes from a series of design studies conducted in Helsinki between 1999 and 2003. In particular, we focus on two studies, *Mobile Image* (1999-2001) and *Radiolinja* (2002). These two studies involved over 5000 multimedia messages. By now, about 20 scientific papers have been written from these data; this paper reports some of the main findings from this body of research.

In *Mobile Image*, we gave a Nokia 9110 Communicator and a Casio digital camera, connected with an infrared link, to four groups of five people (the pilot, male, female, and control groups, the pilot and the control being mixed-gender) for 2–3 months each. Groups were selected so that their technical expertise, access to technology, and gender varied. Photographing became more mundane and ad hoc than in ordinary home mode of photography [cf. 19]. People took images of their meals, dirty plates in the sink, their street, and so forth. Messages were collected as e-mail attachments. For ethical reasons, we did not automatize this procedure, but asked participants to send or forward all their messages to Esko Kurvinen, who was responsible for the project. The Helsinki-based mobile carrier Radiolinja provided free phone service and data connection based on GSM technology.

In *Radiolinja*, we selected three groups from a larger technology and service pilot. The pilot took place in summer 2002, and lasted about five weeks. Each user was given an MMS phone (either Nokia 7650 with an integrated camera or SonyEricsson T68i with a plugin camera). Mixed-gender groups with seven, eleven, and seven members were studied. Out of the pilot, we selected groups to take into account gender difference, terminal types, and the city-countryside axis. Exact numbers are confidential, but we can report that users sent over 4000 messages during the pilot. Over 2000 were unique (the rest being duplicates in group messages, or recycled messages). These data were produced through the Radiolinja system automatically. As in *Mobile Image*, the service was free of charge.

Figure 1 reports weekly messaging frequencies for *Mobile Image* and *Radiolinja* for a period of five weeks (data is comparable for that period). It shows that in both data sets, use follows a downward logarithmic pattern. In *Radiolinja*, routine use settled to 2–3 messages/day/phone for each participant after five weeks. The difference between these two data sets probably relates to the ease of use. While taking a picture and sending it took thirty steps and easily two minutes with the equipment used in *Mobile Image*, the process quickened to less than 30 seconds in *Radiolinja*. Participants knew that they were studied, and were informed about the ethical procedures we used. In particular, we told them how our data was produced, promised not to publish pictures without their consent, and promised to change details of images so that it would not be possible to identify them from our publications. In addition, we have followed standard academic and legal practice and have changed all names and details that could identify people.

---

1) *For Mobile Image* $y = -98{,}434Ln(x) + 179{,}45$, $R^2 = 0{,}9405$; *for Radiolinja:* $y = -835{,}32Ln(x) + 1624{,}4$, $R^2 = 0{,}9655$

ISSN 0085-7130 © Telenor ASA 2005

Multimedia elements in messages

The first question we have focused on is the way in which mobile multimedia messages are constructed. Multimedia messages contain three types of media: images, audio, and text. By far, the most prevalent media type is text (94 % of messages in Radiolinja had text), followed by photographs (87 %) and sound (10 %). People appeared in 56 % of photographs, followed by animals (11 %), scenes (11 %), various types of things (10 %), and food (8 %). This list of topics resembles that calculated by Kinberg et al., although in their study "specific things" were more prevalent [4]. As such, this list resembles ordinary snapshot photography [19]: people capture and share a selected sample of things with their "hiptop" devices.

However, for several reasons, multimedia messaging takes messaging beyond what one finds in photo albums. First, as Ling and Julsrud [6] have noted, people routinely borrow media formats from their environment to construct messages. These include postcards, movie-like stories [10: 55-61], travelogues, posters, TV and newspapers, including tabloid ads. Second, messages are always compiled for particular viewers, which is perhaps best evident in ubiquitous hellos, goodbyes, and signatures in the beginnings and closings of messages. Thus in Message 1, there was not only a greeting in the beginning of the message, but also a "best wishes" element and a signature-like list of names that made the message more personal to the first author. Third, people routinely explain why they send messages in the first place. For example, in a message that has a picture of a neighborhood pond, Laura accounts for sending the message (Message 2), explaining that even though she knows that for some, nature photographs are boring, the scene was so enthralling that she decided to send it anyway. Finally, as we noted above, people routinely do several things in any one message. Thus, while in Message 2 Laura gives this explanation, she also instructs the viewer (Markku) that she is specifically sharing her aesthetic experience with him. At the end of the message, she finally asks his advice on how to create animations (Markku responds to her a few minutes after). The message is "complex" – just like many SMS messages [20: 158-160].

Message 2. Korso 07-15-2002 16:08 Laura to Markku

Oh what a wonderful and lovable scene. I know that many people think that nature photos are boring, but I can't but rejoice when beautiful things can be found even in Korso. Another thing: can you make animations with this phone, and if you can, how?

For these reasons, we believe that to properly describe what takes place in multimedia messages, we need to analyze how they are constructed in detail, and understand them in terms of multimedia elements and their specific constellations in that form in which the recipients find them. Consequently, we have analyzed mobile multimedia as communication rather than, say, in terms of motives [4], genres [6], or even gift-giving [8]. For the same reason, our analysis has focused on methods of action rather than being built on analogies like the notion of "gift." An ethnomethdological perspective provides accuracy, but also detail that opens up new insights for designing services. Avoiding glosses like "humor" typical to content analysis makes analysis more difficult, but also makes it possible to spot message structures that can be used as templates in product development. These structures have the virtue of being grounded in existing, age-old practices.

Responding: Mobile multimedia as interaction

Perhaps the most important observation for our studies is that multimedia messages are something people use to interact with each other. This notion is based on ethnomethodological sociology and conversation analysis [10: 32-35, 11, 16, 21]. The point is not new; it has been made by Taylor and Harper [22] in their study of text messaging and by Berg et al. [8] in a design study on MMS. However, unlike Berg, Taylor and Harper, the Helsinki studies have specifically focused on actual responses to messages, not to mutual obligations on theoretical grounds. As this line of research is not construed from interviews, but from *actual chains of messages*, we have been able to study in detail forms of interaction mediated by mobile multimedia.

The argument has a solid footing in data. For example, we randomly selected 100 messages for closer analysis from the massive *Radiolinja* data. From these messages, we found 43 messages containing a textual formulation that seemed to link them to some previous or subsequent message. Of these 43 messages, we were able to find the other pair for 32 messages. For example, if our sample contained a question, we were most often able to find the reply from the following messages. For the remaining eleven messages, although it first appeared that there was for example a question in the message, it was not treated as such by the recipients – or the answer was delivered by other means. Thus, there is strong evidence that roughly every third message was part of a longer sequence. In addition, messages can link to each other implicitly, by e.g. addressing the same theme as the previous one(s). For example, photos of boy-friends were sometimes replied to with similar photos, without saying in the text that the message should be seen as a reply to the previous one.

MMS is not just a technology for interaction; interaction is an essential phenomenon to be studied if we want to understand mobile multimedia.

Some interaction formats in *Radiolinja* are familiar to anyone from ordinary life. For example, people routinely reply to greetings and hellos, answer questions, and either accept or turn down invitations [10]. Not responding would be considered rude. However, in some cases, the response is optional. For example, a typical postcard contains meaningless, unnecessary details. It is up to the recipient to pick these up and continue messaging, but he may as well choose not to do so.

Sometimes these beginnings lead to longer sequences of interaction. For example, teases have minimally a three-part structure: an opening message which is somehow laughable, a tease which formulates the message as laughable, and a response from the teased person to the teaser [23]. The following example has been analyzed by Kurvinen [13]. Message 3 is from Thomas, who informs six friends that he got engaged to his girlfriend by showing two hands with rings, reporting in the text that it had taken 15 years to get to this point. As Kurvinen notes, announcements of important family events, especially milestones such as marriages, babies born, graduations and funerals effectively call for replies. Not just any kind of reply will do: appropriate replies must be in line with the announcement; joyful events call for congratulations and tragic events invite empathy. The first messages of these pairs set a preference against which subsequent messages can orient to. The reply from Jarkko, one day after the announcement, draws from the visual material of the original to make a teasing comment on rings and engagements (Message 4).

Messages 3–4. Engagement News

Thomas to 6 people /
07-16 2002 16:42

It took 15 years! But good things come to those who wait.

Jarkko to Thomas /
07-16-2002 21:27

Congratulations! Jarkko, Leo and Topi

Messages 5–6. Engagement News Turns into a Teasing Series

Jani to Thomas
07-17-2002 17:12

Screw your ring. Nobody snatches me, except Miss Universe.

Thomas to Jani
07-17-2002 19:25

Well screw you! Just think that I've had more mornings with my woman than you with both of your hands! Hah hah heh hee

In Message 5, Jani teases Thomas by suggesting that he has lost his freedom. To strengthen his point, Jani takes a picture of his hand, which does not have a ring. However, Thomas does not accept this tease as such, but turns it into counter-tease by hinting about the solitary nature of Jani's sex life (Message 6). By closing his reply with a series of textual laughter tokens "Hah hah heh hee", he shows that his message is a tease – and a gleeful one.

The importance of interaction for understanding mobile multimedia is that it explains a good deal of variation in the frequency of messaging [21]. Some messages prompt several responses that are seen in statistics as peaks of activity. For instance, Thomas sent the news of his engagement to "all", that is, to the six people in the group. In response, he first got a series of congratulations from other members. However, as we have seen, Jani's response was untypical, as was Thomas' reply. Jani did not respond to Thomas anymore, but the important thing to pay attention to is that in all, this exchange consisted of 13 messages within two days. This episode explains 14 % of the group's messaging activity in those two days when it took place (N = 91 messages). Thomas' announcement was not a self-standing message, but situated in social action; messaging does not result solely from independent acts of expression, but from sequences in which preceding and subsequent actions relate meaningfully to each other.

Discovering methods of action from others

The final observation to be reported here is that as people use their multimedia devices, their messaging develops with experience. Experience is essentially a social process: different "trajectories" develop depending on what kinds of messages people are exposed to. Precedents are to be found in studies of how SMS use has evolved into innovative user cul-

Messages 7–8. The Giant Green Sociologist

Subject: On the Esplanade / 08-13-1999 15:35

"I'm on the Esplanade! Cool!"

Subject: The Giant Green Sociologist / 08-13-1999 16:36

"STOMP!"

"What was that?"

"STOMP!"

"Run for your lives ... it's ... it's ... the Giant Green Sociologist!"

"STOMP!"

tures [24, 25]. In the course of messaging, people learn new ways to construct messages, i.e. they learn new ethnomethods. Initially, they have a limited stock of methods at their disposal; at a later stage, this stock of methods is larger.

An illustrative case of this development comes from the pilot group of *Mobile Image*, which developed a habit of image manipulation. Message 7 is a typical postcard-like city scene, but the response in Message 8 was far from typical. The response had a Godzilla-sized colleague in the horizon, and it captures the horror felt by the crowd who sees the monster approaching. Also, the picture was digitally shaken to show how pedestrians feel the giant's footsteps.

Of course, this is an extreme case; most inventions are far less conspicuous. As an example of how users discover more ordinary ethnomethods in the course of making messages, we may look at how one group developed a lively riddle culture consisting of 70 messages during the first week of the study. The interesting point is that riddles are *always* based on images; not a single riddle was done with text alone. In fact, every part of a riddle could be visual, not just the opening message of the riddle.

In the group in focus, riddles can be classified into four main groups. The simplest riddles started with a picture of an obscure or an ambiguous object, and the text asked what it was. An example was a picture of a baby's toes shot from the front so that the shape of the foot disappeared. The second typical method was to show a detail so that the whole disappears. Parts of the body, parts of animals, and objects such as carpets worked in this way. In the third method, the image was obvious, but the text made it problematic. For example, in one message, a picture of a sunbathing man was turned into a riddle by asking "Guess what Lars is doing?" This text effectively rules out the obvious reply; what initially seemed evident in the photograph became questionable after reading the text. Finally, a fourth and less prevalent set of methods bordered on satirizing the riddle form. Several messages turned obvious objects – like pizzas and friends – into "riddles" of sorts. If one takes a picture of a glass of red wine, and asks others to guess what's in the mug, the fact that the answer is obvious shows that the riddle is to be taken as a joke.

The first riddles followed the traditional three-part format. However, in the course of *Radiolinja*, the structure of riddles became increasingly more varied. People learned to use:

- *Tips*. As already noted above, the sender may also give tips to recipients.

- *Rewards*. When starting the riddle, a bonus of some kind can be offered. For example in one message a bottle of beer was promised to whoever was able to give the right answer.

- *Side sequences*. Other things that make this structure more complex are side sequences in which the recipient may, for example, ask for a reward if the guess proved to be right. Such inquiries are typically responded to either by promising a reward, or by denying it.

- *Closing suggestions*. Recipients may show the sender that the riddle is getting boring by asking the joker to send the right answer and by making disapproving remarks or gestures about the quality of the riddle.

- *Visual responses*. Recipients' remarks may be visual and lead to visual commentary that takes place aside from the actual textual exchange.

More generally, people learn methods of action and uses of technology from other people. Consequently, their sensitivities for observation change: with new methods at their disposal, new things are considered relevant, funny, and reportable. At certain times, people use technology in certain ways. Later, they use it differently, even when the situation is more or less similar: they simply have different methods at their disposal. The history of use shapes future uses in the domestication of mobile multimedia.

Conclusions and discussion

In a series of studies conducted in the Helsinki metropolitan area, we have explored the use of multimedia technology in mobile phones from a classic ethnomethodological perspective informed by its offshoot, conversation analysis [9, 26]. The reason for building our analysis on this perspective rather than relying on other possible arguments available in literature on mobile phones is that ethnomethodology is capable of producing a rich description of what people do with new technology. Equipped with this description, we are able to get an idea of future uses even though the technology in use might change. As we are dealing with "age-old practices" of ordinary society [22], these practices do not develop nearly as quickly as technology. Thus, our descriptions will inform designers for years to come. Also, it is on this practical ground that larger scale social structures must be built to be successful – including technologies and services.

Most literature on mobile multimedia has tried to explain how it is used: mobile multimedia is treated as a dependent variable to be explained. Our studies in Helsinki have specifically focused on how people interact with each other through multimedia messages. Its basic premise has been that people use ordinary methods of interaction in this new domain. We share this assumption with Berg et al. [8], but have not followed their focus on gift-giving as an example of such an age-old practice. Instead, we have tried to describe methods people use to interact with each other. This analysis differs from other proposed explanations, in particular the motivation-centred argument proposed by Kindberg and his colleagues [4, 27], but less with the cultural argument proposed by Ling and Julsrud [6], which we see as a special case of the more general ethnomethodological perspective.

What if mobile multimedia is treated as an independent variable? What are its consequences? Recently, the French sociologist Carole Anne Rivière has been critical of its consequences. She sees multimedia as intimist and sensational communication that turns real things into spectacle: "Being multimedia tools, they increasingly use intimate play context, which have no rational purpose but rather aim at sensations, and in which the search for immediately shared pleasure is more and more visible" [28: 212]. Koskinen [29] uses less gloomy wording, but still characterizes multimedia as one that promotes a "banal" vision of the world. It supports mutual bonding, but turns the shared everyday into a commentary of small, trivial things rather than, say, politics or religion. If people reproduce "age-old practices" in the new domain through their actions as Berg et al. suggest [8], mobile multimedia contributes to person-to-person interactions rather than to wider institutional structures.

Our observations in this paper are perhaps less pessimistic about how multimedia technology gets domesticated in everyday life. First, people do many kinds of things with mobile multimedia, including many kinds of useful things. Second, we should not be too cynical about mutual entertainment. Is there something wrong in technology that supports fun and mutual rather than commercial entertainment, if it relieves boredom in ordinary life and maintains contacts between friends? Third, we have only begun the work of explicating mobile multimedia. As we see it, using mobile multimedia reproduces old cultural forms, but there are more innovative uses as well. In any case, our perspective shows that mobile multimedia gives people an opportunity to transfer their aesthetic, moral, and humorous experiences into a wireless domain in ways that were not possible using SMS or mobile phones. However, these aspects of life are rooted in ordinary experience rather than in the imaginary worlds of computer games or science fiction – although they may inform users of mobile multimedia.

What kinds of implications do these studies have for service providers and designers of services? The first level concerns how to "prototype" social action. We have consistently treated our studies as prototypes rather than real data. We have given people an opportunity to use (then) futuristic technology to see how they would use it, should it become common one day. We have monitored messaging using an up-to-date social science framework. Our objective has been to produce a picture of what people are going to do with mobile multimedia in the near future. From a design perspective, a picture like this is immensely valuable: building a technology and service roadmap is easier if one knows in advance how technology is going to be used. Of course, we do not know how well the picture we have given describes future uses. Hearsay evidence from other countries suggests that we have not missed the mark totally; in any case, the analysis we have proposed seems to describe the phenomenon better than, say, research that sees digital technology as a world-changing technology in mythical terms [30]. Second, at a more practical level, the utility of our analysis is more questionable. We have proposed several services to our technology partners immediately after our studies were done, and built some specific prototypes to illustrate them as well [31], but we do not know about the possible afterlife of these concepts in company memory. We believe that the approach could be most useful as part of an iterative developmental cycle, not as one-off studies as we have done so far. Unfortunately, this has not been possible, partially because recently mobile operators (at least in Finland) have focused on price-driven competition and market shares instead of e.g. on product and service development. In consequence, multimedia messaging services are still pretty much the same as they were at their launch.

Finally, we are fully aware of some limitations of our studies. Perhaps the main problem is that when we focus on interaction, we can easily lose sight of "larger" social structures. In our analysis, these are relevant only as local achievements – for example, as race, nationality, or gender labels in messages. However, in the long run, a more structural analysis is necessary for research and product development. Secondly, we have focused on a sort of visual chat among young city-dwellers. To generalize from this group is obviously risky, although we have been able to ground our research on interactional patterns that exist for people at large and also outside this domain. Third, institutional context needs to be added, perhaps along the lines proposed by Ling and Julsrud [6], who studied camera phone use at workplaces, or along the lines proposed by Katz [32], who has suggested that people use mobile phones to share religious information and also images of religious objects like the Pope. Finally, due to the experimental nature of our studies, we cannot evaluate the effects of price on multimedia use. Regardless of these limitations, we hope to have shown how design research can contribute not just to technology and service development, but also to social scientific understanding.

Acknowledgements

We wish to thank the Ministry of Trade and Transportation for financial support of the Mobile Image study, and the mobile carrier Radiolinja for inviting us to monitor their technology pilot and for funding the MMS study. In addition, we wish to thank the Nokia Research Center, Tampere, several colleagues at the University of Art and Design, Helsinki, and numerous others who have commented on our presentations and writings. Finally, we are in debt to the people we have studied; they have been generous enough to let us see their messages.

References

1 Silverstone, R, Hirsch, E, Morley, D. 1994. Information and Communication Technologies and the Moral Economy of the Household. In: Silverstone, R and Hirsch E (eds.). *Consuming Technologies. Media and Information in Domestic Spaces*. London, Routledge.

2 Silverstone, R, Haddon, L. 1996. Design and the Domestication of Information and Communication Technologies. Technical Change and Everyday Life. In: Mansell, Rogin and Roger Silverstone (eds). *Communication by Design. The Politics of Information and Communication Technologies*. Oxford, Oxford University Press.

3 Okabe, D, Ito, M. 2004. Camera Phones Changing the Definition of Picture-Worthy. *Japan Media Review*. Annenberg School for Communication, USC. www.ojr.org/japan/wireless/1062208524.php. Accessed April 15, 2004.

4 Kindberg, T, Spasojevic, M, Fleck, R, Sellen, A. 2004. *How and Why People Use Camera Phones*. Consumer Applications and Systems Laboratory. H&P Laboratories Bristol, HPL-2004-216, November 26, 2004.

5 Scifo, B. 2005. The Domestication of the Camera Phone and MMS Communications. The Experience of Young Italians. In: Nyíri, K (ed.). *A Sense of Place*. Vienna, Passagen-Verlag.

6 Ling, R, Julsrud, T. 2005. The Development of Grounded Genres in Multimedia Messaging Systems (MMS) among Mobile Professionals. In: Nyíri, K (ed.). *A Sense of Place*. Vienna, Passagen-Verlag.

7 Licoppe, C, Heurtin, J-P. 2001. Managing One's Availability to Telephone Communication through Mobile Phones. A French Case Study of the Development Dynamics of Mobile Phone Use. *Personal and Ubiquitous Computing*, 5, 99–108.

8 Berg, S, Taylor, A S, Harper, R. 2003. Mobile Phones for the Next Generation: Device Designs for Teenagers. *Proceedings of CHI 2003*, April 5–10, 2003, Ft. Lauderdale, Florida, USA.

9 Garfinkel, H. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ, Prentice-Hall.

10 Koskinen, I, Kurvinen, E, Lehtonen, T-K. 2002. *Mobile Image*. Helsinki, IT Press.

11 Koskinen, I, Kurvinen, E. 2002. Messages visuels mobiles. Nouvelle technologie et interaction. *Réseaux: communication, technologie, société*, No. 112-113, 107–138.

12 Kurvinen, E. 2002. Emotions in Action: a Case in Mobile Visual Communication. *Proceedings of the 3rd International Design and Emotion Conference D+E 2002*.

13 Kurvinen, E. 2003. Only When Miss Universe Snatches Me. Teasing in MMS Messaging. *Proceeding of Designing Pleasurable Products and Interfaces DPPI'03*, June 23–26, 2003, Pittsburgh, Pennsylvania, USA.

14 Battarbee, K. 2004. *Co-Experience. Understanding User Experiences in Social Interaction*. Helsinki, University of Art and Design.

15 Battarbee, K, Koskinen, I. 2004. Co-Experience – User Experience as Interaction. *CoDesign Journal*, 1 (1), 5–18.

16 Koskinen, I. 2005a. Seeing with Mobile Images. In: Nyíri, K (ed.). *A Sense of Place*. Vienna, Passagen-Verlag.

17 Suchman, L. 1987. *Plans and Situated Actions*. Cambridge, Cambridge University Press.

18 Heritage, J. 1989. *Garfinkel and Ethnomethodology*. London, Polity.

19 Chalfen, R. 1987. *Snapshot Versions of Life*. Bowling Green, OH, Bowling Green State University Press.

20 Ling, R. 2004. *The Mobile Connection*. San Francisco, Morgan Kaufmann.

21 Koskinen, I. 2003. User-Generated Content in Mobile Multimedia: Empirical Evidence from User Studies. *Proceedings of International Conference of Multimedia and Expo ICME 2003*, IEEE Publication, Baltimore, MD.

22 Taylor, A S, Harper, R. 2002. Age-Old Practices in the "New World": A Study of Gift-Giving Between Teenage Mobile Phone Users. *Proceedings of Computer-Human Interaction CHI 2002*, April 20–25, 2002, Minneapolis, Minnesota, USA, 439–446.

23 Drew, P. 1987. Po-Faced Receipts of Teases. *Linguistics*, 25, 219–253.

24 Kasesniemi, E-L. 2003. *Mobile Messages. Young People and New Communication Culture*. Vammala, University of Tampere Press.

25 Lorente, S. 2002. Special Number on "Youth and Mobile Telephones". *Revista de estudios de juventud*, 57 (juny).

26 Sacks, H. 1995. *Lectures on Conversations*. Cambridge, Blackwell. Two volumes.

27 Döring, N, Dietmar, C, Hein, A, Hellweg, K. 2005. Contents, Forms and Functions of Interpersonal Pictorial Messages in Online and Mobile Communication. In: *Proceedings of Communications in the 21st Century: The Mobile Information Society*, Budapest 2005.

28 Rivière, C. 2005. Seeing and Writing on a Mobile Phone: New Forms of Sociability in Interpersional Communications. In: *Proceedings of Communications in the 21st Century: The Mobile Information Society*, Budapest 2005.

29 Koskinen, I. 2005b. Managing Banality in Mobile Multimedia. In: Pertierra, R (ed.). *Place, Identity, and Media*. Singapore, Singapore University Press.

30 Mosco, V. 2004. *The Digital Sublime. Myth, Power and Cyberspace*. Cambridge, MA, The MIT Press.

31 Kurvinen, E, Koskinen, I. 2003. Mobile Photo Album: An Experience Prototype. In: Koskinen, I, Battarbee, K, Mattelmäki, T (eds.). *Empathic Design*. Helsinki, IT Press.

32 Katz, J. 2005. Magic in the Air: Spiritual and Transcendental Aspects of Mobiles. In: *Proceedings of Communications in the 21st Century: The Mobile Information Society*, Budapest 2005.

Ilpo Koskinen is Professor of Industrial Design at the University of Art and Design, Helsinki. Trained in sociology and in conversation analysis, his present interests lie in studying mobile and ubiquitous communication technologies. In his opinion, studying these technologies requires that we understand them as human action, which happens to be mediated by technology. His other interests lie in studying how design changes cityscapes, and in interpretive design methodology which (he thinks) provides a useful, though underdeveloped approach to bridge the gap between design and the social sciences. At present, he is writing a book on mobile multimedia, hopefully finished in 2006.

email: ikoskine@uiah.fi

Industrial Designer (MA) Esko Kurvinen has worked as designer, researcher and consultant in the process industry and in the area of mobile communication. His research interests relate to interpersonal communication and social interaction in technology mediated or technology intensive environments and situations.

email: esko.kurvinen@hiit.fi

Infrastructure for a mobile workforce

OLE TOM SEIERSTAD

Ole Tom Seierstad is Chief Security Advisor in Microsoft Norway

Mobile users need more than checking email and voice mail. They need to access important information, acting on that information to the benefit of the business and to collaborate efficiently with customers, partners, and colleagues. This paper presents a solution by Microsoft that enables the mobile users to connect to the corporate network from anywhere, anytime, and by using any device. Security from intrusion and viruses must pervade every part of this vision, from the mobile devices to the software they run, to the infrastructure that supports them.

Introduction

Business communication is more than checking email and voice mail. For mobile users, it is immediately accessing important information; acting on that information to the benefit of the business; and collaborating efficiently with customers, partners, and colleagues. In other words, successful business communications rely on an efficient exchange between people and data, which Microsoft's mobile solutions enable.

In the not too distant past, mobility was merely taking laptop computers on the road. Deployment was simply selecting the best laptop computer and setting up phone lines for dial-up network access. And mobile users had simple access to files and connections to corporate email servers.

Expectations have increased, however, and the Internet has enabled a new vision for mobility. To compete, one must take advantage of this new vision in order to remain competitive in a world where data flows easily, people stay in ready touch, and mobile users are as effective away from the office as they are in the office. To achieve the goal, mobile users need access to their line-of-business (LOB) applications. They need robust access to their email, and they require collaboration tools that help them share and access information. To complete the vision, mobile

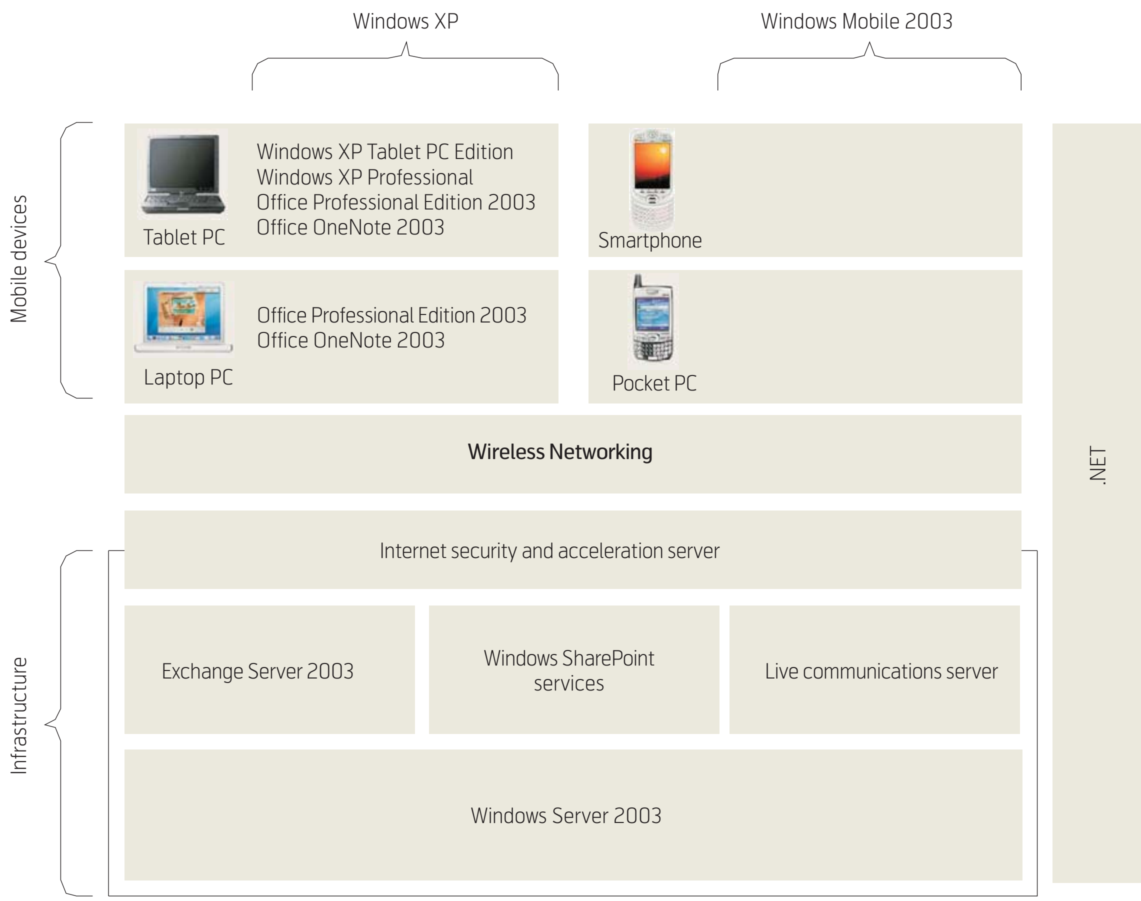


Figure 1 Microsoft mobile vision

ISSN 0085-7130 © Telenor ASA 2005

	Road Warrior	Corridor Warrior	Telecommuter	Data Collector
Profile	• Connects to customers and colleagues	• Spends most time away from desk • Collaborates with colleagues	• Accesses the corporate network irregularly	• Collects and processes data • Delivers enhanced services to clients
Travel	• Mostly external	• Some external travel • Meetings all day	• Accesses network from home or home office	• > 80 % externally
Business Require-ments	• Access to Email, Calendar and PIM • Access to corporate net-work resources and line-of-business (LOB) applications • Data and access security	• Access to Email, Calendar and PIM • Access to information and applications while away from desk • Data and access security	• Access to information and applications while at home office • Data and access security	• Access to corporate network resources and LOB applications • Data and access security
Form Factors	• Laptop PC • Tablet PC • Pocket PC • Smartphone	• Laptop PC • Tablet PC • Pocket PC • Smartphone	• Laptop PC • Smartphone	• Pockets PC • Tablet PC
Mobile Access	• Public wireless • Dial-up • Virtual private networking (VPN)	• Corporate wireless • Remote Desktop	• Cable, DSL • VPN • Remote Desktop Dial-up	• Public wireless • Dial-up • VPN
Examples	• Executives • Consultants • Sales representatives	• Executives • Healthcare workers • Students	• Knowledge workers	• Field service employees in various industries

Table 1 Mobility scenarios

users require the ability to connect to the corporate network from anywhere, anytime, and by using any device. Security from intrusion and viruses must pervade every part of this vision, from the mobile devices to the software they run to the infrastructure that supports them.

Figure 1 shows how to achieve the new vision for mobility from Microsoft.

Mobility scenarios

Even though mobile features offer higher levels of productivity to everyone in an organization, different types of users gain these benefits in different ways. Four types of mobile users are identified, who can benefit from these features. They are road warriors, corridor warriors, telecommuters, and data collectors. Table 1 summarizes these users, and the sections following this one provide more detail about each:

- *Road Warrior*. Professionals who travel frequently and require remote access
- *Corridor Warrior*. Knowledge workers who spend most of their time away from their desks
- *Telecommuter*. Knowledge workers who work occasionally at home
- *Data Collector*. Field service employees who travel fulltime, requiring remote access

Road Warrior

Road warriors are typically professionals who travel frequently, such as sales professionals, spending up to 80 % of their time out of the office. Whether they are corporate executives, members of a mobile sales force, or field representatives; they require technology that enables them to stay connected with customers and colleagues and allows them to work anytime from anywhere. Because road warriors are not within easy reach of the help desk, they need robust hardware and software on the road. And spending most of their time away from the office does not reduce their need for the same applications that colleagues at the office use.

Road warriors have a broad choice of mobile devices from which to choose. They can use a traditional laptop computer, or Tablet PC with wired and wireless networking. In any case, road warriors require access to corporate network resources and LOB applications through dial-up connections, VPN, or wireless public access. They tend to use their handheld devices for calendaring, email, and travel planning, so they can back up their primary computers with a Pocket PC or Smartphones. Their information is with them regardless of whether they're in a taxi, waiting for a flight, or on the go.

ISSN 0085-7130 © Telenor ASA 2005

Corridor Warrior

Corridor warriors are professionals who work in the office but spend most of their day away from their desks. Examples of corridor warriors include healthcare professionals and managers. They travel less than 20 % of the time and are usually dashing through the corridors from one patient or one meeting to the next. They require access to information and applications while away from their desks, ad-hoc collaboration with their colleagues, and a paperless environment.

The primary form factor for corridor warriors is the Tablet PC. They can also supplement traditional desktop computers with Windows Mobile-based Pocket PCs or Smartphones. Regardless, wireless networking is the key to keeping these professionals productive, and wireless Tablet PCs or Pocket PCs help them stay on top of their day. As well, OneNote running on a Tablet PC allows them to take handwritten notes which can be organized and reused efficiently; and InfoPath plugs them into their company's business processes. Adding the collaborative features of Exchange Server 2003 and SharePoint Portal Server further increases the productivity of these professionals. For example, Meeting Workspaces help make meetings more productive by getting information about them to attendees early.

Telecommuter

Telecommuters work at home at least 1 day a week. They are a special case for the IT department, often accessing corporate resources from a home computer by using a broadband or dial-up connection. With the Windows XP Professional feature Remote Desktop, telecommuters can access their office desktop PCs and corporate resources as though they were sitting in front of it. Telecommuters who leave the office only occasionally might take their Windows XP Professional-based laptop PCs back and forth to the office.

In any case, telecommuters either have a desktop computer at home or a laptop computer that they take home with them. They require remote access to corporate network resources using VPN through dial-up or broadband Internet connections. They also connect to their primary desktop computer by using Remote Desktop.

Data Collector

Data collectors are field service employees in a variety of vertical industries. They tend to collect and process data, deliver enhanced services to better serve clients, and are out of the office or away from their desks more than 80% of the time. For example, delivery drivers, warehouse employees, and insurance claim adjusters are examples of data collectors. To help them perform their jobs, a plethora of vertical-market applications are available for wireless Pocket PCs.

In addition to using Pocket PCs to collect data, data collectors can use Smartphones. In any case, they require access to LOB applications through wireless networking and 3G/GPRS.

Wireless networking

Wireless networking has revolutionized mobile computing for all types of mobile users. They are more productive, as they can access corporate network resources by using any public wireless network and can roam the building with their laptop computers and still have network access. When public wireless networks (short-ranged, high-speed 802.1x wireless networks in places like coffee shops, airports, and so on) are not available, they can still access their information by using their Smartphones to connect to the network using mobile operator data services (long range, low-speed data services such as 3G, GPRS, PCS, and so on). Wireless networking means that mobile users can squeeze more work time out of their day by gaining 8 to 12.5 hours per week in added productivity[1]. The following list shows how wireless networking makes road warriors, corridor warriors, and data collectors more productive and improves their quality of life:

- *Road warriors*. Wireless networking can turn idle time into productive time by allowing road warriors to connect to corporate network resources that traditional wired network connections cannot access. For instance, sales professionals waiting for a delayed flight can use their laptop computer to connect to the corporate network by using a public hotspot at the airport. Once connected, they can check email, access product catalogs, collaborate with colleagues via a SharePoint site, and so on. The result is that mobile sales professionals are able to make decisions quicker, since they do not have to delay those decisions until they return to the office or find a wired network connection.

- *Corridor warriors*. By definition, corridor warriors spend more time away from their desks than they do sitting at them. Healthcare professionals are great examples of corridor warriors. Wireless networking keeps corridor warriors connected to the network whether they're in meetings, consulting with patients, or working in the front office. They can use their Tablet PCs to look up patient records, take notes, and send a quick message to a col-

[1] *Empowering Mobile PC Users: The Financial Benefits of Windows XP Professional and Office XP,* Bearing Point, December 2002.

league. And they can use Remote Desktop to connect to their desktop computers, using them as though they were sitting at their desks typing on the keyboard. Wireless networking keeps corridor warriors connected even when wired network connections are not available or are not practical.

- *Data collectors*. Wireless networking is almost a necessity for data collectors. The locations where data collectors typically work are not practical locations for wired network connections. For example, installing wired network connections on a manufacturing floor or garage is not feasible. Wireless networking used in conjunction with devices like Pocket PCs and specialized applications not only makes data collection in these places more feasible but also makes it cost effective. Data collectors in the field, such as delivery drivers, can also rely on wireless networking at public hotspots to synchronize their data when their mobile operator data services connections are not available.

For all types of mobile users, wireless networking allows mobile users to connect to corporate network resources where traditional wired network connections are not available. It enables immediate collaboration regardless of whether a wired network connection is available. Decisions are made quicker because the usual delay of users returning to their desktop computers is eliminated by an instant, wireless connection to the corporate network.

Windows XP

Windows XP Professional has new features and enhancements that make remote and wireless access simple for any wireless user, which in turn provides significant productivity gains for employers. The following list describes the features of Windows XP Professional and the Wireless Zero Configuration service:

- With a wireless network adapter installed, Windows XP Professional searches for available networks to which it can connect. When an available network matches a preferred network, Windows XP Professional connects to it. If there are no configured preferred networks or no preferred networks are found, users can also select a specific network to which they want to connect. Users can prioritize the list of preferred networks – Windows XP Professional stores the list and connects to the networks in the chosen order. Connection management is possible without user intervention, but user interaction is sometimes necessary to choose specific networks or prioritize connection order.

- Automatic configuration makes wireless networking more practical for mobile users. All types of mobile users can easily configure their wireless connections. Wireless networking and simplified network configurations benefit IT professionals, too. Because users are more able to configure their own network connections, they become more self-sufficient. IT professionals no longer need to plan and configure every connection when they deploy the operating system to mobile users.

- As users physically move their wireless computers from room to room, Windows XP Professional automatically remains connected by finding the best wireless AP with which to communicate. When it finds a new wireless access point, it automatically negotiates authentication and authorization with that wireless access point without user intervention, which provides a great experience for the mobile user. A user can move anywhere within a given wireless network and remain *connected* to the network.

Windows XP Professional makes wireless networking better in ways other than provided by the Wireless Zero Configuration service. For example, the operating system includes device drivers for most of the popular wireless adapters. And the device drivers that come with Windows XP Professional fully support Wireless Zero Configuration. In many cases, IT professionals do not need to deploy third-party device drivers for their wireless adapters, and mobile users do not have to download and install device drivers on their phones.

Windows Mobile

Windows Mobile software has support for wireless networking similar to Windows XP Professional. Like Windows XP Professional, Zero Configuration Wi-Fi makes finding and connecting to any 802.1x wireless network easy for Windows Mobile users. When a Wi-Fi-enabled Pocket PC encounters wireless hotspots, Windows Mobile automatically asks users if they want to connect to the network. Afterwards, Windows Mobile saves the connection's settings so that it connects automatically the next time the device is in range of the wireless network.

Messaging and Security Feature addresses business customers' requests for a faster, more direct messaging experience, improved security management, and increased cost efficiency and scalability for their mobile messaging solutions. The Messaging and Security Feature Pack enables IT administrators to better manage and protect information on a device, and includes direct push technology, which helps business users keep their Outlook® Mobile information current by delivering updates quickly and directly to a Windows Mobile-based device from Exchange Server 2003 without requiring businesses to pay for additional and costly servers and middleware.

Enhanced Outlook Mobile Experience

The Messaging and Security Feature will offer businesses the premier mobile messaging solution, going beyond a plain text e-mail model to provide a fast and familiar Outlook Mobile experience. End users will have access to a wide range of business information through the following new features:

- *Windows Mobile direct push technology*. Outlook data – including e-mail messages, calendar and contact information, and new support for Tasks – is pushed directly from Exchanger Server 2003 to a Windows Mobile-based device, providing users with immediate access to Outlook information. There is no need for the additional server infrastructure required by other solutions on the market.

- *Wireless support for contact information*. Support for the ability to look up global address list (GAL) information over the air, a top request from partners, enables business users to access full contact details of co-workers on a Windows Mobile-based device. The feature is broadly integrated throughout the phone dialer, contacts, calendar and inbox applications, providing users with a consistent, productive messaging experience.

Better management and security

While increasing productivity for end users, the Messaging and Security Feature also will improve efficiencies for information technology (IT) organizations. Building on existing capabilities, IT professionals will be able to manage Windows Mobile-based devices the same way they manage PCs and servers through new features that support the following:

- *Remotely enforced IT policy*. Via the console, IT administrators can remotely manage and enforce select corporate IT policy over the air. For example, an IT administrator can mandate a personal identification number (PIN) password to be set for every device, establish recommended and mandatory policies, and draw up lists of users who should be exempt from these policies.

- *Local and remote device wipes*. To better manage sensitive information on a misplaced Windows Mobile-based device, IT administrators can remove all information over the air and reset the device back to its original state. In addition, the administrator can choose to have a device's local memory erased if the correct password is not entered after a specified number of attempts.

- *Certificate-based authentication*. Native support for Secure Multipurpose Internet Mail Extensions (S/MIME), a leading security request from partners, enables Windows Mobile-based device users who are on the go to access their corporate network without using a separate password, and eliminates the need to store corporate login credentials on the device.

Scalable, cost-effective mobile solutions

Beyond device management and security, price and scalability are key concerns for businesses as they consider mobile device deployments. The Messaging and Security Feature will provide a scalable and easy-to-manage mobile messaging solution that gives users a broad choice of device designs and eliminates the need for businesses to outsource their mobile messaging to third parties and purchase additional middleware software or costly servers.

Wireless infrastructure

Wireless networking is beneficial to an enterprise's infrastructure in the following ways:

- *Temporary network connectivity is more practical and less expensive*. Wireless networking makes it feasible to set up, use, and take down temporary networks as required. For example, a company configures temporary wireless networks at trade shows for the benefit of staff and attendees. A more general example is creating a temporary wireless network for the final crunch of a big project, which can be disassembled at the project's completion.

- *Wireless networks are quicker to deploy than wired networks*. Wireless networks are more flexible than wired networks. They are quicker to deploy because one does not have to run cable throughout the building.

- *Wireless networking is more feasible in locations where wired networks are not practical*. Locations such as conference rooms and cafeterias are not practically wired, and wireless networking makes it easy to add a network to those rooms. And traditional wired networking is not practical outdoors, but wireless networking is a perfect solution for an outdoor network. Additionally, older buildings and some types of construction prohibit pulling wires for a traditional network. Wireless networking significantly reduces the cost of networking in those scenarios and enables networking in environments that were not possible. For that matter, wireless networking is cheaper to deploy than cabling a building, and it scales more easily.

Figure 2 Mobility usage scenarios

Wireless security

While Wi-Fi has experienced a rapid growth, the industry has raised a number of security issues. The Wi-Fi standard defines authentication and encryption services based on the Wired Equivalent Privacy (WEP) algorithm. The problem is that WEP provides weak security and does not scale to typical enterprise wireless deployments because key management is almost impossible. Windows XP Professional and the industry's support for the following technologies combine to make wireless networks more secure, however:

- *802.1X*. 802.1X provides authentication-based access control for wireless networks. For more detailed information about 802.1X and supporting protocols. Windows XP Professional ships with support for 802.1X. All major NIC vendors also support 802.1X and most have released Windows drivers that support it. Likewise, all major enterprise access-point vendors are supporting 802.1X.

- *WPA (Wi-Fi Protected Access)*. WPA is a replacement for WEP that provides more robust methods for encrypting data and authenticating network connections. The result is a new level of protection for mobile users who are taking advantage of the Windows XP Professional wireless features.

- *RADIUS*. RADIUS enables centralized authentication, authorization, and accounting for network access. Originally developed for dial-up remote access, RADIUS is now supported by wireless access points. Windows Server 2003 provides Internet Authentication Server (Microsoft's implementation of RADIUS).

- *Active Directory*. For wireless networking, Active Directory® domains contain the user and computer accounts for authentication and the Group Policy settings to deploy computer certificates.

- *Certificates*. 802.1X and related protocols rely on digital certificates to verify the identity of users and computers. For more information about public-key cryptography and the Windows 2000 public key infrastructure (PKI).

Mobile devices

This section describes the four basic types of mobile devices:

- Notebook PC
- Tablet PC
- Pocket PC
- Smartphone

Figure 2 relates the type of mobile user to the mobile devices most suitable for their requirements. Each quadrant in this figure represents a level of productivity and mobility, with the top-right quadrant representing the highest level of both. The sections following this one describe these devices in more detail.

Notebook PC

Notebook PCs (such as laptops, portables, etc.) are the staple computers for most business travelers. Current models contain in a slim form the processing power and conveniences that were limited to traditional desktop computers only a few years ago

Tablet PC

Tablet PCs give mobile workers a powerful, versatile computing experience. The Tablet PC is a fully functional laptop that runs Windows XP Tablet PC Edition. Windows XP Tablet PC Edition is a superset of Windows XP Professional designed to add pen computing to an already rich feature set, not a stripped-down version of the operating system. Windows XP Tablet PC Edition provides new, advanced handwriting and speech recognition, which is surprisingly accurate, enabling the creation, storage, and transmission of handwritten notes and voice data.

Windows XP Tablet PC Edition adds the following to the features of typical laptop PCs:

- *Tablet PCs run Windows XP Professional with specific enhancements for tablet computers*. Users get all the features, reliability, and power of Microsoft's latest operating system, as well as handwriting and speech recognition in a tablet form factor.

- *Tablet PCs can run any application that is compatible with Windows XP Professional*. Users have access to the same tools that they have on their desktop computers in more versatile, tablet computers.

- *Tablet PCs are available in two form factors, pure tablet and convertible, allowing users to choose a model that best suits their needs*. Pure tablets do not have a keyboard and rely on the pen as the primary input device. Convertible tablets have a keyboard and mouse but allow users to rotate the screen to use it like a tablet.

- *Tablet PCs support docking stations and a variety of common accessories*. Users can take advantage of a full complement of peripherals, including a full-size monitor, keyboard, mouse, printers, and external drives.

- *Tablet PCs enable surprise hot-docking as well as fast standby and resume*. Surprise hot-docking allows users to grab their Tablet PCs from their docking stations and get down the corridor in a hurry. Also, Tablet PCs resume from standby mode in about five seconds. This functionality allows users to save power when they are not using their computers without wasting time when they are ready to use their computers again.

Last, pen and ink features extend to ink capabilities so one can use tablets' pens to handwrite text in Outlook 2003, Word 2003, PowerPoint® 2003, and Excel 2003. And OneNote™ 2003 and Tablet PCs empower mobile users to collaborate more effectively by enabling them to take comprehensive notes, access crucial information, and share notes with team members.

Windows Mobile Pocket PC

Windows Mobile-based Pocket PCs give users the freedom to manage their work while on the go. They provide users the choice to communicate how they choose and stay connected to corporate information when they are not in the office. Pocket PCs include familiar Microsoft software, including Pocket versions of Outlook, Word, Excel, and Internet Explorer. And, since Pocket PCs integrate with users' desktop computers, users can take their files with them without having to learn new software.

Here are some examples of what users can do with Pocket PCs, using a form factor that fits in their pockets:

- Wirelessly exchange email and browse the Web;

- Wirelessly exchange messages by using MSN® Messenger;

- Open email attachments and edit them by using Pocket Word, PowerPoint and Excel;

- Access travel information, including local weather, flight status, and maps;

- Make travel reservations, such as car, hotel, flight, and dinner.

ISSN 0085-7130 © Telenor ASA 2005

Microsoft recently released Windows Mobile™ 5 that adds many features that mobile business users need. It includes Zero Configuration wireless, for example, just like Windows XP Professional. It supports Bluetooth, which makes connecting a Pocket PC to other Bluetooth-enabled devices easy. ActiveSync® technology synchronizes users' devices with Windows XP Professional and Exchange Server 2003. Microsoft updated Connection Manager to make managing connections easier for users. And Windows Mobile™ 2003 includes better support for messaging, input, contact management, calendaring, and much more.

Smartphone

Windows Mobile-based Smartphones extend mobile phone technology. Smartphones deliver a rich set of applications, highly integrated telephony functionality, and an open development platform. Mobile phones built on the Smartphone platform provide both voice and text communication and rich wireless data applications in one easy-to-use device, as well as make it a truly personal and powerful mobile phone for work.

With Windows Mobile software, Microsoft is introducing a new phone experience. The company is integrating PDA-type functionality that mobile users rely on to do their jobs into a voice-centric handset that is comparable in size to today's popular cell phones. Smartphones are designed for one-handed operation with keypad access to both voice and data features. They are optimized for voice and text communication, wireless access to Outlook information, and secure connections to corporate networks and the Internet. ActiveSync synchronizes Smartphones with Windows XP Professional and Exchange Server 2003. The Windows Mobile-based Smartphone gives users a choice to communicate via voice or text along with the ability to access information and services so one can stay in touch while on the road.

Mobile collaboration

Everything that this paper has presented to this point has led to collaboration. Providing mobile users tools for connecting to the corporate network and building out the infrastructure to support those connections are the foundations for collaboration. The following list describes the Windows XP Professional and Office Professional Edition 2003 features that enable mobile users to collaborate more effectively:

- *Office Professional Edition 2003*. Office Professional Edition 2003 allows organization and employees to transform information into impact. It includes new and enhanced functionality to help people take advantage of this information, enabling them to communicate more effectively, collaborate with co-workers, and improve business processes.

- *OneNote 2003*. OneNote 2003 enables mobile users to collaborate effectively. It allows them to contribute in an increasingly demanding business environment by taking more accurate, comprehensive notes. It helps mobile users to access crucial information so they can make better decisions and take more effective action. Last, OneNote 2003, shown in Figure 3, gets teams and organizations on the same page so they can work together faster and more easily

- *Windows Messenger*. Windows Messenger is the way for mobile users to communicate with their customers, partners, and colleagues. It keeps users updated with their contacts' online status. Users can collaborate with their online contacts, transfer files, and share applications and whiteboard drawings. Windows Messenger provides users with a great platform for online conferencing and collaboration and also improves relationships with customers and partners, as well as enhances communication among employees.

- *SharePoint technologies*. Windows SharePoint Services 2.0 and Microsoft Office SharePoint Portal Server 2003 enables mobile users to share information whether they are in the office or on the road. When they use it with Office Professional Edition 2003, SharePoint products and technologies provide benefits such as a central repository for documents, an interactive work environment for collab-

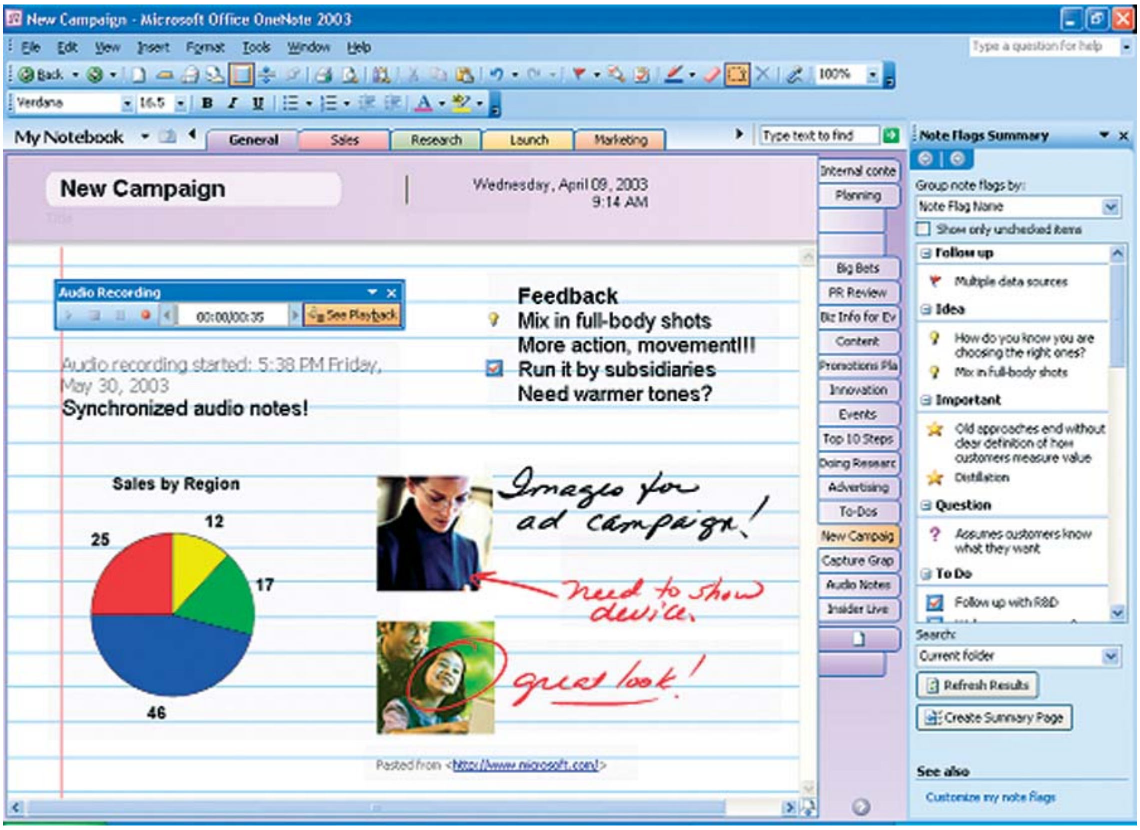


Figure 3 OneNote 2003

oration, and an architecture that allows to separate the Web-hosting processes from the site data-retrieval processes.

Summary

Mobile business users need the ability to work when, where, how, and with whom they want – effortlessly. They need immediate access to information and people so they can act to benefit the business. Their success relies on an effective exchange with their customers, partners, and colleagues, as well as on their data, which the vision that this paper describes enables. This vision has the following attributes:

- *Connected*. To enable mobile business users to stay connected. They can communicate easily and collaborate with customers, partners, and colleagues more effectively.
- *Productive*. To enable mobile business users to work as productively away from the office as in the office.
- *Dependable*. To provide a secure and reliable infrastructure for supporting mobile business users. Dependability is critical on the road; they can rely on these solutions to help secure their information and their network connections ensuring performance at the highest levels.

Source

Source for the article: Jerry Honneycutt, Microsoft Corp.

Ole Tom Seierstad has been with Microsoft Norway for 15 years – his current position is Chief Security Advisor with focus on Microsoft security products and messaging. His previous positions in Micrsoft include windows Mobile Evangelist, Technical Support Manager and head of MSN Norway.

email: oles@microsoft.com

Flexible service integration with BREW

RICHARD SAVAGE

Richard Savage is Director Business Development at Qualcomm Europe

One key challenge device manufacturers are facing right now is how to produce feature rich handsets, at a price level that can be achievable for the mass market. Further, they need to replicate this model on different handsets and different tiers of handsets, providing the operators the common user experience they are looking for. This paper presents the Qualcomm's BREW solution which is open, globally standardised, hardware-independent and consistently deployable on any network and on any mobile device. It gives users an easy to use and customisable user interface and includes a method for discovering, buying, downloading and managing applications on client devices.

This article has been prepared in collaboration with executives from Qualcomm Europe.

Data services is the driving force

Mobile phones have become mini computers in your hand – capable of running a variety of applications. Yet this capability is constrained by market demands that require mass-market phones to be small, inexpensive, light weight and consume little power. These constraints, in turn, have restricted device processing power and memory (storage).

The historic approach to delivering applications to phones has been to shift processing power to servers in the operator's network or to a third-party company via the Internet (known as server-side execution). Portals and mobile games using WAP browsers exemplify this strategy; content and applications are generated by remote servers, passed through the network and displayed in the browser on the mobile phone. The user then keys in responses to choices displayed on the device screen, which are sent back through the network to the server for processing and response.

The weaknesses of the browser-based, server-side approach are clear: high latency and limited interactivity on the mobile device. This makes the most exciting, graphic-rich interactive applications impossible. In effect, with a browser-based solution that requires server-side processing, the mobile phone becomes an unnecessary performance bottleneck and prevents the best applications from being developed because of its inability to process information locally.

However, decreases in size, power consumption and cost of silicon chips have enabled a second approach – putting more processing power on the mobile device. This opens up a new range of applications based on local, or *client-side*, processing.

This is expected to be the predominant growth area for new mobile wireless applications – rendering today's slow server-side approaches virtually irrelevant for many new types of applications. Rather than only using the browser on the mobile phone to run and interact with applications, it becomes a true platform for software applications that can offer a startling array of new consumer data services provided by the network operator.

One key challenge device manufacturers face right now is how to produce feature rich handsets, at a price level that can be achievable for the mass market. Further, they need to replicate this model on different handsets and different tiers of handsets, providing the operators the common user experience they are looking for.

Recognizing these challenges in handset application execution, and the value of a technology model that permits client-side processing of data applications, QUALCOMM unveiled the BREW solution in 2001. The BREW solution was engineered to overcome challenges relating to application execution and specifically address:

- The lack of a unique framework that is easy to replicate on different vendor handsets;
- Programming language inconsistency and fragmentation (Java);
- The possibility of offering the same user experience across different tiers of handsets;
- The pursuance of the user experience independent of new features or standards coming out when the handset is commercially available;
- The possibility to tailor the user experience to the demands of the customer.

 ISSN 0085-7130 © Telenor ASA 2005

Ideally, applications should be scaled perfectly with the various tiers of phone models and make efficient use of the handset's resources and provide a consistent user experience. To succeed at this task requires a complete solution that addresses the needs of all devices – not just the high-end.

Essentially, the BREW Client provides a software connection between the low-level functions of the mobile device and the higher-level applications written by third parties. Because the BREW Client functions as an abstraction layer in the Flash and RAM, BREW software developers do not need to intimately understand embedded systems programming. It provides consistent Applications Programming Interfaces (APIs) available to third-party developers allowing them to focus on the design and creation of the application, not the underlying hardware.

The benefits of this approach are clear: streamlined and efficient software development. With BREW, porting applications from device to device becomes a straightforward task. Time-to-market is vastly reduced, and new applications work consistently from one phone to another.

The BREW Client's easy software porting makes BREW functionality readily available to all handset manufacturers. The BREW platform is air-interface independent and has been ported to a host of air interfaces including CDMA IS-95A, IS-95B, 1x, 1xEV-DO, GSM/GPRS and UMTS handsets. The BREW Client has also been ported on a variety of devices powered by non-QUALCOMM chipsets. Of all available options for client-side processing, the BREW solution is the least expensive and easiest to implement, with the lowest requirements for phone design.

2 Services and application enabling environment

In terms of mobile application and service distribution, the road to profit for mobile network operators has always included providing services their customers find valuable, and to a certain degree addictive. To date, voice has been the leading application on mobile networks but it is clear that non-voice services (data applications) offer an enormous opportunity.

These data applications include instant messaging, games, multimedia, news, entertainment and productivity enhancement tools, to name a few, which appeal to both consumers and business professionals. They also include applications that allow network operators to more effectively communicate with and retain their customers, such as those that customise user interfaces and update handset-dependent network features. New voice-based applications – voice recognition and push-to-talk capabilities – are emerging as well.

As services evolve from voice to data applications, there is also the wireless value chain to consider, one that is growing longer and more complex. A value chain that doesn't offer a viable revenue opportunity for all stakeholders is a chain that will break. There must be profits for application publishers/developers, network operators, device manufacturers and others, in order for new services and useful, relevant applications to proliferate.

While it is possible to develop technology to deliver value, it is a considerably more complex problem to ensure a realistic value chain. The Internet provides a historical example of this concept: protocols such as HTTP provided a method for delivering free World Wide Web applications and services to users, but included no mechanism for reimbursing content providers. The problems of reimbursement and profitability on the Web had to be worked out over a period of years, during which thousands of companies failed and billions of dollars were lost. Overheated publicity and incomplete solutions can stall a market.

For network operators who have to justify expensive network development and spectrum fees, a slow and uncertain approach is not an option. The same is true for application publishers/developers and for manufacturers of devices and hardware – both have limited resources and a broadening range of technologies to support. Any approach to providing new data services must therefore begin with a well-developed plan for revenue sharing along the entire value chain.

The best way for the wireless industry to deliver these applications, in light of these issues, is a solution that:

- Is open, globally standardised, hardware-independent and consistently deployable on any network and on any mobile device;
- Gives users an easy to use and customisable user interface;
- Includes a method for discovering, buying, downloading and managing applications on client devices.

ISSN 0085-7130 © Telenor ASA 2005

3 Enabling the client side

On the client side, the following components are included:

- *A standard Application Execution Environment*. The BREW Client exposes a common set of application programming interfaces (APIs) for standardized development of wireless applications. Figure 1shows where the BREW Client API layer resides on the device.

- *True real-time processing*. With client-side processing, a new range of applications (such as 3D and multi-player games) have been developed and deployed because of the ability of the BREW platform to download and run applications locally on the phone. Even on circuit-switched networks, applications start immediately since they reside on the phone – with no need to wait for a data call to be initiated. True real-time processing extends this advantage, enabling other applications – especially those that are voice-based applications.

- *A more personalised mobile experience – over the air (OTA)*. End users can literally customise their phone with applications and information that suits their lifestyle – and change the applications as often as they wish. One customer might decide to personalise their handset by downloading all business productivity applications (a financial calculator, a stock-tracking application or a streamlined contact database); while another might download entertainment applications and a personalised ring tone. This can all be done over the air – no cables or PC connection required. Network operators can install, recall and update applications OTA. Even in the case of telephony specific applications, there is no need to bring the handset into a store or operator location for servicing.

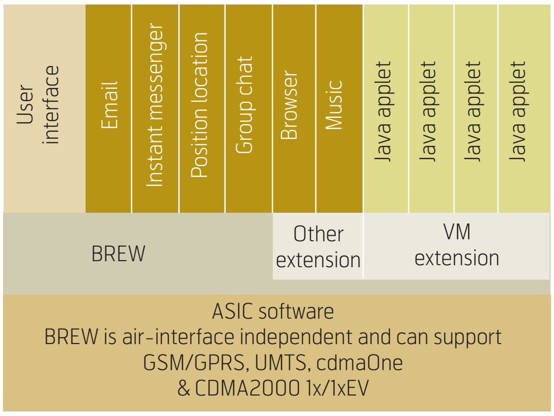


Figure 1 Location of the BREW Client API layer

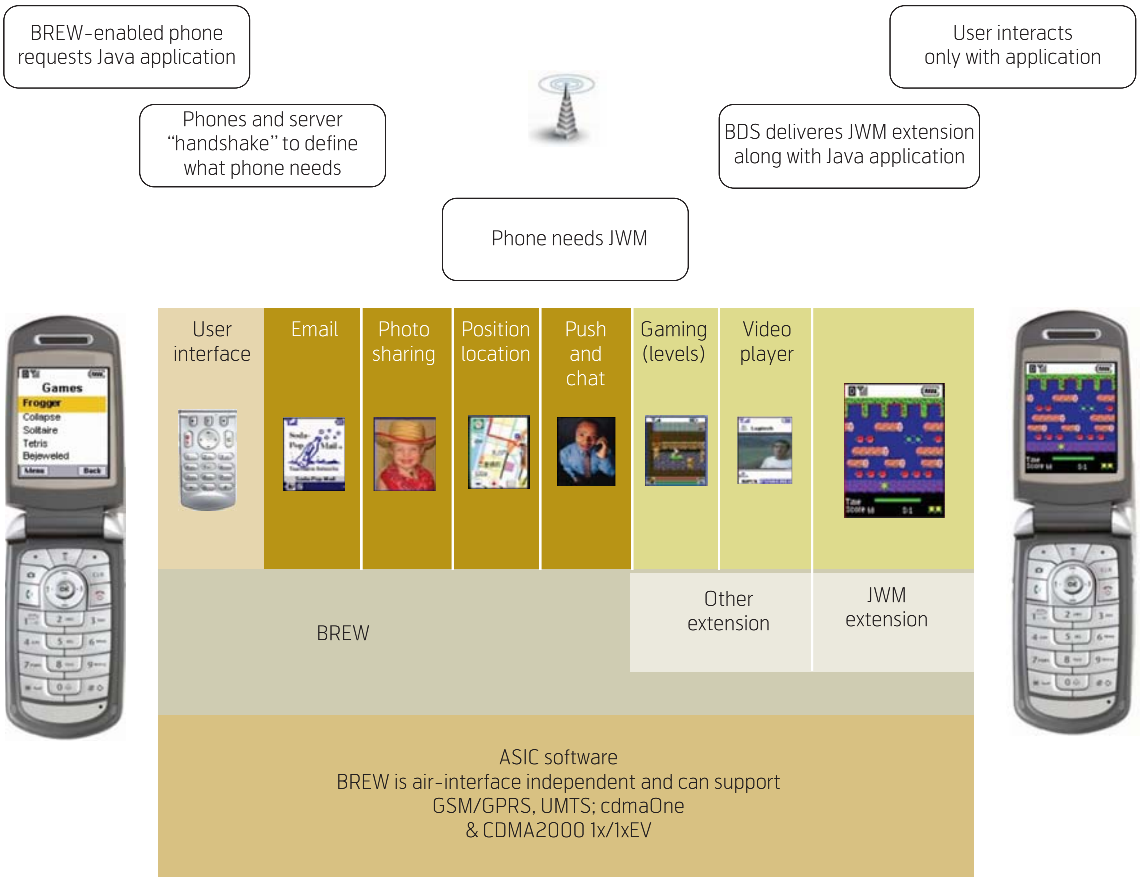


Figure 2 Extending handset capabilities by adding new functionality OTA

ISSN 0085-7130 © Telenor ASA 2005

- *Minimal hardware and software requirements*. The BREW Client is thin, efficient and powerful, and there's no need for a separate virtual machine (VM) for each model of phone. Operators that wish to have both J2ME and BREW on a handset will only need one VM. This opens the door to porting other companies' VMs as well. Client/server-based applications are supported, such as multi-player gaming, as well as client-side applications. New functionality, extensions and virtual machines, can be dynamically added OTA to the handset to extend its capability, as shown in Figure 2.

- *Fast interactivity with information*. With client-side execution, customers can download a database of travel information about a specific locale to their devices in just a few seconds, then search the database or interact with maps quickly and as often as needed. The locally stored information can also be used to enhance position information for personalised location-based services. All customers enjoy equally fast application response times regardless of network bandwidth. The client is tightly integrated with the phone's basic telephony functions and enhances client-side execution. An application can be automatically suspended and then resumed in the event of an incoming call or SMS message.

4 The BREW delivery system

It is not enough to have client-side applications running on mobile phones. It is also critical to have an open, flexible model for discovering, buying, downloading, managing and billing for applications on client devices.

Let us consider a customer with a device that offers these capabilities, and imagine that a new entertainment application has been developed for which this customer is an ideal prospect. To begin, the customer should be able to easily check for new applications from their handset, as well as from their PC, and know they are safe and trusted applications. The network operator's supporting server software should verify that the application will run on the customer's phone (to avoid disappointment) and that memory is available on the phone to download the application.

Application Discovery and Verification of Compatibility
Application Testing at Operator Request
Purchasing (M-Commerce) Process
Downloading
Billing Record Creation
Installation / Execution / Deletion
Over The Air Recall and Update

Table 1 General end-to-end solution requirements

Next, the customer must be able buy or try and then buy the application, which entails several steps for the underlying software. Firstly, they must be able to determine the price and view other information about the application or content (for example, an expiration date on the application or free trial offer). Secondly, they must be presented an advice of charge and then empowered to initiate the download. Finally, a billing record must be created that integrates with the operator's existing billing and customer care systems once the download is completed and successful. To the customer, this process should be easy, self-explanatory and quick – taking just a matter of seconds.

Once the application is downloaded, it must be installed. The installation process should be seamless and invisible to the user. This includes adding it to

Figure 3 The screen shots describe an easy "four click" shopping experience where the customer discovers an MMS application, agrees to the price, prompts the OTA download and begins to use the application

the customer's personal list of available applications and allowing the customer easy access to use the application.

Furthermore, a good customer experience requires certain management capabilities, including the ability to remove the application if needed to regain storage space without loss of related application data or usage rights. Customers must have access to information regarding that storage space which is being consumed by each application. In the event an application needs to be updated or recalled, the network operator should be able to recall and update the application OTA without user intervention, if desired. The general end-to-end solution requirements are summarized in Table 1.

The goal is a good customer experience, and a functional model that does not leave gaps. These requirements demonstrate a need for a platform that is more than just an execution environment. Testing, provisioning, purchasing and customer support all contribute to a complete solution.

5 Customize, personalize and realize

Throughout the world, mobile subscribers are spending ever-increasing amounts of time using their mobile phones. As devices, these 'phones' continue to have more functions that support the mobile lifestyles of their owners. Navigating this increasing mobile phone functionality occupies a large part of the subscriber's time operating the phone.

Figure 4 Examples of customized user interfaces

The user interface has, until recently, been something that is provided on the phone at purchase and is something that cannot be modified.

In marketing terms, the mobile screen and the user interface are intensely valuable real estate. For mobile phone operators, mobile phone manufacturers, brand owners and other third parties to make the most of this valuable user interface, they must have ready access to the people, tools and technology necessary This can be enabled by the use of XML, which has achieved widespread adoption as the standard language for marking up platform-independent content.

The uiOne Delivery System allows operators to offer UI theme updates, content channel/casting and UI content update channels to devices with uiOne capability.

Operators can have more control over the design and update of the user interface, enabling them not only to accurately represent their brand and connect directly to services, but also to keep the user experience refreshed and up-to-date in the following ways:

- *More control over the branding*. Integrating many different aspects of the personality of the brand including colour, style, graphic language, behaviour and imagery into the overall user experience, gives operators the ability to carry their brand identity fully onto the handset.

- *More control over updates*. Control of the user interface, and the ability to selectively update parts of the interface as required.

- *More control over customer segments*. uiOne keeps track of 'who has chosen what interface,' so promotions for that interface will automatically be pushed to interested users, enabling accurate targeting of offers based on clear client preferences without the need for complex data warehouses.

uiOne enables a service store to be provisioned onto a mobile phone, located in an easily accessible place in the top menu of a phone. A service store is a selection of operator service-based products grouped together in a convenient area within the phone's user interface (like stacking products on a shelf). Some services in the store may have an enduring shelf life; others may have a short sell-by date. The store provider uses uiOne to style the service store, showcase the products and refresh the store as and when necessary.

A product in the service store is a link to a mobile network service. When a user selects a product, with one click they are connected to that service. Services can be voice, data or messaging based, all made more accessible to users in the store. With a service store on the mobile handset, take-up of a mobile service is influenced by retailing expertise such as store design and layout, product rotation, promotions and ranging.

More than any other complex electronic product, the mobile phone is an integral part of personal styling. As has been seen with the astronomical growth of downloadable ring tones and wallpaper we seek to dress our phone in more ways than ever previously thought possible.

With uiOne, the user interface can be themed to adopt a particular style or brand. This themed user interface brings together visual designs and interaction models, graphics, animations and sound that replay a theme or brand in a very rich way for users. Ring tone and image downloads form a profitable business for many organisations today. With uiOne, styling and branding the user interface will be an equally profitable business as users seek to restyle the look and feel of their mobile phones.

uiOne themes are a lot more than shallow UI "skins:" they allow the end-consumer to add and expose new key functions and applications of the device by downloading additional components (applications and software libraries) that are bundled with the theme. Rich multi-component themes can be easily downloaded and installed by the end-user without having to go through the hassles of separate purchases and downloads.

The desire for personalisation means that there is a requirement to generate very large numbers of different user interfaces to appeal to different market segments, fashions and the latest developments in the entertainment industries.

6 Future directions

With smaller chipsets and more exciting user interfaces, 3G has the potential to grow into the most used wireless technology in Europe over the next couple of years.

3G terminals are now reducing in size and users will soon be able to personalize their handsets in ways unimagined today. We have yet to fully realize the full potential of 3G and what it can offer. Soon we will consider it a must have.

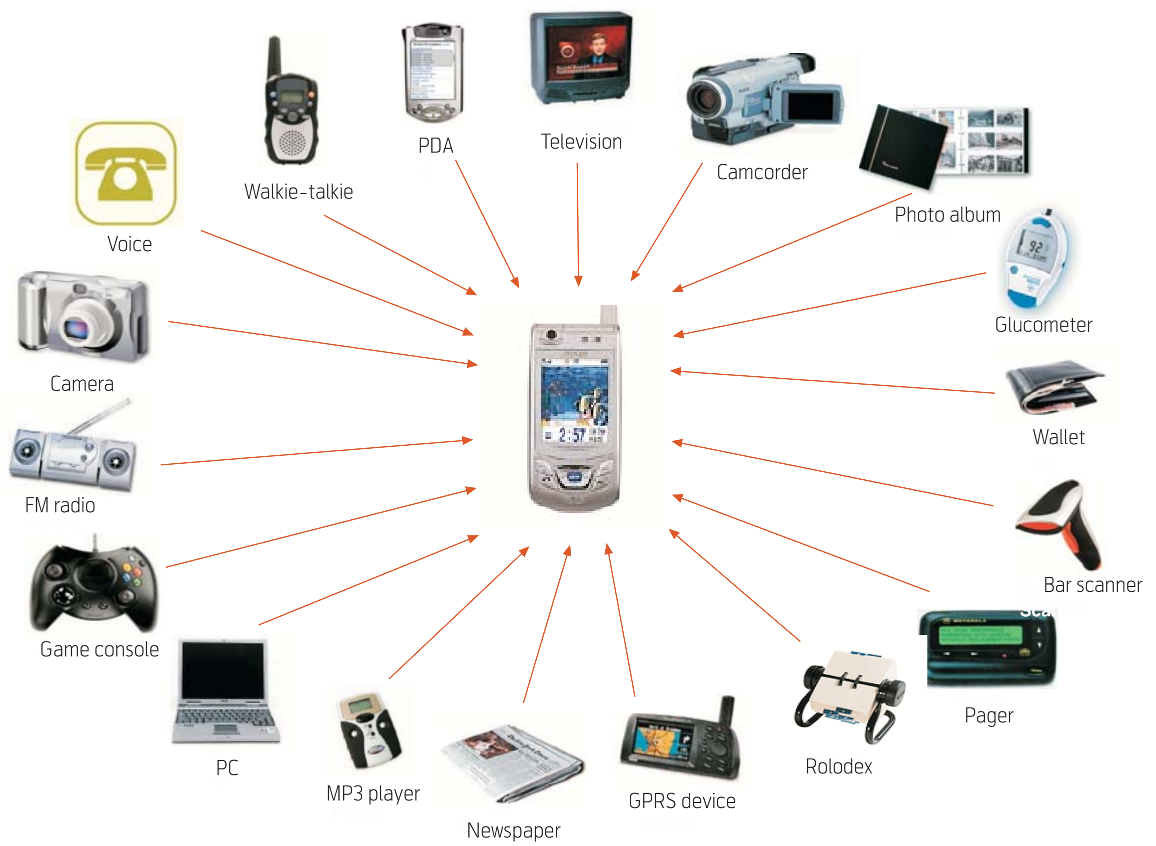


Figure 5 Different consumer electronics devices are converging more and more, in the sense that different functions migrate from one type of device to others

The new air interface extension HSDPA (High Speed Downlink Packet Access) will increase wireless broadband speeds up to a peak rate of about 10 Mb/s, in practice up to 2–3 Mb/s. Such improved data rates will allow the user to download data at an incredibly fast rate. In theory, the faster the data download rate, the cheaper in the end it should be for the operator to send the data. Such data speeds should improve the end user experience enormously.

What can such fast data rates be used for? With a personalized user experience, subscribers will continue to expect more; like data speeds similar to what they are experiencing at home on their ADSL fixed broadband line. If users of home Internet banking could do it easily from their mobile phone as well, they will begin to desire the same experience as they have at home. A technological convergence is also occurring in the terminal world. More and more services can be jammed into one small compact terminal, which gives the subscriber more and more features.

A wide array of services can be integrated into one terminal. Terminals will become smaller and cheaper – and less power hungry. Is such an ideal far into the future or just around the corner? Only time and Moore's Law will tell!

Richard Savage is Director Business Development at Qualcomm Europe, where he is responsible for Scandinavia, Baltic States and Holland. Prior to joining Qualcomm, Richard has held similar positions for the last nine years at Ericsson Telecom AB, and Appload AB. Richard has lived in Sweden since 1996, prior to moving to Sweden he lived and worked in France with several companies including Microsoft, MicroWarehouse and Azlan.

e-mail: savage@qualcomm.com

What's next in mobile telephony and will it succeed?

NIGEL WILCOX

Nigel Wilcox works in the Business Innovation team at Philips' semiconductor division

In the late nineties anyone who suggested that 3G would, in general, still be soft launching in late 2005 and that picture messaging and other new service revenues would be almost insignificant would have been laughed at. The vast profits predicted for do-everything handsets have simply not materialised. So what will the next generation of handsets contain and, more importantly, can they make the networks some money?

Introduction

Back in the late nineties the whole mobile phone industry began setting itself up for 3G, investing vast amounts into network and handset developments and initial services like video telephony. So strong was the thrust for 3G that no one noticed that some companies once generating healthy revenue from GSM had stopped making money, devoting all their resources into 3G and its new services whilst forgetting the current market.

Most manufacturers designed phones that supported every service to some extent and the handset market was segmented mainly by how well each unit performed these functions and how nice they looked. Few subscribers used these new features but the Mobile Network Operators (MNOs) subsidised them in the hope that large new service revenues would soon appear.

We're now at the end of 2005 and the new service revenues are nowhere to be seen. The huge subsidies required cannot continue indefinitely – so what can the next generation of handsets contain to address the industry's problems?

To help predict the future let's start by looking at the past.

Birth of the mobile phone

People have talked to each other since they lived in caves and always will do. Being able to talk to people wherever they are is a clear step forward. The success of the mobile phone is therefore not surprising and the original mobile phone voice functionality is still today the mobile phone's "killer" application. This is unlikely to change.

The GSM mobile phone was a communications revolution as its single common European standard meant mobile phone technology became cheap enough for everyone to own.

The SMS or text messaging revolution that followed is also interesting: At school children write notes in the classroom and pass them to their friends, hopefully without the teacher noticing. Again, SMS is simply an extension of our natural communicative behaviour.

SMS and voice call services go 100% hand in hand. Before mobile phones were commonplace, pagers provided text-messaging functionality. SMS does everything that a pager can, better than a pager can. Critically, with SMS the phone used to call back a message sender is built into the message device: the two devices merge seamlessly.

The mobile develops

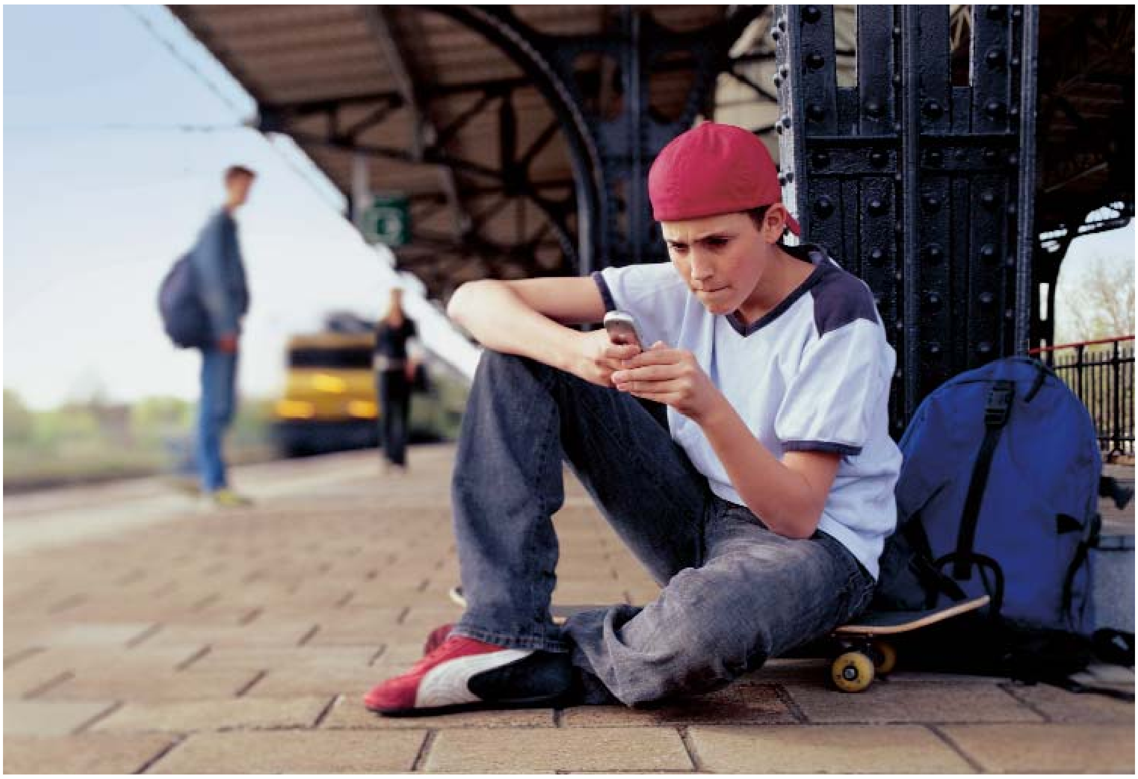
The surprise success of converging pagers with mobile phones started everyone thinking what other applications or services could be combined with a phone. There were "clearly" going to be many areas where vast sums of money could be made by this do-everything communication device. Some of these did indeed make a lot of money; others failed. For example:

WAP

Wireless Application Protocol (WAP) was the first new application to be implemented. It was a little like browsing the Internet but using more efficient transfer schemes for the relatively low bandwidth offered by mobile networks.

At that time the Internet had just become available to everyone and some mobile networks started to promote WAP as "The Mobile Internet". Whilst this was a nice idea it was far too early – the beauty of the Internet is that you can read pages and pages of on-demand information and pictures on a screen. Handsets just didn't have the display capability and showing a couple of lines of text that contained less information than old-fashioned TV teletext was not going to be a hit with consumers.

ISSN 0085-7130 © Telenor ASA 2005



WAP was not the mobile Internet but just a mobile information service. The content on WAP did not reflect this, which was a major factor in its lack of appeal. Whilst WAP has improved dramatically over the last few years, so has traditional internet browsing. Web browsing is also very cheap, so on content and price WAP simply doesn't compete.

Java

As a phone contains all the elements of a basic computer the idea of running applications or games on a phone should be appealing to a large number of PC and game console owners.

The real benefit to running software programs on a phone, however, is that it can substantially increase revenue for the MNOs – This could be through charged-for software downloads or application-initiated communication via the network.

Unfortunately different handsets have different screens, microprocessors and hardware architectures. Program code cannot therefore run directly on every phone.

The Java programming language was invented by Sun Microsystems with the intention of overcoming problems of this nature by being platform independent; i.e. Java applications written for one type of device can be run on any other Java device.

Whilst Java solved compatibility problems to an extent, the mobile phone does not come close to competing with the PC for programming or gaming.

Java-based handsets can compete with dedicated portable gaming devices like the Nintendo Game Boy, however: phone gaming has been a success with a small but dedicated segment of mobile subscribers.

As the components in handsets improve, the gaming experience on handsets will also improve allowing the experience to improve and the market to expand.

MMS

Multi-Media Messaging (MMS) allows picture, sound and video clips to be attached to traditional SMS/text messages. MMS was considered an obvious extension to an already very successful service but usage has been incredibly poor and although some MNOs are reporting a small improvement in uptake, its use is still virtually zero.

Creation of picture messages requires a camera and therefore increases Bill of Material (BOM) costs for handset manufacturers. This translates to significantly increased handset subsidies for the MNOs and so MMS is quite a heavy loss-maker.

Surprisingly MMS *service* usage contrasts sharply with the use of camera phones themselves: people do use the cameras for opportunistic photo opportunities but they simply copy the pictures to their PCs rather

ISSN 0085-7130 © Telenor ASA 2005

than send them to their friends or colleagues via the mobile networks.

Although usage is low practically all networks are still specifying cameras to be provided in all handsets on their networks. This additional cost without revenue can't go on forever but there is perhaps still hope for the success of MMS: The terror attacks on London in the summer highlighted how useful camera phones can be to TV broadcasters who published numbers and web sites for the public to send in their images. The police force did the same to collect intelligence and evidence.

Encouragement by traditional news and entertainment television may be the kick-start MMS needs to become a viable service – but so far no such TV programming has been developed.

Ring tones and wallpaper

Ring tones are very easy to produce and can easily be downloaded over the air to handsets from third party ring tone providers. Billing for this is almost always handled by the network operators who take a sizable slice of the revenue.

These services have been relatively simple to set up and spectacularly successful. Ring tones account for around 90 % of all network downloads to handsets. Simple licensing has helped in the success of these services – in the UK, for example, all fees are collected by one licensing agency making the establishment of a comprehensive ring tone service very straightforward.

The success of this service is probably due to everyone's desire to be different in some small way and project a favourable image of themselves – fashionable, funny, etc. Importantly, phones are clearly naturally good at playing ring-tones and this feature is a seamless and cheap addition to the basic phone.

Music player phones

The Music Player Phone is the latest trend, yet it has not been promoted by the industry anything like as much as camera phones or the mobile Internet were.

The stand-alone portable music player market has been consistently successful since the introduction of the Sony Walkman in the late 1970s. The recoding technology has moved on but the basic concept of a good-quality portable music player remains the same.

Although it is relatively easy for a mobile phone to include playback facilities for digitally stored music the idea of using phones instead of dedicated music players is only now taking off.

First attempts at including this functionality only enjoyed limited success due to storage capacities but this is changing: Practically all new mobiles can now play MP3 and other digital music formats and solid-state memory is becoming cheap enough to include in sufficient quantity to store a large number of tracks.

Most importantly, to make a real success of music phones from the MNOs' point of view requires mechanisms allowing the mobile networks to derive revenue from music sales. Copyright and licence issues that have prevented success in this area are now solved with Digital Rights Management (DRM), and download trigger technologies such as Near Field Communication (NFC) will dramatically help sales.

ISSN 0085-7130 © Telenor ASA 2005

With small improvements in storage and a set of stereo headphones the mobile phone has all the ingredients included in a portable music player – like all other successful mobile applications, merging this functionality into a phone is easy. The portable music player also has a track record of being a successful stand-alone application. Taking all this into account, music playing is very likely to become *the* must-have phone feature in the next couple of years.

New technologies around the corner

So where does the mobile handset go now? – The number of new functions that can be added to a mobile device is enormous. The following are some of those that are likely to be in handsets in the next few years.

NFC

Near Field Communication is a peer-to-peer wireless connectivity system allowing convenient short-range communication between two devices.

NFC moves on from traditional contactless RF ID systems as it can operate in two modes: active and passive. Active mode allows bidirectional communication between two intelligent devices. Passive mode is similar to traditional RF ID where an intelligent device controls communication with a passive NFC-readable tag.

By simply bringing two NFC-enabled devices together they can automatically establish communication between themselves without user configuration.

Any kind of data can be transferred between devices and so NFC provides a number of profit-enhancing opportunities. For example:

Automatic configuration

If a user purchases a new accessory for his device he no longer needs to manually configure it (and probably call customer services in the process) but can simply bring the two devices together. They will automatically share configuration data, pair with each other and start to work immediately.

Automatic music download

A billboard advertising a new music release can include an NFC "Download Now!" logo. Swiping the phone past the logo in the poster transfers a web address to the handset from where the music track is downloaded.

The user will simply see the message "Download [song] now?" and with one click will automatically load the track into the phone's music store for playing later. These NFC readable tags behind the logo require no power supply (deriving power from the RF of the nearby device) and are cheap to incorporate into printed media.

Media licensing

A similar scheme can be used for movie license keys. The license would be transferred to the handset and billed via the user's phone subscription. When the handset is placed on a home video centre later that night, the movie can be automatically downloaded by the video centre and play authorised using the license key from the handset.

Smart keys and ticketing

The potential for access key applications including transport or sports tickets is vast. The subscriber would be able to have these downloaded to his phone via the MNO when purchasing a ticket or hotel room on-line, saving on customer service costs and providing high-levels of security against fraud.

Transaction enabling

NFC can also be used as a payment mechanism, wirelessly and seamlessly transferring payment details rather than using traditional cards or cash.

WLAN

Wireless Local Area Network (WLAN) is not a new concept: Hot spots are now common in big cities and areas where mobile people tend to congregate, such as hotels, airports, etc.

WLAN is an excellent technology for connecting PCs on a network and ideal for allowing PC users who are static for short periods to download their e-mail or surf the Internet. However, it does not easily provide hand-over between cells or have an especially good range but this doesn't matter as laptop users will not usually move when they are using WLAN.

Whilst the charging for use of WLAN hot-spots has so far been difficult, with many networks each requiring their own log-in/credit-card charge, roaming agreements are now being established that allow simple log-in without the need for immediate payment. This makes the use of hot spots far more attractive for the average user.

WLAN connectivity could easily be added to mobile phones. However, a standard mobile phone is not a PC, and WAP may not offer enough to make most users want to surf. Mobile phone WLAN connectivity to the Internet is therefore likely to be only needed in smart phones and GPRS or 3G modems.

There is, however, a killer application for WLAN on *all* mobile phones ...: telephone calls.

UMA

Unlicensed Mobile Access (UMA) is a general term for utilisation of licence-free radio for connection of handsets to the MNO core network for voice and data traffic.

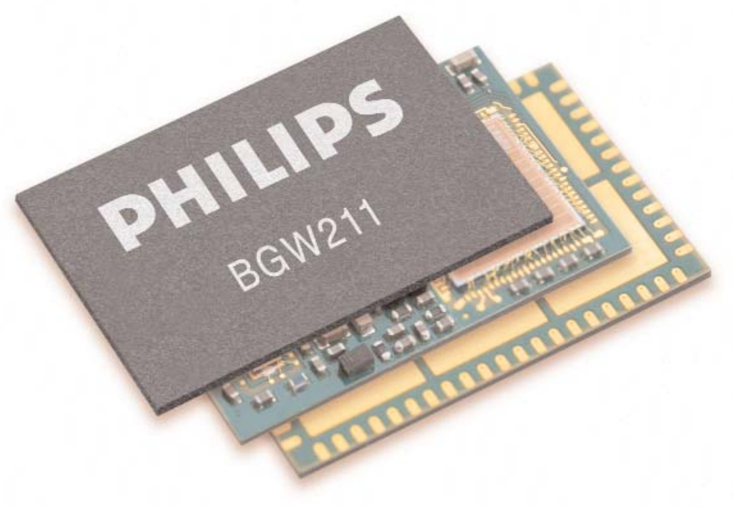
The most common form of UMA utilises WLAN for connection from the handset to a router in the home or office. This in turn is connected via an Internet connection to the mobile network's switching centre.

GSM mobility mechanisms are employed allowing seamless handover between the GSM access network and UMA network. A UMA Network Controller (UNC) is installed in the network that appears to the Mobile Switching Centre (MSC) as just another set of base stations. The handset searches for WLAN "cells" in the same way it does for neighbour cells in a GSM network. The availability of these WLAN cells is reported to the network in the usual way and the network can select this cell for communication with the handset just like any other.

UMA cells can be WLAN hot spots already provided by the network for PC access, WLAN routers in subscribers' homes, or specific UMA WLAN cells. UMA is therefore cheap: the installation of a new UMA cell requires primarily just a WLAN router and Internet connection.

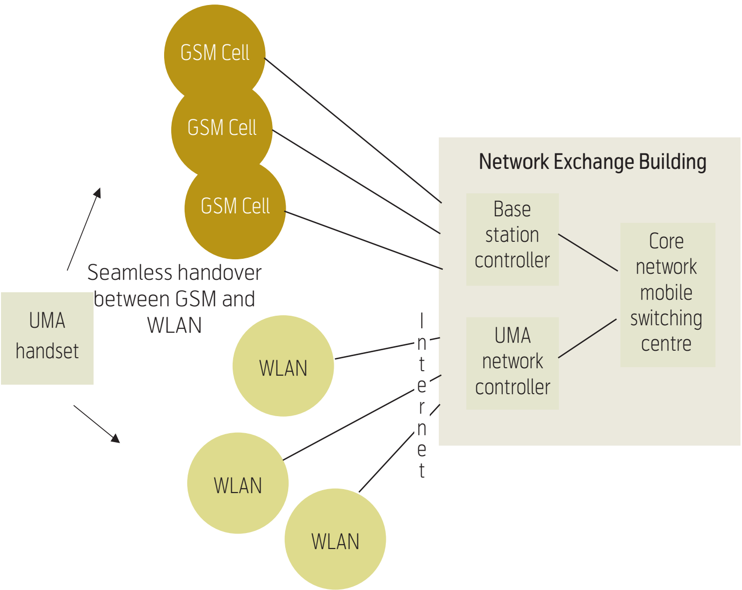
UMA has numerous advantages for the mobile networks, including:

- Very low-cost in-fill of small coverage black-spots;
- Low-cost capacity enhancement in small high-call-density areas (e.g. airport waiting lounges), freeing GSM/3G cells;



- Seamless "one phone" convergence of fixed and mobile handsets in the home/office;
- Transfer of subscribers from the GSM/3G air-interface to WLAN when at home, freeing network capacity;
- Use in aircraft via WLAN (some airlines have Internet access already – e.g. Lufthansa on some of its Atlantic services).

The key benefit of UMA over standard Voice Over IP (VoIP) systems is that it allows the MNO to maintain full control of (and therefore bill-for) all calls. Incoming and outgoing calls are still routed through the MSC. The benefit for the subscriber is that the service is seamless – he can always be contacted on his mobile number, does not have to load any special VoIP software, or need a high-level operating system. He simply initiates and receives calls as normal.



UMA-GSM network architecture

ISSN 0085-7130 © Telenor ASA 2005

There are numerous possible extensions to the UMA idea to allow, for example, zero-cost routing of roaming traffic to subscribers connected to a hotel's WLAN, etc. UMA is also a stepping-stone to IP-based core network services, not a competing technology.

WLAN components are now very small. Philips 802.11g WLAN system, for example, is available as System-in-Package (SiP) taking up only 150 mm^2. The introduction of UMA will therefore have no noticeable impact on physical size. Cost is also very low given the potential up-sides and savings that can be made.

UMA offers packet data transfer in addition to voice – so high speed download of music triggered by NFC tag is a simple additional step.

HSDPA

High-Speed Downlink Packet Access (HSDPA) is an extension to 3G that provides very high downlink data rates from the network to the handset. Headline rates are as high as 14 Mbps but in general most phones will support up to 3.6 Mbps in the 2007/8 time frame.

So far there have been few new services designed specifically for HSDPA but it will provide a mechanism to firmly move already successful *Internet* applications into the mobile domain: high-speed content download to handsets and web-access (on PCs via 3G HSDPA modem), for example.

High-Speed Uplink Packet Access (HSUPA) is an extension to HSDPA allowing fast data transfer from the handset to network. At present this service is some years away and standard GSM/3G bearers will be used in the meantime for uplink data.

TV on mobile

TV "anywhere" is already available on some networks but these services rely on the establishment of a standard video telephony call to a server. The server answers the call with a video of the latest news or weather bulletin, for example. These services are considered to be TV by the subscriber but they are actually download-on-demand video using dedicated bandwidth for that particular subscriber.

Digital Video Broadcasting for Handhelds (DVB-H) is a standard for *broadcast* TV-on-Mobile that is likely to be adopted in large parts of the world. DVB-H uses an IP-based mechanism that can be used for the broadcast of digital TV and anything else networks may want to transmit (e.g. software upgrades).

DVB-H technology is based on terrestrial digital TV standards, works now, and chip-sets will be available in volume from a number of suppliers around the end of 2005. Each country has regulatory issues to overcome but the industry and governments are addressing these.

Whilst DVB-H technology is proven its commercial success depends entirely on the content and subscription models. Putting *traditional* TV content onto mobile phones is unlikely to work (just as the Mobile Internet didn't work): Most people are not going to sit down at night with a beer and watch the football on a phone! Nor will they watch a drama programme – traditional TV is great for this and critically most people watch this kind of entertainment television in their free time.

What subscribers will do is check the news headlines whilst waiting for a train, or catch up on the spectacular crash they missed in the Formula 1 – i.e. short-sharp video content subscribers can watch in the spare moments of the day. Whilst people may also want to watch extended programmes on a train, wireless coverage will have to be ubiquitous to prevent any dropouts or the service would become unusable.

Whilst varying in the exact numbers quoted, different recent studies show that with *mobile-specific* content, a substantial proportion of mobile subscribers would be prepared to pay monthly DVB-H service charges of between three and ten euros. They will also pay more for “premium” content.

Push-to-talk

Two-way radio is very common in some parts of the world, particularly in North America. In Europe it has been less successful probably due to higher population densities and traditionally higher cost. However, the inclusion of a two-way Push-To-Talk (PTT) functionality in a phone, as it already has all the necessary components, is a relatively straightforward addition.

Push-To-Talk over Cellular (PoC) differs from traditional two-way radio in that it uses the cellular network to transfer speech to a server which then relays it to the receiving subscriber(s). This has advantages and disadvantages: It is secure; efficient in its use of bandwidth; requires network control and can therefore be charged-for. On the downside PoC may detract from use of SMS and traditional voice calls and at the same time will not take-off if pricing is too high. The key to its success therefore lies in the billing.

PoC specifications are now complete but, due to delays, competing systems have been established by some networks and different PoC clients have already been included into handsets available now. This is a problem: until all networks adopt the common OMA PoC standard and all subscribers from all networks can communicate together, PoC will not take off. This was the case for SMS – use exploded once inter-network messaging was possible.

There is an education phase to go through too – in America the walkie-talkie is a successful consumer product; in Europe it is not and many people’s first experience of two-way radio may be via PoC. PoC can also be tricky to set up – the use of NFC to pair devices for communication will help this a great deal.

A-GPS

The costs of full hardware Global Positioning Systems (GPS) have been prohibitive for a large part of the handset market. Stand-alone hardware GPS receivers need parallel hardware correlators to process the satellite signals, which are unlikely to fall dramatically in price. Software-based GPS is far cheaper as it uses simpler receiver hardware and uses software algorithms to determine a fix.

The processing power needed to determine the location in software is high. Time-to-fix is determined by the processing power of the decoder’s microprocessor – a slower CPU will require much more time to lock-on and decode the satellite signals. To shorten fix times, or allow GPS functionality to be provided in relatively low power processors, Assisted-GPS (A-GPS) can be used. This employs a fixed land-based server that provides details of satellite location, time of day, etc., on request from the mobile. Once the A-GPS system is locked onto a location, tracking is much easier.

For MNOs GPS allows a number of high value services to be added including dynamic mapping/route finding, traffic information and “find my nearest” applications. These services are already available in many networks but the lack of accuracy without GPS limits their usefulness.

The hardware and software for A-GPS is available now and should start to be included into many more handsets over the next year.

Single Antenna Interference Cancellation

The spectral efficiency of GSM networks is partly limited by frequency re-use in nearby cells and the associated co-channel interference.

Single Antenna Interference Cancellation (SAIC) is a technique for improving downlink performance for all GSM traffic channels by cancelling the largest

undesirable interference signal from local base stations, while processing the signal from the serving base station.

Users of GPRS, for example, often see their bit-rate fluctuate between low and high throughput. This is a symptom of interference from other base stations in the area that are operating on different frequencies. In order to eliminate this interference SAIC passes the received signal through a filter that recognises each signal by analysing its strength and cancelling the one causing the most interference.

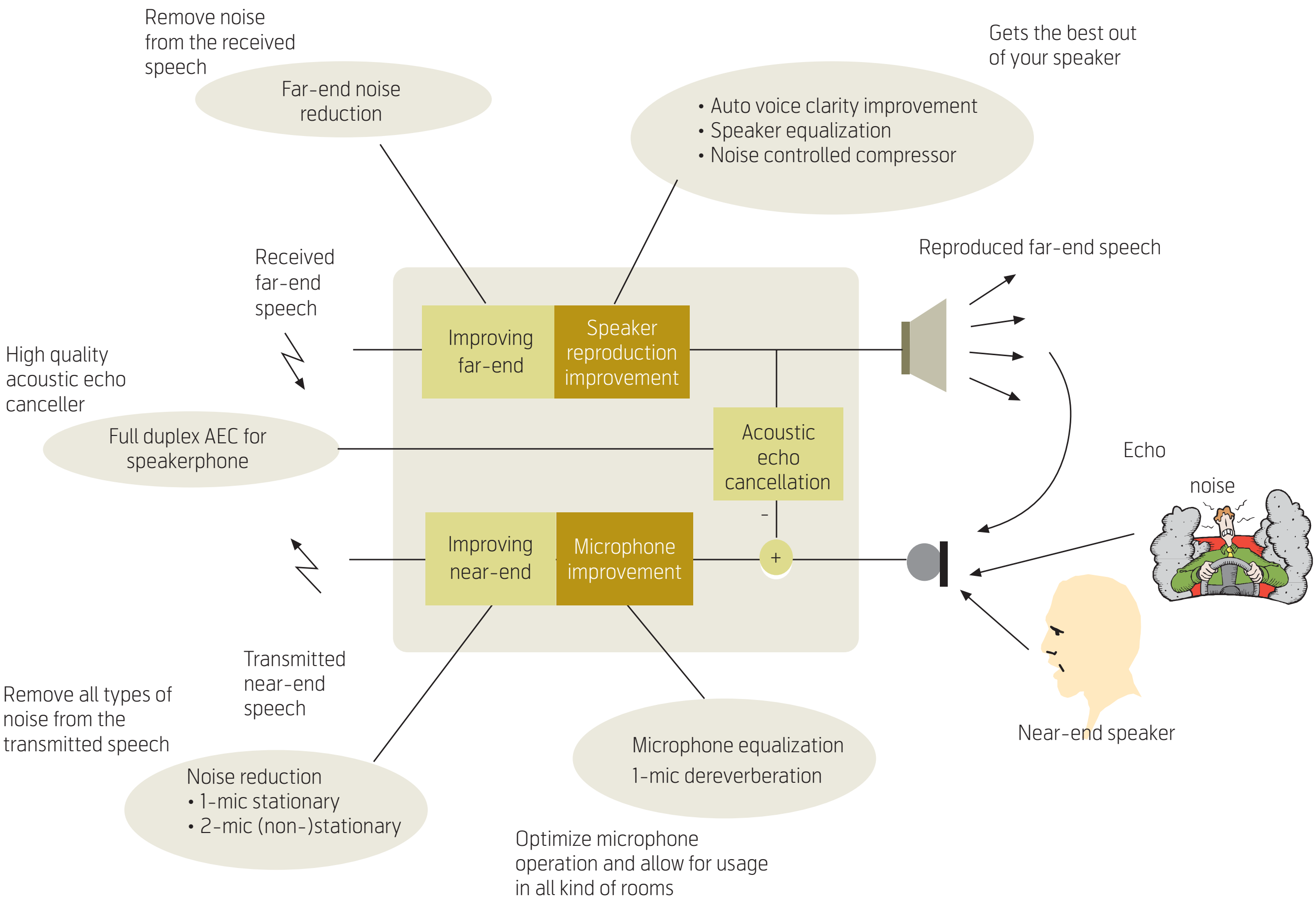
SAIC cancels co-channel interference and adjacent-channel interference caused by the spectral overlap of neighbouring frequency channels. In this way the carrier-to-interference ratio at the handset can be improved significantly, leading to:

- Better voice quality
- Higher data throughput
- Fewer call drops
- Increased network capacity

SAIC allows more aggressive frequency re-use scenarios that could provide more than 40 % capacity gain for GSM voice. Data services are also dramatically improved.

SAIC was recently standardised in GSM as Downlink Advanced Receiver Performance (DARP).

Philips' work on Mono Interference Cancellation (MIC) paved the way for SAIC and the company continues to invest in research in the area: Many advancements are in the pipeline.



Voice quality enhancement

Since SAIC is implemented as software in the handset no network infrastructure modifications are required.

Audio enhancements

Given that the killer mobile application is still voice, a number of new software mechanisms that can improve the sound quality of a call should improve network revenues significantly:

Noise Void

The Noise Void algorithm removes the background noise from the outgoing audio signal when calls are made in noisy environments such as bars or on public transport. Calls that would otherwise be unintelligible can now be easily understood as noise is reduced by more than 20 dB.

Voice Clarity

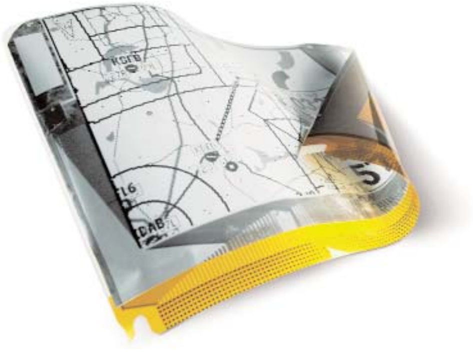
Voice Clarity algorithms work in the opposite way, improving the incoming speech audio intelligibility. The algorithm does much more than simply boosting the volume of speech, as this would just introduce distortion. Voice Clarity also removes the need to change volume levels to suit the environment.

Both Noise Void and Voice Clarity encourage more, and longer, voice calls in noisy environments and do not require any changes in the network.

Concert Sound

Mobile phone speakers and earphones are small and cannot provide concert-like reproduction on their own. However, use of audio signal processing techniques can dramatically improve the sound quality perceived by the listener, without increasing the cost of the headphone hardware.

Concert Sound systems use 3D surround sound processing algorithms that give a natural sound that feels as if it is coming from around the listener rather than between the ears. Coupled with dynamic bass boost techniques that avoid clipping allows small headphones and built-in stereo speakers to produce exceptionally full and life-like sound.



Much improved audio will help drive sales of audio and game downloads into handsets and other portable devices.

Enhancements to Hands Free

Hands Free algorithms that provide the echo cancellation needed for hands free systems have been improved to allow the use of full-duplex speech. Traditional systems use single-duplex that prevents the callers interrupting each other.

Rollable displays

One of the biggest problems in introducing multimedia services has been the limitation of the display. This is not a traditional technology problem – better displays are available – the problem is that the display has to be small for hand-held devices yet needs to be large in order to view multimedia content: A new approach is needed.

As a leader in display technologies, Philips has invested heavily in this area and one solution is the rollable display. This can be "hidden" inside a handset that is rolled-out when larger multimedia content is to be displayed.

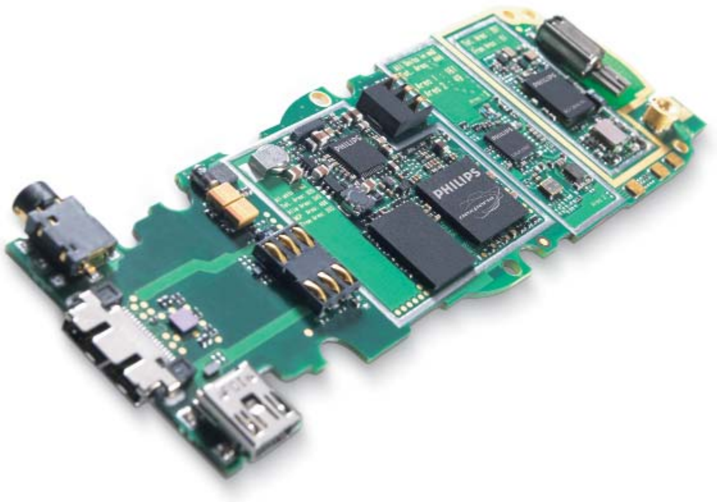
Prototypes exist now for this and allow reasonably large displays to be incorporated in mobile devices. Within 2006 four-level grey-scale 5" diagonal displays of QVGA (320 x 240) resolution will be available in production volumes. Contrast on these devices is better than most newspapers. Colour devices are also on the way but still require significant development and investment.

Aside from rollability and daylight readability, another key feature of rollable display technology is its power consumption: Power is consumed only when the display image changes – once it is set it can be completely disconnected from a power supply.

ULC

It is estimated that approaching 80 % of the world's population live within range of a mobile network yet only around a quarter of the population subscribe to a mobile phone service. There is therefore a vast number of the world's population that could become mobile phone subscribers – some 3.25 billion people. These segments of the world population have earnings that are only a small fraction of those in developed economies: for them to be able to subscribe to a mobile service requires extremely cheap handsets and call tariffs. The challenge is in making low cost handsets and avoiding subsidies that cannot be repaid through call charges.

ISSN 0085-7130 © Telenor ASA 2005



The GSM Association initiated work in this Ultra Low Cost (ULC) area last year by running a tender for low cost handsets. A second round was also completed earlier this year.

In order to get costs down ULC handsets must provide all functionality that is required and nothing that is not. In a market that is driven by voice-calls and SMS messages there is little room for cameras and other expensive additions that will be too costly to use anyway. Other critical factors in ULC markets are:

- Ruggedness – phone has to work;
- Battery life – easy access to mains power is not available everywhere;
- Easy addition of features – the "must haves" of the future, e.g. FM radio or NFC;
- Low-cost brand differentiation.

It should be noted that these simplified handsets are not only useful for ULC markets – they also appeal to the "non-digital generation" in developed economies – e.g. senior citizens and technophobes. These low cost handsets make a compelling case for networks that want to reduce subsidy.

Different countries and brands have different definitions of ULC so silicon vendors must have flexible solutions.

The long-term winners in these markets will optimize hardware partitioning for lowest cost and best performance yet retain design flexibility.

Philips' sub $5 chip set is available in sample form now and will be ready for mass production in the middle of 2006. Together with the other components required in a phone (battery, display, plastics) handset makers can expect to be able to build phones with a Bill of Materials (BoM) around $20 in this time frame. Looking forward this should reduce to $15 by 2008.

Is the subscriber tech'd-out?

The technologies listed above are the hot-topics at the time of writing and there are many more around the corner. But will subscribers use them?

At present people are not using multimedia services partly because they are just too difficult to use and also partly because consumers are overwhelmed by the technologies available on their phones.

Whilst the industry needs to do all it can to make services simple there is also an education and familiarity period which subscribers need to go through before mass adoption will occur.

In general the mobile phone industry can move far faster than consumers can keep up. Transparent features that improve the quality and use of existing services, or help reduce cost should therefore be higher on the priority list of the industry as a whole. The industry also has to be careful not to release technologies too early that are then seen as a failure.

Convergence of technology and features

Convergence is the buzzword of the moment. But what is it? Most consider it to be convergence of mobile, fixed line and Internet connectivity: Always connected, to everyone and everything, everywhere.

As well as being compelling in its ease of use for the subscriber, this "Connectivity Convergence" is also driven by competition. Many mobile operators, for example, also have fixed-line businesses and face fierce competition from cable companies for telephony and broadband revenue. These operators are therefore looking for ways to offer better packages that will prevent subscribers from switching to cable: a combined, simple and seamless service of fixed-line, mobile and broadband all through one subscription is very attractive.

There is a second level of convergence applicable to handsets: "Functional Convergence". This is the combination of the functionality of two or more traditional devices into one.

Many of the services and functionality discussed so far will converge – but getting back to the aim of this article, which will be successful?

Flexible, converged devices that make sense and are simple to use

Connectivity Convergence will happen: Apart from state monopolies (of which there are now very few), networks that don't embrace Connectivity Convergence will lose revenue to their competitors who do offer converged services.

Functional Convergence is more difficult. The mobile industry is notoriously bad at predicting which services will be a success and which will fail. What is clear is that to be successful a service or function must be 100 % suited to inclusion on a mobile device and be as good or better than the device it replaces.

In addition to helping raise Average Revenue Per User (ARPU), the handset industry can increase Average Profit Per User without the subscriber even realising it through the use of capacity and quality enhancing technologies. As many subscribers are overwhelmed by handset technology at the moment this may well be an area to concentrate on for the next couple of years.

The cost of supplying integrated handsets that do everything is prohibitive. Networks are increasingly shying away from high-tech, hard to use, include-everything devices with features that cost a lot but are rarely used.

Due to the long development times of mobile phone chip-sets, to keep handset costs down semiconductor suppliers such as Philips have to provide flexible, core platforms with high-quality additional function blocks that can be added or left-out as necessary.

The functionality included in any device *must* be of excellent quality, or overall revenues will suffer as subscribers stick to their non-converged devices. If great functionality is added well however, the future for new high-quality service profit is undoubtedly very exciting.

Nigel Wilcox is a member of the mobile communications Business Innovation team at Philips' semiconductor division. He is principally involved in handset specification work with the mobile networks, ensuring Philips chip-sets meet the network's requirements, and therefore those of the handset manufacturers Philips supplies. Prior to joining Philips he worked in engineering and project management roles for a number of companies all involved in the development of mobile communications equipment and services. Nigel is a Chartered Engineer and member of the UK's Institute of Electrical Engineers. He holds B.Eng and M.Sc degrees in Electronic Engineering from Warwick and Nottingham Universities.

email: nigel.wilcox@philips.com

ISSN 0085-7130 © Telenor ASA 2005

OMTP – Open Mobile Terminal Platform

LARS WENDT AND DO VAN THANH

Lars Wendt is Senior Advisor in Telenor International

Do van Thanh is Senior Research Scientist at Telenor R&D

This paper gives a brief presentation of the Open Mobile Terminal Platform OMTP Ltd. ("OMTP") (http://www.omtp.org) that has been established to gather the needs of the mobile operator community for mobile terminal platforms and to define and agree on those requirements necessary to deliver openly available standardised application interfaces to provide customers with an improved user experience across different devices.

Introduction

Mobile services have reached higher and higher levels of popularity, but at the same time, the complexity of these services has increased tremendosuly. Mobile operators constantly need to integrate the mobile services they offer seamlessly with the device that deliver the service. Experiences from nearly all operators show that customisation of terminals drives both usage and ARPU.

The operators have started to develop a seamless customer experience across their device portfolio via customised handsets. But the continuously increasing fragmentation of mobile devices undermines this seamless customer experience and might become a threat to the mobile services business.

And underlying all this, the security threats on mobile devices are developing rapidly.

With this setting a number of leading mobile operators came together in 2004 and formed the Open Mobile Terminal Platform group.

OMTP mission

Since the formation, the OMTP group has aimed to define platform requirements necessary for mobile devices to

- deliver openly available standardized application interfaces;
- provide customers with a more consistent and improved user experience across different devices;
- enable individual operators and manufacturers to customize and differentiate their offering.

The main rationale for OMTP is thus to reduce the number of variants, drive consistency, and take unnecessary costs out of the industry (without stifling innovation). Furthermore it is important for operators to let users know that services exist, that they find the service, that they are able to set up the service and lastly, that waits and delays are limited.

Operating structure / organisation

About the OMTP group

The OMTP group was established in June 2004 by eight international mobile operators to collaborate with industry-leading companies in all sectors of the mobile space to promote its goals through open standards and different technologies.

The founding members of OMTP are mmO2, NTT DoCoMo, Orange, SMART Communications, Telefónica Móviles, TIM (Telecom Italia Mobile), T-Mobile and Vodafone. The OMTP group now has about 40 members and participants.

The OMTP group is an open association, with membership or participation available to organizations wanting to adopt OMTP requirements, to assist the process of achieving the OMTP group's goals or simply to follow the latest efforts and results of the OMTP group. Membership is open to mobile operators, whereas participation as a sponsor, advisor or associate is open to any entity. Members, sponsors and advisors are eligible to be actively involved in projects formed by OMTP; associates receive information on ongoing activities. Mobile operators, mobile device manufacturers, software and hardware suppliers interested in participating in the OMTP group can find more information on the OMTP group's website, www.omtp.org.

Who is OMTP open to?

As full members, only mobile operators are allowed, whereas sponsors, advisors and associates can come from software developers, handset vendors and other interested parties.

Members & participants

See Table 1.

ISSN 0085-7130 © Telenor ASA 2005

Is OMTP a standardisation body?

It is important to note that OMTP is not a standardisation body *per se* (meaning that they issue specifications, and subsequently test equipment according to these standards). OMTP has rather opted for an existence where their specifications eventually can evolve into standards, but in the first phase the outcome is handled more like requirements.

Deliverables

When will the first handset be released?

It is expected that the first handsets that have been made according to the OMTP specifications will be released during the second half of 2006.

What will be different with the first release of OMTP compliant handsets?

First editions will only support basic customisation. The first requirements are termed "User Experience". In this context this means that an operator should be able to tailor the look and feel of the device menu structure to its own needs and requirements. This includes setting wallpaper, colour scheme of menus, selection of applications accessible from the top menu list, and applications accessible from the stand-by screen.

Furthermore, it also applies to device management where an operator can download, configure, and update a client-based (off-line) operator menu structure over the air to a device in a seamless and non-intrusive way for the customer (i.e. no repeated questions if the customer wants to accept the updates etc.). The menu structure can call operator defined applications that also can be managed over the air in the same way as the off-line operator menu

In addition, OMTP has defined requirements which limit variations of displays, and thus classified displays into three categories. The same goes for video codecs and memory levels. In addition to this there are specifications regarding provisioning and a logical application framework concept (which is an XML representation of a User Interface). In a follow up release from OMTP requirements regarding CDC (Connected Device Configuration) Java Phone was also issued.

The second edition of specifications from OMTP will be available in the second half of 2006.

Further on the roadmap are issues relating to trust environment, application security and performance benchmarks.

OMTP Membership overview as of 31 August 2005

Members	Sponsors	Advisors	Associates
Amena	EMP	Analog Devices	Hewlett Packard
o2	Nokia	Aplix	Mobile Innovation
Orange/FT	Philips	ARM	Pollex
Hutchison 3G	SavaJe	Axalto	ROHM Electronics Ltd.
KPN	Siemens	Danger	
SK Telekom	STM	Discretix	
Telenor	SUN	Enea	
TEM		Esmertec	
TIM		Freescale	
TMO		Gemplus	
Vodafone		Infineon	
		Microsoft Corp	
		Motorola	
		Palm Source	
		Qualcomm	
		Samsung	
		Sky	
		Skyworks	
		SmartTrust	
		Sony Ericsson	
		Spansion	
		Sprint	
		Surfkitchen	
		Symbian	
		TAO	
		Tegic	
		Texas Instruments	
		TeliaSonera	
		TTPCom	
		TurkCell	
11	7	30	4
		52	

Table 1 Current list of OMTP members and participants

Will the first OMTP phones matter?

For the end user, the differences will most probably be fairly insignificant, since the impact is very much more on the way of interaction between operator and handset provider. But indirectly there will be an impact. It can be expected that a number of operators eventually will start enforcing their procurement divisions to choose OMTP-compliant phones, and there is no doubt that this will have an impact on the market.

Long term goals

The OMTP's ultimate goal is to give operators more control over the device platform and how it can be customised. Operators want customised devices to help differentiate services and have become complaining loudly that device manufacturers do not go

ISSN 0085-7130

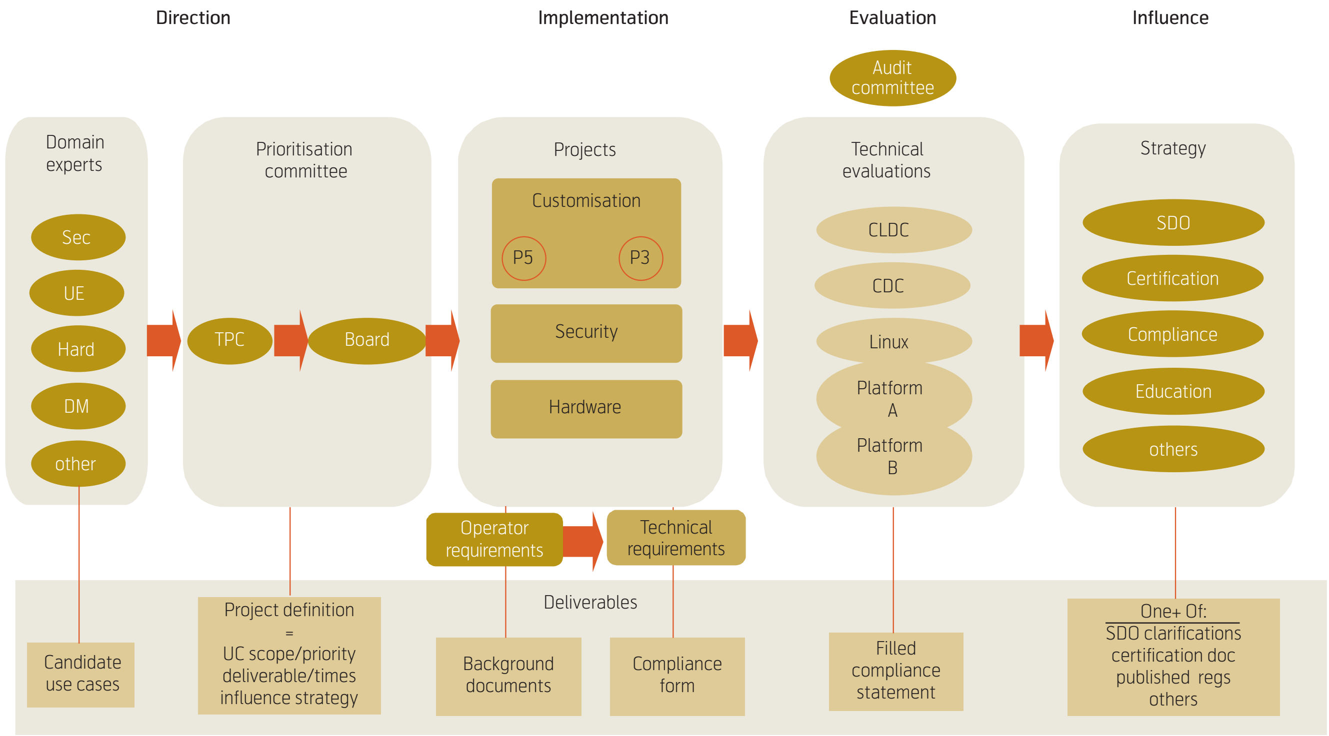


Figure 1 Overall OMTP working process

far enough to support this. The big question is, will the OMTP be able to persuade them to comply? The answer to that is most probably yes – at least to a certain degree. The mobile manufacturers are also in a position where lack of coherent, constructive and streamlined demands from the mobile operators cost them more resources than they are willing to pay for. This means that if mobile operators get more standardised in their demands, it is also a net gain for the manufacturers.

Impact for mobile industry

After one year of operation OMTP can claim to have achieved the following:

- The scope (requirements in the user to terminal dimension) and governance are well established;
- Seven deliverables published, project machinery in good shape;

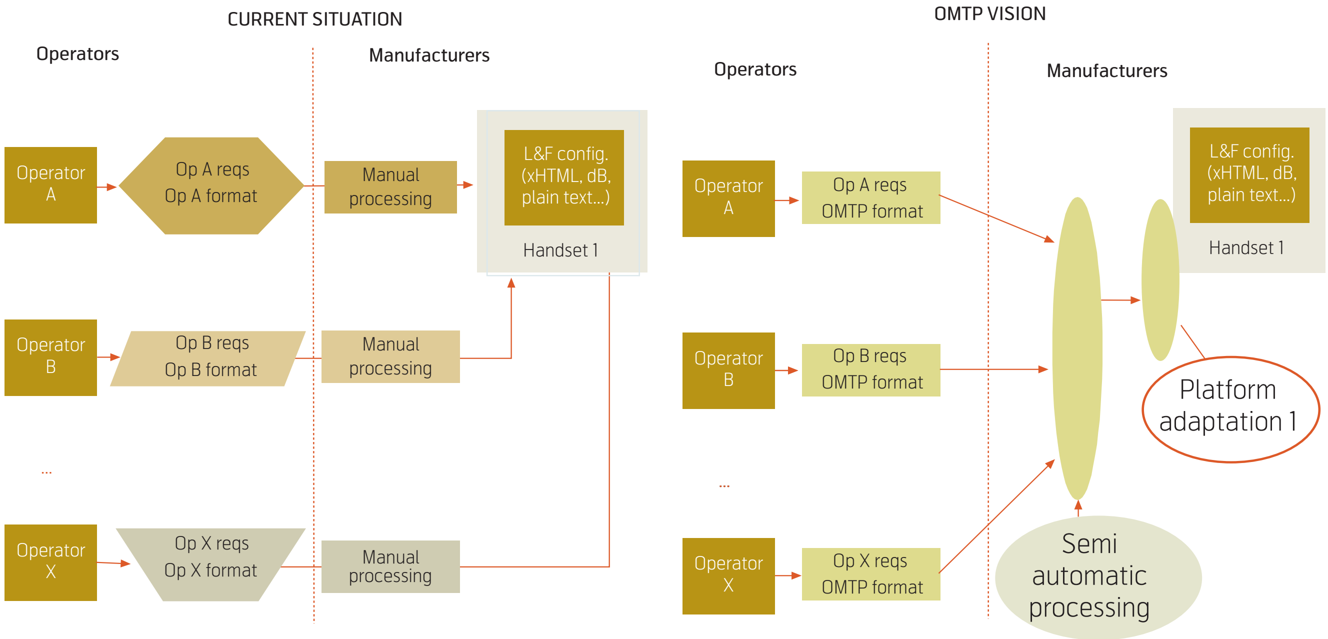


Figure 2 View of OMTP impact

ISSN 0085-7130 © Telenor ASA 2005

- Stable situation with members and participants – Full value chain represented (operators, terminal vendors, hardware vendors and software developers);
- Good support and momentum;
- Methodology from business rationale and use cases to verifiable platform requirements established.

Telenor's role

Telenor is taking a lead role in the OMTP organisation, and utilizes experiences made in its different operations. Some of the first customisation examples like Entry are brought into the set of experiences and are being used to streamline some of the processes we face when interacting with terminal manufacturers. Working together with other mobile giants, like Vodafone, Orange, T-Mobile, NTT DoCoMo and Telefonica, it is recognized that the challenges are very similar.

Summary

There is a strong belief in the mobile industry that operators as well as handset manufacturers need to take a stronger grip on some of the problems which the end users face when using mobile services, and to make the services more intuitive. OMTP is the main vehicle to streamline this, and working together brings out experiences from operators and other industry players alike, which most probably will contribute to a better user experience in the future.

Lars Wendt holds an MSc in Industrial Economics & Technology Management from the Norwegian University of Science and Technology in Trondheim. He is Senior Advisor in Telenor Mobile International in Norway, and has worked internationally for Telenor in both Germany and Hungary.

email: lars.wendt@telenor.com

For a presentation of Do van Thanh, please turn to page 2.

ISSN 0085-7130

Reading emails on mobile phones

DO VAN THANH, IVAR JØRSTAD AND TORE JØNVIK

Do van Thanh is Senior Research Scientist at Telenor R&D

Ivar Jørstad is a PhD student at Norwegian University of Science and Technology

Tore Jønvik is Associate Professor at Oslo University College

The success of the Internet is basically due to two applications, namely the world wide web and the email. Although being a newcomer compared to email, the world wide web has gained more attention due to its ubiquity and its content abundance. The telecom community has put quite a lot of effort in providing access to the world wide web from the mobile phone with the Wireless Application Protocol (WAP) and its successors. This does not necessarily mean that mobile access to email is less compelling. Indeed, the increasing popularity of the Blackberry, which allows the users to send and receive emails on their mobile phones demonstrates the urgent demand for a secure and usable mobile access to email. In this paper, a mobile email solution based on XML Web Services is presented. A brief overview of the email system, followed by a presentation of the challenges in enabling mobile access to email, are first given. The main part of the paper is dedicated to the design of the proposed mobile mail access solution. The overall architecture, the components and interfaces of Email XML Web Service are explained thoroughly. Last but not least, the Mail Web Service Client residing in the mobile phone is presented in detail.

1 Introduction

There is no doubt that mobile users, especially mobile workers, need to communicate with other people, colleagues, partners, customers, etc. The mobile phone, by meeting this real demand has become a highly appreciated device for the users. This is reflected in the increase in number of mobile phones and subscriptions. But, communication in real time or talking directly to people is not sufficient because people may not be available, i.e. busy at a meeting, sleeping, eating, etc. Communication must also include asynchronous exchange of messages. Indeed, the success of SMS (Short Message Service) has affirmed this need. In the Internet, there is another application for message exchange, called *email,* which is even more popular than SMS. Email is much more powerful than SMS since the content is not limited to 160 characters and can carry attachments of very rich format. For professional users, it is really important to be able to read mails while travelling. This is highly appreciated especially in spare time such as waiting at airports, sitting on trains, etc. since it alleviates the work pressure and contributes to the improvement of the quality of life of mobile workers.

Unfortunately, so far there is no satisfactory solution providing emails on the mobile phone because of many issues. First is the problem of firewalls. The mail servers usually reside inside the domain of an enterprise behind firewalls. Due to strict firewall configurations denying direct access from outside the corporate network, email traffic from remote networks is rejected. To overcome this problem, “webmail services” based on e.g. Web (HTTP) and WAP are in use today, as well as VPN (Virtual Private Network) solutions but they all suffer from little flexibility, unsatisfactory user interface and low degree of configurability and efficiency. However, most of the issues are the limitations of the mobile phones. As the name indicates, mobile phones are meant for voice communication and not for data applications. Small displays and limited navigation capability are not suitable for displaying large numbers of emails; emails which individually also may be large in size due to excessive presentation formats and large attachments. Mobile phones also have limited processing capability, storage space and battery life. Last but not least is the plurality and heterogeneity of mobile terminal types that make the standardisation of the user interface difficult.

In this paper we present a mobile email solution based on XML Web Services technologies, which takes into account the limitations of the mobile phones. The paper starts with a brief overview of the email system followed by a presentation of the challenges in enabling mobile access to email. It will then continue with the examination of relevant solutions for mobile access to email. The main part of the paper will be dedicated to the design of our proposed mobile mail access. The path leading to our solution will be presented. The overall architecture of the Mobile Email Access is also described. The components and interfaces of Email XML Web Service are explained thoroughly. Last but not least, the Mail Web Service Client residing in the mobile phone will be described in detail.

2 Overview of the email system

As shown in Figure 1, the current email communication systems consist of two main components, email

ISSN 0085-7130 © Telenor ASA 2005

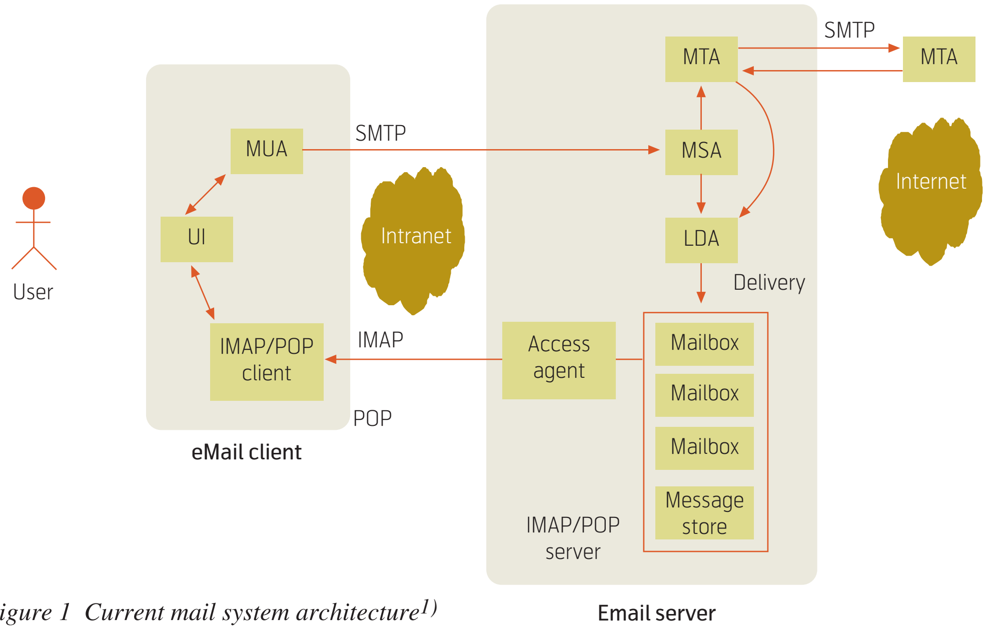


Figure 1 Current mail system architecture[1)]

Server and email Client. They are both compositions of several elements that are making use of several communication protocols, as well as internal service elements. Examples of protocols are SMTP (Simple Mail Transfer Protocol) [1], IMAP (Internet Message Access Protocol) [2] and POP (Post Office Protocol) [3].

To send a mail, the user interacts with the UI (User Interface), which allows her to compose an email. When the user chooses to send the email, the UI will hand over the message to the MUA (Mail User Agent) that will establish an SMTP session with the remote MSA (Mail Submission Agent) to expedite the mail. The MSA can do pre-specified adaptations to the message to comply with the SMTP-IMF standards. Next it delivers the mail either to a local user through an LDA (Local Delivery Agent), or passes it on to an MTA (Message Transfer Agent) which relays the mail to its final recipient(s).

An SMTP mail object contains an envelope and content. The SMTP envelope is sent as a series of SMTP protocol units. It consists of an originator address (to which error reports should be directed); one or more recipient addresses and optional protocol extension material. Historically, variations on the recipient address specification command (RCPT TO) could be used to specify alternate delivery modes, such as immediate display; those variations have now been deprecated.

The SMTP content is sent in the SMTP DATA protocol unit and has two parts: the headers and the body. If the content conforms to other contemporary standards, the headers form a collection of field/value pairs structured according to the Internet Message Format; the body, if structured, is defined according to MIME (Multipurpose Internet Mail Extensions).

The mail will then be sent from MTA to MTA and will finally arrive at the final delivery MTA that deposits the mail in the Message Store through an LDA. This concludes the SMTP transfer of the message.

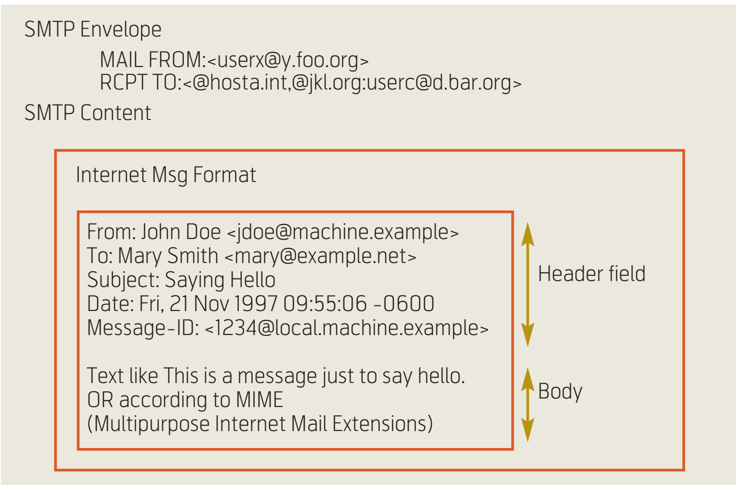


Figure 2 Relationship between SMTP mail object and an Internet Message Format message

[1)] *The current mail server architecture is somewhat simplified to make the reading easier.*

Box 1

```
220 dus12.nta.no ESMTP Exim 3.35 #1 Fri, 09 Apr 2004
15:57:01 +0200
EHLO dus12.nta.no
250-dus12.nta.no Hello jonfinng at localhost
[127.0.0.1]
250-SIZE
250-PIPELINING
250 HELP
MAIL FROM:<jonfinng@stud.ntnu.no>
250 <jonfinng@stud.ntnu.no> is syntactically correct
RCPT TO:<luser@dus12.nta.no>
250 <luser@dus12.nta.no> verified
RCPT TO:<jonfinng@stud.ntnu.no>
250 <jonfinng@stud.ntnu.no> verified
RCPT TO:<jonfinng@dus12.nta.no>
250 <jonfinng@dus12.nta.no> verified
DATA
354 Enter message, ending with "." on a line by itself
X-header: Testmessage
Headers (header part of the message) goes here…
content (body part of the message) goes here…
attachments goes here…
.
250 OK id=1BBwZ4-0000nZ-00
QUIT
221 dus12.nta.no closing connection
```

In order to send and/or relay mail, we must follow the protocol described in RFC2821 (i.e. the SMTP specification). This sequence of commands is often referenced to as the “SMTP Envelope”.

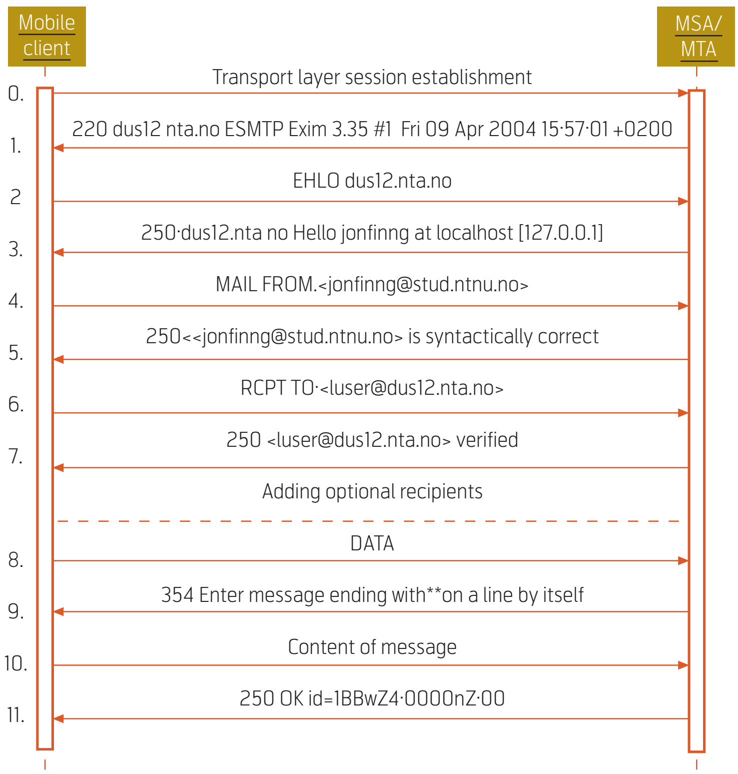


Figure 3 SMTP envelope

SMTP Envelope – excessive message transfers

This is an example of a typical “SMTP Envelope” for sending a normal email to two recipients. It is worth noting the excessive message transfers, see Box 1.

As seen in Figure 3 a minimum of 11 message transfers are needed between the client and server in order to send a single mail. Two extra transfers are introduced for every additional recipient. When sending multiple mails using the same connection to the mail server, a minimum of nine message transfers are needed. This is undesirable; it should be a goal to keep the number of message transfers to an absolute minimum. Since message exchanges introduce extra overhead and delay, it is especially important to keep this number at a minimum when using a wireless link with low bandwidth.

IMF and MIME – Unnecessary headers and complex body structures

IMF (Internet Message Format) and MIME (Multipurpose Internet Mail Extensions) [4] are standards used for representing the actual content of the email. IMF defines which headers are mandatory, and how a standard message should be organized. MIME is an extension to IMF, which defines how complex emails should be represented. This is typically emails which have file-attachments, multiple alternative representation formats or even enclosed email-messages within the message itself.

The IMF representation of an email is quite efficient, and does not introduce any overhead of significance. This makes IMF a good starting point when trying to represent emails adapted for mobile terminals. The problem with IMF is not the format itself, but how it is used. An email usually contains a lot of information enclosed in headers. Most of this information is not relevant to the end user. This information may include a list of mail servers the message has visited on its way to its destination, and several so called “X-headers”[2]. This information is usually not presented in the user-agents and introduces a significant amount of overhead when sending email headers to mobile agents. For an example of the headers in a typical email message see Box 2.

The headers marked in bold are the headers usually presented in a user-agent because they contain the information most relevant to the end user. When stripping out all other headers, the size of the header-element of this email shrinks from 2351 to 323 bytes; significantly reducing this email’s total size. Hence, for mobile terminals with limited bandwidth and

2) *“X-headers” are custom headers for providing extra information about the email. Typically added by user agents, virus and spam-scanners.*

Box 2

```
From thanh-van.do@telenor.com Sat Apr 3 15:37:46 2004
Return-Path: <thanh-van.do@telenor.com>
X-Original-To: jonfinng@homeo.stud.ntnu.no
Delivered-To: jonfinng@homeo.stud.ntnu.no
Received: from merke.itea.ntnu.no (merke.itea.ntnu.no [129.241.7.61])
        by bison.stud.ntnu.no (Postfix) with ESMTP id 214FE321
        for <jonfinng@homeo.stud.ntnu.no>; Sat, 3 Apr 2004 15:37:46 +0200 (MEST)
Received: by merke.itea.ntnu.no (Postfix)
        id 1151913DA83; Sat, 3 Apr 2004 15:37:46 +0200 (CEST)
Received: from localhost (localhost [127.0.0.1])
        by merke.itea.ntnu.no (Postfix) with ESMTP id DABD013DCAB
        for <jonfinng@stud.ntnu.no>; Sat, 3 Apr 2004 15:37:45 +0200 (CEST)
Received: from virus-out-st.online.no (virus-out.ttyl.com [193.212.240.200])
        by merke.itea.ntnu.no (Postfix) with ESMTP id 5B43C13DB64
        for <jonfinng@stud.ntnu.no>; Sat, 3 Apr 2004 15:37:43 +0200 (CEST)
Received: from tns-fbu-22-209.corp.telenor.no ([134.47.162.190] [134.47.162.190]) by
scan.telenor.net with ESMTP for
jonfinng@stud.ntnu.no; Sat, 3 Apr 2004 15:36:55 +0200
Received: from tns-fbu-22-212.corp.telenor.no ([134.47.162.91]) by tns-fbu-22-209.corp.telenor.no
with Microsoft
SMTPSVC(5.0.2195.6713);
         Sat, 3 Apr 2004 15:36:54 +0200
Received: from tns-fbu-2e-004.corp.telenor.no ([134.47.163.148]) by tns-fbu-22-212.corp.telenor.no
with Microsoft
SMTPSVC(5.0.2195.6713);
         Sat, 3 Apr 2004 15:36:54 +0200
Content-class: urn:content-classes:message
X-MimeOLE: Produced By Microsoft Exchange V6.0.6487.1
Subject: RE: Paper MWCN 2004 final version
Date: Sat, 3 Apr 2004 15:36:53 +0200
Message-Id: <375BFFDA619DBC4AA93996BF1F1D23969CD6BB@TNS-FBU-2E-004.corp.telenor.no>
X-MS-Has-Attach: yes
X-MS-TNEF-Correlator:
Thread-Topic: Paper MWCN 2004 final version
thread-index: AcQXAvMvvLDODbn9SeSq5/2eImqLIQCfUkcw
From: <thanh-van.do@telenor.com>
To: <jonfinng@stud.ntnu.no>
X-OriginalArrivalTime: 03 Apr 2004 13:36:54.0486 (UTC) FILETIME=[BA45C760:01C41980]
MIME-Version: 1.0
Content-Type: multipart/mixed;
  boundary=”——_=_NextPart_001_01C41980.B9E57E06”
X-Content-Scanned: with sophos and spamassassin at mailgw.ntnu.no.
X-Spam-Level:
X-Spam-Status: No, hits=1.0 required=3.0 tests=__ORIG_MESSAGE_LINE version=2.20
X-Spam-Level: *
Content-Length: 140839
Lines: 1859
Status: RO
X-Status: A
X-Keywords:
X-UID: 5466
```

client-functionality, only a minimal set of headers should be transferred.

A MIME email’s content is organized in two dispositions: INLINE and ATTACHMENT. The parts marked as INLINE will usually be shown in the mail client when opening the email for reading. The INLINE content parts are encoded as text. The parts marked as ATTACHMENT consist of binary data, and must be opened in an external application or through a plug-in in the mail reader able to interpret that specific attachment. The different parts of the message are separated by custom defined “boundaries”. When a boundary appears in the message, it marks the beginning of a new part. Every part must be identified by a “Content-type” parameter, telling what MIME-type the current part is (e.g application/msword to designate that the part is a Microsoft Word document). Other parameters to a body part may be the encoding and file name of the attachment.

A cut from an email containing a normal text part as INLINE and an attached picture as ATTACHMENT is shown in Box 3.

Box 3

```
******* stripped headers **************
Mime-Version: 1.0
Content-Type: multipart/mixed;
boundary="——=_NextPart_000_2a6d_50c1_182a"
Message-ID: <BAY14-F29YIgoFT3kMU000118b4@hotmail.com>
X-OriginalArrivalTime: 02 Mar 2004 11:54:46.0152 (UTC)
FILETIME=[28483880:01C4004D]
This is a multi-part message in MIME format.
———=_NextPart_000_2a6d_50c1_182a
Content-Type: text/plain; charset=iso-8859-1; format=flowed
____________________________________
MSN Hotmail http://www.hotmail.com
———=_NextPart_000_2a6d_50c1_182a
Content-Type: image/pjpeg; name="teamaa.jpg"
Content-Transfer-Encoding: base64
Content-Disposition: attachment; filename="teamaa.jpg"
/9j/4AAQSkZJRgABAgAAZABkAAD/7AARRHVja3kAAQAEAAAAUAAA/+IMWElD
Q19QUk9GSUxFAAEBAAAMSExpbm8CEAAAbW50clJHQiBYWVogB84AAgAJAAYA
********* stripped binary data ***********
```

The first header element included in this cut tells us that this is a MIME message, and that the mail reader must support MIME to understand this message. When the message contains parts with different content-types the base content type is marked as "`multipart/mixed`" as in this message. One should as well pay attention to the boundaries in this message (`——=_NextPart_000_2a6d_50c1_182a`), separating the two parts of this email.

The first bodypart is of type "`text/plain`" and represents the text in this message. The parameters tell us that the bodypart has characterset iso-8859. The second part of the message is an attached picture with content type "`image/pjpeg`". In this part, the filename and encoding are supplied as parameters.

This example of how to organize the contents of an email is just one of numerous possible ways a mes-

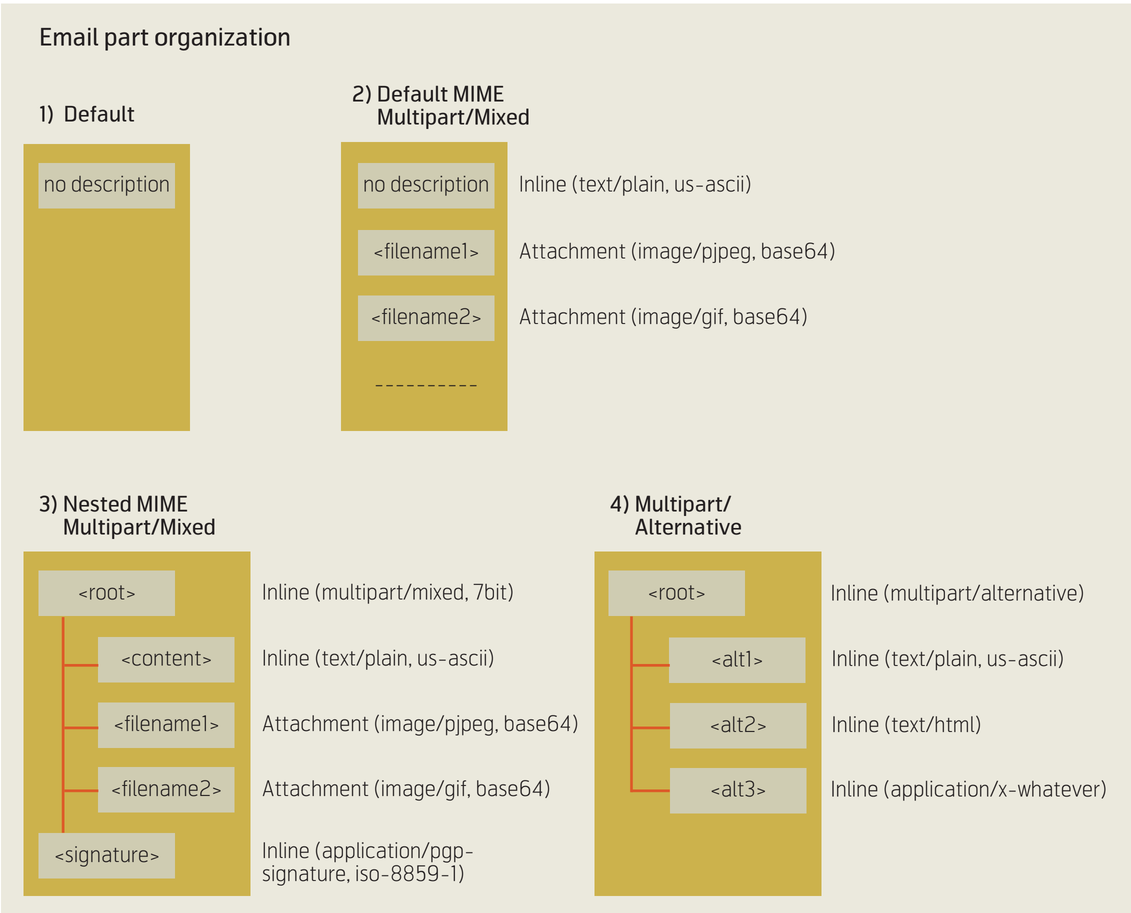


Figure 4 Examples of email structures

sage may be structured. MIME messages also support nested content, and so-called "`multipart/alternative`" body parts (among many other content types). A "`multipart/alternative`" message holds different presentation of the same data in different formats, allowing the client or user to choose which format that is best suited for presentation. Figure 4 shows some examples of how email may be organized in different bodyparts.

MIME is a good and much needed extension to ordinary mail. But the emails' representations may become very complex, and may take time to process on terminals with limited processing capabilities. MIME requires that the end clients have support for showing or opening the attachments sent. In addition, there is no point sending multiple representations of the message itself (multipart/alternative) if the client cannot present it to the user. In such cases, MIME body parts will end up as unnecessary overhead.

3 The challenges of mobile mail access

The access and manipulation of emails stored in the Message Store are enabled by an Access Agent, which implements protocols such as IMAP (Internet Message Access Protocol) or POP (Post Office Protocol). The client application will contain either an IMAP client or a POP client (or both) for the retrieval of emails and mailbox operations. In addition, the email client needs also to be equipped with the functionality required for sending out mail using SMTP. In most cases the clients and server are on an Intranet protected by a corporate firewall, which prevents access to the email server from outside this network. This is done to protect the server from outside attacks and hacker attempts, as well as for stopping unwanted relaying activity, which may include spamming. On the other hand, users are getting more and more mobile and want to access their email accounts also when not being on the corporate intranet. This introduces a conflict of interest between security and availability.

The challenges with today's mail systems when accessed by mobile terminals can be summarized as follows:

- Excessive message exchange;
- Unnecessary overhead in different parts of the email;
- Too complex representations of the emails;
- Unsupported presentation formats;
- Complex implementation of email clients. Different protocol implementations are needed for sending and retrieving mail;
- Problems related to mobile access, due to strict firewall configurations.

4 Current solutions and their limitations

Much effort has been put into solving the problem related to firewall configurations, often denying access to mail operations outside the corporate intranet. Several more or less successful attempts have emerged. The solutions might work well with laptops and stationary computers, but so far, the solutions have severe limitations when being used for providing email to mobile phones.

3.1 Virtual Private Network

In this solution, shown in Figure 5, a VPN client establishes a secured channel from the remote computer through the Internet and across the firewall to the corporate intranet. The client is hence logically connected to the intranet and the user can use a normal email client to access his email using standard protocols.

This solution is very suitable for a PC with sufficient processing and storage capabilities and a fairly stable connection with considerable transfer rate. This solution does not fit mobile phones because of the following issues:

- Most mobile phones are not equipped with a VPN client.

- Mobile phones do not have the sufficient processing capability to equalize the overhead introduced by the encryption and decryption.

- The unstable wireless link may introduce problems to the VPN session.

- Sporadic mail access by the mobile user is not suitable because of long setup time and possible VPN connection timeout.

- Mobile phones may not have sufficient storage for all the mails.

- Mobile phones may not have processing and storage capabilities to host both an SMTP client and an IMAP/POP client.

- Mobile phones have much more limited User Interface capabilities, i.e. small display, limited navigation facility, and small and limited keypad.

- Mobile phones come in a variety of types making the design of user interface difficult.

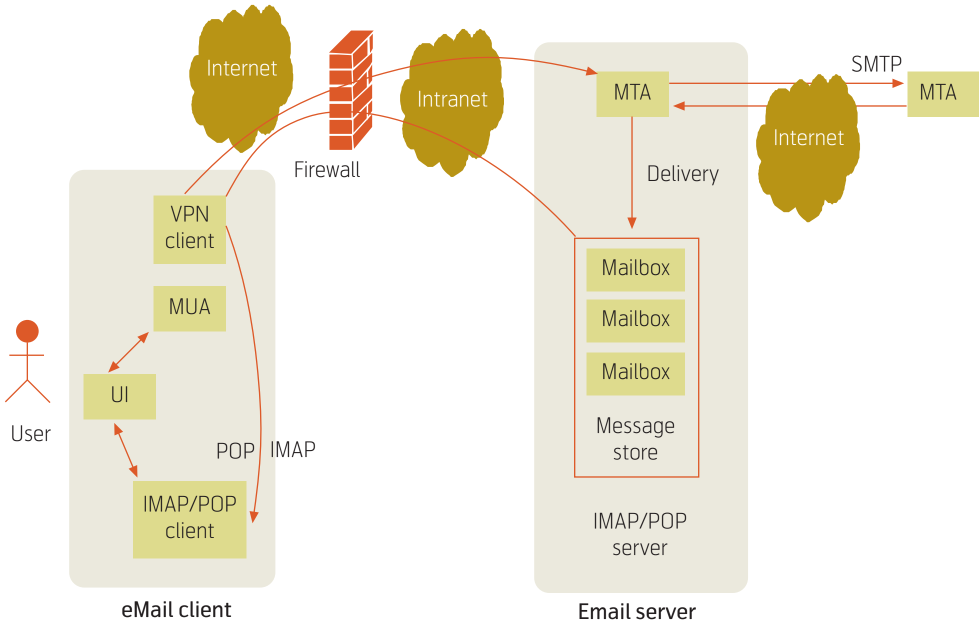


Figure 5 VPN solution

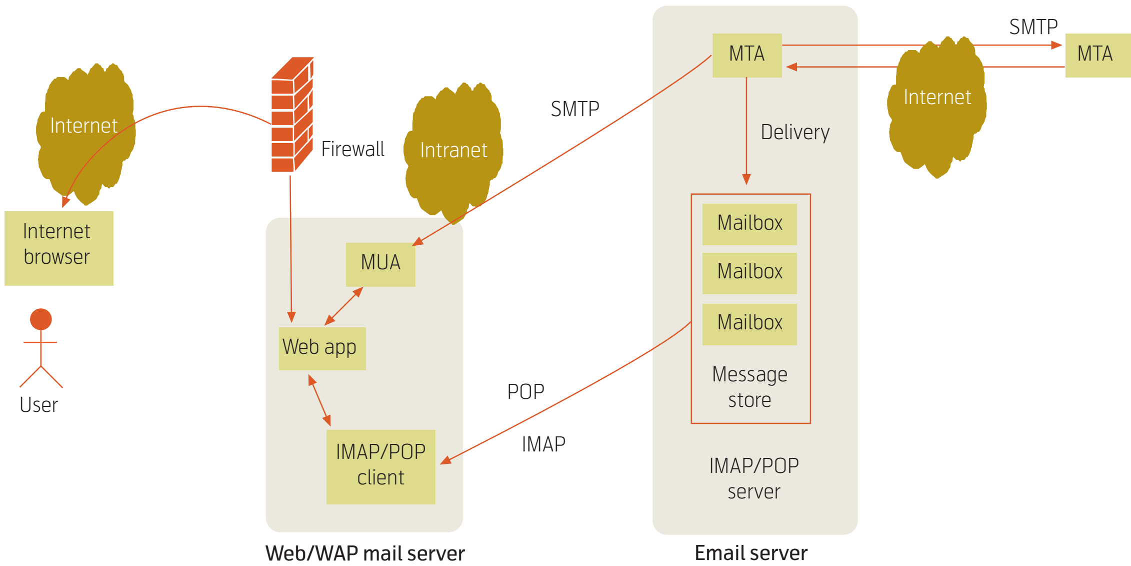


Figure 6 Web/WAP mail overall architecture

3.2 Web/WAP mail

As shown in Figure 6, a Web/WAP mail Server is now introduced. It usually resides in the DMZ (De-Militarized Zone) between the Internet and the corporate intranet. The Web-application server synchronizes with the corporate mail server behind the firewall for the retrieval and sending of mail. Actually, this server contains a Web application with the same core elements as an email client. Both implementations use SMTP and IMAP/POP for communication with the email server. The user can access her mails using a WWW-browser.

The main advantage of this solution is that a mobile user needs nothing more than a WWW-browser to access her email. No additional functionality is required on a mobile phone. The mobile phone can access the user's email account on the webmail server either directly using an HTML/xHTML browser or via WAP using a WML browser. It is worth noting that this alternative applies also for pure web mail services like hotmail, yahoo, online, etc.

The disadvantages are:

- The mails are not stored on the mobile phones but on the Web/WAP Mail server. To read the same mail twice, the user has to access the Web/WAP server again. This is due to the relaxed HTML/WML-standard for which it is difficult to parse out

ISSN 0085-7130 © Telenor ASA 2005

the actual information contents from presentation wrapping.

- The reading of mail may be time consuming and less flexible since the web/WAP pages are generated dynamically on the fly for each access. Caching on the mobile device is difficult and requires much storage capacity, again due to the excessive amount of presentation data supplied together with the information content.

- The mails are not adapted for being rendered in small displays.

3.3 Problems with known solutions

As discussed earlier, the existing solutions for providing mobile access to emails suffer from the fact that they do not take into account the limitations of the mobile phones, i.e. limited processing and storage capabilities, small display, limited navigation capability, small keypads. Neither the limitations of the wireless networks, i.e. unstable connections and high latency, nor unpredictable QoS are properly dealt with.

4 Designing a mobile mail access using XML web services

4.1 Mobile mail access with SMTP and IMAP/POP Web Service

To enable the mail access from mobile phones, the XML Web Service concept is found most suitable due to the ubiquity of the World Wide Web and the ability to traverse firewalls using SOAP (Simple Object Access Protocol) [5]. The most straightforward solution is to expose the whole SMTP and IMAP/POP as Web Services as shown in Figure 7. Each SMTP and IMAP/POP command is mapped to a Web Service method. In fact, each SMTP or IMAP/POP command is encapsulated in a SOAP message and transported to the WS client.

An advantage in this solution compared to Web/WAP solutions is that there are no overhead data specifying the appearance of the content. Compared to VPN, this solution is more robust due to the use of SOAP. SOAP has a more relaxed way of handling sessions. This involves that sessions may survive even if the link goes down for a period of time.

The disadvantage is the numerous functional requirements that are put on the mobile phone. To access and retrieve mails, the WS client must be capable of understanding the SMTP/IMAP/POP commands and communicating properly with the email server. It must therefore be equipped with a MUA (Mail User Agent) with full SMTP support as well as an IMAP/POP client. This functionality is put on top of the SOAP engine.

Other disadvantages are the high number of interactions and also the high amount of downloaded data that are an implication when implementing this solution. This is definitely not suitable in wireless networks with limited bandwidth.

IMAP and SMTP are not pure request-response protocols. They also include the possibility of server-originated messages. If full compliance with SMTP and POP/IMAP should be achieved, the clients must be able to listen for method invocations from the

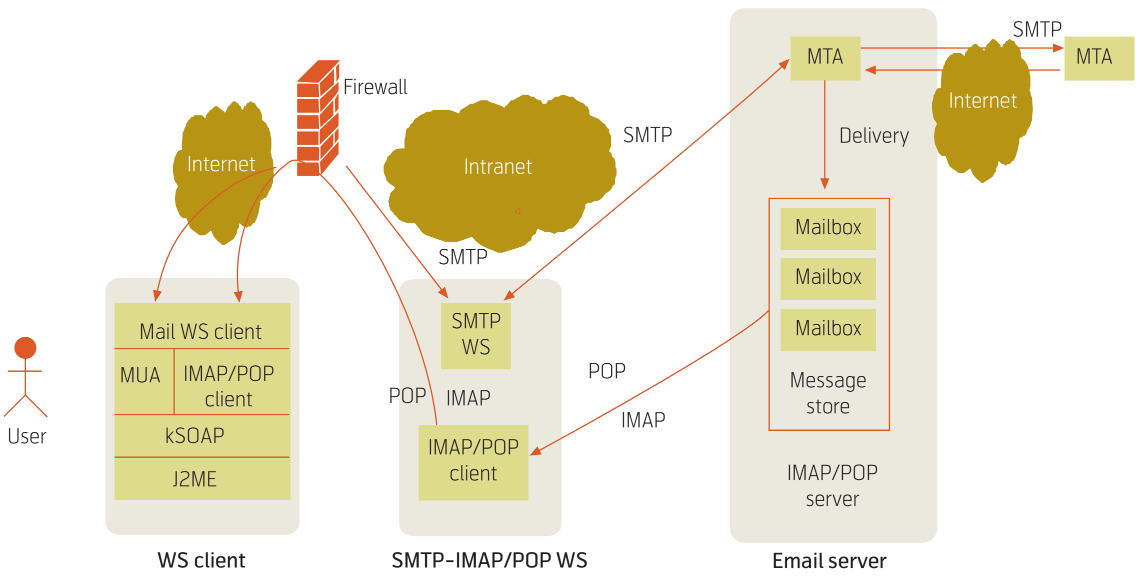


Figure 7 A straightforward SMTP, POP, IMAP mapping to XML Web Service

ISSN 0085-7130 © Telenor ASA 2005

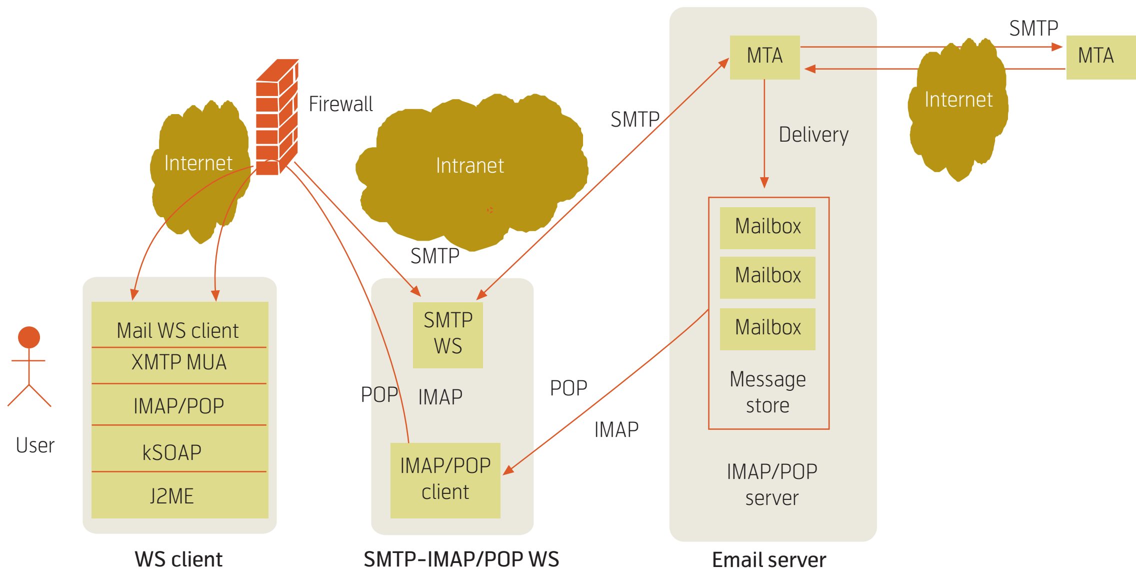


Figure 8 XML Web Service making use of XMTP

server. In this case server originated method invocation implies a full IMAP/POP Web Service implementation on the client. This is infeasible due to limited processing and storage capacity on mobile terminals. They do not have a static internet address, and the connection is frequently shut down. In other words, mobile phones were not designed to work as servers.

4.2 Mobile Mail Access with XMTP

The use of XML Web Services is taken a bit further by the introduction of XMTP (XML MIME Transformation Protocol)[6]. XMTP is a protocol for mapping IMF or MIME messages to an XML representation.

When using XMTP, we are able to transport the entire message in XML (see Figure 8). This makes it easier for the client to parse the message. An XSD (XML Schema Definition) can be used to interpret the different fields of the message, making presentation of the mail in the client potentially faster.

XMTP contains no functionality beyond the IMF message mapping. This means that when using XMTP, all message exchanges and overhead must remain as in the previous approach. XMTP may in itself introduce some amount of overhead due to the verbose XML notation.

There is very little gain using XMTP instead of direct mapping, it even adds more complexity on the server side by introducing a conversion routine between SMTP and XMTP. An approach using XMTP is therefore turned down as suitable for usage with mobile terminals.

4.3 Mobile Mail Access with XMMAP

In this section, a protocol called XMMAP (XML Mobile Email Access Protocol) is introduced. The goal of this protocol is to solve some of the major problems related to email access from mobile terminals, and to overcome the limitations and insufficiencies of the previously proposed solutions. A Telenor patent application[3)] has been filed to the Norwegian Patent Authority in 2004.

As we could see from the message sequence chart in Figure 3, a minimum of 11 messages is required between the client and server just in order to send an email. This is highly undesirable when using a mobile terminal. If the link goes down in the middle of a message sequence, the entire procedure has to be repeated. Additionally, every message transfer introduces unnecessary overhead from underlying protocols. On top of this we have the considerations regarding overhead in headers and representation as discussed in the section on IMF and MIME.

By introducing XMMAP we reduce the number of message transfers to two for most IMAP/POP/SMTP operations. As shown in Figure 9, we are down to one message from the mobile client to the web service, and one message in return.

[3)] *Telenor patent: A system, method and protocol format for mobile email communication – Inventor: Do van Thanh, Jon-Finngard Moe & Eivind Sivertsen – Norway application no: NO 2004 2233, filed 28 May 2004.*

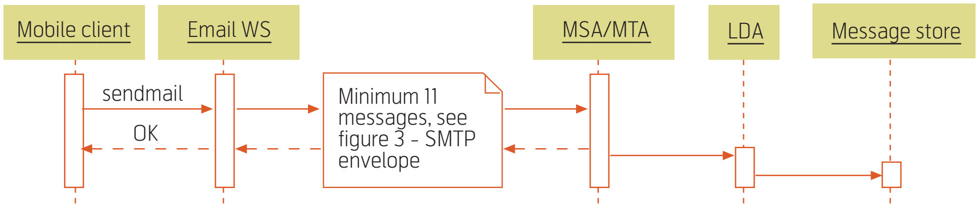


Figure 9 Sending an email using XMMAP

POP and IMAP do not have as many message transfers as SMTP. Often only one request and response is required per invocation. On the other hand, these protocols require that the user authenticate before invoking any operation on the account. This introduces session management and extra delay and overhead.

The major benefit when using XMMAP is that it both combines and simplifies the functionality and information given by the MTA (SMTP), Access Agent (IMAP/POP) and the representation format (SMTP-IMF/XMTP). This is achieved by mapping the information into an XML format and tying different parts of the format to specific requests and responses to SOAP methods.

Like XMTP, XMMAP utilizes the strength of XML to simplify parsing of the received information. The terminal can use an XML-parser to retrieve the information from the XMMAP-message as well as to construct new XML documents. This makes implementation of an XMMAP-compatible mail client a very simple task compared to a full-scale email client supporting SMTP with a wide range of MIME-types, as well as the client implementations of POP and IMAP.

XMMAP introduces a new concept of messaging. While traditional protocols rely on frequent message transfers with well-defined operations, XMMAP is more flexible, making it possible to do more operations in one exchange.

Box 4

```
<Message>
   xmlns=''http://www.finngard.org/ 2004/03/xmmap/''
   xmlns:web=''http://www.w3.org/ 1999/02/22-rdf-syntax-ns#''
   web:about=''mid:107840631700 2232@lycos-europe.com''>
   <BoxNumber mailbox=''INBOX''>12</ BoxNumber>
   <Headers>
            <From>John Doe &lt; johndoe@telenor.com&gt;</From>
            <To>&lt; finngard@finngard.org&gt; </To>
            <Subject>XMMAP protocol, suitable for PDA?</Subject>
            <Date>Fri, 2 Mar 2004 12:23:12 -0700</Date>
   </Headers>
   <Flags protocol=''imap''>
        <Seen>1</Seen>
        <Answered>0</Answered>
   </Flags>
   <Body charset=''ISO-8859-1''>
   Hi! Do you suggest using the XMMAP format for PDA's as well as mobile phones?
   </Body>
   <Attachments>
        <Attachment content-type= ''application/x-ms-word''
            encoding=''base64''>
            <AttachmentNumber>1</ AttachmentNumber>
            <Filename>pda_description. doc</Filename>
            <Size>1345</Size>
            <Content>/9j/4AAQSkZJRgABAQEAYABgAAD//gAcU29mdHdhcmU6IE1pY3
                     Jvc29mdCBPZmZpY2X/2wBDAAoHBwgHBgoICAgLCgoLDhgQDg
                     0NDh0VFhEYIx8lJCIfIiEmKzcvJik0KSEiMEExNDk7Pj4+JS
                     5ESUM8
                 ....
            </Content>
        </Attachment>
   </Attachments>
</Message>
```


ISSN 0085-7130 © Telenor ASA 2005

XMMAP is in its basic form a representation of an entire mail account, spanning everything from login credentials to flags in a specific message. This makes it possible to use sub-parts of the XMMAP-format to represent different parts of a mail account for different purposes. This is especially useful when utilizing XMMAP to invoke methods on the mail server. When invoking a method, only the relevant subparts of the formats are sent, thus avoiding unnecessary overhead.

4.3.1 XMMAP data format

As mentioned, XMMAP is a representation of an entire mail account. This section gives a brief description of the format including several examples. We begin straight at the point by giving an example of how the most important part of an account, a single message itself, is represented in XMMAP, see Box 4.

The format is very much self-explanatory as it contains elements from the mentioned known protocols (SMTP, XMTP, POP3 and IMAP). XMMAP maps perfectly into a standard SMTP-IMF message, with or without MIME extensions, and vice versa. In addition to this, an XMMAP-message supplies the essential information given by both POP3 and IMAP mail access protocols.

The <Message> element is the root of a MIME/RFC2822 message. A <Message> element contains <BoxNumber>, <Headers>, <Flags>, <Body> and <Attachments> elements.

4.3.2 Explanation of elements in XMMAP

namespaces

The namespace links to a dated URI, which holds relevant XML Schemas for describing the current usage of the XMMAP protocol.

web:about attribute

This attribute is adopted from the XMTP-format, and contains the message identifier URI.

<BoxNumber> element

A <BoxNumber> represents the message's number in a mailbox. The actual mailbox may be given as parameter when not implicit from surrounding context.

Box 5

```
<Body charset=''ISO-8859-1''>
   MESSAGE COULD NOT BE REPRESENTED IN TEXT/PLAIN!
</Body>
<Body charset=''ISO-8859-1'' content-
type=''text/xml''>
   &lt;xml&gt;
   &lt;Element&gt;Data&lt;/ Element&gt;
   ....
</Body>
```

<Headers> element

The <Headers> element contains standard RFC 2822-headers with local names representing the header names. Parameters are represented by child elements.

One major point when using XMMAP is to not transfer each and every header contained in the original SMTP-IMF-message. Instead it should be aimed at using only a limited set of headers transferring information relevant to the end user; e.g. an SMTP message often contains several "Received" and "X" headers. This information might be relevant when backtracking message-paths and for system specific processing tasks (such as firewalling and spam filtering). But this information is not relevant to the end user. To save storage space and transmission capacity on the mobile terminal, as many header elements as possible should be stripped from the message. This is done by the SMTP->XMMAP gateway before delivering the message to the mobile terminal.

<Flags> element

The <Flags> element supplies the current flags of an email as child elements. When retrieving previously set flags with XMMAP, the set flags are transmitted and the tag holds the value "1". When setting flags, a value of "1" means that the flag should be set, and "0" means that the flag should be unset. The flags are mapped directly from the IMAP specification.

<Body> element

The body element contains the body of the email message. The current character set is given as a parameter. If no parameter is given, "us-ascii" is assumed as in SMTP-IMF. The content of the <Body> element is always represented as "text/plain"-MIME type. If the original message supplies optional ways of representing the message, all but one alternative are stripped from the message by the SMTP->XMMAP gateway. If the body is only represented in for example "text/html", the HTML-tags are stripped. The results after stripping are supplied as "text/plain". This is done to reduce the amount of data transmitted, as well as supplying the data in a format as simple as possible for use with mobile terminals.

Optionally the representation alternatives can be supplied and the corresponding MIME types given as parameters to the <Body> element. If the client cannot represent a body part, an error message will be supplied and shown to the user.

An example of alternative representations of <Body>-element is found in Box 5.

<Attachments> element
This is only relevant if the message contains attachments, which are supplied within the Attachments-element. Each attachment is contained within an Attachment-element. The MIME type and encoding are given as attributes. Encoding may be omitted, base64 is then assumed. In addition, an <Attachment> element contains <Filename>, <Size> and <Content> elements, containing the filename, size of the file and binary encoded data, respectively.

Attachments may be large, and in many cases the user is not interested in downloading these, but only in receiving the message-text and attachment-description. In such cases the <Content> element may be left empty (<Content/>). Even if the <Content> element is empty, the user receives information about the file type, file name and size of the file. From that information the user can decide whether he wants to download the attachments later or not.

The described elements are sufficient for representing an email message, including elements both from SMTP-IMF representation format as well as flag information used in access protocols. But in order to represent an entire account, some additional elements are needed. A representation of an entire account using XMMAP is shown below.

Account representation using XMMAP
See Box 6.

<Account> element
The Account element represents an entire mail account with all its meta-information as well as content. The account contains <Mailboxes>-elements. Account elements can also include login credentials. This can be <UserName>, <PassWord>, <Protocol> and <Port>. The <Account> element is necessary only when logging on to an account, and when receiving a list of available mailboxes from the server.

<Mailboxes> element
An account may contain one or more mailboxes; these are listed within the Mailboxes-element-tag as MailBox-elements.

Box 6

```
<Account xmlns=''http://www.finngard.org/xmmap/''>
  <UserName>jonfinng</UserName>
  <PassWord>l337s3cr37</PassWord>
  <Protocol>IMAPS</Protocol>
  <Port>443</Port>
  <MailBoxes>
     <MailBox>
        <BoxName>INBOX</BoxName>
        <Unread>12</Unread>
        <Total>23</Total>
        <Messages />
        <MailBoxes>
           <MailBox>
              <BoxName>INBOX.old </BoxName>
              <Unread>0</Unread>
              <Total>10</Total>
              <Messages />
              <MailBoxes />
           </MailBox>
        </MailBoxes>
     </MailBox>
     <MailBox>
        <BoxName>Work</BoxName>
        <Unread>3</Unread>
        <Total>34</Total>
        <Messages>
           <Message xmlns= ''http://www.finngard.org/2004/04/xmmap/''
              xmlns:web=
                  ''http://www.w3. org/1999/02/22-rdf-syntax-ns#''
              web:about= ''mid:107840631700223 2@lycos-europe.com''>
              ...
           </Message>
        </Messages>
        <Mailboxes />
      </MailBox>
   </MailBoxes>
</Account>
```


ISSN 0085-7130 © Telenor ASA 2005

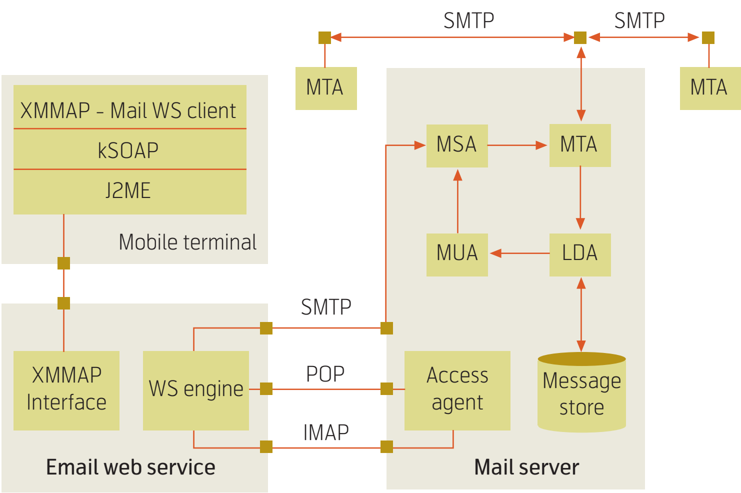


Figure 10 Email Web Service Architecture using XMMAP

<MailBox> element
Information is often wanted on how many messages a mailbox contains and how many of them that are new. One way of retrieving this information is to download all messages in a box and check which one of them that has the right flags set. This is not clever to do on mobile terminals with limited storage and processing capacity. IMAP and POP have functionality that offers this information without needing to download all messages.

The Mailbox-element in the XMMAP-format offers information usually provided by POP and IMAP. The element can also hold full representation of Mailboxes with content. <BoxName>, <Unread>, <Total>, <Messages> and <Mailboxes> are sub elements of a Mailbox-element. BoxName, Unread and Total contain the name of the mailbox, the number of unread messages in it and the total number of messages respectively.

<Messages> element
The Messages-element can contain one or more of the messages of this folder. It can also be left as an empty tag if this information is not wanted in the current request. A Mailboxes-element within a MailBox-element contains sub-mailboxes of the current mailbox if such exists.

loginMobileTermXMMAP	loginProfileXMMAP
getMailboxesXMMAP	getHeadersAsXMMAP
getNewHeadersAsXMMAP	getMessagesAsXMMAP
setFlagXMMAP	deleteXMMAP
sendMailXMMAP	logoutXMMAP

Table 1 XMMAP methods

4.3.3 XMMAP defined methods
The XMMAP data format is useful for representing an account in a minimalist way, and may also be well suited for storage purposes. On the other hand, the data format lacks the coupling to specific methods related to mobile mail access. This coupling is achieved by defining SOAP methods using XMMAP messages as parameters. The set of methods shown in Table 1 are implemented so far. More details can be found in Box 7.

4.3.4 Architecture overview
Figure 10 shows the overall architecture and its different software layers and interfaces. The Web Service Client (WS Client) needs support for SOAP and J2ME as with the other proposed solutions. Also, the client in this solution needs a Mail User Agent (MUA) capable of XML document creation and parsing, as well as for sending and receiving XMMAP messages. In other words: No support for any other mail protocols is needed. This makes the client implementation much simpler compared to traditional solutions.

Every operation on the client requires only *one* XMMAP message sent to the web service, and one message in return. The WS Engine interprets the message client request, does the required communication with the IMAP/POP/SMTP server, and converts the result into an XMMAP message sent back to the client.

Figure 11 shows the interaction between some of the components from Figure 10 when a mobile client sends an email using the web service. We see how the SMTP interaction is stripped down to one message for the client.

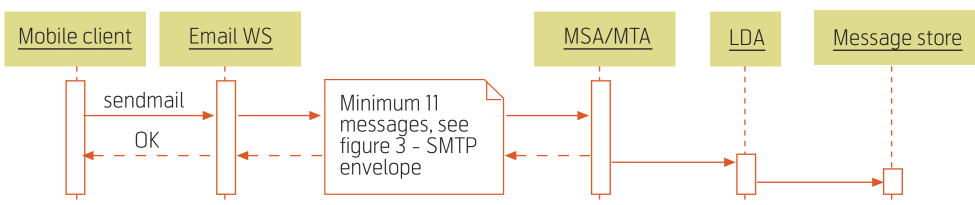


Figure 11 Messaging between participating components when sending an email using XMMAP

ISSN 0085-7130 © Telenor ASA 2005

4.3.5 Internal Web Service Functionality

As mentioned briefly earlier, the web service will adapt the mail messages for mobile terminals like mobile phones and PDAs. This is achieved by removing unnecessary email headers. Alternative representations of the same content (e.g. both plain text and HTML representations of the message body) are reduced to one. Attachments are by default kept back until the user specifically asks to get them. These actions keep the amount of data transmitted and the number of message exchanges to a minimum.

A summary of the internal functionality offered by the XML Web Service (see Figure 10):

- SMTP-IMF to XMMAP gateway and vice versa
- XMMAP to IMAP/POP gateway and vice versa
- Authentication of users
- Holding and managing the user profiles
- Session management for all interfaces and related active connections
- Content adaptation. This includes everything from tag and attachment stripping to picture resizing
- Optional: Local caching of messages and attachments
- Provide necessary interfaces (see section 4.3.6 Interfaces).

4.3.6 Interfaces

The XML Web Service should provide these interfaces:

- SMTP interface. The XML Web Service will do the necessary communication with the MSA/MTA when sending email. In order to make this work, it has to do all communication with the MTA using the SMTP protocol.

- POP/IMAP interfaces. Necessary for retrieving mail, and manipulation mail account. Should also have TLS/SSL support.

- SOAP/XMMAP interface. This interface is used for communication with the client, and is thoroughly described in this document.

- Optional: SOAP interfaces. One or more interfaces for communication with other XML Web Services. Examples may be special content filtering services such as spam and virus filters.

The XML Web Service should have a fast connection to all interfaces so there will be minimal delays in web service-to-mail server communication. All interfaces should provide support for encrypted communication using SSL/TLS.

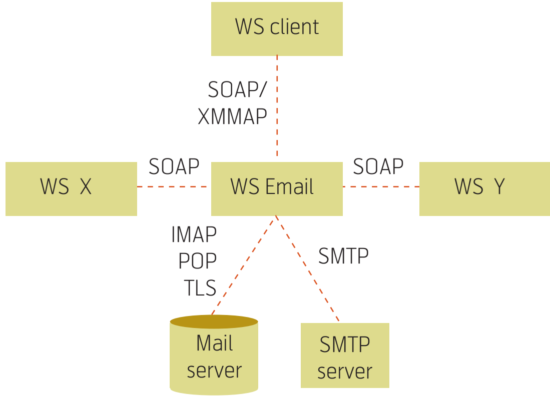


Figure 12 Web service interfaces

5 The Mobile Mail WS Client

The client side of the email Web Service system is realised by an application developed for Java 2 Micro Edition (J2ME) enabled handheld devices. In most cases, this would be a cellular phone. The application contains all the fundamental functionality of a mail client. In addition to the Web Service interface, the client will interface the user through a Graphical User Interface. Since most of the functionality of the service is covered by the earlier sections this section mostly considers the runtime environment of client, the graphical user interface as well as how access to the Web Service functionality is acquired.

5.1 Java 2 Micro Edition (J2ME) overview

The email client application for Mobile phones is implemented in J2ME, Java 2 Platform for Micro Edition. J2ME is a Java platform for small consumer and built-in devices like Mobile Phones and PDAs. These devices offer limited memory and CPU power, which makes efficiency an important issue. J2ME is a set of standard Java APIs, but it is a scaled-down version of the Java platform in order to be more efficient when used on a wireless, handheld, or other small devices.

The platform offers the benefits of Java technology and provides the opportunity to write applications with standard Java development tools. These applications will suit a wide range of devices since the J2ME is deployed in a wide range of Mobile Phones, PDAs and other products supported by Java.

The configuration suited for resource constrained devices like cellular phones is the Connected Limited Device Configuration (CLDC)[7]. CLDC is further released in two versions, 1.0 and 1.1. Most mobile phones nowadays support version 1.0, which does not accept the use of floating point numbers.

The CLDC version 1.0, as specified by Java Specification Request (JSR) 30, is targeted towards devices with the following capabilities:

- 160 KB to 512 KB of total memory budget available for the Java platform
- 16-bit or 32-bit processor
- Low power consumption, often operating on battery power
- Intermittent network connection, possibly wireless and limited to a bandwidth of 9600 b/s or less.

The 160 KB memory budget is derived from the minimum hardware requirements, as follows:

- At least 128 KB of non-volatile memory available for the Java Virtual Machine and CLDC libraries;

- 32 KB of volatile memory for the Java runtime object memory.

Sun's original Virtual Machine (VM) for CLDC was known as the KVM, which stood for Kauai Virtual Machine, sometimes also known as the Kilo Virtual Machine. The CLDC VM is, apart from a few differences, compliant with the Java Virtual Machine Specification and the Java Language Specification.

The libraries available are typically split into two categories: those defined by CLDC and those defined by a profile and its optional packages such as Mobile Multimedia API (MMAPI) and Wireless Messaging API (WMA). Figure 13 shows how these components fit together.

Networking on CLDC devices has been streamlined so that the programmer does not have to fully understand the underlying device capabilities. The Generic Connection Framework (GCF) has been created, streamlining the implementation of networking within applications. This also helps provide a smaller footprint.

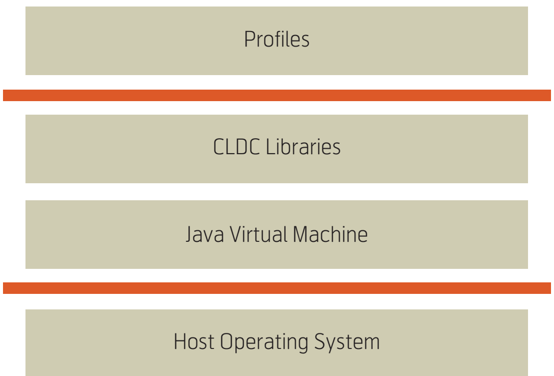


Figure 13 The high-level software architecture of a complete J2ME platform based on the CLDC configuration

The Mobile Information Device Profile (MIDP)[8] combined with CLDC provides a focused platform for mobile information devices, such as mobile phones and entry-level PDAs. MIDP provides the vertical integration required to make the Java runtime environment applicable to these devices by providing direction for the base environment provided by CLDC.

Currently there are two specialised profiles for CLDC-based J2ME platforms; the work with a third one has recently been initialised. They are defined as the Mobile Information Device Profiles (MIDP). The profiles are specifications of APIs, which means that they are collections of functions and functionality. In the first version, MIDP 1.0, all basic elements for a Graphical User Interface are included. However, as the mobile phones get faster and the displays get larger, MIDP 1.0 does not offer enough. The need of something better is present. This leads to the "Gaming Profile" MIDP 2.0. This profile gives the developer far more freedom. To make it short you can say that MIDP 2.0 is a greatly extended edition of MIDP 1.0, including parts that were missing in the previous specification like Over-the-Air (OTA) download of applications.

By default J2ME does not allow use of the native file system of the handset. To compensate for this J2ME has a built-in database system called the Record Management System (RMS). RMS may be compared with a relational database, where each row is identified by a key and can store an array of bytes. Although it is very simple, it can be used to improve many applications where persistent storage of data would be beneficial.

The Java Specification Request (JSR) 172 [9] is Java's Web Service standard for J2ME. Before this specification was released developers were forced to make use of third-party packages in order to use Web Services, or make their own. JSR-172 combines an XML parser and access to remote Web Services in one package. To develop a Web Service client using JSR-172 is fairly easy. JSR-172 is based on static stubs. These stubs are generated from a WSDL. A stub consists of all server calls or methods defined in the WSDL. Request and request classes, or collections of variables, are generated as well. These capabilities are provided to the J2ME platform:

- Access to remote SOAP

- XML-based web services

- Subset of JAX-RPC 1.1 functionality

- Separate deliverable from XML Parsing optional package. The web services API optional package should not depend on the XML parsing optional package.

- It MUST be possible to deliver the web services optional package independent of XML parsing.

- Provide access to web services from J2ME, no server capabilities. This JSR will not define web service endpoints for the target devices. This functionality may be addressed in a future version of the specification.

- Parsing XML data
  - Strict subset wherever possible of JSR-063 JAXP 1.2 functionality.

The mobile phone that has been used to run the described test application did not natively support the JSR-172 specification, but support was added by the developers by integrating Sun's reference implementation with the application. Cellular phones with built-in JSR-172 support are currently starting to emerge, and Nokia is even working on a developers' guide describing how to develop Web Services based on JSR-172 for Nokia phones.

5.2 Client architecture

The interface between the client and the server is entirely specified by the WSDL of the already deployed email Web Service. The communication towards this service is done using XML over the Web, using SOAP encoding.

5.2.1 Client functionality and components

This section illustrates the functionality of the implemented email Web Service client using Use Case diagrams as part of the Unified Modeling Language (UML). Figure 14 displays a high-level use case diagram for the email client.

This high-level diagram puts the functionality into different categories; mail management, contact list and configuration. The dashed line indicates a dependency. The dependency illustrates that functions in "Mail Management" component need to know the account settings in order to operate properly.

The mail management component

The management of the email in IMAP and POP3 is carried out in about the same way for the user as a usual mail client. The functions supported can be seen in Figure 15. The user is provided with the most common features for management of mail, such as sending and retrieval of mail.

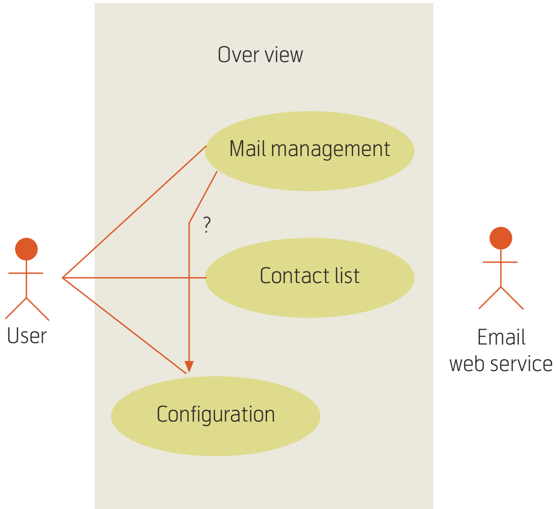


Figure 14 High-level use case diagram displaying the three categories of functionality supported by the client

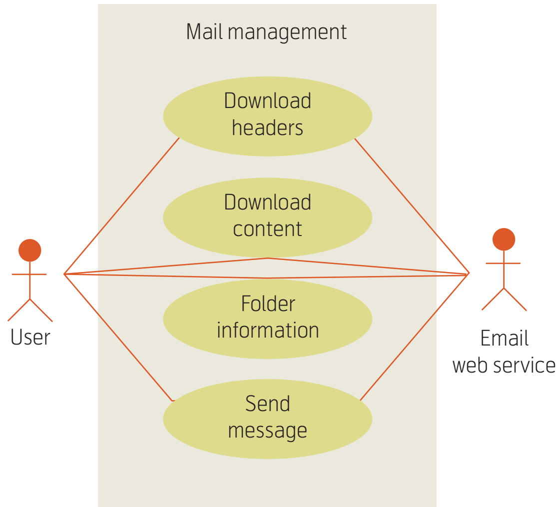


Figure 15 Functions of the mail management component

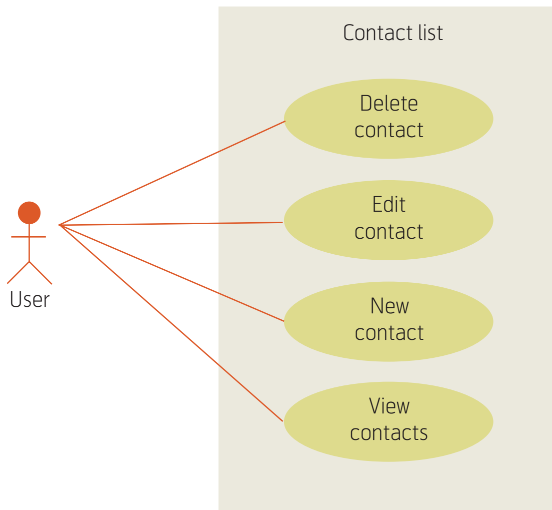


Figure 16 Functions of the contact list component

ISSN 0085-7130 © Telenor ASA 2005

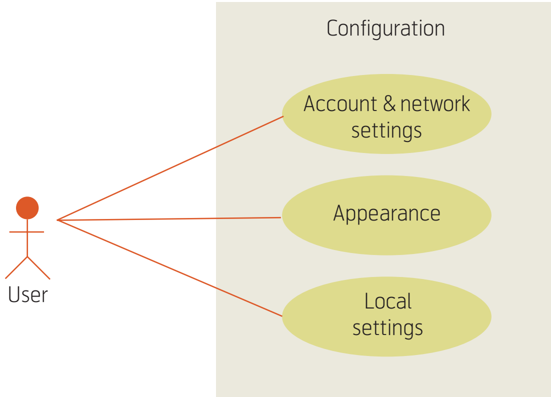


Figure 17 Functions of the configuration component

Figure 18 The main menu of the email client

Figure 19 The Inbox of the email client

Figure 20 Displaying an overview of new incoming messages

The contact list component

To make the email client user-friendly, it was necessary to implement a contact list. All normal functionality for management of contacts is available for the user – adding, editing and deleting, as illustrated in Figure 16. The list supports immediate creation of mail by clicking on a contact. In addition, it is also possible to select multiple contacts simultaneously, in order to send the same email to several recipients.

The configuration component

The configuration component provides persistent storage of all settings (see Figure 17), e.g. mail account settings, network settings, settings regarding the appearance of the client and other local settings.

The Graphical User Interface

The Graphical User Interface is important; in fact it might be just as important as the application runs logically correct. Everyone has tried a poorly designed GUI, and everyone knows that it really does matter how things look and feel; people quickly stop using these applications. J2ME lacks a bit when it comes to design of fancy GUIs, at least when using items. Using default items such as lists and text fields gives little flexibility when it comes to colours. In fact, there is no common way to set the background colour of these items. This may be solved by using licensed third-party compiler packages, such as J2MEpolish. By taking advantage of the possibilities the Canvas class gives, you can make attractive menus and lists.

All menus or lists in this application have basically the same design. Logically they are equal, but visually they differ slightly. All lists feature the same functionality. What makes these lists better or more appropriate for this application is their ability to expand the selected item. Figures 18–20 show a few rendered screenshots of the custom list.

6 Conclusion

In this paper, we propose to provide email functionality on mobile terminals using XML Web Services, which has proven to be reasonable. No other solution offers this combination of standardized protocols, adaptation and presentation flexibility, low bandwidth consumption as well as simple interfaces and implementation. This makes this solution ideal for the complex world of heterogeneous mobile terminals.

Email may be accessed from anywhere even outside the firewall, from any terminal without introducing complex authentication procedures or IP-tunnels using VPN solutions. Thanks to SOAP, sessions can survive temporary connectivity failures and even change of underlying protocols.

ISSN 0085-7130 © Telenor ASA 2005

As SOAP implementations on different mobile terminals become more mature, new functionality such as XML-based encryption and PKI-based authentication may be added to extend the service. With these extensions the concept of providing email to mobile terminals with XML Web Services will become an obvious and powerful solution.

The proposed solution was implemented and tested at the Telenor R&D's laboratory. The email Web Service was implemented by two Master students from the Norwegian University of Science and Technology (NTNU). The mobile mail client was implemented on J2ME by two BS students from the Oslo University College. The prototype is working quite well and meets our expectations.

References

1 Klensin, J. (ed.) 2001. *RFC2821: Simple Mail Transfer Protocol*. IETF, April 2001. online: http://www.ietf.org/rfc/rfc2821.txt

2 Crispin, M. 1996. *RFC2060: Internet Message Access Protocol – Version 4rev1*. IETF, December 1996. online: http://www.ietf.org/rfc/rfc2060.txt

3 Myers, J, Rose, M. 1996. *RFC1939: Post Office Protocol – Version 3*. IETF, May 1996. online: http://www.ietf.org/rfc/rfc1939.txt

4 Borenstein, N, Freed, N. 1993. *RFC1521: MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies*. IETF, September 1993. online: http://www.ietf.org/rfc/rfc1521.txt?number=1521

5 Gudgin, M et al. (eds.) 2003. *SOAP Version 1.2*. W3C, June 2003. online: http://www.w3.org/TR/soap/

6 Borden, J. 2000. *XML MIME Transformation Protocol (XMTP)*. The OpenHealth Group, 2000. online: http://www.openhealth.org/documents/xmtp.htm

7 Sun Microsystems Inc. 2003. *JSR139: Connected Limited Device Configuration. Specification, Version 1.1, Java 2 Platform, Micro Edition (J2ME)*. Java Community Process (JCP), March 2003. online: http://jcp.org/en/jsr/detail?id=139

8 Sun Microsystems Inc. 2000. *JSR37: Mobile Information Device Profile. JCP Specification, Java 2 Platform, Micro Edition, 1.0a*. Java Community Process (JCP), December 2000. online: http://jcp.org/en/jsr/detail?id=37

9 Sun Microsystems Inc. 2004. *JSR172: J2ME Web Services*. Java Community Process (JCP), October 2003. online: http://jcp.org/en/jsr/detail?id=172

Box 7 – XMMAP methods

loginMobileTermXMMAP and loginProfileXMMAP

These methods are used for authentication and are mandatory for retrieval of mail, since remote IMAP and POP servers require authentication. It is recommended to use these methods for authentication regardless of proceeding operation. This is because authentication of users largely prevents misuse of the service, as well at offering better back tracing possibilities when undesired usage is detected.

When invoking one of these two methods a session is established by the Web Service. It is recommended that this session is a SOAP session, but it may also be based on underlying protocols such as HTTP. The Web Service also manages a connection to the mailserver related to each session. The session fetches a previously stored profile, or creates a new one. The profile, among other information, holds custom adaptation properties for the mobile terminal used in this session.

The difference between these two methods is that "loginProfileXMMAP" uses a pre-stored profile, while "loginMobileTermXMMAP" takes more input parameters which are needed to create a new profile for present and future use. These parameters are typically related to the mobile terminal as screen size and maximum number of colours supported. These parameters are not part of the XMMAP-message itself.

The response of an invocation of these methods is a session- and profile ID in addition to the XMMAP-message (see example below). The XMMAP message is a list of the mailboxes in the requested account.

 ISSN 0085-7130

Request message example

```
<Account>
   <UserName>jonfinng</UserName>
   <PassWord>secret</PassWord>
   <Host>imap.stud.ntnu.no</Host>
   <Protocol>IMAPS</Protocol>
   <Port>443</Port>
</Account>
```

Response message example

```
<Mailboxes>
   <Mailbox>
      <BoxName>INBOX</BoxName>
      <Unread>2</Unread>
      <Total>23</Total>
   </MailBox>
   <MailBox>
      <BoxName>INBOX.old</BoxName>
      <Unread>0</Unread>
      <Total>4</Total>
   </MailBox>
</Mailboxes>
```

getMailBoxes

This method retrieves the mailboxes from an account when already logged in. The message exchange is quite similar to the login-methods. The request can however be stripped down to `<Account />` or completely omitted since the user is already logged in, and the account information is stored within the session.

getHeadersAsXMMAP and getNewHeadersAsXMMAP

These methods are used to fetch headers from messages in a mailbox. It is often desirable to only fetch information about unread mail to avoid flooding the screen on the mobile terminal with information about old emails. Therefore, a method called "getNewHeadersAsXMMAP" should be implemented in addition to "getHeadersAsXMMAP". Which headers to send and which to strip off must be set as part of the user profile.

Request message example

```
<Mailbox>
   <BoxName>INBOX.old</BoxName>
</Mailbox>
```

Response message example

```
<Messages>
   <Message web:about=''mid:1078406317002232@lycos-europe.com''>
      <Headers>
      <To>me@here.com</To>
         <From>someone@there.com</From>
         …
   </Headers>
</Message>
<Message>
   …
</Message>
</Messages>
```

getMessagesAsXMMAP

As the name suggests, this is the message for retrieving mail message content. The request message must provide the name of the mailbox and message number(s). Note that the content of the eventual attachments are not initially sent; only the information describing them. The user can then choose to download them later using the "getAttachmentsAsXMMAP" method.

Request message example

```
<Mailbox>
   <BoxName>INBOX</BoxName>
   <Messages>
      <Message>
         <BoxNumber>12</BoxNumber>
      </Message>
      <Message>
         <BoxNumber>21</BoxNumber>
      </Message>
   </Messages>
</Mailbox>
```

Response message example

```
<Messages>
   <Message>
      <Body>
         …
      </Body>
      <Attachments>
         <Attachment
content-type=''application/x-ms-word''
encoding=''base64''>
            <AttachmentNumber>1</AttachmentNumber>
            <FileName>pda_spec.doc</FileName>
            <Size>1240211</Size>
         </Attachment>
         <Attachment
content-type=''application/pgp-signature''
encoding=''base64''>
            <AttachmentNumber>2</AttachmentNumber>
            <FileName>signature</FileName>
            <Size>202</Size>
         </Attachment>
      </Attachments>
   </Message>
</Messages>
```

getAttachmentAsXMMAP

This method can be invoked to get the content of an attachment. The client will have received a description of the attachment(s) from a call to the getMessagesAsXMMAP method, and uses the attachment number(s) to identify which attachment(s) within which message(s) it wants. Only the attachment(s) will be returned, the body content is omitted.

Request message example

```
<Mailbox>
   <BoxName>INBOX</BoxName>
   <Messages>
      <Message>
         <BoxNumber>12</BoxNumber>
         <Attachments>
            <Attachment>
               <AttachmentNumber>1</AttachmentNumber>
            </Attachment>
            <Attachment>
               <AttachmentNumber>4</AttachmentNumber>
            </Attachment>
         </Attachments>
      </Message>
   </Messages>
</Mailbox>
```

Response message example

```
<Messages>
   <Message>
      <BoxNumber>12</BoxNumber>
      <Body/>
      <Attachments>
         <Attachment
content-type=''application/x-ms-word''
encoding=''base64''>
            <AttachmentNumber>1</AttachmentNumber>
            <FileName>pda_spec.doc</FileName>
            <Size>1240211</Size>
<Content>/9j/4AAQSkZJRgABAQEAYABgAAD//gAcU29m
dHdhcmU6IE1pY3Jvc29mdCBPZmZpY2X/2wBDAAo
HBwgHBgoICAgLCgoLDhgQDg0NDh0VFhEYIx8lJC
IfIiEmKzcvJik0KSEiMEExNDk7Pj4+JS5ESUM8
...
</Content>
         </Attachment>
         <Attachment
content-type=''image/jpeg''
encoding=''base64''>
            <AttachmentNumber>4</AttachNumber>
            <Content>
            …
            </Content>
         </Attachment>
      </Attachments>
   </Message>
</Messages>
```

setFlags

This method can be used to set message flags on the IMAP/POP server. The method returns nothing, except from error messages if something goes wrong. Note: setting the DELETED flag with setFlags marks the message as deleted, but keeps it in the mailbox. For permanent deletion, see the delete method.

Request message example

```
<Mailbox>
   <BoxName>INBOX</BoxName>
   <Messages>
      <Message>
         <BoxNumber>12</BoxNumber>
         <Flags>
            <Seen>1</Seen>
            <Answered>1</Answered>
         </Flags>
      </Message>
      <Message>
         <BoxNumber>4</BoxNumber>
         <Flags>
            <Deleted>1</Deleted>
         </Flags>
      </Message>
   </Messages>
</Mailbox>
```

delete

This SOAP method marks a message as deleted, or deletes it permanently (also called expunging) from the IMAP/POP server. Whether or not to expunge is given as a parameter to the method call. The default is not to expunge. The method returns nothing, except from error messages if something goes wrong.

Request message example

```
<Mailbox>
   <BoxName>INBOX</BoxName>
   <Messages>
      <Message>
         <BoxNumber>14</BoxNumber>
      </Message>
      <Message>
         <BoxNumber>15</BoxNumber>
      </Message>
   </Messages>
</Mailbox>
```

sendMail

This is the method to use for sending mail from the mobile client. It is sufficient with one message exchange compared to SMTP's minimum of 11 exchanges. If using the IMAP protocol, the Web Service sends the final message to the user's "sent" box and marks messages as "Answered" in the current mailbox if the message is a reply (given by a 'ReplyNumber' element together with the headers). Setting several `<To>`, `<Cc>` or `<Bcc>` headers in the request message adds multiple recipients in XMMAP. These headers are parsed by the Web Service and converted to SMTP on its outgoing interface. A RFC2822 compliant message is also created by the Web Service from the information received before it is sent to its final destination by SMTP.

Request message example

```
<Message>
   <Headers>
      <ReplyNumber>12</ReplyNumber>
      <From>Mobile Finngard &gt;finngard@xmmap.com&lt;</From>
      <To>Jomar Jalla &gt;jomar@home.com&lt;</To>
      <To>Britt Jalla &gt;britt@home.com&lt;</To>
      <Subject>Hold yer horses, picture coming up</Subject>
   </Headers>
   <Body>
      Here it is.. :)
   </Body>
   <Attachments>
<Attachment content-type=''image/png''
encoding=''base64''>>
         <Filename>onthebeach.png</Filename>
         <Size>12343</Size>
         <Content>
/9j/4AAQSkZJRgABAQEAYABgAAD//gAcU29m
dHdhcmU6IE1pY3Jvc29mdCBPZmZpY2X/2wBDAAo
HBwgHBgoICAgLCgoLDhgQDg0NDh0VFhEYIx8lJC
IfIiEmKzcvJik0KSEiMEExNDk7Pj4+JS5ESUM8
...
         </Content>
      </Attachment>
   </Attachments>
</Message>
```

This message does not need any response if the Web Service could successfully send the mail to its final destination. A SOAP fault describing the error is returned if the mail for some reason could not be delivered.

logout

To log out, the server needs no more information than the session id. Hence the logout call is a SOAP message with empty body. This will close any connections to the user's mail server and delete the session from the web service server.

Summary

The methods briefly described here are only a limited subset of the most important methods that can be implemented by coupling SOAP and XMMAP for email access on mobile terminals. There is in principle no limitation as to what additional methods that may be implemented. The XMMAP format is flexible, and can be extended by new elements in any part of the format if found convenient. There is no fixed order in how the current messages are sent after the user has logged in. When defining new methods, this principle should be followed in order to keep every message independent from both previous and succeeding messages.

For a presentation of Do van Thanh, please turn to page 2.

For a presentation of Ivar Jørstad, please turn to page 21.

Tore Jønvik is Associate Professor at Oslo University College, Faculty of Engineering. Jønvik holds an MSc in Physics and a PhD in Informatics from the University of Oslo. He is guest researcher at Telenor R&D and is working in two Eureka-Celtic projects, ADPO and Fidelity.

email: tore-erling.jonvik@telenor.com

Fetching home music – Sending photos home

DO VAN THANH, IVAR JØRSTAD AND DO VAN THUAN

Do van Thanh is Senior Research Scientist at Telenor R&D

Ivar Jørstad is a PhD student at Norwegian University of Science and Technology

Do van Thuan is Lead Scientist at Linus AS

More and more people have mobile phones. More and more people have PCs and broadband connection to the Internet. It is only a matter of time before they require access to their home computers from their mobile phones. This paper proposes an architecture based on XML Web Services allowing remote access to the user's home network file system. A global system architecture is provided and details of the mapping between the network file system and an XML Web Service are depicted using XML Schema Definitions.

1 Introduction

Imagine a passenger waiting at the airport. He may want to listen to music. His mobile phone can function as an MP3 player but unfortunately, he forgot to load the songs that he has selected and saved in his PC at home. It would be nice if he were able to access to his home PC and fetch the tunes that he wants. Imagine an executive travelling on a business trip who experiences some magnificent scenery. She takes pictures using a 2 Megapixel digital camera on her mobile phone. She may want to keep these pictures and send them to her home computer.

Indeed, more and more households are acquiring broadband Internet connections to their home. On the home network they may have several computers running various services and hosting private documents and files. It is hence quite desirable to be able to access these services and documents from outside their home via stationary computers or mobile devices like cellular phones or PDAs.

The permanent connection to the Internet paves the way for a new type of service provisioning where services are partly located in the home network of the users and accessible from a cellular phone. Potential services can be content based like browser bookmarks, address books, MS Office documents or similar, but it can also be applications rendering more dynamic services to the user (giving an output, provided some input).

Unfortunately, due to the threats from the Internet, most home networks need to have firewalls. In addition, Network Address Translation (NAT) is also used. Firewalls and NAT make the access to the home system from the outside world difficult. This paper proposes an architecture allowing remote access to home file systems from outside devices which can be a PC, PDA or mobile phone. Without compromising security, the proposed solution is simple and requires only minimal technical skills to take into use. All that is needed is to download and install an application on the computer where services reside (in the home network), as well as to download and install a client on the device in use.

2 The evolving home network

A. Firewall and routing problems

Only a few years ago, most of the home connections to the Internet were modem-based. Only when needed would the user initiate the dial-in process and hook up his PC to the Internet. The situation has changed dramatically with the emergence of broadband Internet connection possibilities like cable TV, xDSL, etc. These technologies do not only offer higher access bit rate to the users but also make the user home network "always on". The users do not have to dial up anymore but can just go right away on the Internet whenever they like. Unfortunately, a coin always has a reverse side. The home networks are now exposed to attacks in the same way as enterprise networks and some protection is necessary. Home networks, similar to enterprise networks, most often employ a router with embedded firewall to connect the home LAN to the Internet through e.g. an ADSL or cable TV connection. The firewall protects the home networks against intruders but at the same time prevents all access from outside even by the owners themselves.

To allow incoming access to the home network, this firewall/router must be bypassed and the communication must be forwarded to the appropriate computer on the home network. It might be necessary to open up several TCP ports (e.g. 138, 139 and 445) and direct the incoming traffic on these ports to the computer that is sharing resources. This is often a complicated task that might be too difficult for the common user.

There is a need for a solution that allows legitimate accesses to the home network but at the same time preserves security.

ISSN 0085-7130 © Telenor ASA 2005

B. Multi computer and user support

Not long ago, most households had only one computer and one unique user – usually the head of the family. The situation has changed dramatically and nowadays households have often several personal computers connected together by a LAN. The number of users is extended to include all the household's members. Networked file systems are used to allow users to share computers and disk space. A user on one computer can also access files stored in a disk located on another computer.

The most common network file system in existence today is the Common Internet File System (CIFS) [1] published by Storage Networking Industry Association (SNIA), a non-profit trade association dedicated to ensuring that storage networks become complete and trusted solutions across the IT community. CIFS is based on the Server Message Block (SMB) and NetBIOS protocols. The first attempt to define the NetBIOS was initiated in 1984 [5]. In 1988 this protocol had evolved to the Server Message Block (SMB) protocol. The history of SMB is long and complicated, and the standardization has been undertaken by several different actors, among them Microsoft. Eventually, SMB has evolved into CIFS, which is the standard network file system used by Microsoft operating systems today.

For other systems than operating systems developed by Microsoft, a software suite called Samba has since 1992 been developed to allow for example Linux clients to access an MS Windows file or print server. It can also let the Linux computer act as a CIFS file or print server, thus it allows MS Windows clients to access these shared resources.

CIFS could be used directly by a remote terminal (e.g. a computer) to access resources on the home based LAN. However, for mobile devices like cellular phone or PDA, access to the home network is not possible since there is no CIFS client. The implementation of a CIFS client might be challenging due to the physical limitations of the mobile devices such as processing capabilities, storage, battery life, etc., not to mention that the protocol is complicated with many details that must be coped with in an implementation.

Another major challenge already mentioned is the firewall issue. The home network often connects several personal computers to the Internet, computers belonging to different users (e.g. family members). Access to each of these computers from remote locations, by the owner, should be possible. Moreover, the configuration for forwarding packets in an ordinary firewall is static and cannot differentiate between the sources of the request. Even if the firewall could decide the destination of the packet based on the source address of the request, it would not have been an adequate solution, because the Dynamic Host Configuration Protocol (DHCP) is often used. Thus, the IP address of the client device will most likely change over time. The multi user support could therefore only be supported in a higher layer; i.e. the application layer.

It is hence desirable to have a solution that enables remote accesses to all the home computers and files by legitimate users from mobile devices.

3 Existing solutions

Currently, Virtual Private Networks (VPNs) are used to allow access to file systems in remote locations from e.g. laptop computers by employees that are traveling (often referred to as road warrior configuration) or to access corporate networks from their home. VPNs can be a solution to allow home users to access their home networks.

One of the advantages of using this solution is that it provides a secure tunnel between two locations, across the public Internet. This tunnel is realized by strong cryptography, both for authentication of users and for encryption of content that travels across the tunnel. Also, as suggested by the name, when using VPN the networks connected by the tunnel will appear as one network by the user, thereby allowing easy access to well-known features of geographically separated networks. The technology is often used for interconnecting branch offices of enterprises. However, this solution has a couple of drawbacks when considered in the light of mobility of novice users of Information and Communication Technology (ICT).

VPN solutions are often complex to install and configure, and often requires in-depth knowledge of computer and networking technology. This means that ordinary users cannot do this task themselves. It is therefore not common for home based Local Area Networks (LAN) to employ such solutions, and therefore it is not possible to access resources residing in these networks from remote locations.

VPNs are based on strong encryption to allow a secure tunnel between devices across the public Internet. For resource constrained devices, this strong encryption is often not possible due to the amount of required processing power and memory needed to perform the mathematical calculations. Also, the strong cryptography used in VPNs might be an unnecessary overhead for the type of users targeted in this paper. Instead, more lightweight cryptography can be applied, if found necessary.

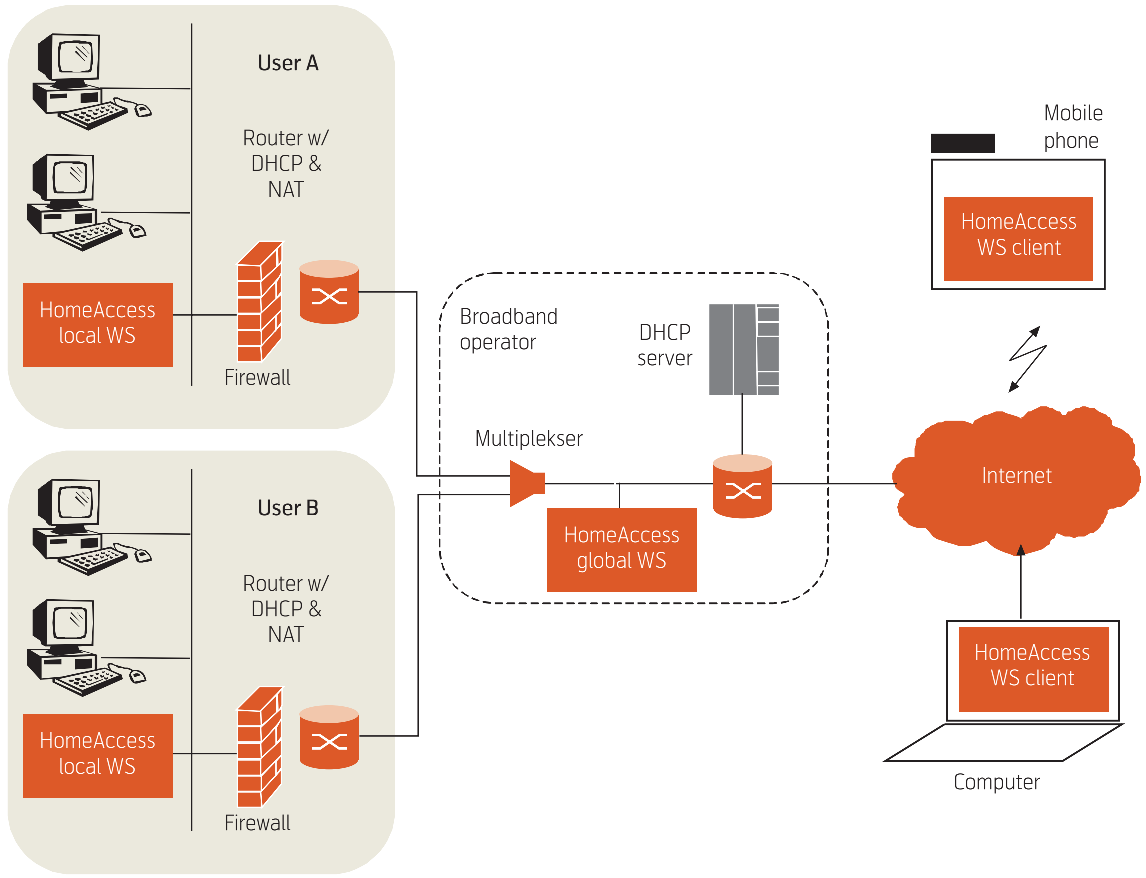


Figure 1 Mobile Home Access architecture

4 Overall architecture

Mobile Home Access is a solution that will allow access to any files or services residing on computers on a LAN behind a firewall, typically a private Local Area Network (LAN). The solution as shown in Figure 1 consists of three components:

- A Home Access Local Web Service that is installed on the PC to provide access to files and services;
- A Home Access Global Web Service addressable by a global IP address;
- A Home Access Web Service Client that is installed on the terminal(s) used to access files and services.

4.1 Home Access Local Web Service

The role of the Home Access Local WS is to expose the relevant operations of the native file system on the World Wide Web such that mobile client can use them to access files and services located within the home network.

As shown in Figure 2, the Home Access Local Web Service has three interfaces:

- The Native File interface
- The Web Service File interface
- The Administrative interface.

In addition to the file access functionality the Home Access Local Web Service also has the functionality to periodically query the IP address of the home network and upload it to the Home Access Global Web Service or a defined globally accessible storage area.

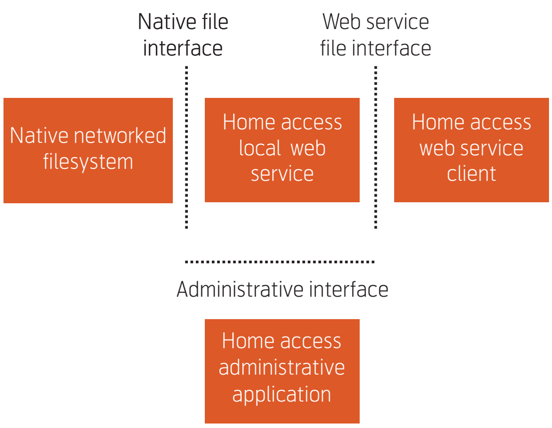


Figure 2 Interfaces in the Mobile Home Access system

ISSN 0085-7130 © Telenor ASA 2005

The Native File interface

At the Home Network, there could be several users sharing several heterogeneous computers and peripheral devices. It is assumed that the Home Network is equipped with a networked file system that allows the users to view and to access remote files located on other computers from one computer. Examples of such a networked file system can be Sun Network File System (NFS) [3][4] or Common Internet File System (CIFS) [5]. However, only CIFS will be considered further since it is incorporated in Microsoft Windows which is installed in most private households.

CIFS is a file sharing protocol. Client systems use this protocol to request file access services from server systems over a network. It is based on the Server Message Block (SMB) protocol widely in use by personal computers and workstations running a wide variety of operating systems. The protocol supports the following features:

- File access
- File and record locking
- Safe caching, read-ahead, and write-behind
- File change notification
- Protocol version negotiation
- Extended attributes
- Distributed replicated virtual volumes
- Server name resolution independence
- Batched requests
- Unicode file names.

Although CIFS is independent of the transport layer, the common transport today across this interface is through NBT (NetBIOS over TCP) or across raw TCP connections.

Raw TCP transport

This mode of operation is the simplest, where SMB messages can be sent immediately to port 445 on the server. Based on the response to a TCP connection request to this port, and possibly a reply to the SMB message, either RAW transport can be used further or an NBT session can be initiated as described below.

NBT transport

In this mode, additional messages must be sent to gain access to the file server and to initiate a session. Also, a valid NetBIOS name of the file server is needed in the message that requests a new session. The need for NBT support completely relies on the servers that should be supported and whether these are expecting correct NBT semantics.

The Web Service File interface

Since the goal is to enable file and service access from outside devices, especially mobile phones, which have several limitations, the requirements on the Web Service interface are as follows:

- It should support both regular computers and limited mobile devices;

- It should be capable of adapting itself to the device type;

- For mobile phones with physical limitations in terms of storage, processing, small display, etc. the operations/methods should be rich such that only few operations are required to accomplish a task.

To cope with these requirements the Web Service File interface consists of the following sub-interfaces:

- Authentication interface
- Administration interface
- Tunnelling interface
- Reduced Mapping interface.

Before allowing access to home files and services, it is important that the identification, authentication and authorisation are carried out properly. The Web Service File interface must have an Authentication Interface.

The Tunnelling interface is more suitable for access from a remote personal computer which has a CIFS client installed. The Reduced Mapping interface is intended for mobile devices with limited capabilities.

Authentication interface

This interface controls identification, authentication and authorization to shared resources.

IAUTHMustAuthenticate(Challenge) – This method is used to notify the client that it is required to authenticate itself prior to accessing any resources through the service access point. This method can be used as a response to any type of request from an unauthenticated client.

IAUTHAuthenticateRequest(Credentials) – This method is used by the client to request authentication by providing proper credentials.

IAUTHAuthenticateResponse() – This method is used by the service access point to notify the client about the outcome of the authentication process.

The Administrative interface

Access to administration methods requires successful authentication through the interface described earlier.

The administrative interface can be used from remote clients as well as from clients on the home network, which could be an administration application.

The Administrative Interface allows a user to specify:
- What directories on the home computer(s) should be made accessible to remote devices.

In addition it allows a System Administrator to configure:
- User accounts.

IADMListHosts() – Lists all hosts on the home network

IADMListUsers() – Lists all registered users

IADMListDirectoriesOnHost(String host) – Lists all accessible directories on the specified host

IADMSetAccessRights(URI resource, Int accessrights) – Sets the specified access rights on the specified file or directory

IADMGetUserConfiguration(String user_id) – Retrieves the specified user's configuration

IADMSetUserConfiguration(String user_id, Configuration c) – Sets the specified user's configuration

Each user's access rights to resources and preferences are controlled through two methods *(IADMGetUserConfiguration* and *IADMSetUserConfiguration)*. By defining a generic method which passes the configuration as a parameter to the Home Access Local Web Service, maximum flexibility is achieved, and new features can easily be added later on.

A configuration contains at least the following definitions:

1 Definition of access rights (readable, writable or both) to specific files and directories;

2 For each resource it should be possible to define which component of the resource is available (offset, length etc.);

3 Definition of access rights and format/presentation from specific device or group of devices;

4 Definition of access rights from specific IP-addresses, subnets or domains;

5 Definition of access rights by specific users and groups of users.

Tunnelling interface

In tunnelling mode, a complete CIFS message is encapsulated in a Simple Object Access Protocol (SOAP) message by the Home Access Local Web Service using binary attachments. At the Web Service client side (on the terminal), the CIFS content is extracted from the SOAP message and exposed through a CIFS server. This way, an ordinary CIFS enabled browser (e.g. Windows Explorer) can be used to access the remote file system.

There are basically two approaches to embedding binary data into SOAP messages:

The first approach is to embed binary data as Base64-encoded data. The most appropriate way of doing this is to embed the data within an *<xsd:base64>* element. However, with current JSR-172 implementations, the *<xsd:base64>* element is not properly supported. The solution is to embed the data within an *<xsd:string>* element. Using base64 encoding results in an increase in size of 1/3 of the original size. For SMB messages containing only signalling information, this might not be a problem, but for the messages containing file contents it is. By using compression before base64-encoding, it is possible to avoid too much overhead for some types of content.

The other approach is to use SOAP with Attachments (SwA) [6], but this is not yet supported by all SOAP platforms. It is however supported with JAX-RPC [7] through SAAJ [8]. Utilising one of the referenced specifications, SwA will be supported by all SOAP platforms in the near future.

Also, the client application exposing the file system is required to authenticate itself towards the service access point before access to the remote file system is granted and the file system can be exposed. Except for this authentication process, all other commands follow the network file system protocol semantics.

The tunnelling interface has two methods:

ITUNReqCommand(CIFSAttachment) – Transports a complete request command from client to host with network file system.

ITUNResCommand(CIFSAttachment) – Transports a complete response command from host with network file system to client.

Reduced mapping interface

Every CIFS message can be replaced by a corresponding SOAP message. In theory, each field of a CIFS message could be mapped into an entry of a SOAP message by the Home Access Local Web

Service. At the client side (terminal), the SOAP message is parsed and the original CIFS message reconstructed and exposed through a CIFS server. However, such a full mapping scheme introduces a lot of overhead and it is not sure that the mobile device is capable of receiving and processing all the data that it gets. A reduced mapping scheme is more efficient and has the following advantages:

- Reduces the content of each message;

- Reduces the number of total messages;

- Reduces the requirements on the mobile device; there is no need for a complete CIFS client on the device, but only a client which supports the specified methods;

- Transferred block size can be adapted to fit the available resources on the handset, as well as to fit the processing capabilities of the parser on the client.

Only the most important parts of native network file system messages are mapped to an XML format. In addition, a set of management interfaces that are used between the client and the service access point, are defined. These interfaces control connection establishment towards shares, as well as maintenance of sessions and teardown of connections.

IACCListResources(URI uri, String pattern, Boolean recursive) – Lists all resources on the specified URI matching the specified pattern. If pattern is left empty, all resources on the specified URI are listed. Setting recursive to true allows this method to be used for searching for specific named resources throughout the entire tree defined by uri.

IACCReadResource(URI uri) – Reads the contents of the specified resource as specified in the user configuration described previously. This method incorporates several methods of the network file system, such as protocol negotiation, session setup etc.

IACCWriteResource(URI uri, WriteSpecification ws) – Writes to the specified resource the content specified by ws (e.g. create/offset/append, data, length etc.). This method incorporates several methods of the network file system, such as protocol negotiation, session setup, etc.

4.2 Home Access Global Web Service

The Home Access Global Web Service is required for the three cases:

- Dynamic global IP address
- Permanent local IP address
- Dynamic local IP address.

It collaborates with several Home Access Local Web Services belonging to different users. It must have a list of users to serve. Before establishing the connection of a Home Access Local Web Service of a user, sufficiently strong authentication must be carried out.

It has the following functionality:

- Discovery and updating of the home network's current IP address;

- Relaying the method requests from the Home Access Web Service Client to the Home Access Local Web Service;

- Exchange of endpoint information (IP addresses and port numbers) between client and server, so as to allow easier NAT/firewall penetration.

It has the following interfaces:

- The Native File interface
- The Web Service File interface
- The IP update interface.

The first two interfaces are the same as the ones defined for the Home Access Local Service.

The IP update interface has the following method:

IUpdateIP(user_id, IP address) – To update the IP address of the specified user.

IGetCurrentIP(user_id) – Returns the current IP address of the specified user.

4.3 Home Access Web Service Client

There are two types of Home Access Web Service Client:

- Tunnelling Client
- Reduced Mapping Client.

The Tunnelling Client will use the Tunnelling interface to interact with either the Home Access Local Web Service or the Home Access Global Web Service. This Client is suitable for regular PCs. It incorporates also a CIFS server such that a regular CIFS client like Windows Explorer can be used to access the remote files and services. In fact, the Tunnelling Client acts much like a CIFS proxy, since it is exposed as a CIFS file system itself, and is a one-to-one mapping of the CIFS file system on the server side.

The Reduced Mapping Client will use the Reduced Mapping interface to interact with the Home Access Local Web Service and Home Access Global Web Service. This Client is suitable for mobile devices such as mobile phones or PDAs (Personal Digital Assistant). It also incorporates a file browser and a user interface (UI) which are designed for devices with limited display and navigation ability.

5 Experiences with existing implementation

The current implementation of the Mobile Home Access system is currently subject to a performance evaluation. The evaluation is performed using the Nokia 6680 cellular phone connected to the Telenor Mobil 3G network. The preliminary results of this evaluation are as follows.

With the initial implementation of the transfer mechanism, an effective throughput rate of 16 kbit/s was achieved. This seemed incredibly low, and it was uncertain if this was due to the overhead introduced by XML processing and SOAP encoding/decoding, base64 encoding/decoding combined with the latency of the cellular network. Since the system is based on a request/response mechanism, it seemed likely that transferring small amounts of data in each request/response pair would increase the total cellular network latency substantially. By increasing the amount of data transferred in each request/response from 3000 bytes to 22,000 bytes, the throughput rate increased to around 81 kbit/s. When content transferred is not already compressed (e.g. Word documents and similar), effective throughput rates of up to 155 kbit/s have been reached. In theory, the amount of data transferred could be increased above 22,000 bytes, but with slightly larger amounts the current XML parser breaks down.

The estimated processing time spent in the mobile device after receiving a chunk of data and before sending the next request for data was for a block size of 22,000 bytes 1.3 seconds. The measured time spent in the XML parser itself was 1058 ms for a block size of 22,000 bytes. As illustrated by these numbers, too much time is spent in processing received data in the handset. With support for SOAP with Attachments, the processing time spent in the XML parser would be dramatically reduced, and the throughput rate substantially increased. The implementation and preliminary testing of this approach is currently being done, but since SOAP with Attachments is not supported by JSR-172, these increased throughput rates will not be achievable by native libraries in the handsets in the near future.

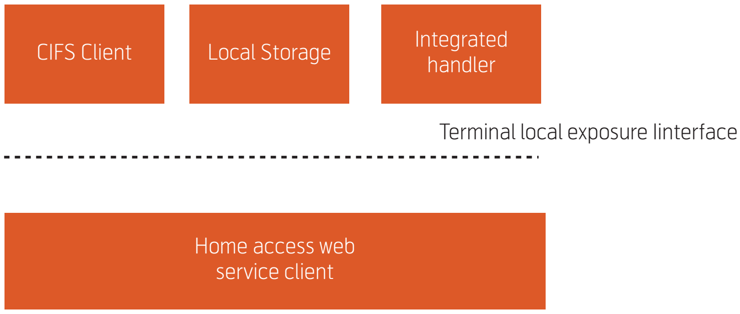


Figure 3 Exposure mechanisms on the terminal; the Terminal Local Exposure interface

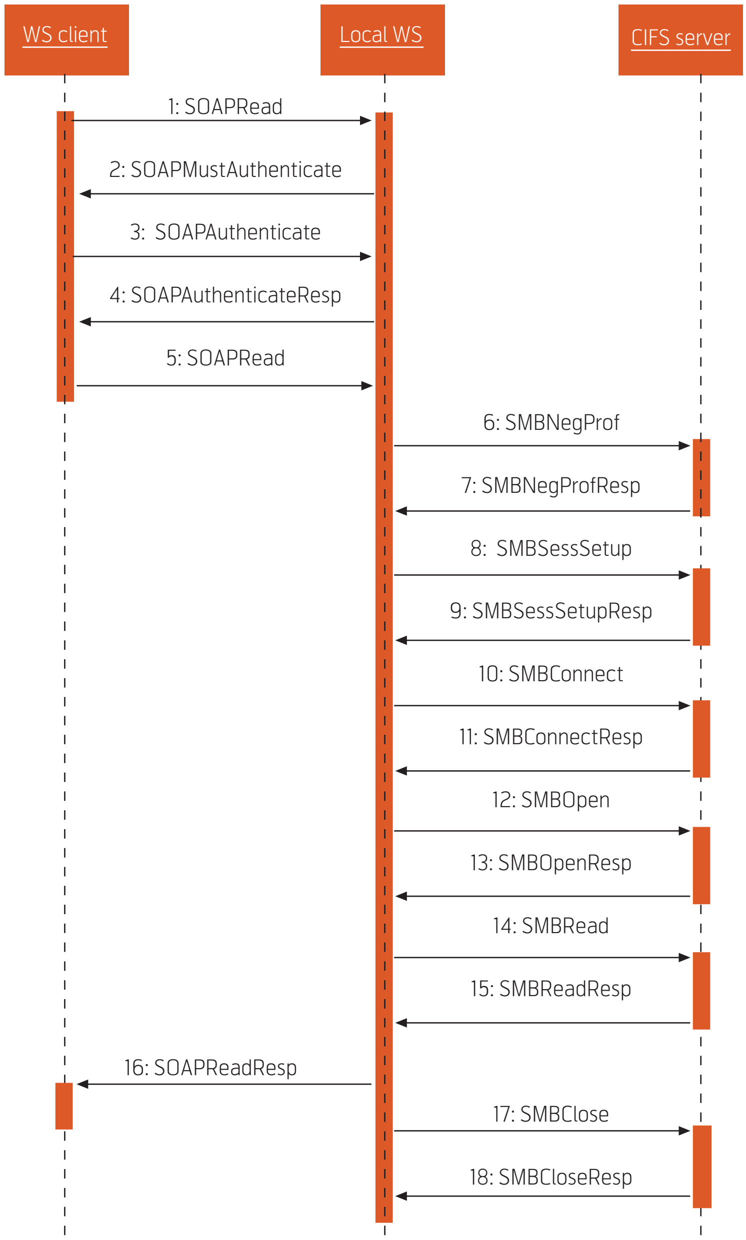


Figure 4 Call sequence for reading a remote file

ISSN 0085-7130 © Telenor ASA 2005

Conclusion

This paper proposes an architecture for providing access to a private home network file system from a remote terminal which prevents the security of the home system. The proposed solution is based on the Web Service concept. For the mapping between the local file system and the Web Service, three alternatives are presented but only two are recommended. The first one, Tunnelling mode is completely compatible for remote personal computers with installed CIFS client like CIFS enable browser. The second one, Restricted mapped mode, reduces the number of method calls and makes the implementation of a client for mobile terminals feasible.

References

1 SNIA CIFS Technical Work Group. 2002. *Common Internet File System (CIFS)*. Technical Revision: 1.0, February 27, 2002.

2 Shearer, D. 1996. *History of SMB Project*. Available online: http://www.samba.org/cifs/docs/smb-history.html

3 Callaghan, B. 1996. *RFC 2054: WebNFS Client Specification*. Sun Microsystems Inc., October 1996, available online: http://www.ietf.org/rfc/rfc2054.txt

4 Callaghan, B. 1996. *RFC 2055: WebNFS Server Specification*. Sun Microsystems Inc., October 1996, available online: http://www.ietf.org/rfc/rfc2055.txt

5 Hertel, C R. 2004. *Implementing CIFS*. Pearson Education Inc. (ISBN 0-13-047116-X)

6 World Wide Web Consortium (W3C). 2000. *SOAP Messages with Attachments*. Available online: http://www.w3.org/TR/SOAP-attachments

7 Sun Microsystems Inc. 2002. *JSR101: Java API for XML-based RPC, JAX-RPC 1.1*. June 11, 2002, available online: http://jcp.org/en/jsr/detail?id=101

8 Sun Microsystems Inc. 2003. *SOAP with Attachments API for Java (SAAJ) 1.2*. October 21, 2003, available online: http://jcp.org/en/jsr/detail?id=67

For a presentation of Do van Thanh, please turn to page 2.

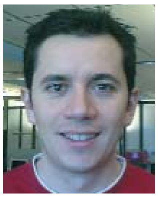
For a presentation of Ivar Jørstad, please turn to page 21.

Do van Thuan is currently Lead Scientist at Linus AS – a Norwegian firm specialising in custom development and system integration for telcos and the process industry. After finishing his studies at the Institute of Informatics, University of Bergen, in 1984, he joined Norsk Data where he worked first with COBOL and FORTRAN compilers, then with OS command interpreters and last with SQL processors for databases. Since then he has been involved with several large projects developing information systems and process/production control systems. These days he is member of two European projects, ADPO – Personalised Adaptive Portal Framework, and Fidelity – Circle-of-Trust based on Liberty Alliance's Specification. His research interests are distributed systems, component technology and software design methods.

email: t.do@linus.no

The role of RFID in the mobile phone

JUAN CARLOS LOPEZ CALVET



Juan Carlos López Calvet is Research Manager in Telenor R&D

The purpose of this article is to give a general overview of the RFID technology and the role it will play when included in mobile phones. The article gives an overview of the RFID market, the current status, the future of RFID and the markets that are driving the technology through the hype curve. It then goes through a description of why it makes sense to include RFID in the mobile phone and what is motivating the different players pushing the technology.

The article also includes a technology overview of the RFID and NFC technologies and finalises with a short conclusion of how mobile operators and other players are positioning themselves in order to take advantage of the opportunities provided by the inclusion of RFID/NFC on the mobile phones.

1 Introduction

As with any new promising technology, RFID is going through a hype where huge expectations are created followed by just as huge disappointments. The technology fans see a sea of opportunities while others fear that the technology will affect their privacy. But what is actually all this hype about?

RFID is in its most simple form just a bar code that can be read from a distance, and some of the most obvious advantages of this technology are:

- Objects do not need to be passed individually through a laser scanner.
- Many objects can be read at the same time from the same reader.
- RFID tags can hold more information than bar codes and therefore users or companies can benefit from richer information.

Based on these advantages, the applications that the technology has are almost endless, like locating and tracking an individual product in real time through production, supply chain and warehouse, contactless payments, positioning, personalization and identification just to name a few.

Along with all the opportunities come questions like: Are people going to monitor each item I'm wearing when I enter a store? Can anyone read the contents of my wallet from a distance? Am I going to be monitored and located everywhere I go?

And the answer is no, the technology is not there, neither for the advantages nor the concerns, but in very specific cases it can already provide a clear benefit for users and companies.

1.1 RFID today

Although RFID can be used in many areas, it is supply chain management that is currently driving the RFID hype due to the implications of the Wal-Mart and Department of Defence Mandates.

1.1.1 The Wal-Mart Mandate

Linda Dillman, Wal-Mart's CIO, publicly announced in June 2003 that the retailer would ask its top suppliers to tag pallets and cases beginning in January 2005. The news created a storm in the retail and consumer packaged goods industries.

Wal-Mart receives roughly 1 billion cases per year from its top 100 suppliers. Dillman said that the company would not be tracking every single case from the top 100 suppliers by January 1, 2005, but rather that it would ramp up over time. "Our goal is to track all pallets and cases." [7]

Dillman also indicated that Wal-Mart would start deploying EPC (Electronic Product Code) technology in the United States and would quickly move to implement it in Europe and then in the rest of its overseas operations.

Because of its size, Wal-Mart will have a major effect on how RFID technology is adopted. Consider a few facts. Wal-Mart's annual sales are greater than the combined sales of the entire semiconductor industry. Wal-Mart's sales are greater than the gross domestic product of Turkey. Wal-Mart imports more goods from China ($14 billion) than Japan does ($10 billion). And it employs more people than Ford, General Motors, Exxon Mobil and GE combined [10].

1.1.2 US Department of Defence RFID Mandate

Following in Wal-Mart's footsteps, the US Department of Defence asked its top 100 suppliers to put RFID tags on shipments.

 ISSN 0085-7130

The DOD suppliers are among the largest companies in the world. The top 100 include Boeing, General Dynamics, Lockheed Martin, Northrop Grumman and Raytheon. By endorsing EPC technology for tracking goods in the military supply chain, it could help spread EPC technology throughout the manufacturing sector because the military's top 100 suppliers also have divisions that make commercial airplanes, electronic components, trucks, ships and other products [9].

1.1.3 EPC and the "Internet of Things"

The EPC network, using RFID tags, will enable computers to automatically recognize and identify everyday objects, and then track, trace, monitor, trigger events, and perform actions on those objects. The technology will effectively create an "Internet of Things."

The industries and media are focusing on this "Internet of Things" for manufacturing, retail, transportation, health care, life sciences, pharmaceuticals, and government, offering an unprecedented real-time view of assets and inventories throughout the global supply chain. Once the products have unique identification codes and with mobile phones that can read such codes (like NFC mobile phones), the concept of the "Internet of Things" will not only be used for industry process, but it will also migrate to the consumer market, giving end users the possibility of accessing product information from their mobile phones.

The electronic product code uniquely identifies objects and facilitates tracking throughout the product life cycle, so when a reader scans an RFID tag on a case of milk, the tag will answer with an EPC number for the inventory management system, letting it know exactly which case of milk was just scanned.

The data structure of an electronic product code is described in the table below.

Electronic Product Code

01	0000B92	00056A	0004325F0
Header	EPC Manager	Object Class	Serial Number
0-7 bits	8-35 bits	35-59 bits	60-95 bits

- *Header:* This field tells the reader what type of number follows, in order to understand what the number represents, for example: a military unique identifier number or a complete EPC structure.

- *EPC Manager:* This field identifies the company or the company entity.

- *Object Class:* This field is a unique numeric identifier that refers to a specific product in an inventory or in a catalogue.

- *Serial Number:* This field is the specific instance of the object tagged. In other words, when an interrogator reads this object it knows exactly which unique object has been read and not only what type of object.

2 RFID market overview

2.1 Supply chain management

Due to the size of Wal-Mart and its 100 top suppliers, the media has been focusing on the use of RFID on the Supply Chain, but the reality is that it is early days and most of the RFID deployments for the supply chain are still in the pilot phase.

Europe has some interesting cases like the Metro Group, which is using RFID step by step throughout the entire process chain (http://www.metro-group.com.tr), and Marks & Spencer, who tested the ability to check stock deliveries and count stock quickly in stores using RFID tags. But in Scandinavia the picture is different. Most of the retail industry has just migrated to unified bar codes and are not expecting to do any investments for the use of RFID in the supply chain for the next three or four years.

Nevertheless, some niche areas could see the benefits of RFID, like fresh goods; e.g. monitoring the fish temperature as it goes through the supply chain. Such areas could give industries like retail a good opportunity to test the benefits of the RFID technology without having to make huge investments to include their whole line of products.

2.2 Asset management

There seems to be a common perception that there are other areas that will make use of the RFID technology before supply chain management, and the best candidate is Asset management.

By using active RFID tags companies have the possibility to better track and monitor their assets throughout a supply chain, in an enterprise, and dispersed into the field. This area has been commercial for some time now and due to the influence of Telecom Operators it got the name M2M or machine-to-machine.

From the point of view of RFID middleware providers like IBM and Oracle, RFID is only a small part of their M2M middleware offering, meaning that an M2M platform that is used to monitor the temperature of refrigerators can also be used to track goods going through the supply chain, for the middleware platform it is just a matter of handling different types of signals and giving them a business meaning.

The reason why asset management is more mature is because the cost of investment in technology in order to track expensive or important items is minimal compared to the value of loss or malfunction of such items, and this is the same reason why the supply chain is not yet mature, since the cost of monitoring a package of milk is much higher than the value that it brings. Some areas that could instantly benefit from M2M could be airports, oil platforms or any other industry that needs to track and monitor expensive assets (see Figure 1).

2.3 Workflow management

Workflow management is the area where the NFC and RFID technologies converge and is also gaining a lot of momentum. Mobile Workflow Management makes use of mobile phones enabled with RFID/NFC readers in order to automate information retrieval and data capture in the field. Some example applications are:

- *Service Information:* Get up to date service information by touching an item.
- *Automatic Search:* Touch a tag to automate extensive search strings.
- *Work order Generation:* Touch a tag attached to the machine and a new work order is available from a web page.
- *Failure Reports:* Touch a tag and get linked to a page to enter the reason for failure.
- *Materials Used Report:* Touch a tag and get linked to a page to enter the materials used.
- *Recording Travel Expenses:* Attach a tag to the dashboard of a car. Touch and enter beginning and ending mileage to have time stamped data for a travel expense form.
- *Time and Date Stamping (ex. Security):* Attach tags to sites that are visited by security guards. Get accurate time stamps and proof of work done.
- *Time and Attendance Recording:* For example, personnel could use personal tags to touch a single phone in the office, at a construction site or field office to clock in and out.
- *Automatic Meter Reading:* Touch a tag attached to a meter and the meter readings are recorded in the phone. Time stamp and tag information is available automatically.

The Norwegian company AD Columbi has been offering these types of services for some time using bar code scanners attached to mobile phones and is currently collaborating with Nokia for the usage of

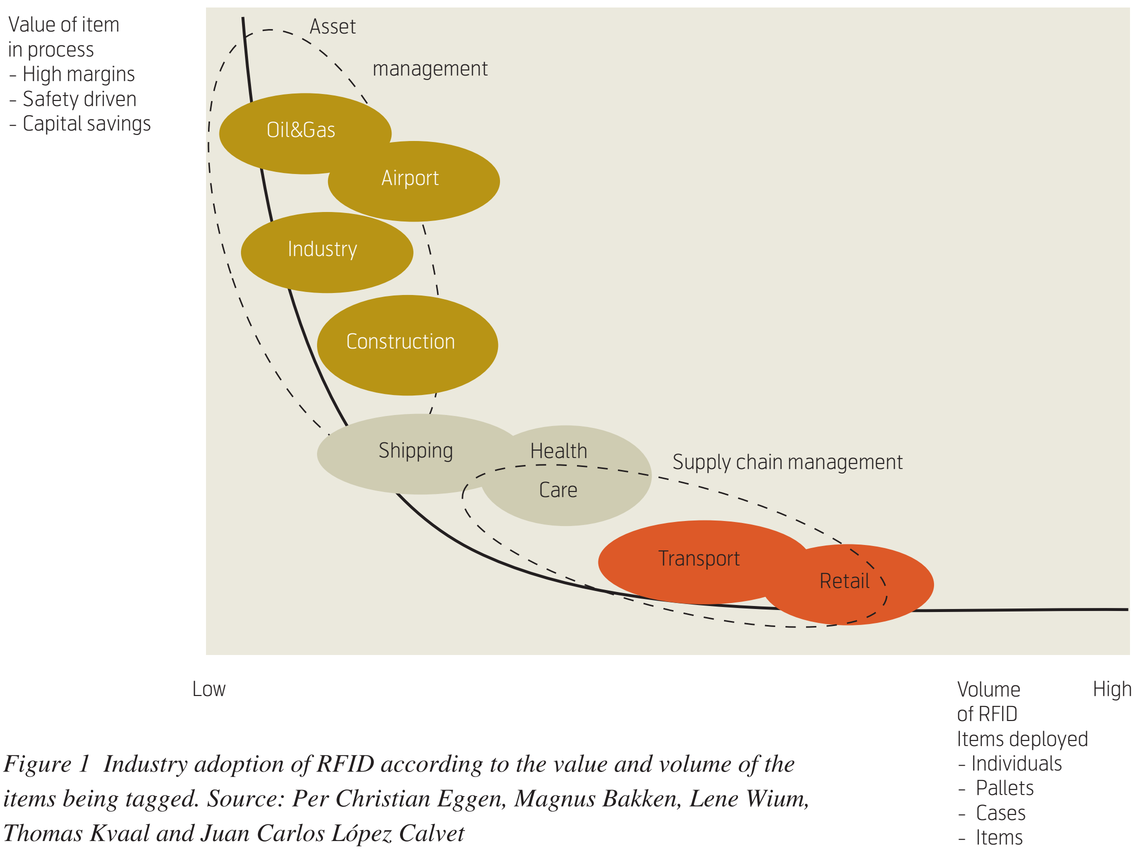


Figure 1 Industry adoption of RFID according to the value and volume of the items being tagged. Source: Per Christian Eggen, Magnus Bakken, Lene Wium, Thomas Kvaal and Juan Carlos López Calvet

ISSN 0085-7130 © Telenor ASA 2005

Global potential	Billion/year	RFID leadership
Library	0.1	Singapore
Museums, art galleries	0.1	Europe
Laundry	1	Europe
Tires	1	Europe
Military items	2	US
Blood	2	Europe/US
Test tubes	2	Europe/US
Archiving paperwork	2	US
Drugs	30	US
Books	50	Japan
Postal	650	Europe
Retail items	10,000	Europe/Japan/US

Table 1 Potential maximum yearly sales by 2020 or earlier for different types of item level RFID tagging [Source IDTechEx]

RFID/NFC enabled phones for customers like ISS Norge that uses the solution to inform their workforce about jobs to be done in a specific place by pushing the necessary information to the phone and register when the job is done by touching an RFID tag.

3 RFID in the future

3.1 EPC Generation 2

For most of 2004 there has been a lot of focus on the second-generation UHF standard of EPCGlobal[1], but there are only few people who really understand how this new generation differs from the first-generation EPC standards (Class 0 and Class1). The biggest difference between these two standards is that there is now a single global protocol instead of the two (Class 0 and Class 1) that EPC previously had. The implication of a unified protocol is that readers will be able to read tags regardless of where they are manufactured.

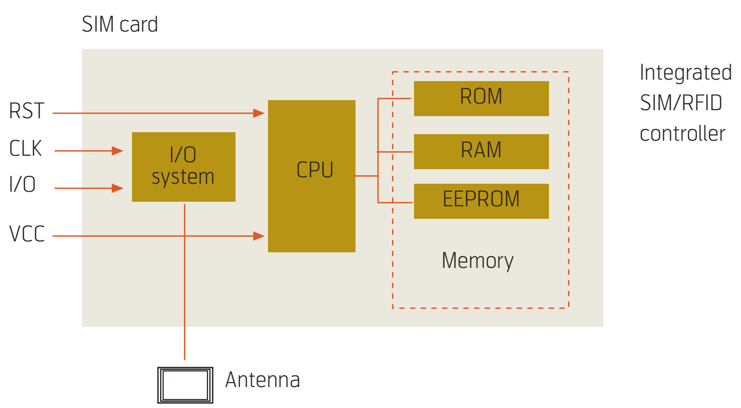


Figure 2 Sketch of an integrated RFID/SIM controller inside a SIM card

Another important aspect of the UHF Gen 2 protocol is that it uses the available radio spectrum more efficiently, providing a much better performance in Europe than any other UHF protocol.

Finally and probably the most important addition to the new standard is that a Gen 2 reader will now be able to read active tags with sensors, which moves RFID from a pure supply chain management approach to a more general asset managing approach which includes a wider range of markets.

3.2 The facts about product level tagging

Item level is a term that usually refers to fairly small items in high volume. Item level tagging offers far more benefits than case and pallet tagging. They include crime reduction, error prevention and brand enhancement as well as cost, service and so on. It is only with item level tagging that the consumer will clearly see benefits.

Others with their own priorities are proceeding very rapidly to trial and roll out item level tagging with excellent paybacks, usually employing the well proven 13.56 MHz frequency where the environment and production quality are less of a problem.

There are fresh food items tagged in Botswana and Japan, survival equipment in France and cigarettes, videos and even artificial logs in the US. Supermarket items have been tagged in the Philippines. Most are trials of course, but gas cylinders and beer barrels tagged in Denmark, the UK, France and elsewhere have long been full rollouts with excellent paybacks of one year or so. And this is a worldwide phenomenon. Their objectives vary greatly of course, and the list changes by the day. Those that quietly get on with item level tagging are increasingly reaping rewards.

From Table 1 it is possible to conclude that while tagging pallets does increase supply chain visibility, item level tagging satisfies specialized needs that bring direct benefits.

1) *EPCGlobal is leading the development of industry-driven standards for the Electronic Product Code (EPC) to support the use of Radio Frequency Identification (RFID).*

4 RFID and the SIM card

In March 2004 Telenor R&D filed a patent application for a SIM card with embedded RFID capabilities. The idea was to bring all the benefits of contactless cards into the mobile phone using the SIM card as a storage device. By doing so the mobile phone will be able to function as an electronic ticket, ID, key or credit card.

4.1 Merging RFID and GSM services

One of the main purposes of including the RFID technology into the SIM card was to be able to offer an alternative for the contactless card infrastructure being installed in different parts of the world (with the first beneficiary being contactless ticketing).

Including RFID into the mobile phone gives contactless cards the advantage of being able to be installed, updated and cancelled over the air through the services offered by GSM.

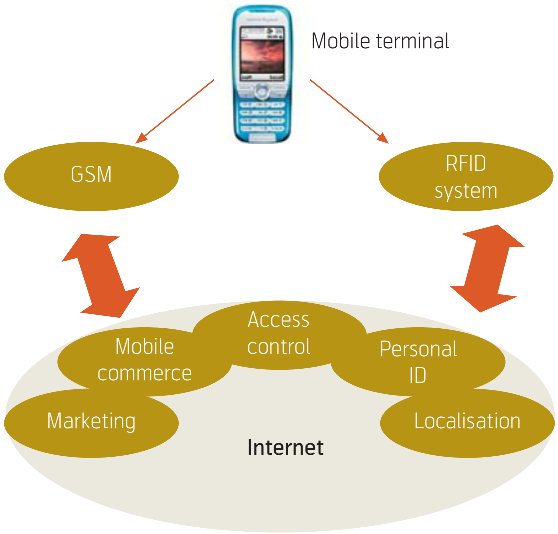


Figure 3 Merging of RFID and GSM services

4.2 The challenges of RFID in the SIM card

One of the biggest challenges of RFID in general is that it does not work that well when it is attached to metal or glass because the radio field gets distorted. Therefore, putting an RFID transponder in the heart of the mobile phone (the SIM card) is not an easy task, because the SIM card is usually placed under the battery and is surrounded by many different materials that disturb the RFID radio field.

In order to solve this challenge companies like Giesecke & Devrient made an adaptation of a Siemens mobile phone that connected an external antenna on the cover of the mobile phone to their dual interface SIM card (this type of SIM card has contactless capabilities, but need an external antenna).

Although solutions like this could offer many of the capabilities envisioned by Telenor R&D, we did not believe that the mass market was going to adopt solutions that were not fully integrated with the mobile phone and we did not believe either that mobile phone manufacturers would change their hardware designs to include an antenna for such SIM cards. Telenor R&D therefore opted to follow the NFC technology instead.

5 Near Field Communication (NFC)

5.1 NFC introduction

Near Field Communication is a short-range wireless technology optimized for communication between various devices without any user configuration. The goal of the NFC technology is to make two objects communicate in a simple and secure way just by putting them close to each other. NFC can also be used to start other communication technologies like Bluetooth and WiFi by exchanging the configuration and session data.

5.2 The NFC Forum

In April 2004 the NFC Forum was established by the three founding members Nokia, Philips and Sony, with the vision of bringing the physical world and the electronic world closer together. This was not a new vision, but was rather the Auto-ID lab's "Internet of Things" vision, which was brought to the consumer market. By placing tags and readers into objects, users would be able to interact with them by touching them with other NFC devices (e.g. mobile phones).

5.2.1 A winning team: Philips, Sony and Nokia

The initial founders of the NFC Forum saw the variety of opportunities that this technology would bring, but they also had an adequate background, which is why they were able to form a winning team.

Philips offerings are so broad that you could find their technology in almost any electronic device, but they provide two key elements into the NFC technology: The MIFARE standard, which is the most used contactless ticketing and payment standard in Europe with approximately 80 % of the market and the NFC chip itself.

Sony is also one of the biggest players in consumer electronics, but when it comes to the NFC technology they provide the FeliCa standard, which is the other biggest contactless ticketing and payment standard, mainly used in Asia.

 ISSN 0085-7130

And finally Nokia, who is one of the most dominant mobile phone manufacturers in the world. Because of their vast penetration of devices in the consumer market they are the perfect partner to bring this technology to the masses. The main reason why Nokia was able to bring this vision to reality before other mobile phone manufacturers is that they were already active in integrating the RFID technology on their mobile phones targeted at industry usage, such as mobile workforce management.

5.2.2 The NFC Forum grows quickly

From April to September 2005 another 44 members have joined the Forum from a wide variety of industries. There are four levels of membership, but the founding members remain in the same industries with the addition of VISA and Master Card, who together dominate the payment market.

Without having the insight on why these partners have joined the Forum, one could probably guess their motivation by grouping them into the following segments:

- *Mobile phone providers:* Nokia, Motorola, Samsung Electronics, Philips and Microsoft;

- *Semiconductor manufacturers:* Philips, NEC, Renasas and Texas Instruments;

- *Payment:* MasterCard International and Visa.

The fast growth of the Forum plus the relevance of the players involved give credibility to this emerging technology. Most of the players that Telenor has had contact with indicate that they are serious about their efforts towards the NFC technology and that we can expect devices to be launched with full support of their companies starting from 2006.

5.3 NFC applications

In order to emphasise the idea of close proximity, Philips (http://www.semiconductors.philips.com/markets/identification/products/nfc/index.html) has divided the NFC applications into the four following areas:

- *Touch and go*
  Applications such as access control or transport/event ticketing, where the user only needs to bring the device storing the ticket or access code close to the reader. Also for simple data capture applications, such as picking up an Internet URL from a smart label on a poster.

- *Touch and confirm*
  Applications such as mobile payment where the user has to confirm the interaction by entering a password or just accepting the transaction.

- *Touch and connect*
  Linking two NFC-enabled devices to enable peer-to-peer transfer of data such as downloading music, exchanging images or synchronizing address books.

- *Touch and explore*
  NFC devices may offer more than one possible function. The consumer will be able to explore a device's capabilities to find out which functionalities and services are offered.

5.4 NFC and the mobile phone

NFC adds intelligence and networking capabilities to the phone and creates many new opportunities to add product and service capabilities to the handset like digital transactions and sharing in very close proximities.

One of the main technological advantages of an NFC is that its chip can act as reader, but it can also act as a card, and is backward compatible with the contactless card standards. Therefore it makes a mobile phone an ideal device for making payments, since with a mobile phone you can control the interaction with the card (no one can read your card if you do not want them to), the mobile operator can update or cancel the card remotely and it is a device that users are already carrying.

When acting as a reader, an NFC mobile phone has the possibility to exchange data with other NFC devices, but most importantly it can trigger the download of content related to a specific object like a movie poster. For example: if a user walks by a movie poster, by just touching the poster it will trigger the browser, then the phone will automatically download information related to that movie, in what theatres it is being screened, and it will give the user the possibility to purchase tickets if desired.

5.5 The first NFC mobile phones are launched

On 9 February 2005 Nokia introduced the world's first Near Field Communication (NFC) product for payment and ticketing which was an enhanced version of the Nokia NFC shell for the Nokia 3220 mobile phone. It was the first mobile phone to deliver all the services envisioned by the NFC Forum, which are: service discovery, payment and ticketing.

5.5.1 Why the shell concept?

New technologies like NFC offer many exciting possibilities, but nowadays it is not easy to change a line of production just because a new technology is supposed to bring great benefits. In order to test the new capabilities of the NFC technology, Nokia decided to use one of their models in production (Nokia 3220) that had the capability of communicating with the mobile phone cover, and by reusing this phone they were able to create a shell or cover with an embedded NFC chipset.

Users or companies willing to test the NFC capabilities are now able to buy an off-the-shelf Nokia 3220 with an extra NFC shell. But such as solution is only expected to be used in the first pilots and once the technology is proven we can expect mobile phones with full NFC integration offered not just by Nokia, but other mobile phone manufacturers like Samsung, SonyEricsson and Philips by 2006.

5.5.2 Why is it taking so long?

The main reason for the delay of NFC mobile phones is standardization. The NFC technology involves mainly semiconductor, SIM and mobile phone manufacturers. And all of them have to create standard interfaces to be able to communicate with each other. Once the interfaces are standardized it usually takes one year for phone manufacturers to go from design to production of a new mobile phone.

5.5.3 The role of the SIM card for the NFC technology

With the introduction of the NFC technology into the mobile phone, the SIM card takes a more important role not just for telecom operators, but also for payment, ticketing and SIM card providers.

When NFC functions as a contactless card, it requires a place to store critical information such as ticket numbers, credit card accounts or ID information. This storage place could be basically anywhere in the mobile phone (RAM), but since the SIM card has storage capacity and already offers a high level of security, there seemed to be a common agreement between all the companies involved in the NFC Forum to use the SIM card for storage of critical and sensitive information.

The only problem is that communication between the NFC chip and the SIM card does not exist today and therefore has to be standardized. This is one of the main reasons why NFC mobile phones have not yet been launched. The communication between the SIM card and the NFC chip requires a high-speed transaction time in order to offer a real alternative for today's ticketing and payment system. Users would never accept a new ticketing solution that is not easier or faster than the already available solutions offered by contactless plastic cards.

5.5.4 First NFC pilot efforts

Rhein-Main Verkehrsverbund (RMV), the public transport network operator for Frankfurt's greater area in Germany, launched a mobile ticketing project in early 2005 in the city of Hanau.

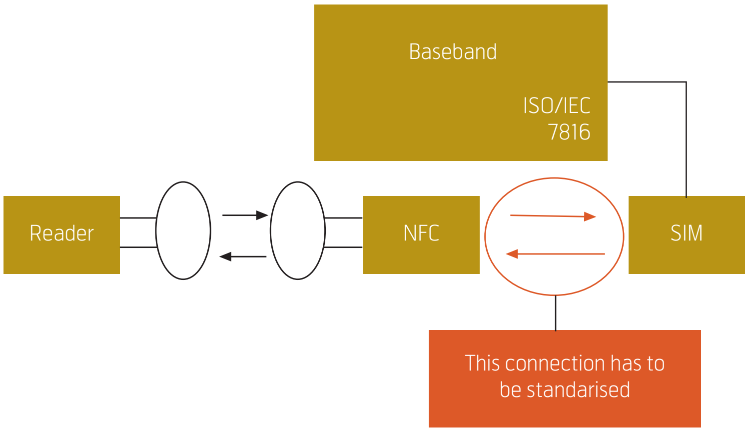


Figure 4 Using the SIM card as the storage unit for NFC transactions in a mobile phone

The project was the first live ticketing application based on NFC using mobile phones as a contactless ticket for an already existing ticketing infrastructure. The solution was tested in Hanau allowing customers to buy, store and use tickets with a Nokia 3220 NFC mobile phone.

Since the pilot used the first version of the NFC Nokia phones, it was not able to store the electronic ticketing information on the SIM card, but instead it used an integrated smart card controller in the phone.

The aim of the project is to provide partners (Nokia, Philips and RMV) with practical experience of NFC mobile phone ticketing that are compatible with already installed contactless smart card infrastructures.

Telenor R&D and Nokia are currently joining efforts to test the NFC technology together with the most relevant transport systems in Norway, since there is a national effort to migrate ticketing systems all around the country towards a contactless card infrastructure where travellers can use the same card in any kind of transport (train, ferries, subways and busses).

6 Market drivers for NFC

6.1 Contactless smart cards

Contactless smart cards have now reached a mature state with over 540 million cards on the market. There are two types of cards that have been dominating the arena: The FeliCa card developed by Sony and the MIFARE card developed by Philips based on the ISO 14443-A standard.

ISSN 0085-7130 © Telenor ASA 2005

ISO 14443 became a standard in 2001 and has two standard components (A and B). The cards come in different shapes and sizes, but they must comply with either A or B while readers and point-of-sale terminals must comply with both. The standard works on the 13.56 MHz frequency and several major transport systems in Asia have been in public service for more than a year. Other cities are in the process of switching to this new technology primarily to reduce maintenance costs, increase system reliability, security, provide greater convenience for users, and improve the processing speed, especially for busses.

Contactless smart cards can also be used to substitute magnetic strip cards for payment and authentication; they provide all the security features of smart cards plus the advantage of being contactless.

6.2 Ticketing

The main driver for contactless smart cards is that it has already an installed infrastructure of contactless ticketing systems around the world. Cities like Hong Kong and London were early adopters of this technology; Hong Kong with a solution called Octopus and London with a solution called Oyster. Most European countries are also migrating their ticketing systems towards contactless cards including Norway, which has a national effort to use one card that will be compatible with every transport ticketing system.

Not only are transport systems benefiting from the high-speed transaction time of contactless cards, but also event venues like football stadiums are applying this technology due to the drastic reduction of queues.

The football world cup arranged in Germany next year will bring massive attention from the media since it will only use contactless tickets in every venue. The tickets are personalized and unique to each person and cannot be transferred. Such a big event will trigger a wave of adoption of contactless ticketing systems around the world.

We can conclude that contactless ticketing systems are not just a hype, but are here to stay at least for the next ten years, and with such a big installed infrastructure the probability of using mobile phones as contactless tickets increases dramatically.

6.3 Service discovery

NFC service discovery is the concept of interacting with (tagged) objects in the real world through the use of an NFC enabled devices. The example most commonly used is that of a movie poster with an RFID tagged attached to it: when the user touches the poster with the NFC mobile phone, it can automatically download information about the movie, ringtones, logos or even purchase tickets. The main idea is to provide easy access to information related to the tagged object.

6.3.1 Industry will adopt NFC/RFID mobile phones first

This concept has many uses ranging from marketing – as the poster example – to workflow management for industry process automation. Nokia has clear plans to be a leader in this market by launching the first RFID/NFC mobile phone for the industry segment with their 5140 model.

The Nokia RFID phones are designed specifically for industry sectors ranging from security, services and utilities to health care and public administration. Some application examples include meter reading, maintenance and reporting of task completion.

Telenor R&D and the Norwegian company AD Columbi, who specializes in automation of processes like task completion through the use of mobile phones, have been cooperating in order to bring the first RFID pilots to the Norwegian market.

The main reason why the industry is more mature to adopt this technology is that companies specializing in the mobile workforce have a total control over which mobile terminals their mobile workers need in order to perform their tasks and are therefore not dependent on the penetration of this type of device in the market.

6.4 Payment

Contactless payment systems are penetrating the market just as quickly as the ticketing systems, and the most important success of such adoption is the agreement between credit card issuers to use a single RFID chip standard (ISO/IEC 14443). Since American Express, MasterCard and Visa are already rolling out their contactless payment cards for consumers, merchants are also moving quickly to upgrade their terminals to be able to handle RFID-based transactions. The standard agreed on by MasterCard, American Express and Visa is the same standard that is supported by NFC mobile phones and therefore fully compatible.

In contrast to ticketing systems, the US have been quicker to adopt payments through contactless cards, since they have a stronger credit card culture than Europe. MasterCard, Visa and American Express have been running pilots showing positive results for merchants, indicating that users who pay by contactless cards tend to purchase more items compared with consumers paying by cash.

Although the use of mobile phones in contactless payment systems could be a straightforward step technology-wise, there are some concerns surrounding privacy that must be addressed in order for the mass market to adopt this technology.

6.5 Privacy concerns

Users are increasingly more concerned about the privacy of information, and when it comes to RFID cards or phones that hold sensitive information such as bank accounts or credit card numbers, they are afraid that a thief will be able to read the card with a mobile RFID interrogator and steal all the data stored in the card. Such fears could be the biggest barrier for adoption of RFID based payments. Such concern might be due to all the negative media coverage that surrounded some of the early adoptions of RFID for supply chain management like Gillette, who had to cancel their RFID pilot due to a very aggressive movement towards the protection of privacy.

But the truth is that RFID payment devices offer a higher level of security than traditional payment cards. The ISO 14443, which is the standard adopted by credit card issuers (MasterCard, Visa and American Express) allows the account information in the card to be encrypted, giving each company the possibility to use a different encryption method and keys.

Unlike RFID for supply chain management, this standard is specified for very short-range communication (in the range of 10 cm) making it difficult to read from a long distance. Even if someone is able to read the data with a specialized long-range interrogator, the attacker will still have to crack the encryption algorithms to be able to make some sense of the data acquired. The concept behind using such a range is to guarantee that the card is read only when the user wants it to be read, just as it is today with contact or magnetic stripe cards.

When the cardholder waves the card by an RFID payment terminal, it turns the encrypted number into a digital signature, which is passed through the payment network and then decrypted to authorize the transaction. To further protect the account information, the digital signature changes each time a card is read. So even if a thief were to somehow access the digital signature, it could not be used to make another transaction.

The main challenge is therefore for card associations, banks and merchants to send a clear message to consumers that contactless payment systems are actually more secure than today's cards if they want consumers to comfortably adopt the technology.

7 Technology overview

7.1 RFID basics

RFID (Radio Frequency Identification) is a means of storing and retrieving data through electromagnetic transmission to an RF compatible integrated circuit. This RFID tag or integrated circuit is usually a single solid-state memory chip.

7.1.1 RFID system components

A micro controller or PC, a reader or interrogator, and a transponder or tag are the most basic components of an RFID system.

7.1.2 RFID readers

Readers or interrogators are used to identify objects by communicating with the tags using a wireless RF link. An interrogator is then used to "scan" the tag. The interrogator (controlled by some type of host computer) transmits an RF signal out to the RFID tag(s). In order for these tags to be read by the interrogator they must be presented in a defined RF area of saturation, known as an RF portal, or RF field of view. This RF signal first activates the RFID tag(s), and then interrogates each tag based on criteria received back to the interrogator from the first RF transmission.

7.1.3 RFID tags

RFID tags or transponders are basically a source chip with an antenna. When these tags are within the field of an RFID reader or interrogator they react by sending the information stored in their memory. This information could be an Electronic Product Code (EPC) number that gives the interrogator a unique identification number for that tag, meaning that the interrogator will not only know what type of object is in the field, but it will know specifically which unique object is in the field. There are two types of RFID tags, which mainly differ in tags that have a battery and tags that do not have a battery.

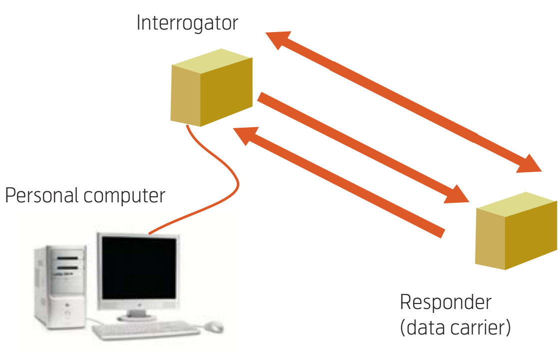


Figure 5 Basic elements of an RFID system

ISSN 0085-7130 © Telenor ASA 2005

7.1.3.1 Active RFID tags

Active tags have a source of power, usually in the form of a battery. The advantages of active tags are many, but the most obvious one is the ability to increase the range of readability through the use of an extra source of power.

Active tags can also be used to store status information like temperature, pressure, light, etc. This information can provide valuable information about the tagged objects since companies will be able to know not only where the object is at every step of the supply chain, but also how the object is handled over time.

The disadvantage of these types of tags is the price, which makes them affordable only when the value of the data received is bigger than the investment required in order get such data.

7.1.3.2 Passive RFID tags

Passive tags are tags without any source of power. These types of tags use the principle called inductive coupling to transfer the energy from one circuit (such as a conductive antenna and associated circuitry) to another by means of mutual inductance between the two circuits. It basically uses the energy provided by signals sent by the reader to send a signal back with the information stored on the tag.

These tags are much cheaper than active tags and are commonly used in pallets to increase visibility in the supply chain. Although the price of these tags is considerably lower, it has not yet reached the point where every single product can be tagged, but with increased adoption and economies of scale the price of tags is expected to reach the critical point for them to be used at item level.

7.1.4 RFID software

The last basic component in any RFID system is the software that controls the interrogators in order to synchronize when and where tags are read, filter the received data (since tags and readers communicate more than one time when they are within each other's field) and gives the data meaning for business processes.

7.2 NFC technical overview

7.2.1 ISO 14443

The ISO 14443 is the international standard for 13.56 MHz identification cards. This standard was initially designed for payment and ID cards and it was later modified to include contactless cards. This standard is most commonly used for payment and ticketing systems due to the higher security and faster communication that it provides. There are two types of ISO 14443 readers, A and B. They basically differ in speed, signal modulation, coding format and anti-collision method.

NFC is compatible with the broadly established infrastructure based on ISO 14443-A (Philips MIFARE technology), ISO 14443-B, as well as Sony's FeliCa card used for electronic ticketing in public transport and for payment applications.

NFC devices can operate in a reader mode that allows communication with a wide variety of contactless smart cards or RF transponders (tags). NFC devices can also work in a card emulation mode, which enables the NFC device to act as a smart card towards smart card readers, such as public transport and point of sale terminals.

7.2.2 NFC Philips transmission modules

Philips is currently providing two NFC transmission modules called pn511 and pn531, which are the modules that the first Nokia implementations are using. The following section is taken from the Short Form Specification of both modules, which can be downloaded from Philips website (http://www.semiconductors.philips.com/markets/identification/datasheets/).

The two chips provided by Philips (pn511 and pn531) are highly integrated transmission modules for contactless communication at 13.56 MHz. These transmission modules utilise an outstanding modulation and demodulation concept completely integrated for different kinds of passive contactless communication methods and protocols at 13.56 MHz.

The transmission modules support three different operating modes:

- Reader/writer mode for FeliCa and ISO 14443-A cards;
- Supports Card interface mode for FeliCa and ISO 14443-A/MIFARE in combination with secure μC;
- IP-1 mode.

Enabled in reader/writer mode the module internal transmitter part is able to drive a reader/writer antenna designed to communicate with ISO 14443-A /MIFARE or FeliCa cards and transponders without additional active circuitry.

The receiver part provides a robust and efficient implementation of a demodulation and decoding circuitry for signals from ISO 14443-A compatible cards and transponders. The digital part handles the complete ISO 14443-A framing and error detection

(Parity & CRC). The modules support contactless communication using MIFARE Higher Baud rates up to 424 kbit/s in both directions.

In the reader/writer mode the transmission modules support the FeliCa communication scheme. The receiver part provides a robust and efficient implementation of the demodulation and decoding circuitry for FeliCa coded signals. The digital part handles the FeliCa framing and error detection like CRC. The modules support contactless communication using FeliCa Higher Baud rates up to 424 kbit/s in both directions.

Enabled in card mode the transmission modules are able to answer to a reader/writer command either in FeliCa or ISO 14443-A/MIFARE® card mode. The modules generate the digital load modulated signals and in addition with an external circuit the answers can be sent back to the reader/writer.

Additionally the transmission modules offer the possibility to communicate directly to several NFC enabled devices in the NFC IP-1 mode. The NFC mode offers different communication baud rates up to 424 kbit/s. The digital part handles the complete NFC framing and error detection.

8 Strategy

8.1 Operators

The SIM card has always been the most important asset of GSM mobile operators, but regardless of many efforts to increase its value towards the user, operators have not been able to reap the benefits of having such an important asset in the heart of every GSM mobile phone.

The introduction of the NFC technology brings the unique opportunity to easily connect the SIM card with the physical world. All the visions of mobile payments can finally be realized without cumbersome WAP menus or slow SIM toolkit applications.

With the adoption of NFC, the SIM card will increase its value dramatically by storing the end user's most valuable information: credit card numbers, bank account numbers, personal IDs, plane tickets, bus tickets, bonus cards, etc. This gives operators the unique opportunity to offer a "real state" type of business to third parties such as credit card issuers, banks, transport companies, etc.

It is a win-win situation between mobile operators and third parties that offer some kind of contactless card services, since operators increase the value of their SIM cards and implicitly their relation with their customers; and third parties reduce their administration costs of delivery, storage, security and management of contactless cards.

8.2 Mobile phone manufacturers

For mobile phone manufacturers the main reason to adopt the NFC technology is the increased overall value of mobile phones. Mobile phones are already used by almost everyone throughout developed countries, but with the integration of NFC the mobile phone cam become the single most valuable asset for consumers, since it will not only be used as a means of communication but can potentially replace wallets, keys, tickets and IDs.

8.3 Semiconductor manufacturers

For semiconductor manufacturers like Philips it is just a matter of numbers. The mobile phone is by far the most purchased consumer electronic product, and the interest of mobile phone manufacturers towards this technology creates a big market for those in the NFC semiconductor business. The inclusion of the NFC technology in other consumer electronics like TVs, radios, MP3 players, etc. will also increase the opportunities for semiconductor manufacturers to sell their products.

8.4 SIM manufacturers

SIM card manufacturers, together with mobile operators have constantly been trying to increase the value of the SIM card, but the lack of cooperation with mobile phone manufacturers has rendered most of the efforts fruitless, since there is no guarantee that SIM toolkit applications developed by operators or SIM manufacturers will work in every single mobile phone on the market.

The value of the SIM decreased so much that questions were raised on the real need to include the SIM in GSM mobile phones. But with so much momentum and support from all the different industries interested in the introduction of the NFC technology the SIM card has guaranteed its place in the mobile phone as the security storage unit for sensitive information for the NFC technology.

8.5 End customers

The most important aspect of the adoption of the NFC technology in the mass market is the security concerns that have been raised by the hype of RFID in general. If the industry is able to manage such concerns by sending a clear and consistent message regarding the security advantages of the NFC technology, end users will surely adopt the technology due to the benefits that the technology can bring to their everyday life.

ISSN 0085-7130 © Telenor ASA 2005

References

1 ABI research. *Near Field Communications*. Published Q3 2004. http://www.abiresearch.com

2 Forrester Collection. *RFID: The Complete Guide*. Spring 2005. http://www.forrester.com

3 Philips Semiconductors. *Near Field Communication PN511-Transmission module (Short Form Specification Revision 2.0)*. February 2004.

4 Philips Semiconductors. *Near Field Communication PN531-mC based Transmission module (Objective Short Form Specification Revision 2.0)*. February 2004

5 *Potential of the RFID technology for Telecom Operators*. Eurescom Project Report, January 2004.

6 *Guide to Smart Card Technology*. Cyberspace Center, Hong Kong University of Science and Technology, November 1998.

7 Wal-Mart Expands RFID Market. *RFID Journal*, 18 Aug 2003. http://www.rfidjournal.com/article/articleview/539/1/26/

8 Wal-Mart Draws Line in the Sand. *RFID Journal*, 11 June 2003. http://www.rfidjournal.com/article/view/462/1/1/

9 U.S. Military to Issue RFID Mandate. *RFID Journal*, 15 Sept 2003, http://www.rfidjournal.com/article/articleview/576/1/1/

10 The Wal-Mart Factor. *RFID Journal*, 17 March 2003. http://www.rfidjournal.com/article/articleview/346/1/27/

11 *Sony Ericsson M2M wireless communications*. http://www.sonyericsson.com/images/spgc/CWS31AFW_12781_36_0_4002.pdf

12 *Nokia Field Force Solutions*. http://www.nokia.com/nokia/0,,55738,00.html

13 *Near Field Communication Forum Launches Four Technical Working Groups*, 28 June 2005. http://www.nfc-forum.com

14 Item level RFID gathers pace. *IDTechEx*, 7 Feb 2005. http://www.idtechex.com/products/en/articles/00000140.asp

15 *NFC Technology*. Philips, 31 Oct 2005. http://www.semiconductors.philips.com/markets/identification/products/nfc/technology/index.html

16 Giesecke & Devrient. *(U)SIM card from G&D combines mobile and contactless technologies*. 31 Oct 2005. http://www.gi-de.com/portal/page?_pageid=44,94279&_dad=portal&_schema=PORTAL

Juan Carlos López Calvet is Research Manager in Telenor R&D. He studied Telematics at ITAM in Mexico City with a specialization on Elliptic Curves Cryptography. During his work as researcher at Telenor R&D he has filed a total of six patents in different areas ranging from security to e-commerce. He has won two second places in Telenor's business idea competition "Seed" and currently works on bringing emerging mobile technologies to the market.

email: juan.calvet@telenor.com

ISSN 0085-7130 © Telenor ASA 2005

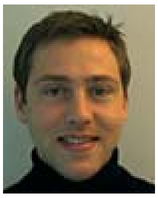
Future management of mobile phones

DO VAN THANH, ANNE MARTE HJEMÅS, ANDERS BJURGÅRD AND ØYSTEIN CHRISTIAN LØKEN

Do van Thanh is Senior Research Scientist at Telenor R&D

Anne Marte Hjemås is Research Scientist at Telenor R&D

Anders Bjurgård is Head of Business Logic & Terminals in Telenor Nordic



Øystein Chr. Løken is Product Manager for Device Management in Telenor Nordic

The increasing diversity and complexity of mobile phones call for an efficient Device Management System. Unfortunately, although there are currently quite a lot of effort and activities around device management, there is not yet a satisfactory one. This paper presents a vision of a future device management system elaborated by Telenor Nordic Mobile. It starts with an overview of the standards related to device management. Before describing the future Device Management System, the limitations of current systems are presented. Last but not least, the architecture of a future Device Management System is introduced and explained.

1 Introduction

Today's mobile phones are far more advanced than only a few years ago. The services that are available are not just plain old telephony anymore but rather complex and diverse. The user is faced with a broad range of devices and settings, and as the mobile phone converges to become a mini PC, the user will surely be overwhelmed by options, application settings, configurations, and services. He/she will be reluctant to install new services when there were problems with the first one. More advanced and open devices often mean more bugs, and an increasing risk of attacks by malicious viruses. However, the user is not the only player having problems. The mobile operator is also affected. Indeed, more problems also mean more pressure on the mobile operator's customer service. Due to the variety of devices and applications, the training of call centre agents could be a problem. On the other hand, the agents cannot expect the user to know what version of the operating system they have on their device. Some users may not even know what their device manufacturer or model is. Trying to figure this out before starting to solve the problem takes time and frustration. Many challenges are coming, and being able to guide the users through them will be an enormous added value.

The ideal solution to all the problems described above is to have a system that can perform remote and automatic configuration and management of the mobile phones with insignificant user's effort and minimal intervention from customer service. Such a system is called Device Management System in the mobile communication jargon. Unfortunately, although there are currently quite a lot of effort and activities around device management, there is not yet a satisfactory one. This paper presents a vision of a future device management system elaborated by Telenor Nordic Mobile [13]. It starts with an overview of the standards related to device management. Before describing the future Device Management System, the limitations of current systems are presented. Last but not least, the architecture of a future Device Management System is introduced and explained.

2 Standards for device management

There are many different protocols and technology for device management that support different functionality, and there will probably always be more than one protocol, but most standardization bodies and vendors look at SyncML DM as the most promising protocol that all other protocols merge into. OMA's SyncML DM supports all the functionality needed within 3–5 years, and is flexible for change. SyncML will be considered in more details in the following sections.

2.1 SyncML

"SyncML is a new industry initiative to develop and promote a single, common data synchronization protocol that can be used industry-wide. Driving the initiative are Ericsson, IBM, Lotus, Motorola, Nokia, Palm Inc., Psion and Starfish Software. The SyncML initiative is supported by several hundred leading companies in the industry, and additional companies are welcome to join the open initiative and participate." [1]. SyncML joined the Open Mobile Alliance (OMA) [2] in 2002 and the SyncML protocol became hence an OMA standard.

Figure 1 gives an overview of the SyncML Protocol Architecture. SyncML provides a common interface for synchronization that supports several different transport protocols and physical media. This will help achieve the goal that any mobile device should be able to synchronize with any data repository. The SyncML Protocol does not dictate the internal representation of data, but rather how to exchange data in an interoperable way. The protocol is designed to be extensible, allowing developers the freedom to define new data formats or protocol primitives.

 ISSN 0085-7130 © Telenor ASA 2005

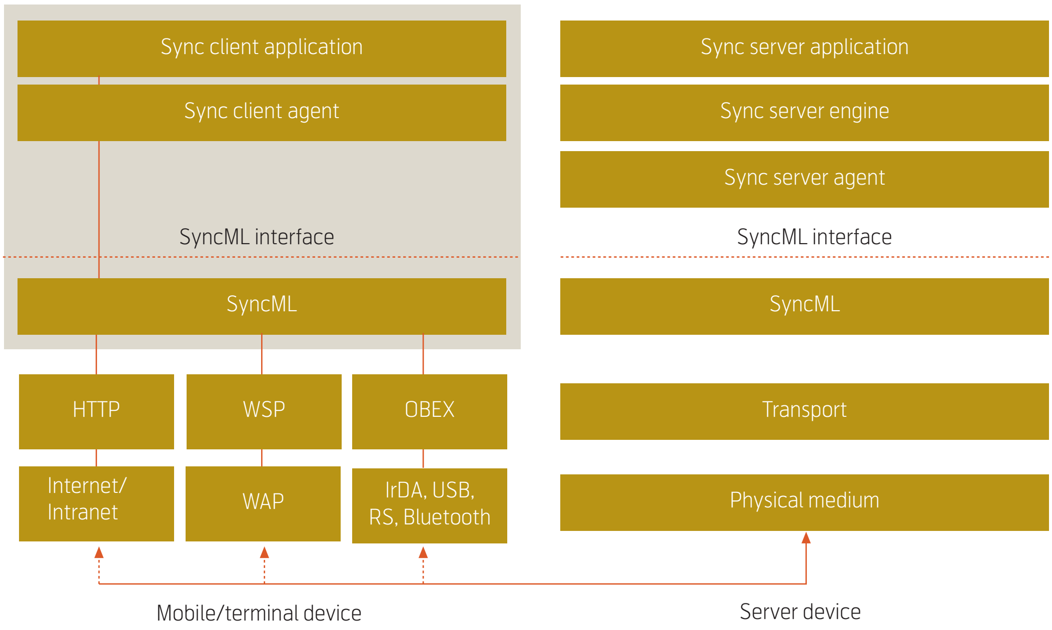


Figure 1 SyncML Protocol Architecture [2]

SyncML is based on XML and provides an XML DTD allowing representation of the information required for device management and data synchronization. This provides the means to represent commands, data and also metadata. The Device Management and Data Synchronization protocols describe how to exchange messages conforming to the DTD.

Being tailored for mobile devices with potentially low bandwidth, a binary encoding is defined. This encoding is based on work by the WAP Forum; WBXML (WAP Binary XML).

The SyncML specification contains the following main components:

- An XML-based representation protocol;
- A synchronization protocol (SyncML DS) and a device management protocol (SyncML DM);
- Transport bindings for the protocol;
- A device description framework for device management. [6]

2.2 SyncML DM

The Device Management Working Group continues the work done by the SyncML initiative and WAP Forum concerning Device Management (DM). Of most importance here is the SyncML Device Management Protocol.

The SyncML Device Management Protocol [3] is a management protocol that uses the SyncML Representation Protocol. Uniquely addressable management objects are used to reflect the various aspects of device settings and run-time environment. Device settings may be read or written using commands that operate on the management objects.

Figure 2 gives an overview of the two phases of the SyncML DM Protocol – authentication and management. Authentication may be initiated by the user himself, or possibly by some form of notification sent from the server. The SyncML DM Protocol specifies several notification methods. Either way, the authen-

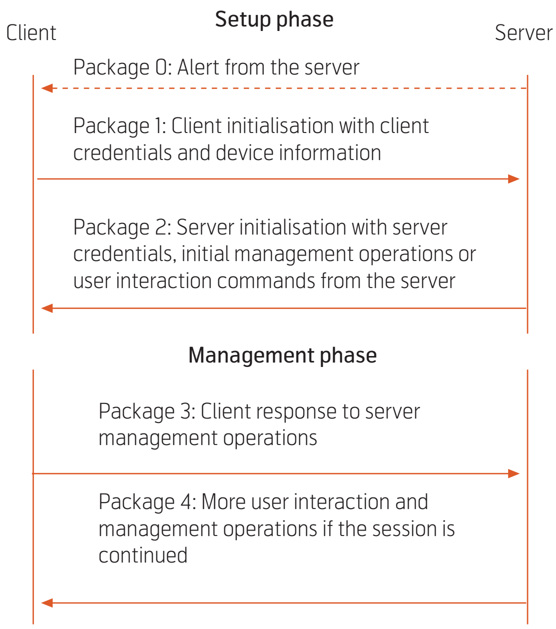


Figure 2 SyncML DM Protocol Parts [4]

ISSN 0085-7130 © Telenor ASA 2005

tication phase ensures that the client and server can establish a trusted relationship. Also, the authentication reply, Package 2, from the server may contain management operations to be carried out by the client and may end the device management session if no further communication is needed. One or more iterations of the management phase then follow. [4]

SyncML Device Management Bootstrap is the initial process in which an un-provisioned device is provided with the necessary settings to start the ordinary device management process described above. SyncML DM Bootstrap is not intended to be used for other purposes.

Bootstrapping can be done either as a customized bootstrap or a server initiated bootstrap. Figure 3 gives an overview of customized bootstrap. The device is provided at the factory with the necessary information required to engage in a management session. This is not very flexible, and requires that the manufacturer knows details about the operator's network.

Figure 4 shows a server-initiated bootstrap. By way of automatic detection, the server initiates the management session.

SyncML DM promises to simplify device management by providing a common protocol across different devices. Of course, settings for each device will still be determined by its capabilities. Mobile phone manufacturers can expose any type of device settings, including runtime environment. This will provide the capability to upgrade software or firmware and provide bug fixes.

SyncML DM v1.1.2 became an approved enabler release in January 2004. An approved enabler release has passed the candidate status and has associated interoperability test cases generated by OMA.

An example from [6] provides a more hands-on feeling of how the protocol appears (Figure 5 and Figure 6).

The protocol contains necessary authentication support, alerts, error codes, etc. For those interested, [6] is a recommended document.

2.3 SyncML DS

The Data Synchronization Working Group continues the work done by the SyncML Initiative regarding Data Syncronization, DS. The SyncML Data Synchronization protocol specifies how the SyncML Representation Protocol is to be used for synchronization of data.

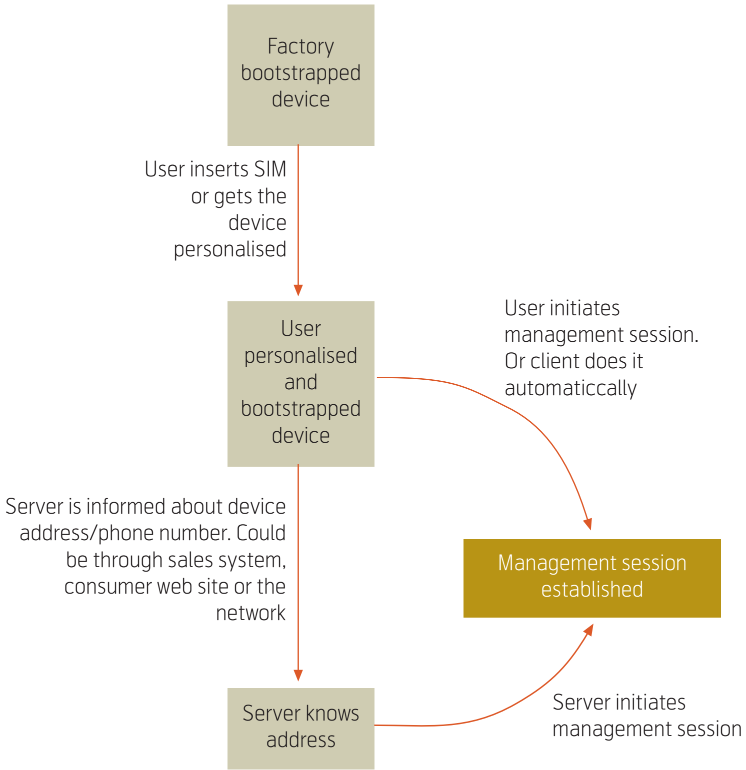


Figure 3 Customized bootstrap [5]

Synchronization of data between two devices requires that a change log be kept. This allows the server and client to figure out which changes to the data are most recent. SyncML DS does not pose any restric-

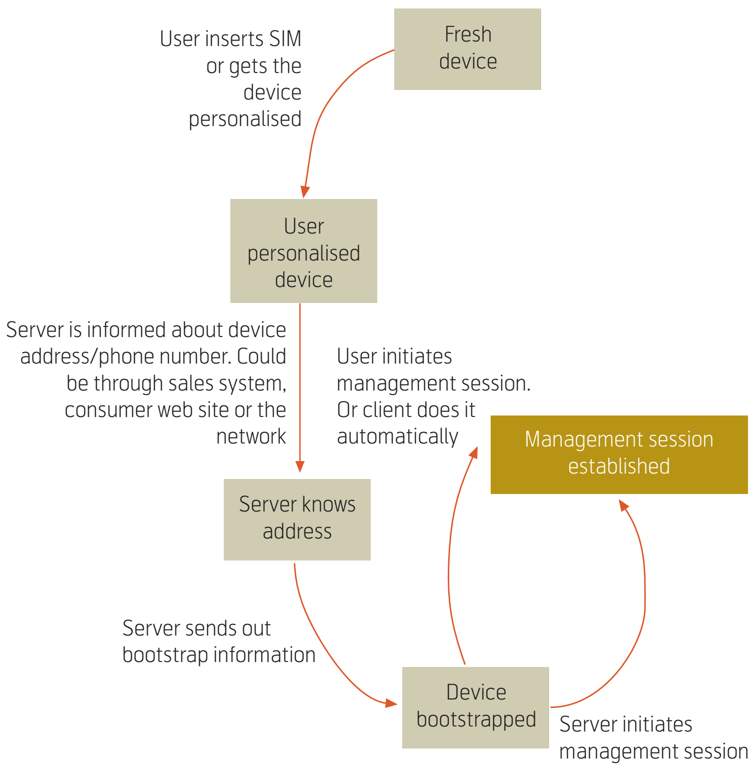


Figure 4 Server initiated bootstrap [5]

ISSN 0085-7130 © Telenor ASA 2005

```
<SyncML xmlns='SYNCML:SYNCML1.1'>
    <SyncHdr>
        ...blah, blah...
    </SyncHdr>
    <SyncBody>
        <Status>
            <MsgRef>1</MsgRef>
            <CmdID>1</CmdID>
            <CmdRef>0</CmdRef>
            <Cmd>SyncHdr</Cmd>
            <Data>200</Data>
        </Status>
        <Alert>
            <CmdID>2</CmdID>
            <Data>1100</Data> <!-- User displayable notification -->
            <Item></Item>
            <Item>
                <Data>Your antivirus software is being updated.</Data>
            </Item>
        </Alert>
        <Get>
            <CmdID>3</CmdID>
            <Item>
                <Target>
                    <LocURI>./antivirus_data/version</LocURI>
                </Target>
            </Item>
        </Get>
        <Final/>
    </SyncBody>
</SyncML>
```

Figure 5 An example of initiation of an anti-virus update

```
<Add>
    <CmdID>2</CmdID>
    <Meta>
        <Format xmlns="syncml:metinf">b64</Format>
        <Type xmlns="syncml:metinf">
            application/antivirus-inc.virusdef
        </Type>
    </Meta>
    <Item>
        <Meta>
            <Size xmlns='syncml:metinf'>37214</Size>
        </Meta>
        <Target><LocURI>./antivirus data</LocURI></Target>
        <Data>
            <!–Base64-coded antivirus file -->
        </Data>
    </Item>
</Add>
```

Figure 6 An example of adding an anti-virus software update

tions on how this change log is implemented, as long as the devices are able to specify which data items have changed. Figure 7 shows the basics of data synchronization. The device, in this case a mobile phone, will send a record of all changes since the previous synchronization took place. The server will then check this record against its own data and decide which data to keep, change or delete. The server will then send a reply to the client with its own set of modified data. [7][8]

The advantage of SyncML DS is that it allows synchronization of any data that the device manufacturer decides to expose, using a common framework. Also the most important mobile phone manufacturers are supporting and implementing this protocol.

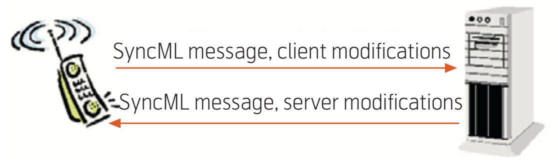


Figure 7 Synchronizing mobile client with server [8]

3 Current Device Management Systems

Current Device Management Systems are still rather primitive and limited. Some operators have a deal with some distributors to pre-configure devices with MMS, WAP and Internet. The distributors get a provision for a successful configuration. The end users can obtain the settings themselves via WEB, SMS, WAP or IVR solutions. In most cases the users need to specify their phone type and also their phone number in some cases. The SMS interface is used a lot, and the Web interface has even higher and more increasing usage than the SMS interface. They can also call customer service, but the process is no different except that the call centre agent performs the same operation. In some cases the agents have to ask the customers what manufacturer and device model they have. They will then receive an Over-The-Air (OTA) message that sets up the device. For technical reasons, they have to do this for each setting: MMS, WAP and Internet.

The OTA management functions offered today are quite often limited setups of CSD, GPRS, MMS, ISP, Bookmarks and email settings, and SyncML DS.

OTA can be requested in the following ways:

- SMS sent by customer

- Mobile operatorWEB-sites: for example telenormobil.no, djuice.no, djuice.se, telenormobile.se

- Mobile operator WAP-site: for example wap.telenormobil.no

4 Future Device Management System

4.1 Players of the Future Device Management System

In addition to the customer who is the end-user, a future Device Management System will interact with several other important players that have different needs and expectations. The identified players for the future DMS are as follows:

- *Customer/device user:* The customer is the service subscriber and device user. This should be one of the most important users to consider. A future DMS will satisfy many customer needs. The customer wants:
  - Services to perform flawlessly from day one with minimum interaction
  - Backup and synchronization of personal data
  - Device independent operation. The services follow the user, not the device
  - A faultless and safe device
  - Customized and personalized devices and services

- *Corporation:* A corporation is also a very important customer. A corporation will gain in addition to the customer:
  - Improved management of mobile devices
  - Bulk configuration
  - Presence information
  - Branding and corporation menus

- *Operator:* The operator provides the network and its resources. The operator can also host and administer the solution. Example: Telenor Nordic Mobile and NetCom. The operator wants:
  - One solution for all devices
  - Enhanced service level and several administration levels adapted to customer service, net division, DMS administrator and corporate users
  - Increased service uptake
  - Decreased customer care costs

- *Service Provider:* The Service Provider develops and provides value added services, and is responsible for the provision and continuous availability of subscribed services. Responsibilities are management, administration and operation of services. The service provider relies on the operator by means of the physical network and resources. The needs are much the same as those of the Operator.

- *Content Provider:* Also called third parties. The content provider in this case delivers services through the service provider. It could be Java games or third party applications. Content providers rely on the service provider for provisioning and billing. Benefits are:
  - A standardized tool for application configuration
  - Faster content delivery
  - Improved content adaptation

- *Device Manufacturer:* The manufacturer produces the actual devices and has knowledge of terminal capabilities, known bugs, updates, etc. The manufacturers have many benefits of a standardized Device Management.

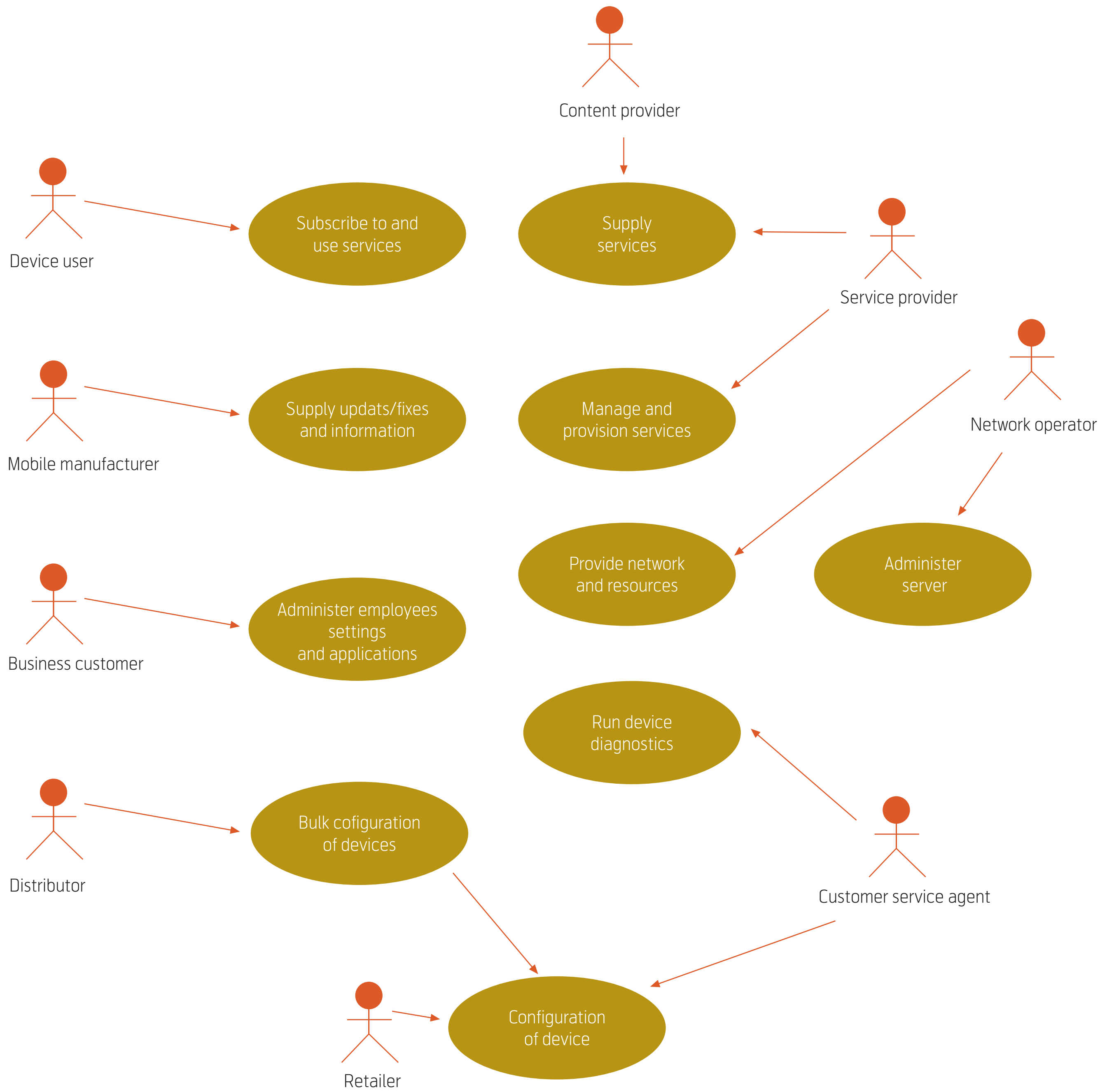


Figure 8 Players and business use cases

- Only one solution to support in the devices
- No need to call back devices with firm/software errors, can perform remote updates

- *Customer Service Agent/Operator's Customer Service:* A customer Service Agent talks to the device users, provides information and helps them solve their problems. Benefits are:
  - Instant knowledge of device characteristics
  - Quicker service
  - Better tools to aid customer

- *Retailer:* The retailer sells the devices to the end user. They can pre-configure the phone for the user and perform repairs etc. Benefits:
  - Easier configuration

- *Distributor:* The distributor can have a deal with a Network Operator to pre-configure a volume of devices. Benefits:
  - Bulk configuration

All the players need a user interface to the system. The different players have different access. Some players have access to similar functionality, but access to a subset of information. We prioritize cus-

ISSN 0085-7130 © Telenor ASA 2005

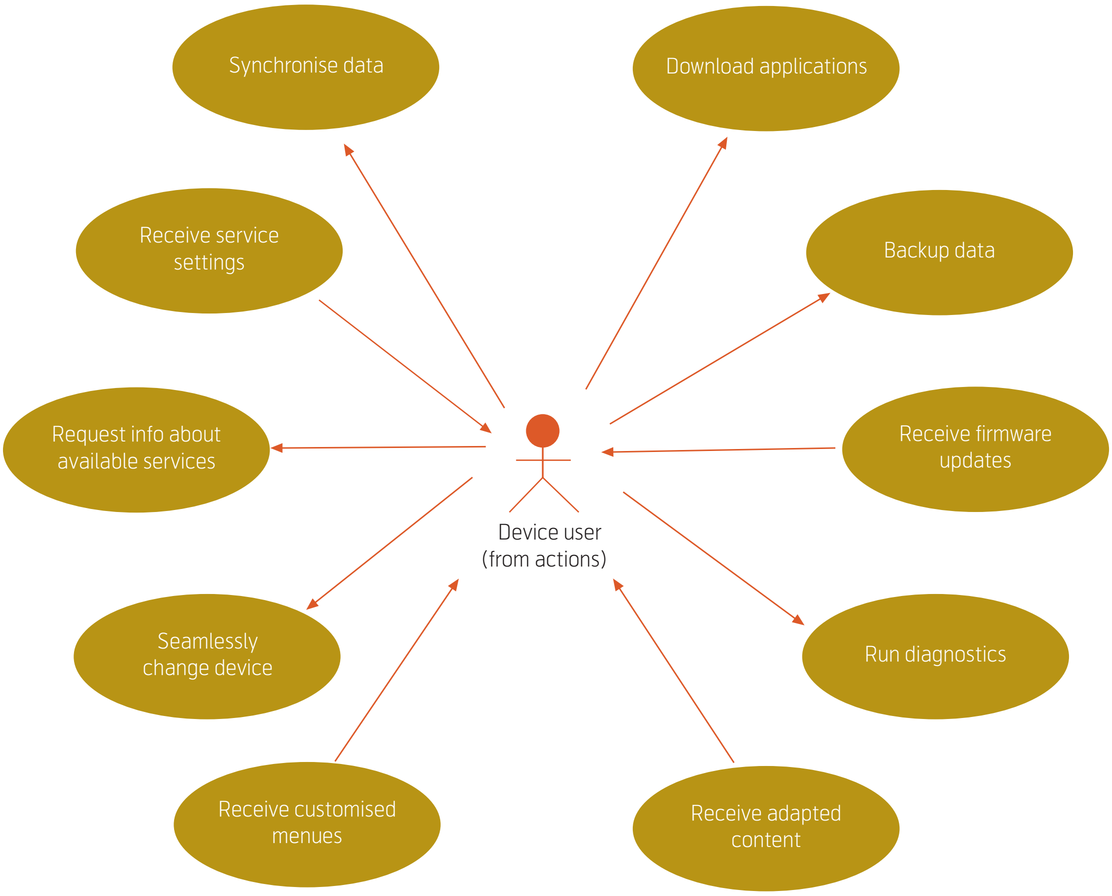


Figure 9 Use cases for device user

tomer angle, but the other players are involved in many of the use cases.

Figure 9 shows some use cases for the device user. "Seamlessly change device" means that the data and services follow the user, not the device.

4.2 Features of the Device Management System

There are four features that we will probably see in most Device Management Systems and devices within three years. The features present in the Device Management System are:

- Remote Configuration
- Software and Firmware updates
- Mobile Diagnostics
- User Data Management
- Personalisation

4.2.1 Remote Configuration

Remote Configuration includes provisioning of key parameters to configure the handset and activate the service remotely, setting initial configuration information in devices, and modifying or reading operator parameters. The interaction with the user can be fully automatic, with user input, or through a session. How automatic the configuration is, should be based on the nature of the data. If it is configuration for basic services like MMS, it should be automatic. For services that have personal configuration parameters, there may be a need for some kind of user interaction.

Bulk configuration must also be enabled and is a useful operation that can customize groups of devices, for example updating the electronic payment software package on the SIM card of all prepaid customers. An administrator of the Device Management system can manage branded services, menus and other UI customization remotely. If the service requires certain software or firmware updates, we rely on the functionality in the next section. By providing automatic download of service menu for specific services to the phone, the usage and user experience of the services will most likely increase. Users with advanced WPABX solutions can then easily use functionality like call transfer, add/remove themselves from hunting groups etc. without having to remember cryptic key codes.

4.2.2 Software and firmware updates

Device management can also perform software operations, i.e. upgrading software packages or installing new applications on the terminal to provide for new services. This includes both native and Java applications. Downloading software will be more attractive when the future high-speed connections are available. There will also be remote software management where the functionality for the Administrator will be [12]:

- Application and component inventory
- Install, delete, update, check, stop, kill, start application and/or component
- Remote reboot.

All with the proper user consent and authorization, of course.

Software operations also include installation of software patches/bug fixes for operational improvement, i.e. firmware. Firmware is programming that is inserted into programmable read-only memory, thus becoming a permanent part of a computing device. Firmware is created and tested like software. When ready, it can be distributed like other software and, using a special user interface, installed in the programmable read-only memory by the user. Firmware is sometimes distributed for printers, modems, and other computer devices [9]. In short: computer programme instructions incorporated in a hardware device that cannot be changed under computer program control are defined as firmware (e.g. the operating system of a device).

Updating firmware can improve performance, fix bugs or add functionality. Many still believe that the mobile phone is secure from viruses and worms, but there are already security holes in some devices where worms can spread via Bluetooth, and SMS can be sent without the user noticing in Java applications [12]. Firmware updates are delivered as binary packages replacing the malfunctioning part of the software image. The mobile manufacturer delivers the firmware update packages to the operator, which distributes the updates to the relevant handsets. The updates include a security voucher from the mobile manufacturer ensuring secure handling.

4.2.3 Mobile diagnostics

It is important to be able to read as much information as possible from a device, e.g. service settings, device model, a list of installed or running software, version information, hardware configurations, etc. A service can then be adapted to device capabilities and user preferences. The system can also listen for alerts sent from a device, and invoke local diagnostics on it. The user will be notified of the diagnostic session and can give permanent or temporary rights for certain administration actions. The user can also revoke the rights. The possibility of automatic detection is also sorted under this functionality. Automatic detection means that the network detects the (new) combination of device and SIM-card as soon as the terminal is turned on or comes within range. This feature is very important for many of the user scenarios that are discussed later.

4.2.4 User data management

User data management includes backup and synchronization of user data, system and service settings etc. Standards like SyncML DS, for synchronization of mobile terminals, has been in the works for many years, and is already included in the new, advanced devices. The user can now synchronize data anywhere, anytime. The backup could be either content aware or content unaware. No content awareness means that a raw, binary image is copied to a server. The data cannot be restored to another device model. Content aware backup means that the user can select what will be backed up and the data can be restored to another device model. Data like calendar, contacts etc. could be synchronized to a USIM card so that the possible bad consequences of borrowing a phone are reduced. The customer only needs access and the correct settings to a reliable server.

4.2.5 Personalization

If the laws and regulations allow it, the service provider could gather information about the user and use personalization to offer services that the user might be interested in. The service offer could be an extra item in the operator specific menu; either a concrete offer or a menu option that fetches new service offerings. Personalization uses information that is either obtained from or provided by the user. A good example is the online book store Amazon.com which now has expanded to selling DVDs, toys and much more. The site gathers information about the user based on what the user buys, views or searches for. Based on that information and what other users with similar preferences have bought, the user will be presented with offers that the personalization engine has predicted to be interesting. The recommendations, based on a complex set of rules, are often quite accurate.

This could also apply to telecom services. If Nick has ticked that he is interested in subject A, or has downloaded Application A, he could receive recommendations about similar applications. Or maybe many users that have downloaded Application A also have downloaded Application B. The system could then recommend Application B to Nick.

4.3 Architecture of a future Device Management System

The architecture of a future Device Management System is shown in Figure 10.

DMS server

The DMS server contains the actual business logic for Device Management. The DMS server's main responsibility is to keep track of existing terminals and their general attributes, OTA standards (proprietary and SyncML DM), handle requests and produce statistics. All non-OTA interfaces to external systems should be standardized and controlled/owned by the operator. Device Management is developing fast, and it is vital that it is possible to change vendors or at least keep them on their toes. When detecting a new IMEI-MSISDN combination the DMS should add or update an entry in the Subscriber-Terminal DB, perform a mobile diagnostic and register and validate settings etc. on the device. If any updates are necessary an update session will be initiated towards the device with or without user interaction.

There are also advantages of having the terminal database outside the DMS server. The operator has more control, but the vendor should have an interface to the database in order to keep it updated. This would probably also mean more server customization for the vendors if the operator wishes to use another server.

Having all device related information and operations in one place fits the CSF-philosophy. Using the DMS for all device operations fits the LEGO principle. The question is how many LEGO parts the system should be made of? The terminal database could be one separate self-serviced LEGO-part, but it is an integral part of the DMS and the data is not operator specific. Regardless, the other systems should only use the DMS interface to obtain device information without having to know whether the database is in the server or in a separate LEGO part used by the server.

Storage of IMEI in the HLR / automatic detection

An infrastructure in the network for detecting SIM-IMEI combinations must be present. Several vendors offer a solution inspired by Vodafone's 3GPP-proposal to store the IMEI in the HLR. This is a simple and elegant solution for simple applications, but increasing the subscriber profile size too much would decrease the number of subscribers to be served by an HLR. HLRs are expensive and difficult to update. They are system critical components, and changing their interfaces is a long process through standardization forums. However, storing only an IMEI number, or even just the TAC (Type Allocation Code) is suffi-

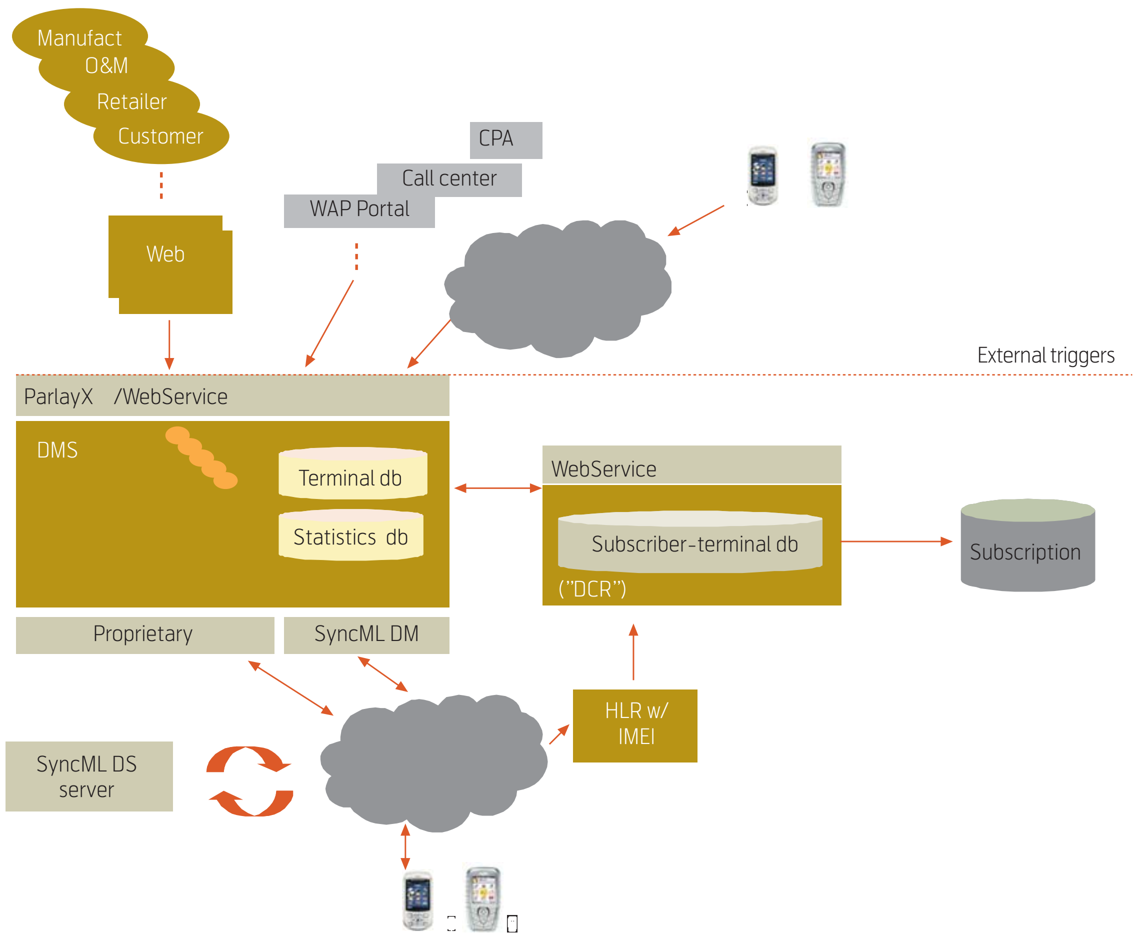


Figure 10 Architecture overview

cient for most of the basic usages like MMS or very performance critical applications. It is also necessary to detect new or unregistered combinations of IMEI-MSISDN, and the HLR is a good candidate. There must also be a trigger that notifies the DMS or the subscriber-terminal DB (depending on architecture) of new IMEI-MSISDN combinations.

Subscriber-terminal

A customer can have more than one subscription. We assume that the services follow the subscription, and hence the device data will be coupled with the phone number (MSISDN). The subscriber-terminal relation is perhaps the most difficult issue. Where should the relation be stored and what should be stored? A simple subscriber-IMEI relation might be present in the HLR within 2005, and the DMS already keeps a database of information about the terminals. An application could look up the subscriber's device type in the HLR, and then find more information in the terminal database. However, this is not extendible if in the future one should wish to extend the subscriber-terminal relationship with more information, e.g. what version of the virus software the customer has. This information could be requested direct from the device when needed, but the most flexible option would be to have a separate, extendible database to store detailed subscriber-device specific information. This is important for applications where the action is not triggered from the subscriber; for example, if one wants to automatically send virus software updates to all customers who have a specific device-operating system combination. It is very likely that the combination of device, firmware and software is not as obvious in the future as it is today. Today, device management is in most cases not as real-time critical as call handling, which also uses the HLR. Updating the HLR is an expensive task, and having a separate database for more high-level user data will provide for more flexibility and will be more future-proof. It is impossible to predict 100 % what type of information that is useful to store about a subscriber in 3–5 years. This database is most important for network-triggered actions and statistics. By separating this database from the DMS, the system will be more flexible and maintainable. The content of the database could be very Telenor specific, and integrating too much with the DMS would tie Telenor to one DMS vendor.

The subscriber entries could be populated either by a simple mirroring process, or by an event-based method that uses automatic detection to detect valid subscriptions in the network that do not already exist in the database. If it is a Nordic subscription, the DMS-db could be updated with the new subscription. With the event-based method there should be a clean-up process making sure that devices that have not been registered in the network for *n* days, should be deleted. Regulations might even require such a process. The database should be handled by the DMS, but should be a separate, DMS-vendor-independent component. It could function as an extended HLR.

The subscriber-terminal data can be stored in a Lightweight Directory Access Protocol (LDAP) structure. LDAP defines how data should be accessed, but not how it is stored. Descriptive information is stored in the attributes of an entry, where each attribute describes a specific type of information. An example could be:

```
msisdn: 555 555 555
name: John Doe
imei: 347398473
os: Symbian v 7.0
```

Subscription database

Another difficult issue is how to integrate DMS with the subscription information. It is important to know if the subscriber is eligible for GPRS, which is a relevant problem at several operators. There is one of two statements that must be broken: No dependencies or no double storage. If we strive for minimum dependencies, we mirror the subscription database, or parts of it, to the Subscriber-Terminal Database (STdb). This would mean significantly more data in the STdb, also it will be more prone to error because of the delay between the mirror update. Event-based updates would solve this problem, but demands that the subscription database is designed to send events to listeners. Either way, duplicate data will be stored in two different databases. Another solution would be to have DMS or a process in the Subscriber-terminal to query the subscription database directly. Since the Subscriber-Terminal db probably will be more permanent and stable than the DMS, the queries could be handled from the db, thus the DMS would not have to know where the subscription data is. Today there is much stress on the subscription database, and this connection is not always stable. Today there is no integration with subscription data in the Norwegian DMS solution, but there is a clear need for it in the future. This integration must be taken into account when designing the subscriber-terminal database.

Web

Today the Web solution is tightly integrated with the DMS. The ideal would be to separate web from server so that the web solution uses the same interface as the other systems. The web interface will get more complex as more roles with more options get involved. Examples of this are web interfaces for web administrators in a company, operator administrators, company employees and end-users. Skilled web developers is a must-have, and there are many qualified third parties that are cheaper than the DMS vendors.

J2EE

It is interesting to see that there is some J2EE-standardization work in progress regarding Device Management. Although this is too immature to be considered yet, one should be open to the possibility of great savings by integrating certain DMS functions into existing J2EE-infrastructure.

SyncML DS

Synchronization of PIM data is a very simple and cheap solution to put into operation. The only thing you need is a server that runs, e.g. JBoss with sync4J. However there are costs related to storage space, backup systems and other O&M tasks. We have put the SyncML DS server in the picture, because it should be readily available to Telenor's customers. If Telenor does not host it, there should be an agreement with a third party that Telenor offers a default set-up for.

5 Conclusion

The increasing complexity and diversity of the mobile phones are the fast growing challenges to mobile operators, and an efficient Device Management System is the best tool to solve it. The vision for a future Device Management System (DMS) is to lower the service threshold for the customers and make sure that the average user only has to think content, not technology. The most important functionalities like Remote Configuration and Mobile Diagnostics, including automatic detection, must be implemented.

Reference

1 *OMA Homepage*. (2004-04-15) [online] – URL: http://www.openmobilealliance.com

2 *Building an Industry-Wide Mobile Data Synchronization Protocol – SyncML White Paper, Version 1.0*. (2004-04-16) [online] – URL : http://www.openmobilealliance.org/tech/affiliates/syncml/whitepaper.pdf

3 *SyncML DM: A SyncML Protocol for Device Management*. (2004-04-16) [online] – URL: http://www.openmobilealliance.org/tech/affiliates/syncml/syncml_devman_business_cases_whppr.pdf

4 *SyncML Device Management Protocol, Version 1.1*. (2004-04-16) [online] – URL: http://www.openmobilealliance.org/tech/affiliates/syncml/syncml_dm_protocol_v11_20020215.pdf

5 *SyncML Device Management Bootstrap*. (2004-04-16) [online] – URL : http://www.openmobilealliance.org/tech/affiliates/syncml/syncml_dm_boot_v11_20020215.pdf

6 *SyncML Representation Protocol, Device Management Usage, version 1.1*. (2004-04-16) [online] – URL: http://www.openmobilealliance.org/tech/affiliates/syncml/syncml_dm_represent_v11_20020215.pdf

7 *SyncML Representation Protocol, Data Synchronization Usage, version 1.2, Draft Version 14-Jan-2004*. (2004-04-16) [online] – URL: http://member.openmobilealliance.org/ftp/Public_documents/DS/2003/OMA-DS-2003-0548R02-SyncRepresentDataSyncUsage.zip

8 *SyncML Sync Protocol, version 1.1*. (2004-04-16) [online] – URL: http://www.openmobilealliance.org/tech/affiliates/syncml/syncml_sync_protocol_v11_20020215.pdf

9 *What is*. (2004-03-29) [online] – URL: http://whatis.techtarget.com/definition/0,,sid9_gci212127,00.html

10 Toroi, Teemu. *The SyncML Road Ahead – Application Development and Device Management*. SyncML Initiative, 2002-01-30.

11 Nokia. *Smart Messaging Specification, Revision 3.0.0*. 2000-12-18.

12 *ALD homepage*. (2004-06-14) [online] – URL: http://www.thebunker.net/release-bluestumbler.htm

13 Hjemås, A M, Brecke, H, Rønning, M, Eldhuset, K. *Mobile Device Management In The Future*. Fornebu, Telenor R&D, 2004. (Telenor R&D report R 20/2004)

 ISSN 0085-7130

For a presentation of Do van Thanh, please turn to page 2.

Anne Marte Hjemås has been working as researcher in the Service Platform unit of Telenor R&D, Trondheim since 2000. She holds an MSc degree in Technical Cybernetics from the Norwegian University of Science and Technology (NTNU) and has worked with system architecture and development for eight years. She has been involved in several commercial projects as well as research projects. Her current research interests are system architecture, service platforms for telecoms, device management, IM and Presence.

email: anne-marte.hjemas@telenor.com

Anders Bjurgård is Head of Business Logic and Terminals, and has been working in Telenor Mobile's unit for Product Development since 1997. He has been technical responsible for several product launches within the Consumer and Corporate area for mobile products. From 1995 to 1997 he was Technical Project Manager for Telenor/Digifone in Ireland with the responsibility for Service Platform and Core Network from bid to launch of the GSM net. His main interest today is the evolution of mobile services into all-IP services.

email: anders.bjurgard@telenor.com

Øystein Christian Løken has been working with terminals and device management related areas in Telenor Mobile Nordic since 2001. He holds a Bachelor degree in Cybernetics from Buskerud University College, Norway, a Bachelor degree in Microelectronic Engineering from Griffith University, Australia, and a Master degree in Information Technology from Bond University, Australia. He has been involved in several commercial projects in Telenor Nordic, such as the commercial launch of MMS and 3G, to more device management projects, such as the Telenor Entry project. He is currently product manager for the Telenor Nordic DMS platform and responsible for Telenor Nordic variant implementation.

email: Oystein-Christian.Loken@telenor.com

From personal mobility to mobile personality

MATTHIAS WAGNER, MARKO LUTHER, ROBERT HIRSCHFELD,
WOLFGANG KELLERER AND ANTHONY TARLANO

Matthias Wagner is Project Manager at DoCoMo Euro-Labs

Difficulties in using and accessing new services have been the most frequently mentioned reasons for slow service acceptance in the past. For future mobile services to succeed, it is therefore critical that users are able to get intuitive and convenient access to the services they personally need in a given situation or context. We introduce the concept of Mobile Personality, which allows the mobile user to develop her own online personality in terms of personal preference, usage and service profiles over time, as well as the offered services to acquire a unique proactive behavior. This vision of adaptive personalized services is essentially based on the advanced profiling and personalization concept, context-aware computing as well as flexible and evolvable service support middleware. Through a practical use case and a detailed explanation as well as an interrelation of the essential enablers, this paper gives an insight into the foreseen transition from personal mobility to mobile personality.

Marko Luther is Researcher at DoCoMo Euro-Labs

Robert Hirschfeld is Senior Researcher with DoCoMo Euro-Labs

Wolfgang Kellerer is Head of Research group at DoCoMo Euro-Labs

Anthony Tarlano is Researcher at DoCoMo Euro-Labs

1 Introduction

Systems beyond 3G (B3G) are considered to encompass heterogeneous access networks to provide a best availability of mobile connectivity to the customers. These systems are not only considered to integrate several network platforms, they also strongly encourage the vision of a substantial richness of services and applications. On the other hand – with the complexity of services constantly increasing – we can only grasp a vague impression of how most end-users will soon be confronted with a broad variety of services and multiple ways to combine them. Difficulties in using and accessing new services have already been the most frequently mentioned reasons for slow service adoption in the past. This might even be more so for novel multimedia-type services and context-aware applications in the future. As a consequence, it is critical that users are able to get intuitive and convenient access to the services they personally need in a given situation, environment or context.

This work is motivated by ongoing research activities at DoCoMo Euro-Labs which are focusing on key concepts to enable personalized service provisioning in systems beyond 3G. In projects such as [7] or [11] we have identified, together with European operators, manufacturers and vendors, several system concepts and application domains that promise to be key drivers of future mobile communications systems. We advocate that an effective use of future services can only be achieved through adequate personalization concepts and proactive service advertisement and adaptation. It will essentially be new concepts for service provisioning and deployment together with new service paradigms that attract customers and bring mobile communication beyond voice applications to the mass market. In the following we sketch our concept of an evolving virtual personality, which is not limited to users, but can also be applied for other entities of a mobile system to implement adaptive personal services that develop their own initiative spawning a Personality Space (see Figure 1). Over service deployment- and usage-time, such a "Mobile Personality" allows different users to develop their personal preference profile and typical service usage patterns according to their personal demands, as well as the offered services to acquire unique proactive behavior.

2 Towards personalization in systems B3G – a glimpse of the future

Personalized services and applications are considered to be among the most compelling features for mobile communication systems. They promise high customer benefit through selecting specific services from a rapidly increasing diversity of mobile service offerings, and adjusting these services to their individual needs. Taking the example of a business traveler, in [27] and [12] we have already sketched some of the key personalization issues in future mobile communication systems through application scenarios. In the following we revisit these scenarios particularly with regard to an evolving Mobile Personality.

Let us consider a sample user, named Michael, who plans and takes a business trip from Boston to Paris including many different steps and various responsibilities. Setting up all the necessary preparations is a complex matter and manually finding adequate services can already be quite time-consuming. But communicating personal requirements and preferences to many services – that might not even be feasible to fulfill Michael's needs at all – will definitely become tedious. For the sake of simplicity, let us assume that in Michael's case the personalization and planning tasks for the business trip "only" consists of

 ISSN 0085-7130

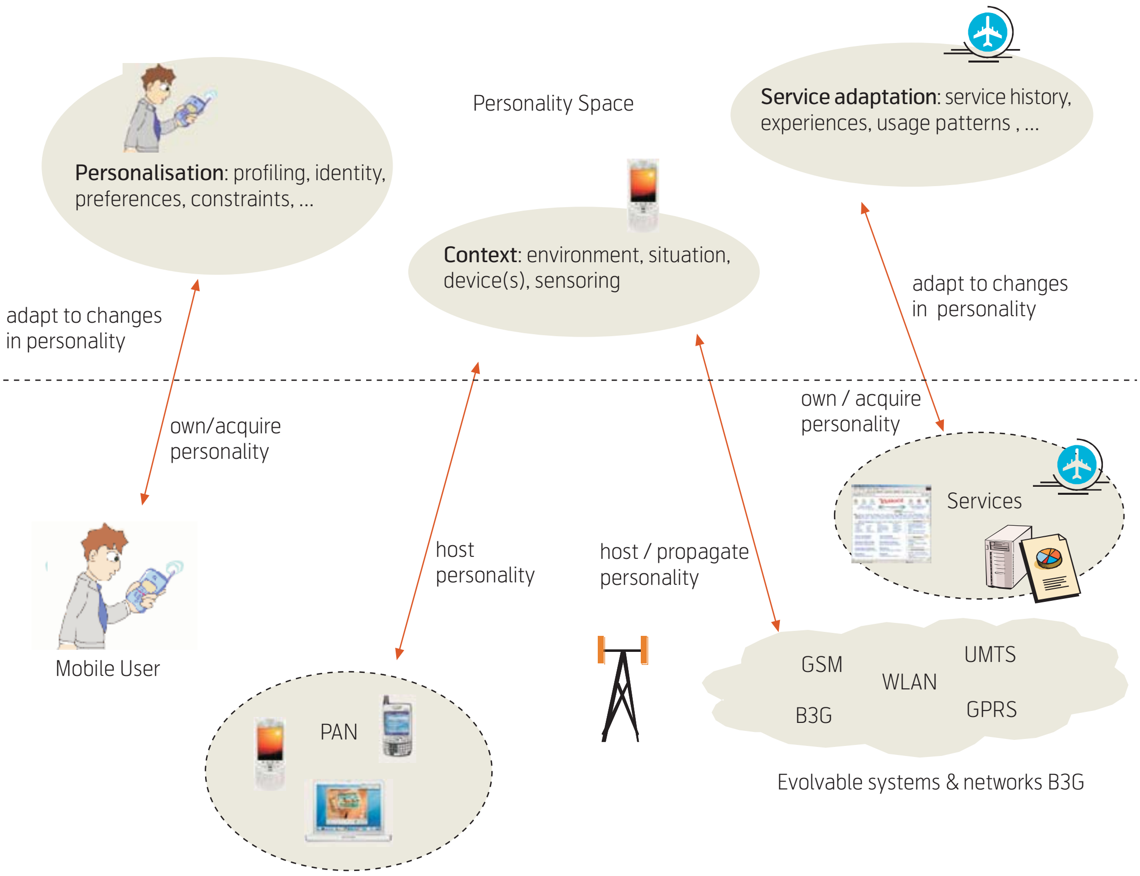


Figure 1 Vision of mobile personality in systems beyond 3G

- setting up the necessary transportation
- guidance to the meeting place
- adaptation of services and user devices to changing environments.

Obviously, Michael's trip will start with a flight from Boston to Paris. A flight booking agent, e.g. initiated through an interactive Web portal and later monitored on a mobile terminal, might try to buy a plane ticket in accordance with Michael's personal preferences. In this case, all available airline services need to be discovered that allow performing the task of booking a flight. This could be done either manually through the travel portal or by the activated travel agent. In either case, a flexible and intuitive way to model and express user preferences should be supported. In Michael's case, his user profile might contain hard constraints on the departure and arrival date and the class of the ticket, e.g. Business Class, together with soft constraints on the preferred airline ("I prefer to fly Air France to Delta Air Lines"), the type of flight ("non-stop") and the route ("as fast as possible"). Preferences associated with Michael are stored in his personal profile which can either be stored on his primary personal device or distributed in the network [1].

In the same way his air ticket is booked Michael's travel agent might also take care of reserving a rental car at his destination. Having successfully booked a flight Michael finally arrives at his travel destination. He heads straight for the airport's rental car center where he picks up his reserved car just by identification and authorization via his mobile phone. His phone transparently and automatically initiates a discovery of available services. The in-car equipment synchronizes with the phone and automatically adjusts mirrors, seat and heating to Michael's personal preferences. In addition to his convenience-driving settings, his phone discovers the car's built-in navigation system. The address of the meeting is immediately transferred and the appropriate maps are loaded by the navigation system. Using a local traffic information service, the navigation system chooses a route and is able to predict the arrival time. Since the in-car system signals that there is enough time before the meeting starts, Michael decides to get some cash in the local currency. He accesses an ATM locator service, which shows him the way to the ATM closest to his current location that is able to charge his credit card at the lowest costs. Once Michael has selected an ATM the navigation service has to adjust the route and collaborates with another service to find a nearby parking space.

Finally, Michael is guided to the meeting. Through his situation-aware communication environment, his preference of only being disturbed during meetings in the case of emergencies is activated. At the meeting room, the settings of Michael's communication devices are thus automatically adapted as the session starts. There is no need to explicitly switch to another device profile anymore. Even if there is no internal device profile available that meets the situation requirements, parameters from an external profile are transferred temporarily while carefully respecting emergency settings. During the meeting, Michael is able to transfer his video streaming session from his laptop to the built-in screen and conference system in the meeting room. In addition, other services like a nearby printer and a video transcoder are discovered which support project work. All these devices are initialized with Michael's settings derived from his personal profile. Since an important partner in the meeting is called away to an emergency, the meeting has to be re-scheduled on that day taking the schedules of all meeting parties into account. Meanwhile, Michael might be interested to meet with other researchers from the company. A scheduler service, available through the corporate WLAN, allows for short-term arrangements. It displays the availability of staff and administrative information such as room number or telephone, which are derived from the users' online profiles.

3 Enablers of mobile personality

Various system and implementation aspects have to be considered in engineering future service provisioning platforms that support the above scenario. The scenario already reveals a number of requirements that are essential for the realization of personalized service provisioning towards a Mobile Personality. In particular we consider the following requirements to be most important:

- Modeling the user through advanced profiling;
- Perception and modeling of the environment (context awareness);
- Supporting a user-centered service discovery and service selection process based on the modeled information;
- Processing the modeled information and supporting their effect on the service execution including service adaptation;
- Flexible service support middleware that allows profile propagation and service transfer as well as personalization-driven service composition.

In the following we will discuss the main building blocks for mobile personality.

3.1 Advanced personalization concepts

In our scenario, the "personality" of a user is reflected by the set of personal profiles associated with him and passed to other users, network nodes and service providers. User modeling and profiling beyond device independence – covering user preferences and wishes – are essential for supporting such a virtual mobile personality. Recently profiling standards have been established for describing service delivery context: the Composite Capabilities / Preferences Profile (CC/PP) created by the W3C [25], the User Agent Profile (UAProf) created by the Open Mobile Alliance [18] (formerly WAP Forum) as well as the Generic User Profile (GUP) put forward by the 3GPP [1]. They specify an XML – and (in the case of CC/PP and UAProf) an RDF [16] based framework – to address common needs for device independent service access. Although they provide an interoperable basis for meta data descriptions of profile information, current profiling languages are not yet adequately suited for advanced profiling needs.

As shown in the scenario, semantic-based and co-operative service discovery and selection are integral parts of proactive services; i.e. user needs and wishes will be identified as complex tasks, which are typically further divided into simpler sub-goals and heavily relate to the user's context and environment. For user profiling languages there is a lot to learn from Knowledge Engineering and the database world where the taxonomy or organization of profile elements is often referred to as schema or ontology. We advocate that the design of future profiling languages for personalization can particularly benefit from the current approaches of the so-called Semantic Web [6] where the layering of content descriptions has a similar quality: on top of XML, RDF provides a simple yet coherent structure for the expression of basic semantics and a foundation for different Web Ontology languages with a varying level of expressiveness [19]. For services to be proactive, the sub-goals of a user's tasks have to be further explored and subsequently matched to adequate services. Matchmaking can be achieved with the help of advanced discovery and selection mechanisms [3][4] and autonomous intelligent agents [24] that search the user's service environment according to his personal preferences.

3.2 Context-aware computing

Context awareness is an attribute of a service that is capable of accessing, interpreting and manipulating knowledge of its environment and to adapt the service behavior accordingly [15]. For the provisioning of intelligent services, having their own personality, we are developing a context spaces agent service architecture that allows acquiring, managing and processing context information based on agent technology [24]. Here, a significant attribute of an intelligent agents environment is the existence of other intelligent agents with the possibility for social communication between agents [17]. Thus, allowing agents to exchange context information and experience reports of service usage as well as to adapt to services to user profile/preference information and vice versa.

Introspection capabilities are a key requirement within context-aware computing and specifically in the context spaces architecture. The construction of a context space provides the capability of accessing knowledge of the environment for logic reasoning purposes. To enable introspection in the architecture, accessible data attributes of the environment are modeled and constructed as a context tree structure. Agent services using logic, such as dependency rules or logic expressions, can use the context space as data source for personalization input. Since the resulting context tree structure is an overlay data model designed to remain independent of the original storage model, it can be used for both transient and persistent data originating from data storage repositories or broadcast sensor data.

3.3 Flexible and evolvable service support middleware

Service support middleware has to be capable of efficient session management including profile management and service mobility. The service session control as core part of the service support middleware has to provide advanced service provisioning functions including those described above. We regard application-layer signaling to act as a coordination facility supporting information exchange between the acting entities. To support advanced personalization concepts, we advocate a proxy-based service signaling approach [13]. Here profile management, service session management and resource control are realized in separate interacting proxy servers that are on the end-to-end service signaling path. This concept is substantially enhanced to enable a preference-based session management [12]. Preference-based session management makes use of profile information already in the proxy servers (e.g. network edge nodes) that are traversed by the service signaling messages and not only in the target application.

In addition, regarding user requirements and emergent services, future systems have to be most flexible. To date, nearly all service architectures center on required functionality within a strictly layered system structure. In [21] we describe an architecture that supports adaptation and evolution in systems B3G in general and on a middleware level in particular. Such an adaptable environment allows making personalization effective for service execution by a flexible introduction and exchange of functional components, such as context manager or service selector. Programmable platforms on all system layers form the basis for the component management.

Adaptation and programmability of services should not be limited to content adaptation as it is in most of today's systems. Instead it will involve modifications to behavior (service logic) [9], service interaction and signaling as well. How dynamic service adaptation can benefit from dynamic programming environments such as [8] is further explained in [10].

4 Preference-based service discovery on mobile devices

Using Web-based services has already become an integral part of our everyday life. Semantic Web technology and the advent of universal and mobile access to Internet services will only add to the broad range of existing services on the Web and provide addi-

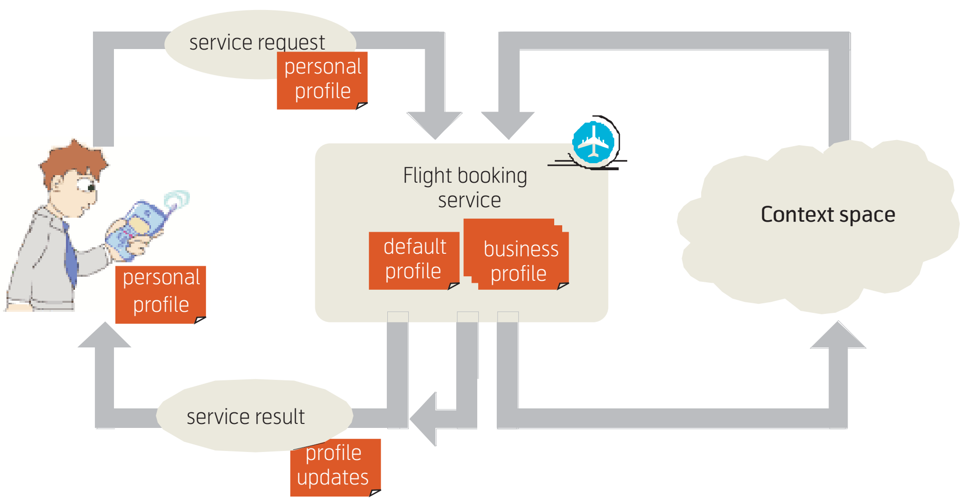


Figure 2 Cooperative service execution and context spaces

tional features like knowledge-based, location- or context-aware information. On the other hand, so far little work has been done to explicitly account for aspects of mobile computing in semantic service frameworks. Whereas much of the work in Semantic Web services discovery and composition concentrated on the functionalities of the services, contextual information, personal preferences and more generally personalization are more pressing challenges in the mobile computing arena. In order to manage an increasing amount of mobile services, it is essential that Semantic Web services standards explicitly support the needs of developers and users, such as the discovery and selection of services they personally need in a given situation or context.

In an early case study we have implemented MobiOnt and MobiXpl as a semantic toolbox to explore mobile user-centered services on the Semantic Web [6]. Our vision is to take full advantage of future complex service offerings on limited client devices and to handle the need for personalized service discovery in mobile environments. At this point MobiOnt and MobiXpl are early prototypes that are realized and can be demonstrated as plug-ins to the Protégé knowledge workbench [22]. MobiXpl emulates different commercially available handsets, whereas MobiOnt encapsulates central preference-based matchmaking mechanisms. Implementations of MobiOnt as a central network component and MobiXpl as a Java-based client running on an actual phone are currently implemented.

4.1 A Concrete usage scenario

We study an extension of the usage scenario published in [28] that addresses a future mobile Internet radio scenario. Internet radio has become increasingly popular in recent years with boosting numbers in Web radio stations and subscribers [20][23]. In this context, personalized access to content is particularly important to accommodate varying technical as well as personal user needs and preferences. In our testbed we have modeled Internet radio stations as Web services with varying service characteristics. Radios channels are described using an Internet radio ontology (a fragment of the ontology is shown in Figure 4) that consists of concepts that describe and classify Web radio services in terms of program format, origin, audio format characteristics and a time-based classification of streamed audio content. This service ontology is then used for preference-based service discovery. Note that our Internet radio use case is only one of many possible applications for the MobiOnt.

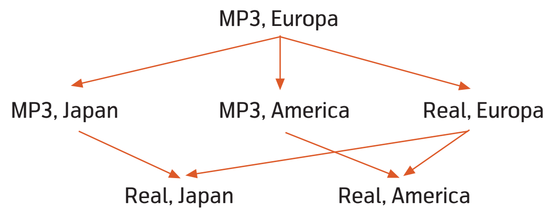


Figure 3 A user-defined preference ordering

4.2 User preferences

While browsing the service ontology, service concepts with key relevance to the user can be selected and combined to preferences. In our preference framework [14], these (partially) ordered feature sets are directly handled without the use of any explicit quality or ranking values: user preferences are introduced as a special relation with the semantics of considering some object (or class) A superior to another object (or class) B ("I like Music channels better than News stations"). Preferences indicate constraints that a service should fulfill to best meet its requirements. On the other hand, even if none of the indicated preferences are met, a match can still be possible. To manage multiple user preferences complex preferences can be inductively constructed from a set of base preferences by means of preference constructors [14][28].

Figure 3 shows an example of a combined preference from the radio scenario. Here a user has indicated that she generally prefers radio programs from Europe to those from Japan or America. Still, the two latter choices are her preferred choices over any other available program. Due to the technical capabilities of her player, she also prefers MP3 encoding to Real. Further, she has specified that both base preferences are equally important to her.

4.3 Cooperative service discovery

User preferences constructed during preference building define a service request that ultimately needs to be mapped to the underlying service ontology. MobiOnt therefore implements a flexible discovery algorithm that can be extended through different strategies. The goal of service discovery is to retrieve those service instances from the ontology that represent the best matches to given preferences.

The implemented preference-based service matching is performed along the lines of the determined preference order to implement cooperative behavior: if the search for a perfectly matching radio station fails, the initial query is gradually relaxed along the path of the (complex) preferences until a next-best match can be found. Thus, if in our example from above during service discovery no match could be found in European

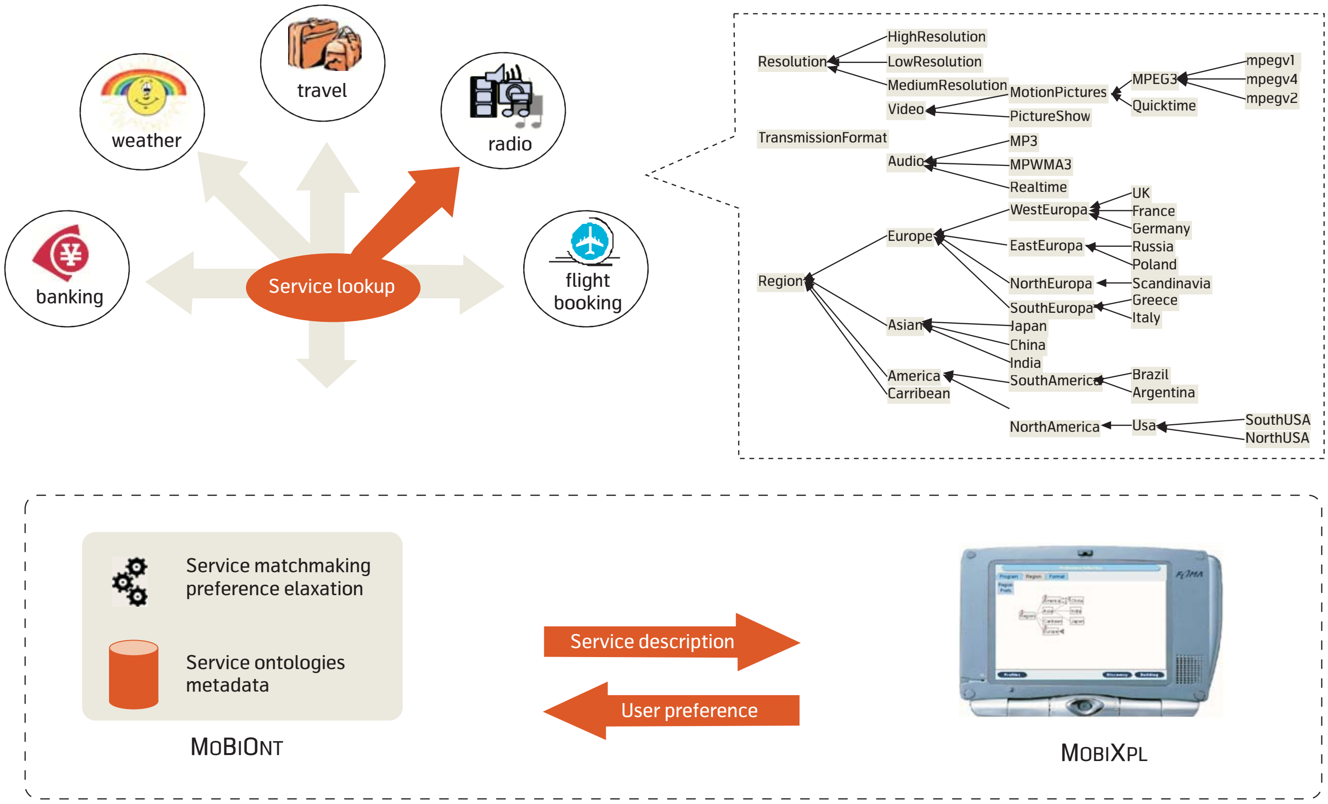


Figure 4 MobiOnt / MobiXpl – A testbed for mobile semantic-based services

programs in MP3 encoding, the next discovery step consists of trying to match radio stations that broadcast Japanese or American programs in MP3 or European programs in Real. If neither of these two second-best choices are available, any other program is matched. Further implementation and application aspects as well as selective ontology browsing and preference building and mapping are further explored in [4][5].

4.4 mobiXpl – a user interface for preference-based discovery

Parts of the Internet radio ontology are carefully exposed to the user through MobiXpl, the graphical frontend to our framework (cf. Figure 5). MobiXpl emulates different mobile terminals and consists of a mobile ontology browser with support for individual user views as well as an intuitive interface to user preferences. The idea is to only display selected concepts and sub-ontologies depending on the user's

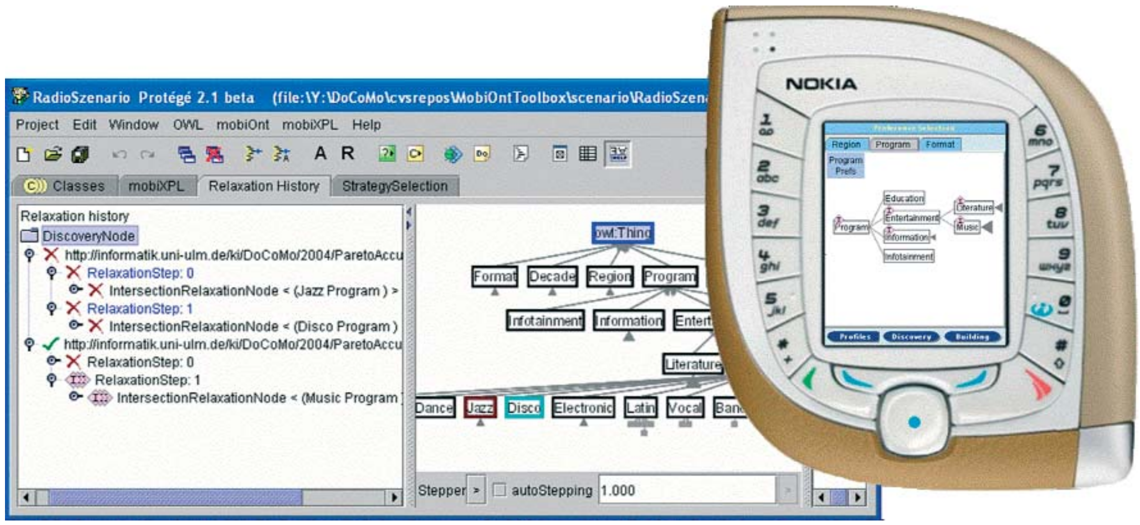


Figure 5 MobiXpl – exploring ontology-based service catalogs on the mobile device

ISSN 0085-7130 © Telenor ASA 2005

experience level and usage profile. While browsing the service ontology, concepts that circumscribe services with key relevance to the user can be selected and combined to user preferences. Subsequently, these preferences are used during the service discovery to implement cooperative behavior: if the search for a perfectly matching radio station fails, the initial query is gradually relaxed along the lines of the determined preferences until a next-best match can be found. Both application aspects, selective ontology browsing and preference building and mapping are fully explored in [28].

5 Towards a mobile personality

Expressing user preferences, wishes and dislikes in an intuitive way is crucial for the flexible provisioning of mobile services. With the vision of Mobile Personality in mind, we have developed essential concepts and solutions for the flexible matchmaking of evolving user preferences and services; namely,

- an algorithm for preference-based negotiation and interaction with services [3];

- a mechanism for the use of semantically rich services descriptions together with service usage patterns in aggregated service catalogs for cooperative service discovery [4][5];

- early prototypes that leverage preference-based matchmaking in semantically rich service catalogs for the mobile user.

In cases where the perfect match of the user's preferences to the available services is not possible in the given context the next nearest match has to be provided in a cooperative fashion. The outlined approach assumes that user preferences are modeled in terms of hard and soft constraints. Hard constraints model user preferences that definitely need to be matched during service discovery and selection, whereas soft constraints represent parts of a user profile or a user request that can be relaxed during the matchmaking to the available services. Our profiling concepts comprise a notion of usage patterns to express preferences of user groups and typical service invocation patterns. For example, a general preference from the travel domain could be that "everyone prefers a short traveling time" (i.e. a departure date with maximum proximity to the arrival date is preferred). The iterative adaptation of these patterns to evolving user needs plays a key role in the concept of Mobile Personality. As indicated in Figure 2, user preferences are passed to the service provider for service matchmaking and execution. If no match between the personal profile and the default execution profile of the service is available two basic conflict resolution strategies (and combinations thereof) are applicable: on the one hand, the cooperative service execution might decide to relax user's preferences until a match can be made. On the other hand, the user might be associated with a typical group pattern, e.g. "Business User", resulting in the enrichment of his request with additional preference data from the pattern.

Beyond the one-time usage the lifecycle of profile data is crucial in the Mobile Personality paradigm. For instance, when an intelligent agent reserves a rental car at Boston airport on behalf of our sample user Michael for the first time, he will do this based on some base preferences. However, on a revisit to Boston, Michael's user profile will already contain a usage history of services together with an experience report. Again, the left-hand side of Figure 2 illustrates this through the cycle of profile usage and profile update. Based on the user's profile and/or usage history – and maybe additional usage histories, e.g. from Michael's colleagues or users with similar preferences and interest, a more informed decision can be made. In the case of Michael it is conceivable that the choice of car rental agency, formerly price-based, will change based on experiences with the agency in case of a car breakdown or due to unanticipated technical problems with adapting his mobile terminal to the in-car navigation system.

Referring to Figure 2, we show a more detailed prospective of adaptive service interaction and the use of a context space for introspection and interaction by intelligent service agents: in the diagram, Michael initiates the flight booking service using a service request containing meta-data representing his personal profile. Upon receiving the service request, the flight booking service processes the service request and promptly returns the service result and the profile update. In addition to sending the service result to Michael, the flight booking service annotates the context space by writing the updated profile to the context space. Service agents with a similar interest in Michael, using the introspection capabilities of the context space, observe the annotation of the space and use logic to determine if the updates in the profile can be used for personalization. Through agent social interaction, service agents such as a group of car rental agents reason that additional profile update should be made available to Michael. The profile updates are annotated into the context space and the flight booking service is notified for asynchronous delivery to Michael.

6 Summary

The evolution towards systems and services B3G bears the risk of increasingly confronting the customer with technology features instead of service aspects. We claim that similar problems in using and accessing new services have already led to slow service adoption in the past. As a consequence, users might not be able to fully understand novel services and to benefit from future applications. In this paper – with customer acceptance and ease-of-use as the most important success factor for forthcoming telecommunication systems in mind – we have presented our vision and first steps towards supporting advanced personalization concepts under the umbrella of a "Mobile Personality".

The Mobile Personality paradigm is characterized through personal services, user preference and service profiles that evolve over time according to changes in the user's environment, context or the service offerings themselves. The main building blocks of this personalization concept are: advanced personalization concepts, context-aware computing, as well as highly flexible and evolvable system architectures. Our ongoing work has emphasized generic approaches to context management as well as service discovery and selection that feature extendable semantic descriptions of services, users and devices.

7 References

1 *3GPP. Generic User Profile*. http://www.3gpp.org

2 Ankolenkar, A, Burstein, M, Hobbs, J et al. DAML-S: Web Service Description for the Semantic Web. In: *Proceedings of the International Semantic Web Conference (ISWC'02)*, Sardinia, Italy, LNCS 2342, Springer, June 2002.

3 Balke, W-T, Wagner, M. Towards Personalized Selection of Web Services. In: *Proceedings of the International World Wide Web Conference (WWW2003)*, Budapest, Hungary, May 2003.

4 Balke, W-T, Wagner, M. Cooperative Discovery for User-centered Web Service Provisioning. In: *Proceedings of the 1st International Web Service Conference (ICWS2003)*, Las Vegas, Nevada, June 2003.

5 Balke, W-T, Wagner, M. Through Different Eyes – Assessing Multiple Conceptual Views for Querying Web Services. In: *Proc. of 13th Int. WWW Conf. (WWW'04)*, New York 2004.

6 Berners-Lee, T, Hendler, J, Lassila, O. The Semantic Web. *Scientific American*, 284 (5), 34–43, 2001.

7 *EURESCOM Project P1203*. http://www.eurescom.de

8 Hirschfeld, R. AspectS – Aspect-Oriented Programming with Squeak. In: *Architectures, Services, and Applications for a Networked World*, LNCS 2591, Springer 2002.

9 Hirschfeld, R, Wagner, M. PerspectiveS – AspectS with Context. In: *Proceedings of the OOPSLA 2002 Workshop on Engineering Context-Aware Object-Oriented Systems and Environments*, Seattle, WA, November 2002.

10 Hirschfeld, R, Wagner, M, Kellerer, W, Prehofer, C. AOSD for System Integration and Personalization. In: *Proc. of the AOSD Workshop on the Commercialization of AOSD Technology*, Boston, MA, March 2003.

11 *IST MobiLife*. http://www.ist-mobilife.org

12 Kellerer, W, Wagner, W, Balke, W-T. Preference-based Session Management. In: *Proceedings of the 8th International Workshop on Mobile Multimedia Communications (MOMUC2003)*, Munich, Germany, October 2003 (to appear).

13 Kellerer, W, Berndt, H. Next Generation Service Session Signaling. In: *Proceedings of TINA 2002*, Petaling Jaya, 2002.

14 Kießling, W, Köstler, G. Preference SQL – Design, Implementation, Experiences. In: *Proc. of the Int. Conf. on Very Large Databases (VLDB'02)*, Hong Kong, 2002.

15 Naghshineh, M. Context-Aware Computing. *IEEE Wireless Communications Magazine*, Special Issue, October 2002.

16 Lassila, O, Swick, R R. *Resource Description Format: Model and Syntax Specification*. W3C Recommendation, 1999.

17 Odell, J et al. Modeling Agents and their Environment. In: *Proceedings of the AOSE Workshop of AAMAS*, 2002.

18 *OMA*. http://www.openmobilealliance.org, 2003.

19 McGuinness, D L, van Harmelen, F. *OWL Web Ontology Language Overview*. W3C Working Draft. http://www.w3.org/TR/owl-features/. 2003.

20 *NullSoft SHOUTcast*. http://www.shoutcast.com.

21 Prehofer, C, Kellerer, W, Hirschfeld, R, Berndt, H, Kawamura, K. An Architecture Supporting Adaptation and Evolution in Fourth Generation Mobile Communication Systems. *JCN Journal of Communications and Networks*, 4 (4), 2002.

22 *Protégé Ontology Editor and Knowledge Acquisition System*. http://protege.stanford.edu/

23 *Radio Locator*. http://www.radio-locator.com/.

24 Talano, A. Context Spaces Agent Architecture. In: *Proceedings of the EURESCOM Summit*, Heidelberg, 2002.

25 *W3C. Composite Capabilities/Preferences Profiles*. http://www.w3.org/Mobile/CCPP, 2002.

26 *W3C. Web Services Activity (Architecture Domain)*. http://www.w3.org/2002/ws/, 2002.

27 Wagner, M, Balke, W-T, Hirschfeld, R, Kellerer, W. A Roadmap to Advanced Personalization of Mobile Services. In: *Proceedings of the International Conference DOA/ODBASE/CoopIS (Industry Program)*, Irvine, CA, October 2002.

28 Wagner, M, Liebig, T, Noppens, O, Balzer, S, Kellerer, W. mobiXPL – Semantic-based Service Discovery on Tiny Mobile Devices. In: *Proc. of Workshop on Semantic Web Technology for Mobile and Ubiquitous Applications* (in conj. with ISWC'04), Hiroshima, Japan, 2004.

Matthias Wagner is the Project Manager of the Evolutionary Systems project at DoCoMo Euro-Labs. His group is concerned with adaptive software as well as user guidance and contextual intelligence for the next generation of mobile systems. In his research he mainly focuses on proactive services and personalisation issues. Currently, he acts as a work package coordinator and a project management team member of the IST integrated project MobiLife. He is also responsible for coordinating DoCoMo Euro-Lab's activities in and contributions to the World Wide Web Consortium (W3C). Prior to joining DoCoMo, he worked for the department of Artificial Intelligence at the University of Ulm and the department of Databases and Information Systems at the University of Augsburg. Matthias Wagner has numerous scientific publications in international journals, conferences and workshops. He holds a Master degree and a Ph.D. degree in Computer Science.

email: wagner@docomolab-euro.com

Marko Luther is a Researcher in the Evolutionary Systems project at DoCoMo Euro-Labs where he is focusing on context-aware computing and ubiquitous services. Methods researched and applied include formal ontologies, qualitative reasoning and distributed semantic computing. Marko Luther has an extensive background in Artificial Intelligence and Formal Methods. Prior to joining DoCoMo, he worked for the department of Artificial Intelligence at the University of Ulm and has numerous scientific publications in international journals, conferences and workshops. He holds a Master degree and a Ph.D. degree in Computer Science.

email: luther@docomolab-euro.com

Robert Hirschfeld is a senior researcher with DoCoMo Euro-Labs, the European research facility of NTT DoCoMo Japan. There he works on infrastructure components for next generation mobile communication systems with focus on dynamic service adaptation and context-oriented programming. Prior to joining DoCoMo, he was a principal engineer at Windward Solutions in Sunnyvale, California. Robert Hirschfeld received a Ph.D. in Computer Science from the Technical University of Ilmenau, Germany.

email: hirschfeld@docomolab-euro.com

Wolfgang Kellerer is heading the Ubiquitous Services' Platforms research group at NTT DoCoMo's European Research laboratories in Munich, Germany, focusing on capabilities of next generation mobile service support platforms. Research topics include service signaling architectures, context-awareness, ubiquitous communications (peer-to-peer, sensor networks), and cross-layer system design aspects. Dr. Kellerer is currently serving as the elected Vice Chairman of the Working Group 2 (Service Architecture for the Wireless World) of the Wireless World Research Forum (WWRF). He has published over 50 papers in respective journals, conferences and workshops in the area of service architecture and mobile networking. He is also active in European IST projects such as IST MobiLife and IST Simplicity. Before he joined DoCoMo Euro-Labs, he worked at the Institute of Communication Networks at Technische Universität München (TUM), Munich, Germany. In 2001 he was a visiting researcher at the Information Systems Laboratory of Stanford University, California, US. He holds a Master and a PhD degree in electrical engineering and information technology from TUM.

email: kellerer@docomolab-euro.com

Anthony Tarlano is member of the Object Management Group (OMG) Agent Special Interest Group (Agent-SIG) focused on the specification of Software Agent Technologies and has recently held positions as co-chairman of the Agent SIG and as a member of the OMG Architecture Board. His current research activities at DoCoMo Euro-Labs are focused toward the realization of Context Aware Computing Services for 4th Generation Wireless Systems using Technologies such as, Peer-to-Peer Protocols, Middleware Object Systems, Multi-Agent Systems and Dynamic Transformation Processing Systems. His formal educational background includes studies in Computer Science and Data Communications Engineering at the University of Mississippi, School of Engineering and Research Center for Telecommunication. In addition to his educational experience, Mr. Tarlano has contributed to variety of technically cutting-edge projects as a member of the research and development technical staff at Harris Corporation, Litton-FiberCom, and Bell Labs in Murray Hill, New Jersey

email: tarlano@docomolab-euro.com

The Virtual Device – The future mobile phone?

DO VAN THANH AND IVAR JØRSTAD

Do van Thanh is Senior Research Scientist at Telenor R&D

In this paper we introduce a novel mobile device concept, which considers all the autonomous devices on the user 's Personal Area Network (PAN) as one big "Virtual Device" having multiple input and output units and providing a coherent and surround interface to the user. Such a concept is quite valuable to the user since it allows him to manage all his devices as one. The paper also proposes and describes a PAN middleware that enables the formation of a Virtual Device on the PAN and that is capable of handling the dynamical presence of devices and the diversity of device types. In order to identify the necessary capabilities and functions of such middleware, two PAN configurations, namely Isolated PAN with multiple Open devices and Networked PAN without Open devices are considered thoroughly.

Ivar Jørstad is a PhD student at Norwegian University of Science and Technology

1 Introduction

There is no doubt about the popularity of the mobile phone. Both the number of mobile phones and the number of subscribers world-wide are not only increasing but also at a faster pace. The mobile phone has become the companion of many people. They always bring it with them. They speak, watch and listen to their phone. The abilities and functions of the mobile phone are surely very valuable for the user. However, the current physical form of the mobile phone, although very light and handy, is not always the most suitable one. The driving user would find it difficult to talk on phone while driving with one hand. A hands-free or headset allowing the user to speak on the phone without having to hold the handset is definitely a better form. It allows the user to have both hands free. To write on the tiny keypad of the mobile phone may be an exciting challenge for youngster but it is not appealing to all people. The sounds and pictures offered by the mobile phone are good enough for the mobile user but cannot be compared with the pictures and the surround sounds of the home theatre system.

It is hence desirable dynamically to be able to move the input functions, e.g. microphone, keypad, navigation, pointing, etc. and the output functions, e.g. display, loudspeaker, etc. of the mobile phone to external devices.

This paper presents a novel concept called Virtual Device that aims to realise the mentioned demand. The Virtual Device incorporates all the autonomous devices in the vicinity of the user and allows him to use them as one big "Virtual Device" [1] with distributed multiple input and output capabilities. The paper starts with a thorough description of the Virtual Device concept with the clarification of its usefulness. Next, it will explain how the concept can be realised.

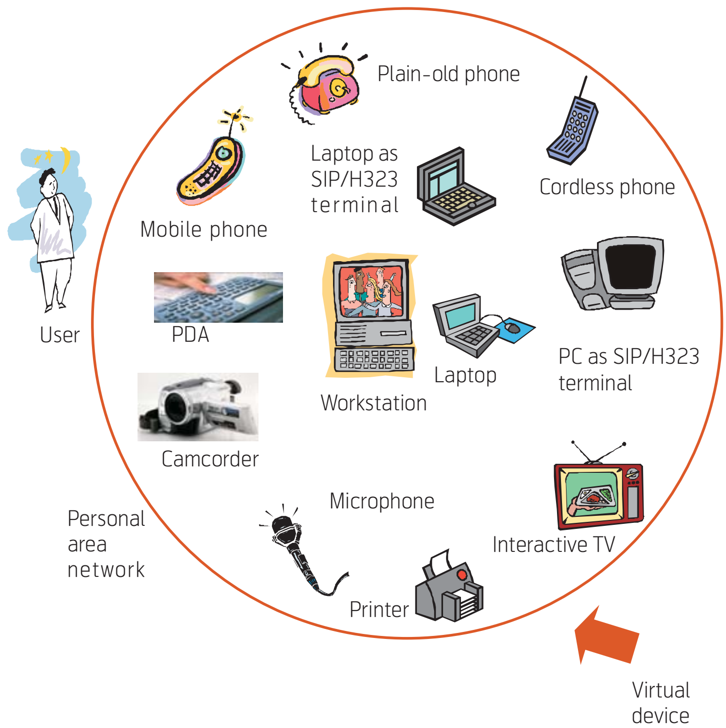


Figure 1 The Virtual Device on PAN

2 The Virtual Device concept

The Virtual Device is a novel concept, which considers all the autonomous devices surrounding the user as one big "Virtual Device" having multiple input and output units and providing a coherent and surround interface to the user [2]. With the Virtual Device concept, the user is no longer communicating using one phone but a combination of multiple phones combined with other electronic devices both portable and stationary.

Such a concept is quite valuable to the user since it allows him to manage all his devices as one (see Figure 1).

 ISSN 0085-7130 © Telenor ASA 2005

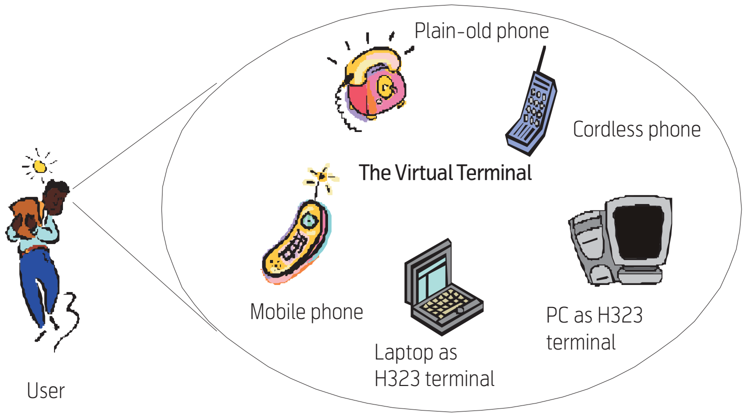


Figure 2 A virtual terminal comprising several communication devices

2.1 Device unification and coordination

The most powerful feature of the Virtual Device is the ability to combine mobile and stationary devices. Since many communication devices, e.g. mobile and fixed phones, may have a limited user interface, the combination of these devices with various other devices with different capabilities into a "Virtual Terminal" as illustrated in Figure 2, may result in much better user interfaces and hence enhance the offered services.

During a service session, a user may want to use one or more elementary devices with just input and output functionality in addition to his communication device. Examples of such elementary devices can be a big screen display, loudspeakers, a microphone etc. In other cases the user may want to reroute the voice or data streams coming in to his mobile phone to other, more convenient devices such as a fixed phone, a PC etc. Another task for the Virtual Device might be to send the contents of a file from the mailbox to a nearby printer to provide a hardcopy of incoming text, pictures etc.

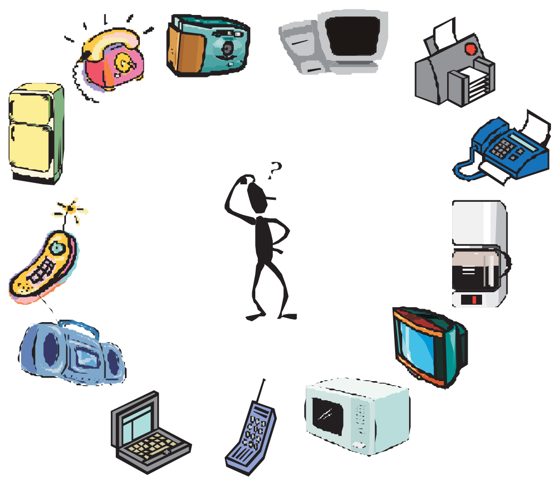


Figure 3 Handling several devices might be difficult for a non-technical user

The Virtual Device will also help the user in utilizing stationary devices on visiting places, e.g. printers, TV screens, etc., and to coordinate them with the mobile devices carried along. Whether or not stationary devices should be available to visiting users should depend only on the permission from the owner of the devices and not on the configuration or technical interoperability of the devices. As long as the visiting user has permission to use stationary devices, the Virtual Device will be responsible for unifying and coordinating the mobile and stationary devices so that they can work together in a seamless way.

With the Virtual Device concept, all the different communications devices, computing devices and other electronic and peripheral devices will be considered as one big terminal with multiple input and output capabilities. The user terminal will no longer be an integrated and recognisable device but a set of distributed devices that allow access to certain services, e.g. multimedia communication services, various computing services, etc.

2.2 Uniform device management

When the user is confronted with several different devices, the management of all the different devices might be complex and time consuming, and it can be difficult for a non-technical user to master it all (see Figure 3). For example, there will be several terminal profiles and several user profiles the user has to relate to. From the user's point of view, it would certainly be desirable to have one unique user profile specifying for example his home, office and travelling environments.

The Virtual Device will allow the user to define, add and remove the devices that are included in the Virtual Device at any time. The user's many terminal profiles and user profiles on different devices will be replaced by a unique user profile defined once and for all in the Virtual Device. Whenever this profile is set up or updated, it will automatically be accessible from all other devices contained in the Virtual Device. In this way, the user can set up and modify his personal preferences for all his devices at one place. For example, the user might have different profiles for different environments – at home, in the office, travelling, out of reach, etc. In each of these profiles, he may configure the composition of the Virtual Device according to what devices are available in the different environments. He may also specify which of the profiles should be active at any given time in different ways, e.g. according to a timetable or a location table or he can specify directly by manually informing the Virtual Device and thus override any existing time- or location table previously in effect.

The different profiles can also be set up to specify different personal preferences when it comes to subscription of services, user interfaces on particular device types etc., so that the user will be offered the same services and experience the same look and feel regardless of where he is and what particular device he is using. The only requirements are that the device in question supports such services and such user interfaces.

The Virtual Device will offer a dynamic configuration depending on which personal service environment is in effect and considering what services are available at any time. For example, the Virtual Device can have one configuration for the home environment specifying which devices are available in the evenings when the user is at home and another configuration for the office environment specifying what devices are preferably used at work.

Another inconvenience avoided when using the Virtual Device is the repetition of the same tasks for many devices. Instead of defining his profile and preferences on every device, for example enabling voice-answering service both on his cellular phone and his plain old telephone, he can define it once and for all in the Virtual Device. Also, the address list defined for one device will automatically be available on other devices and there will no longer be a need for the user to define and update a distinct address list for each device.

2.3 Cost reduction and increase of convenience

Other advantages of the Virtual Device concept are the cost reduction and the increase of convenience. Devices can share resources like processing power, battery power, memory, data, applications, network access points and user interfaces. The user does not have to purchase the same resource twice and can hence save money. It is also more convenient for him to carry as little equipment as possible when moving.

2.4 A truly user centric concept

The proposed concept will be user centric instead of device centric in that it will be possible to address a person directly instead of addressing one of his devices (Figure 4). With the Virtual Device, the user will be accessible anywhere, anytime and regardless of what communication device he has at his disposal.

The Virtual Device will be responsible for handling available devices, which may include a variety of personal or public devices. The user should keep his Virtual Device of what devices he can be reached by at any time, but this does not necessarily have to trouble the user. It can be done automatically according to

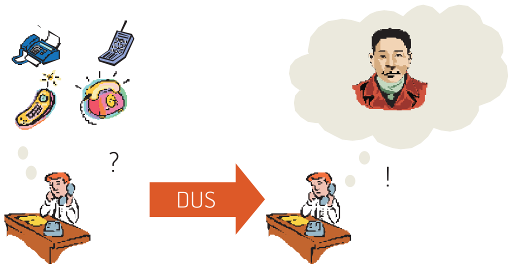


Figure 4 Addressing a person instead of a device

some predefined settings, for example default registration, a timetable or a location dependent table. Another solution is to make use of service discovery protocols to detect automatically all nearby services and devices. The Virtual Device will know which device it will be most appropriate to contact the user by at any time. Anyone wishing to communicate with him will dial up his Virtual Device, which in turn will contact the user. In other words – the Virtual Device will function as the user's personal secretary. Incoming calls will be transferred to the most appropriate devices at any given time, and it will receive messages for the user when he does not want to be disturbed.

The Virtual Device will offer personal mobility to the end user in that there are no fixed relations between the user and any specific device. Any relations both with respect to devices and subscriptions are dynamically set up and taken down by the Virtual Device. In this way, the different identities of the different devices will become transparent, but the identifier belonging to the Virtual Device will stay unchanged.

3 Related works

In the project EU IST-2001-34157 PACWOMAN, there is a different definition of the Virtual Device which consists of a Master device and several Basic, low cost, low power radio terminals. Each device can be considered as one device on the PAN. The PACWOMAN Virtual Device corresponds to the idea introduced in Wearable Computing where a main PDA is controlling and communicating via wireless links with miniaturised peripheral devices. The Virtual Device in this paper, on the other hand, consists of all devices of various types within and beyond the PAN. The Eurescom project P1101 "Always-On" Device Unifying Service, defines a Virtual Terminal concept that consists of networked devices without local connectivity between each other. No clear PAN is defined. The focus is on the transfer of voice and data sessions between devices, e.g. to move a conversation from a mobile to a fixed phone or vice versa.

ISSN 0085-7130 © Telenor ASA 2005

The Virtual Device in this paper may be confused with Jini from Sun Microsystems. Jini enables services and resources offered by devices over a network to be shared by users. No attempt is made to exploit the benefits of integrating devices together as defined by our Virtual Device.

4 Realisation of the Virtual Device concept

The Virtual Device concept can never be a reality without the emergence of local short-range communication technologies such as Bluetooth [3], WLAN [4], IEEE 802.15 [?], etc. that enable local connectivity between devices. The devices belonging to a person will form a private Personal Area Network (PAN) where devices appear and disappear dynamically. Unfortunately and most importantly, connectivity does not necessarily mean communication, and many challenging issues must be resolved before devices can collaborate and together form a Virtual Device. This paper proposes and describes a middleware that enables the formation of a Virtual Device on the PAN and that is capable of handling the dynamical presence of devices and the diversity of device types. But, before proceeding with the description of the Virtual Device middleware let us start with a review of the architectures of the computer since they are the fundament of the Virtual Device concept.

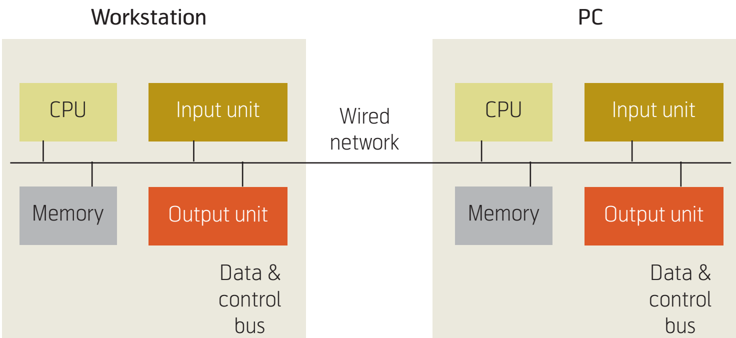


Figure 5 Distributed computing

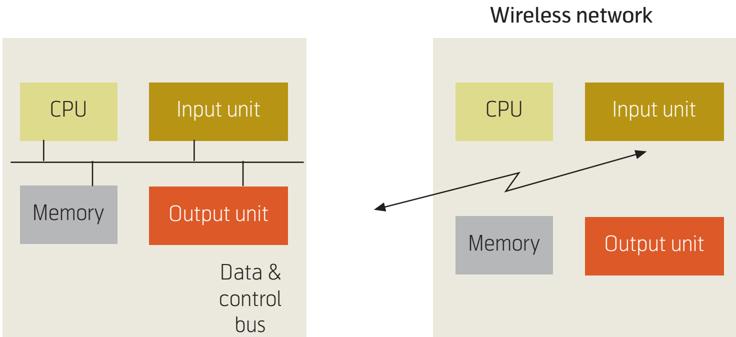


Figure 6 Wireless Distributed computing

5 The Virtual Device – the natural and inevitable evolution

The computer, from its infancy until now has a constant evolution and has experienced several paradigm shifts:

- Mainframe and monolithic architecture
- Distributed system and network system
- Wireless distributed system and personal area networks.

From the 50s till the 80s, the computing world is dominated by mainframes and minicomputers. They have a monolithic architecture composed of:

- A Central Processing Unit (CPU)
- A Memory system (Volatile Storage)
- A Persistent storage
- An Input and Output unit which manages series of terminals.

The heart of the mainframe is the CPU but the main vein is the bus system. The bus system is used to convey both the control information necessary for the control and synchronisation in the system and the data that is exchanged between the units. Generally speaking, it is possible for the bus to convey both operations and data streams although, at the implementation level, a bus system may consist of two separate parts, one for control and one for data.

During the 80s, microprocessors, workstations and PCs made their appearance and in a very short time forced the mainframes to retreat. However, they would not have such a success without the breakthrough in data communications. With Ethernet, higher data rate can be exchanged between computers in the Local Area Networks. The Distributed Computing paradigm is introduced and the goal is to build a huge mainframe based upon independent computers. The transparency concept is aimed at hiding the complexity created by the communication distribution such that developers can concentrate on the development of their applications [5][6][7]. The following transparencies are defined:

- access transparency
- location transparency
- relocation transparency
- migration transparency
- persistence transparency
- failure transparency
- replication transparency
- transaction transparency.

To realize the Distributed Computing paradigm a software layer, or more precisely, a middleware is introduced between the applications and the infrastructure. A well-known example of such a middleware is CORBA (Common Object Request Broker Architecture) promoted by the Object Management group [8].

With the emergence of short-range wireless technologies like Bluetooth and WLAN, the natural evolution is to replace the wired network with the wireless network to obtain the wireless distributing computing paradigm as shown in Figure 6.

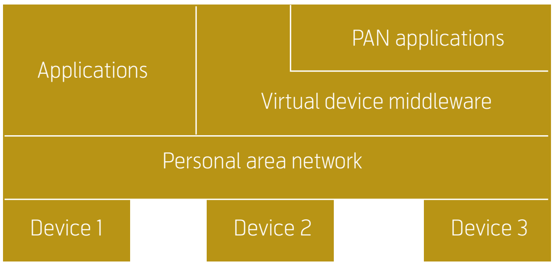


Figure 7 The Virtual Device architecture

6 The Virtual Device middleware

Although sharing some similarities with the traditional Distributed Computing the Wireless Distributed Computing also has major differences and to realize the Virtual Device requires yet another type of middleware (Figure 7). The main difference is that all the transparencies except Access Transparency are undesirable. Indeed, one of the goals of the Virtual Device is to be in control of the location of the applications and to do relocation, migration, duplication, etc. The other differences are the heterogeneity and dynamics of the devices on the PAN. In the traditional Distributed Computing, the computing nodes are supposed to be stable and equal or at least having the minimum of capabilities e.g. processing, storage, communication, etc.

Indeed, on the PAN there is a variety of devices such as:

- Communications devices (fixed, mobile & IP phones)
- Computing devices (computer, PC, laptop, PDA, etc.)
- Peripheral devices (printer, scanner, digital camera, camcorder, etc.)
- Electronic devices (TV, stereo equipment, loudspeakers, etc.)
- Electronic appliances (fridge, espresso machine, washing machine, etc.)
- On-body or Off-body
- Mobile or stationary
- With or without network connections
- With or without local wireless connections.

It will be necessary with a careful and unambiguous definition of device types but for the time being it is sufficient to consider the device types at a coarse level as follows:

- *Primitive device:* Simple device that cannot operate alone but is slave to other devices, e.g. earphone, microphone, display, etc.
- *Open device:* Device allowing the installation and execution of PAN middleware and other applications, e.g. PDA, laptop, workstation, etc.

- *Closed device:* Device that can operate stand-alone but is like a black box; i.e. its internal structure and functions are totally hidden and communicates with the environment through only well defined interfaces. For instance, a digital camera is an advanced device having processor, memory, input and output units but does not allow the installation and execution of foreign applications. It communicates with the rest of the world via interfaces like serial, parallel, USB, etc.

In order to identify and define the capabilities and functions in the PAN Virtual Device Middleware (PVDM) let us consider two PAN configurations:

- Isolated PAN with multiple open device
- Networked PAN without open device.

6.1 Isolated PAN with multiple open devices

In this PAN configuration, there are multiple Open devices, multiple Closed devices and multiple Primitive devices on the PAN. Figure 8 shows a PAN with two Primitive devices: microphone and earphone, two

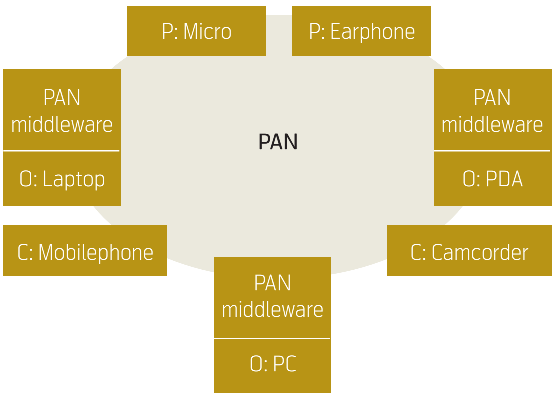


Figure 8 A PAN with multiple Open devices

ISSN 0085-7130 © Telenor ASA 2005

Closed devices: Mobile phone and Camcorder, and three Open devices: PDA, laptop and PC. The PAN Virtual Device Middleware should have the following capabilities and functions:

A. Device profile

The PVDM should be able to detect and to recognise what are the devices present in the PAN, what type they are and what capabilities they have or what service they are offering, e.g. printing, sound input, text displaying, etc. The definition of device types must be logical and unambiguous. It must be logical to address both composite devices and inherited devices. It must be unambiguous to guarantee the identification of a device type. The definition of device types should be standardised.

B. Device discovery and updating

To support the PAN's dynamics, where devices appear and disappear, the PVDM should be equipped with mechanisms for Device Discovery and Updating. One method is to carry out polling periodically. Since Primitive devices may be tied to their master and not visible to others, the PVDM should also ask Open and Closed devices for Primitive devices which are connected. In a second method, the PVDM only runs the device discovery once and relies on triggering from the network layer to update its device set. In a third method, the changes are only updated when an attempt to reach a device upon request from an application fails. The best options will be determined based upon the operational scenarios of Activity 2. The PVDM should also store information about the present devices. It is necessary to have a naming convention, e.g. PDA 1, Mobile 2, etc. The network addresses (IP, Bluetooth, etc.) and protocols for each device must also be stored.

C. Device input and output redirection

The PVDM should be able to redirect an output stream from one device to the input stream of another. For example, the sound output stream from the PDA can be redirected to an earphone and the sound input redirected to a microphone. It must intercept the output stream addressed to a device at an output port, perform the necessary transformation and send it to another device on the PAN. An API (Application Programming Interface) will be defined for this functionality.

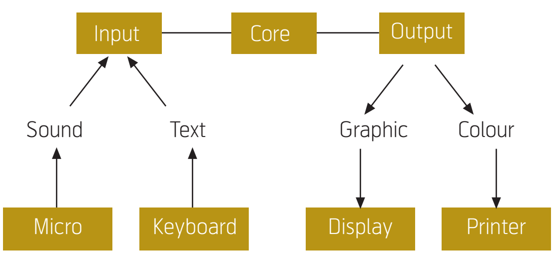


Figure 9 PAN Application structure

D. Application/service input and output redirection

The input and output redirection should also be done at the application/service level, i.e. one application can be redirected to one device while another is redirected to another device. However, this capability is more difficult since it depends on the feature of the operating system and may also require a new application structure as shown in Figure 9. An application consists of three components: Core, Input and Output. The Core component contains the main logic of the application and should remain invariant. The Input and Output Components should be specified according to the services supported such as sound input, text input, graphic display, colour printing, etc. If a device offers a service matching the one required then it can be used for the application.

E. User profile

The user and owner of the PAN should be reserved the right to define the following:

- What devices are allowed to participate in her PAN when present within the PAN's coverage?
- How should the devices be used at a particular location or at a particular time or according to a certain timetable. How should the redirection between devices occur?
- Which devices should be internal and which should be global and visible to the outside world?
- Which applications can execute and what are their settings?
- How should the applications be distributed among the devices, i.e. how should the components be distributed among devices according to criteria such as quality, price, battery level, etc.?

In addition, the user must have the ability to alter, remove, and add devices and applications at any time and anywhere. These hard requirements induce a serious challenge on the design of the User Profile structure and the architecture of the system around it.

F. Interface to the user

Since the user, as owner of the PAN, should be the one having the right to decide everything, there is a need for an application for PAN Control that offers interfaces allowing him/her to communicate and control the PAN Virtual Device. As with other applications, the PAN Control should also be structured as Core, Input and Output and should support a variety of input and output services.

G. Interface to applications and services

With the emergence of the PAN, a brand new type of PAN-based applications is born. This new type of applications will actively take advantage of the unique properties of the PAN: device dynamics and resource sharing. A typical example of PAN-based applications is an application which focuses on the control of existing applications, their parallel execution, their flexible and dynamic composition, and the distribution of their input and output. The PVDM should provide an Application Programming Interface (API) allowing the application access to the PAN capabilities and functions. It is also necessary by careful study to decide which technology, e.g. Java, CORBA, XML Web service, etc. should be used to implement such an API.

H. Optimal resource sharing

In addition, it is desirable to achieve better collaboration and more optimal resource sharing between the Open devices. There are several alternatives to achieve improved Resource Sharing as follows:

1 *Distributed Operating System:* Such an alternative demands the implementation of the Distributed OS on all of the Open device types. With rapid growth in type of mobile devices this could be difficult.

2 *Distributed Computing:* The traditional Distributed Computing middleware must be extended with functionality to cope with the dynamics of devices.

3 *XML Web Services:* The devices are more loosely coupled in this alternative and it can cope better with device dynamics [9][10].

4 *Application distribution, coordination and control:* the PVDM performs the distribution, coordination and control of applications based on the Core, Input and Output application structure. The distribution of the input and output components is already taken care of by the Application Redirection Function described earlier. Concerning the Core, it could only be moved and resumed to a compatible Open device.

The functionalities of the PVDM are depicted in Figure 10.

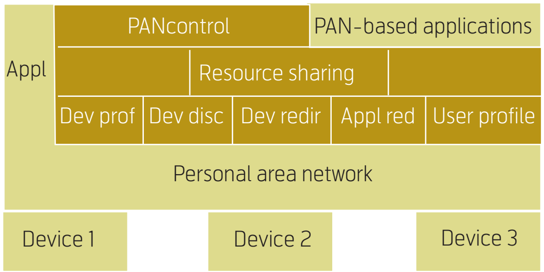


Figure 10 PVDM for PAN with multiple Open devices

6.2 Networked PAN without Open device

In this configuration the PAN does not have any Open device but only either Primitive or Closed devices. Usually without an Open device, the PAN will collapse but fortunately there is a way to remedy the situation if there is one device with network connection. As shown in Figure 10, the mobile phone allows communication with a computer running the PAN Virtual Device Middleware (PVDM). In this case the PAN can function very well because it is logically equivalent to the configuration Isolated PAN with unique Open device if the Computer on the network is considered as a PAN device. This configuration is however very interesting since it opens for several relevant usage scenarios. Indeed, the network computer can be a home PC or an office workstation that the user always leaves behind when traveling, but it can still participate and give support to her PAN. This network computer can also be provisioned and managed by a Service Provider that offers PAN service to the users.

It is essential to consider the role of the network operator. It is interesting to investigate the newly associated business model.

7 Conclusion

In this paper we introduce a novel concept, which proposes to realise a Virtual Device based on autonomous devices in the Personal Area Network. We also propose a PAN middleware necessary for the realisation of the Virtual Device. The capabilities and functions of the proposed PAN middleware are identified and explained. However, we are still in the

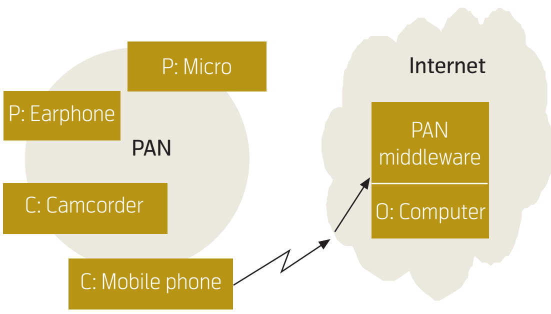


Figure 11 A PAN without Open device

ISSN 0085-7130 © Telenor ASA 2005

analysis phase. In order to obtain a running PAN middleware, the design, implementation and testing should be carried out and many challenging issues such as the device discovery, the design of the user profile, the choice of appropriate approach for resource sharing, etc. must be treated. Last but least, the PAN will only take off with the wide availability of short-range wireless technologies that consume less energy.

References

1 Do van Thanh, Jønvik, T, Vanem, E, Tran Dao van, Audestad, J A. The Device Management Service. *Proceedings of The IEEE Intelligent Network Workshop 2001 (IN2001)*, Boston, USA, May 6–9, 2001. (ISBN 0-7803-7047-3)

2 Vanem, E, Tran Dao van, Jønvik, T, Do van Thanh. Extending VHE with the Device Unifying Service. *Proceedings of The IEEE International Conference on Comunications ICC 2002*, New York, USA, April 28 – May 2, 2002. (ISBN 0-7803-7400-2 (softbound) ISBN 0-7803-7401-0 (CD-ROM))

3 *Bluetooth*. https://www.bluetooth.org/

4 *IEEE*. http://standards.ieee.org/getieee802/portfolio.html

5 ITU-T X.901 | ISO/IEC 10746-1. *Open Distributed Processing Reference Model Part 1 – Overview*. 1998.

6 ITU-T X.901 | ISO/IEC 10746-2. *Open Distributed Processing Reference Model Part 2 – Foundations*. 1996.

7 ITU-T X.901 | ISO/IEC 10746-3. *Open Distributed Processing Reference Model Part 3 – Architecture*. 1996.

8 Object Management Group. *CORBA V2.0 – The Common Object Request Broker: Architecture and Specification, Revision 2.0*. July 1995, Updated July 1996.

9 *Executive Overview: Web Services*. The Forrester Brief, Forrester Research Inc., October 2001.

10 Cerami, E. *Web Services Essentials*, 1st edition. O’Reilly Media Inc., 2002.

For a presentation of Do van Thanh, please turn to page 2.

For a presentation of Ivar Jørstad, please turn to page 21.

IP-based resource discovery in fixed and mobile networks

PAAL E. ENGELSTAD AND GEIR EGELAND

Paal E. Engelstad is Research Scientist in Telenor R&D

Geir Egeland is Research Scientist in Telenor R&D

This article introduces the most common IP-based naming services and service discovery mechanisms. The reader is made familiar with the most common naming services used on the Internet and how names can be resolved in mobile wireless ad hoc networks. Further, the reader is introduced to service discovery mechanisms for an ad hoc network where traditional service discovery mechanisms might be deficient, and where service discovery is particularly important since the availability of services is dependent on the network dynamics.

1 Name resolution

As human beings, we prefer to remember the name of a computer. Computers, on the other hand, prefer to address each other by numbers, which on the Internet is 32 bits or 128 bits long, depending on whether IPv4 or IPv6 is used. This is one reason why we need a naming service that can handle mapping between computer names, which we humans find convenient to remember, and between network addresses (i.e. numbers), which computers deal with. Another reason is that according to the Internet model, an IP address does not identify a host, such as a web-server, but a network interface. Although the host makes changes to its network interface or network attachement, it is convenient for the users and applications that the name of the host remains unchanged. As such, keeping different identifiers at different layers helps keep the protocol layers more independent and also reduces problems associated with layering violations.

In a middleware perspective, naming services is also a question of keeping higher layer names of entities independent of their lower layer identfiers and their actual locations. Here, the naming service is not only a way of helping the users of applications; it is just as much a help for the software developer.

In this section, the reader is made familiar with the most common naming services used on the Internet and how names can be resolved in mobile wireless ad hoc networks.

1.1 An architecture for naming services

A fundamental facility in any computer network is the *naming service*, which is the means by which names are associated with network addresses, and network addresses are found based on their names. For example, when you use an electronic mail system, you must provide the name of the recipient to whom you want to send mail. If you want to access a web site you must provide its URL, which again is translated into the network address of the computer hosting the web site.

To give some examples, the Domain Name System (DNS) [1][2] maps the host name of the University of Oslo's public web server, which is www.uio.no, to the IP address `129.240.4.44`. Another example can be a VoIP system that maps a SIP identifier to an E.164 number, e.g. `URI sip:dave@my_telecom.com:5060` is translated to `+47 904 30 495`.

The association between a name and the lower layer identifier of a network entity is called a *binding*. Some naming services, such as DNS, also have the possibility to do reverse mapping, e.g. to map an IP address to a corresponding higher layer name, and to do mapping from one higher layer name to another. Some naming services, such as DNS, also have the possibility to do reverse mapping, e.g. to map an IP address to a corresponding higher layer name, and to do mapping from one higher layer name to another.

In the following section, we will describe a generic model for naming services, which will serve as a reference model for the remaining part of this article.

1.1.1 A generic model

The process of looking up a name in a computer network consists normally of the following steps: First, the binding between higher layer names and lower layer addresses must be registered in the network. This procedure can be referred to as *registration*. The registration is normally done only once, and the binding is provided through some administrator authority. The bindings are normally registered with a server that holds a binding cache. An example of such an entity is a DNS server. With a strict authentication regime, the registration can be done automatically, as illustrated in message 1 of Figure 1-1.

Second, the network entities that desire to resolve a name must be informed of the corresponding binding that is registered in the network. This procedure can be referred to as *name resolution*. There are two different principle approaches to name resolution, namely *push* and *pull*:

 ISSN 0085-7130

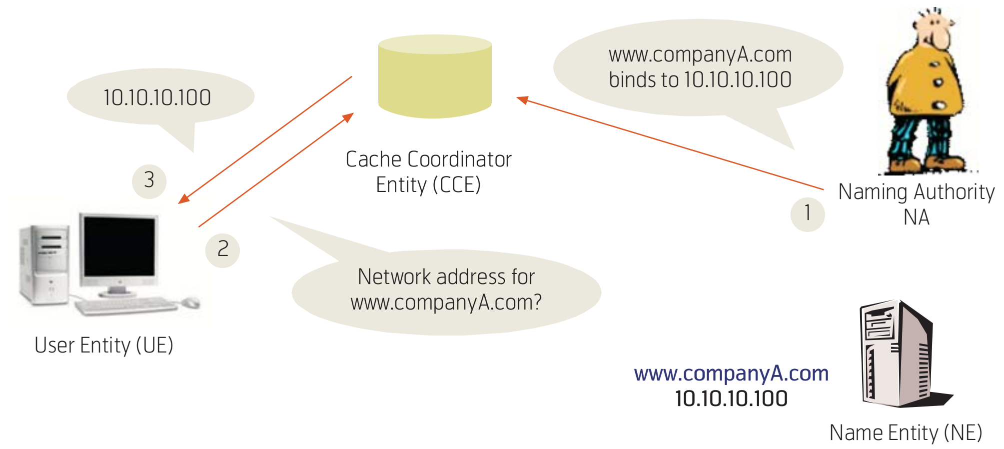


Figure 1-1 A Cache Coordinator Entity (e.g. the enterprise DNS server) is updated with a new binding for its Named Entity (e.g. a public web server)

- *Push approach:* The bindings are proactively broadcast to all network entities that might need to use the bindings for name resolution some time in the future.
- *Pull approach:* A network entity that desires to resolve a name to a network entity issues a request on demand at the time the binding is needed.

Due to scalability issues the *pull* model has been chosen for the Internet, and this model will be our focus in the following sections.

The generic model involves four networking entities for the registration and name resolution procedures.

These are:

1 A *User Entity* (UE) represents the network entity issuing a request to resolve a name.

2 A *Named Entity* (NE) represents the network entity that a name points to.

3 A *Naming Authority* (NA) is authorized to create a binding between names and addresses of Named Entities (NE).

4 A *Caching Coordinator Entity* (CCE) does intermediate storage of bindings.

User Entity (UE)

The role of the *User Entity* (UE) is to resolve the mapping between a name and the lower layer identifiers of an NE by retrieving a binding from the naming service. The UE may be software components or actual end-users that want to look up a specific name. In most cases the UE will offer a low level functionality directed towards system components.

In the DNS naming system, requests are issued by the *resolver* in the computer's operating system. The initiative to activate the resolver can come from an end user typing a web address in a browser, or from an application needing to access a binding. The request will end up at the entity caching the binding for the name requested, and the binding containing the resolved identifiers will be returned to the resolver. This is illustrated by messages 2 and 3 in Figure 1-1.

Named Entity (NE)

The *Named Entity* (NE) is the network entity (e.g. host or computer) identified by a name. For communication, it uses the network interfaces referred to by the lower layer identifiers of a binding.

Naming Authority (NA)

The *Naming Authority* (NA) is the authority or system of authorities permitted to assign names to NEs and bindings between the names and the lower layer identifiers of the NE. This is normally only configured once, but might require to be updated if parameters of the network configuration changes. For example, an ISP can perform the administrative task of configuring their DNS server to map the network address of a customers public web server to the network address assigned to the customer, or the network administrator of an enterprise network will configure the company's local DNS-server to map a computer's name to a fixed network address.

There exist solutions that enable an NE to update its own binding directly with a Caching Coordinator Entity (described below). This requires that the Caching Coordinator Entity can authenticate the NE. With no authentication, it would be possible for someone to insert false information into the caching entity, and in the worst case impersonate another network entity by hijacking its binding. The NE must

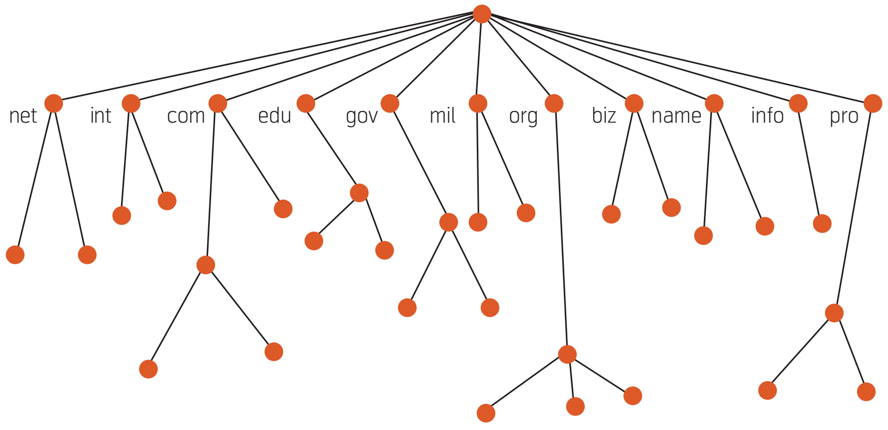


Figure 1-2 The structure of the DNS name space

also be authorized to automatic updates to its binding, and the node thus also has the role of an NA.

Caching Coordinator Entity (CCE)

The primary task of the *Caching Coordinator Entity* (CCE) is to act as a cache for name bindings. The CCEs are important for the efficiency of the naming service in the presence of a large number of UEs and NEs. Normally the NE knows the location of the CCE it is supposed to register its bindings with. The UE, which somehow has to retrive this information, does not normally know the location of the CCE where the binding is located. For the DNS system, the location of the local CCE, i.e. the local DNS server, is normally provided dynamically to the UE, by for example mechanisms such as DHCP. Here, the UE normally uses the local CCE to be able to locate and retreive a binding stored at another CCE on the Internet.

1.1.2 The Domain Name System

The DNS' distributed database is indexed by domain names. Each domain is basically just a path in an inverted tree, called the *domain name space*. Figure 1-2 illustrates the structure of the domain name space.

The practical operation of the DNS system consists of three modules:

1 The *DNS resolver* that generates DNS requests on behalf of software programs;

2 The *recursive DNS server*, which searches through the DNS in response to queries from resolvers, and returns answers to those resolvers;

3 The *authoritative DNS server*, which responds to queries from recursive DNS servers.

The DNS resolver acts as the UE described above, while the DNS servers act as CCEs. The registration is normally a manual process and the Named Entities normally do not take part in the name resolution process (unless they use DNS Secure Dynamic Updates [5]).

1.2 Resolving Names without the use of DNS servers

In some situations it is not feasible to make use of the DNS system to resolve name bindings, and in some cases the DNS system might not even be available. For example, consider a spontaneous setting where people in some airport lounge want to make use of their laptop's WLAN feature and connect to each other to exchange music or other information they may find interesting (Figure 1-3). Without any DNS service, they will have to identify themselves using the network address, which for human beings is not very attractive.

If it was somehow possible to define a separate name space in addition to the DNS system, and if some mechanism could advertise and resolve these names in such a spontaneous setting, users could search and identify names in a more human-friendly way. *Multicast DNS* [6] and *Link Local Multicast Name Resolu-*

Figure 1-3 Computers communicating without a DNS server

tion [7] are two competing solutions addressing this scenario. These will be described in the following.

1.2.1 Multicast DNS name resolution

Multicast DNS (mDNS) is a way of using familiar DNS programming interfaces, packet formats and operating semantics in a small network where no conventional DNS server has been installed [6]. In short, the mechanism works by enabling a node to search for the network address of the computer with the name *X* by sending a multicast DNS message asking *"Does anyone know the network address of node X?"* If a node with the name *X* is present on the network, it will respond by returning a DNS response containing information about its network adress. Multicast DNS is a part of the Mac OS X® operating system, where its implementation is called *Rendezvous*.

1.2.2 Link Local Multicast Name Resolution

Link Local Multicast Name Resolution (LLMNR) is a peer-to-peer name resolution protocol focused on enabling resolution of names on the local link [7]. LLMNR utilizes the DNS packet format and supports all DNS formats, types and classes. LLMNR is not intended as a replacement for DNS, and as a result, it is only used when a DNS server is either not available or is not providing an answer to a query.

LLMNR differs from mDNS in many ways. First, LLMNR is an IETF standards track specification, while mDNS, which is used in Apple Rendezvous, is not. LLMNR is designed for use only on the local link, while mDNS also offers site-wide usage. Furthermore, mDNS sends multicast responses as well as multicast queries.

1.3 Name resolution in ad hoc networks

Mobile ad hoc networking started out from military research on Packet Radio Networks. In the late 1990s, however, the issue was included as a working group item of the Internet Engineering Task Force (IETF). The goal of the IETF was "to develop a peer-to-peer mobile routing capability in a purely mobile, wireless domain. This capability will exist beyond the fixed network (as supported by traditional IP networking) and beyond the one-hop fringe of the fixed network." [8].

1.3.1 Characteristics of ad hoc networks

A mobile ad hoc network consists of mobile routers – often simply referred to as "nodes". They are free to move about arbitrarily and wireless technology is used for direct communication between the nodes. Due to the dynamic nature of the wireless media and the arbitrary mobility of the nodes, the network forms a random, multi-hop graph that changes with time. The network is an autonomous system that may operate in isolation, or it may optionally have gateways that connect it as a "stub" network to a fixed network infrastructure. Since a node is not necessarily in direct radio range with any other node in the network, the nodes must participate in the routing process and be willing to forward packets on behalf of other nodes in the network.

An ad hoc network is a network that is created spontaneously, without support from the existing fixed Internet infrastructure. The network might be formed when people equipped with their portable PCs come together at conferences, or it can be used to network a user's personal wireless devices into a Personal Area Network (PAN). Ad hoc networks may also be formed during emergency situations when legacy network infrastructures are unavailable or damaged. Yet another application of ad hoc networking is on a military battlefield in places where fixed network infrastructures are unavailable or not feasible to use.

The salient characteristics of ad hoc networks include not only the dynamics of the network topology. In addition, the links are bandwidth constrained and of ever-changing capacity. Furthermore, the nodes often rely on energy-constrained batteries to move about freely, and energy conservation is an important design goal for many ad hoc networking technologies.

Due to the dynamic networking topology and the fact that nodes might enter and leave the network frequently, it is often also assumed that an ad hoc network is without any pre-existing infrastructure and that it is difficult to maintain an infrastructure in such a dynamic environment.

Due to the lack of pre-existing infrastructure, it is anticipated that direct peer-to-peer communication between nodes will be popular on ad hoc networks. This means that any node may in principle operate as a server (e.g. Web server or SIP-server) and be contacted directly by other MANET nodes. Any node may also operate as a client and contact other servers available in the network.

A Mobile Ad hoc Network that is equipped with IP-based routing is normally referred to as a "MANET". There are two principal different approaches to routing in MANETs, namely reactive and proactive routing. These two approaches will be detailed below, however, with most emphasize on reactive routing.

1.3.2 Reactive and proactive routing protocols

The routing protocols in mobile ad hoc networks can be divided into two different approaches:

- Reactive
- Proactive

A reactive routing protocol has no prior knowledge of the network topology, but finds a route to a given destination on demand. A proactive routing protocol, on the other hand, tries to always have a complete updated picture of the network topology.

Reactive routing protocols are normally preferred when nodes are highly mobile, when only a subset of nodes are communicating at any one time, and when communication sessions last for a relatively long time. Proactive routing protocols, on the contrary, are preferred for lower levels of mobility and when communication is random and sporadic. Reactive routing is not well known to most people, probably because routing on the fixed Internet is proactive by nature. As a consequence, we will summarize the features of reactive routing in the following:

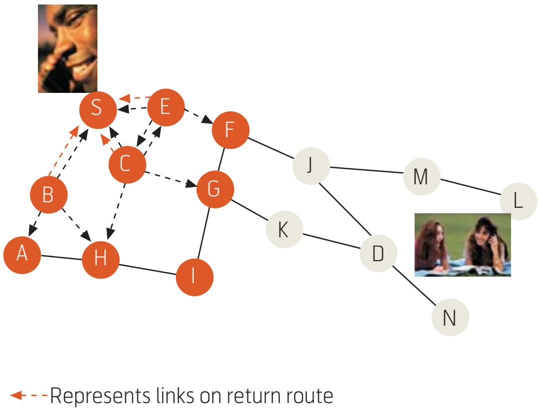


Figure 1-4 Route Requests (RREQs) in AODV

A number of reactive routing protocols have been proposed over the years. The most widely studied and popular proposals include the Ad-hoc On-demand Distance Vector (AODV [9]) routing protocol and the Dynamic Source Routing (DSR [10]) routing protocol.

Reactive protocols allow source nodes to discover routes to an IP address on demand. Most proposals, including AODV and DSR, work as follows: When a source router requires a route to a destination IP address for which it does not already have a route, it issues a route request (RREQ) packet. The packet is broadcast by controlled flooding throughout the network, and sets up a return route to the source (Figure 1-4).

If a router receiving the RREQ is either the destination or has a valid route to the destination IP address, it unicasts a Route Reply (RREP) back to the source along the reverse route. The RREP normally sets up a forward route from source to destination. Thus, the pair of RREQ and RREP messages set up a bi-directional unicast route between source and destination. Once the source router receives the RREP, it may begin to forward data packets to the destination. (The acronyms RREQ and RREP are borrowed from AODV.)

Most protocols let routes that are inactive eventually time out. If a link becomes unavailable while the route is active, the routing protocol normally implements an algorithm to repair the route. Often the router upstream to the link breakage would send an error message upstream towards the source.

AODV is a protocol that stores state information in the network. Routers that receive RREQs set up the return routes in the route tables as backwards pointers to the source router, while RREPs that are propagated back to the source along the reverse route leave forward pointers to the destination in the route tables. The Dynamic Source Routing (DSR) protocol, on the other hand, does not rely on routing state in the network. Instead, DSR uses source routing. The RREQ collects the IP addresses of all the nodes that it has passed on the way to the destination. The destination subsequently sends by source-routing a Route Reply back to the source of the request, providing it with the source route to the destination.

1.3.3 The importance of name resolution in MANETs

Name resolution is an important feature in an ad hoc network, since addresses may change relatively frequently due to the network dynamics, e.g. when nodes enter and leave the network. Furthermore, it is often impossible to use the address as a well-known identifier, since IP addresses for nodes on the MANET will normally be auto-configured at random, and nodes may also need to change addresses due to addressing conflicts. Devices and resources should instead be identified by stable and unique higher layer names (e.g. Fully Qualified Domain Names).

If a MANET is connected to the Internet, a MANET node may use the existing mechanism for name resolution on the Internet, namely DNS, to look up the IP address of another MANET node. However, in most scenarios the MANET will not always be permanently connected to a fixed infrastructure, and the DNS infrastructure on the Internet might be unavailable. Relying entirely on the DNS on the Internet would not be a robust solution to name resolution in the MANET.

One option would be to introduce a DNS infrastructure into the MANET. However, DNS is designed with a fixed network in mind, and has a relatively static, centralized and hierarchical architecture that does not fit well to MANETs.

Without a name resolution method in place, MANET users cannot easily use the applications that are developed for fixed networks for local communication on the MANET. In the following, solutions to name resolutions in ad hoc networks will be explored. The focus is mainly on name resolution in reactive MANETs, because name resolution in proactive MANETs might be a less challenging task.

1.3.4 Architectures for resolving host and service names in ad hoc networks

For name resolution in ad hoc networks, a MANET node may hold the same role as presented above for fixed networks. A node may act as a User Entity (UE) that wants to resolve a name, a Named Entity (NE) that wants to make its services available to other MANET nodes, and/or a Caching Coordinator Entity (CCE) that holds a central repository for cached bindings and assists other UEs and NEs with name resolution. A binding maps a name to an IP address(es) and possibly a port number(s) that the UE may subsequently use to contact the NE.

There are three name resolution architectures that need to be considered for ad hoc networks:

1 *Distributed architecture*
2 *Coordinator based architecture*
3 *Hybrid architecture*

These name resolution architectures will be described in the following.

Distributed architecture

As shown in Figure 1-5, this architecture contains no CCE. Instead, a UE floods the *Name Resolution Request (NREQ)* throughout its surroundings in the network (1). The flooding can be limited by a *flooding scope* parameter. Each NE responds to a *NREQ* for its own names (i.e. no name caching is allowed) with a unicast *Name Resolution Reply (NREP)* (2).

Coordinator based architecture

Certain nodes in the MANET are chosen to be Caching Coordinators (CCEs), a role quite similar to a DNS server. The interaction between User Entities, Named Entities and Caching Coordinators are illustrated in Figure 1-6. The CCEs announce their presences to the network by periodically flooding CCE Announcement messages (1). The flooding can be limited to a certain number of hops, determined by the *Coordinator announcement scope* parameter. Due to the dynamics of ad hoc networks, the Named Entities must be allowed to register their bindings automatically with the CCE. Hence, a Named Entity that receives CCE Announcements unicasts *Name Registration* messages to register its bindings (i.e. names and associated IP addresses) with CCEs in its surroundings (2). A User Entity that has received CCE Announcement messages may unicast a *Name Resolution Request (NREQ)* to a selected CCE to discover desired services (3). The CCE finally responds with a unicast *Name Resolution Reply (NREP)* (4). The selected CCE is often referred to as an affiliated CCE.

Hybrid architecture

This architecture combines the two architectures described in previous sections. UEs within the Coordinator announcement scope of one or more CCEs will register their bindings with them. They must however also be ready to respond to flooded NREQs. When a UE unicasts an NREQ to its affiliated CCE in line with the *Coordinator based architecture* (Figure 1-6), the CCE responds with a positive or negative NREP. However, if there is no CCE in the UE's sur-

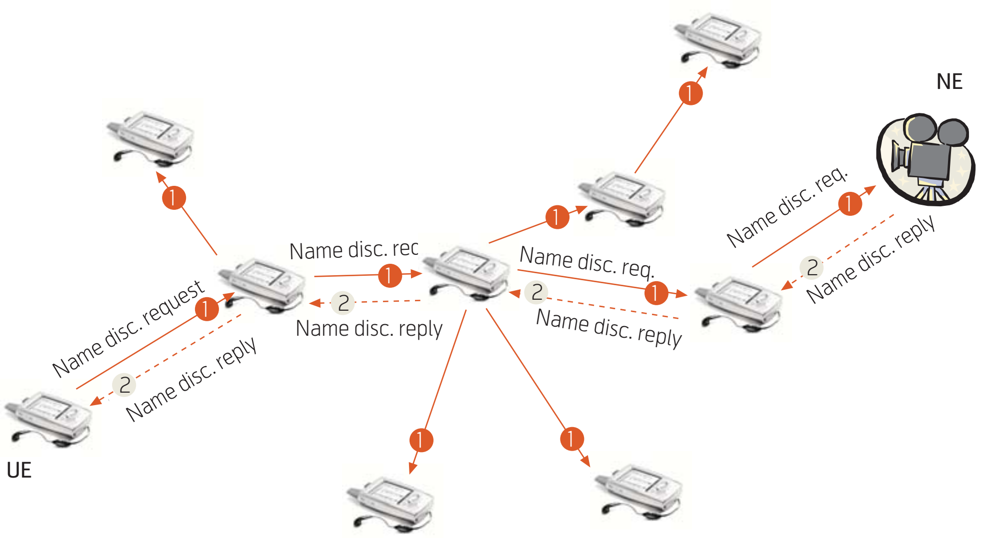


Figure 1-5 The Distributed architecture with User Entities (UE) and Named Entities (NE)

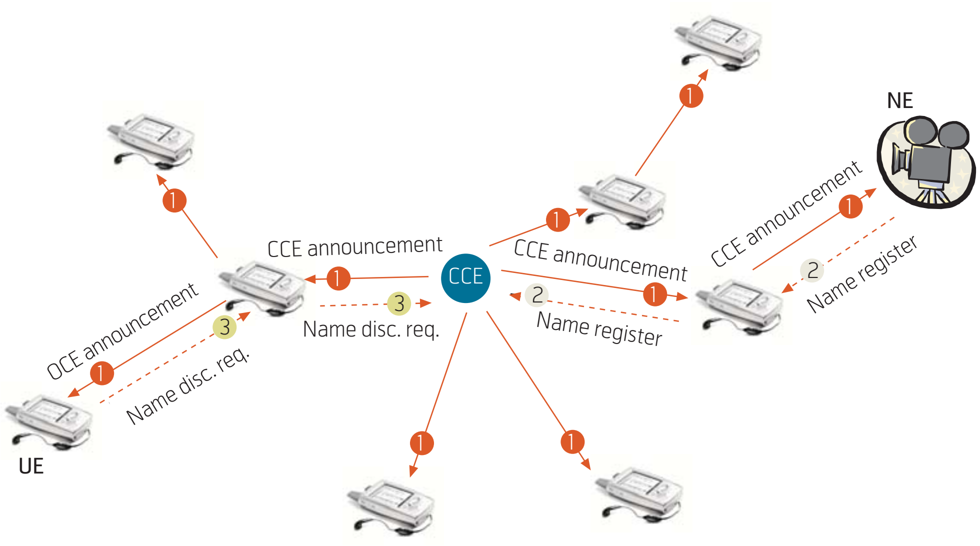


Figure 1-6 The Coordinator based architecture with User Entities (UE), Caching Coordinator Entities (CCEs) and Named Entities (NE)

roundings or if the affiliated CCE returned a negative NREP, the UE will simply fall back to the *Distributed architecture* (Figure 1-5). Both CCEs and Named Entities may respond to a flooded NREQ with a positive NREP that matches the requested service.

Intermediate node caching

An additional alternative – or a supplement – to the three architectures is to use intermediate node caching. The name resolution may for example follow the distributed architecture, but intermediate nodes are allowed to cache bindings found in NREPs that they are forwarding. Later, when receiving an NREP for a cached binding, the intermediate node resolves the name on behalf of the NE according to the cached binding. The bindings should contain a lifetime value that controls for how long a binding should be kept valid in a cache.

1.3.5 Emerging principles for name resolution in reactive ad hoc networks

Many ad hoc routing protocols are designed to conserve the scarce networking resources by reducing the need for and the negative impact of system-wide flooded broadcasts. Flooded broadcasts not only exhaust the available bandwidth on the network and reduce the scalability in terms of number of nodes accommodated on the network. Broadcasts also consume battery power of all networked devices.

Reactive routing protocols, such as AODV, are designed to reduce to the furthest extent the need for system-wide flooded broadcasts associated with route discovery. Although route discovery is efficient in terms of reducing the number of flooded broadcasts from two to one, name resolution that is not optimised with respect to the route discovery would not work efficiently with reactive routing. The process of contacting a node on the MANET would require two or three broadcasts, as illustrated in the left side of Figure 1-7. Two broadcasts are necessary for the initial name resolution, because the User Entity first has to flood a name resolution request. The reply returned by a node that can resolve the name also requires flooding, since the node does not have a route to the node that issued the request. Finally, the User Entity will have to flood a regular RREQ to find a route to the resolved IP address.

Alternatively, if the node resolving the name to an IP address is the Named Entity of the name (and not a node that has cached the name binding), the specification might mandate that the reply be returned by unicast to the User Entity. Then, before replying, the node must first flood an RREQ to discover and set up a unicast route to the User Entity and send the name resolution reply by unicast along this route. It would then be possible to reduce the number of flooded broadcasts from three to two, because the User Entity already has a route to the resolved IP address as a result of the name resolution process when it contacts the Named Entity. Further details are provided in [11][12].

It is however possible to reduce the number of flooded broadcasts to one, as illustrated in the right side of Figure 1-7. The solution is to use routing messages as carriers for name resolution. First the User Entity floods the name resolution request (NREQ). By piggybacking the NREQ on a route request (RREQ) packet, a "return route" to the User Entity would be formed as part of this flooding. By also piggybacking a name resolution reply (NREP) on a route request (RREQ) packet, the NREP is sent by unicast along the "return route" to the User Entity. The

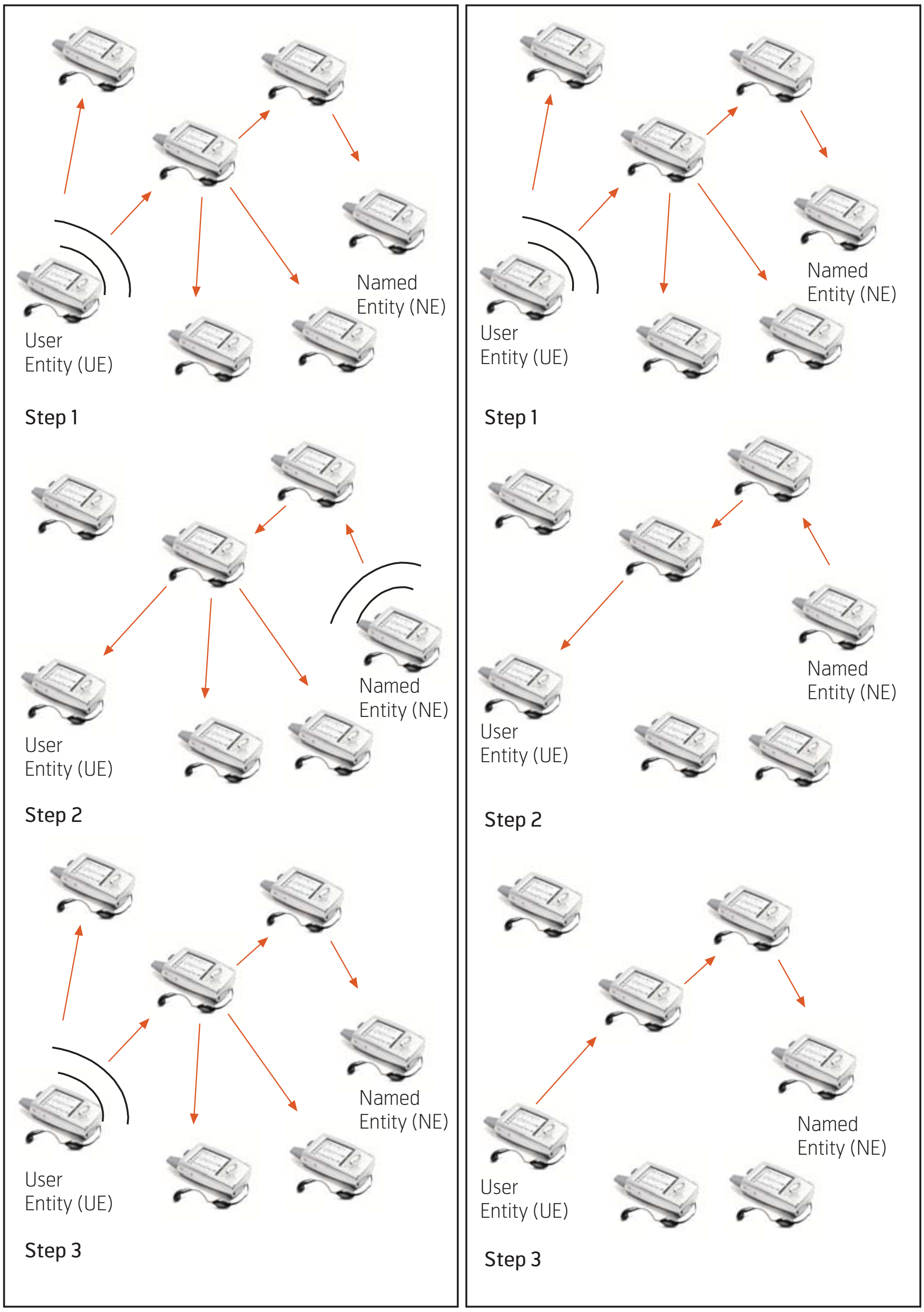


Figure 1-7 Name resolution without optimization (left) and with optimization (right)

RREP also ensures that a "forward route" is formed as part of this transmission. When the User Entity finally contacts the service at the IP address of the resolved name, the service request is unicast along the forward route that was put in place by the RREP. In summary, only one flooded broadcast is required in total. The idea of using routing messages as carriers has been proposed for name resolution in [11]. In fact, the same mechanism can also be used for service resolution, as we will see in the next part on service discovery below.

Needless to say, this broadcast issue is a smaller problem in proactive ad hoc networks, since all unicast communication (including the NREP) can be sent along unicast routes established beforehand by the routing protocol. However, the broadcasting of the NREQ might benefit from reusing the efficient flooding capabilities (e.g. using Multi-point relays) that are built-in features of many proactive protocols, including the *Optimized Link State Routing* (OLSR) protocol [13].

1.3.6 A proposal for name resolution in reactive ad hoc networks

Overview

A mechanism for name resolution in MANET is proposed in [14]. It is mainly targeted at users that can supply their MANET node with a Fully Qualified Domain Name (FQDN) from the globally unique DNS name space. The user may have control over some part of the DNS name space or may have received the FQDN from an organization that they belong to or subscribe to. The proposed name resolution scheme shares similarities to the Link-Local Multicast Name Resolution (LLMNR) protocol and multicast DNS protocol for local-link name resolution presented above. The mechanism proposed for ad hoc networks specifies compressed message formats allowing for bandwidth-efficient name lookups. As an option, it also specifies message formats that reuse the format of DNS messages, allowing for name lookups that are fully compatible with DNS.

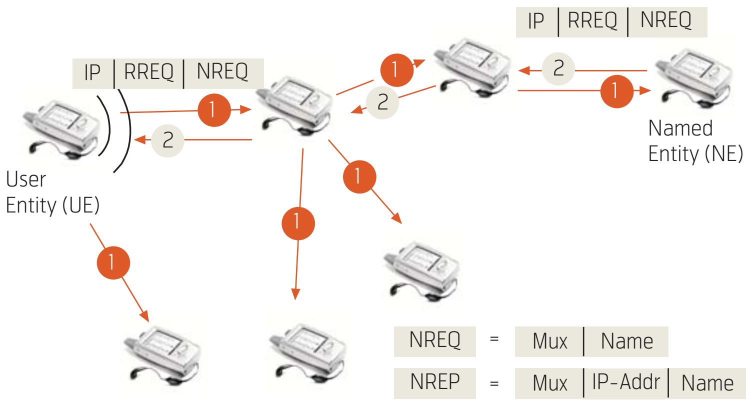


Figure 1-8 A Name Resolver (NR) floods a Name Resolution Request (NREQ), carried by a Route Request (RREQ) header, throughout the network (1). A Name Server (NS) process with the requested name-to-address mapping unicasts a Name Resolution Reply (NREP), carried by a Route Reply (RREP) header, back along the reverse route formed by the RREQ (2)

Name Resolution Requests and Replies

The proposed scheme uses the distributed architecture presented above, with no intermediate node caching. No Caching Coordinators are allowed, and instead, only User Entities and Naming Entities are present on the MANET. When an NREQ is broadcast by flooding throughout the MANET, each node with a Named Entity processes the request. By carrying the NREQ as an extension to an RREQ (Figure 1-8), the number of broadcasts required for name resolution is reduced as explained earlier in this article. Hence, a return unicast route to the User Entity of the request is already in place for a node that wants to respond to the NREQ.

The destination IP address contained in the RREQ, indicating the address to which a route is searched for, is set to a pre-defined value. This can be a zero address, a broadcast address or a pre-assigned multicast address to which no node can cache a route. Hence, intermediate nodes without a valid address mapping for the requested name will not respond to the RREQ part of the message.

The NREP is carried as an extension to an RREP message (Figure 1-8). The User Entity sending of the NREP will normally include its own IP address as destination IP address in the RREP message to ensure that a forward route is formed.

By carrying the response in an RREP message, a responder that is identified by the name that is searched for, can supply the User Entity with the resolved IP address in addition to a unicast route to that IP address. Hence, the User Entity does not have to issue an additional broadcast to discover a route to the resolved address when it subsequently tries to contact that address.

Interaction with External Networks

MANETs might be connected to external networks through Internet Gateways (IGWs). An IGW is a MANET router that also is a host or a router on an external network (with Internet connectivity). The IGW may have access to a conventional DNS server over the external network and it may also provide other MANET nodes with access to the external net-

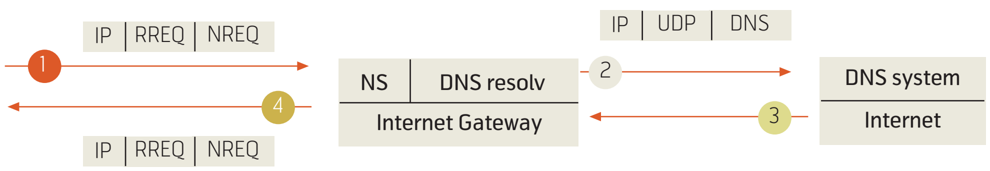


Figure 1-9 An Internet Gateway (IGW) which receives a Name Resolution Request (NREQ) on the MANET (1) may try to resolve the requested name using the Domain Name System on the Internet (2). A successful response (3) may be injected as a Name Resolution Reply (NREP)

ISSN 0085-7130 © Telenor ASA 2005

work. The scheme proposes to use each IGW as a DNS proxy, as shown in Figure 1-9.

The main advantage of using the IGW as a DNS proxy is that there is only one name resolution scheme used on the MANET and that nodes can resolve names in one single process.

Response Selection

A flooded NREQ might result in reception of multiple NREPs. If the Named Entity present in the MANET has registered its name in the DNS, both the Named Entity and each IGW present in the network may return an NREP. Furthermore, if the NREP contains a DNS SRV resource record to resolve a non-unique service name (as explained later in this article), many Named Entities present in the MANET may respond to the same NREP. To deal with the possibility of multiple responses, the User Entity should wait for some milliseconds to collect responses that might arrive. The proposal includes response selection rules that ensure that a response from a Named Entity present locally on the MANET will always have preference over responses arriving from other nodes, such as IGWs, since a local response might be more reliable and up-to-date. Furthermore, a direct route through the MANET should normally have preference compared to a route that goes through external networks. If the User Entity has multiple addresses to select from after applying this selection rule, it should select – as a secondary selection rule – an IP address to which it has a route that is preferred by the routing protocol. That means that it will normally select an IP address to which it has valid routes and select the IP address that is the fewest hops away from the User Entity.

2 Service discovery

The only information two end-points communicating over the Internet need to know, besides their own configuration, is the network address of the end-point they communicate with. Any of the two end-points might offer a wide range of services such as email, ftp, http etc., but there exists no defined way for the other end-point to find out which of these services are being offered. Instead, the users themselves have to remember the name of the computer offering the services and to know in advance the TCP or UDP port numbers associated with the set of services desired or to rely on the well-known port numbers.

Would it not be easier if services and the associated IP addresses and port numbers could be searched for and discovered dynamically? This was indeed the case for Macintosh computers that used AppleTalk networking. The Mac user did not require any assistance from a network administrator or even a complicated manual to locate services. Service discovery worked automatically and was operated by using a simple interface.

This section will introduce the reader to various service discovery mechanisms that enable the same type of functionality that the AppleTalk protocol offered [15]. Further, the reader is introduced to service discovery mechanisms for an ad hoc network where traditional service discovery mechanisms might be deficient, and where service discovery is particularly important since the availability of services is dependent on the networks dynamics.

2.1 A generic model

Users typically want to accomplish a certain task, not query a list of devices to find out what services are running. It makes far more sense for a client to ask a single question: *"What print services are available?"* than to query each available device with the question, *"What services are you running?"* and sift through the results looking for printers. This latter approach, which is called *device-centric*, is not only time-consuming, it generates a tremendous amount of network traffic, most of it useless. On the other hand, a *service-centric* approach sends a single query that generates only relevant replies. This is what service discovery is all about.

The process of discovering services on a computer network, is similar to the process of looking up a name described in the previous chapter, but instead of asking *"Who has this name?"*, service discovery involves the question *"Who offers these services?"*. Normally three entities, or agents, are needed in a service discovery architecture for a computer network. These are:

1 A *User Agent (UA)* that represents the network entity issuing a request to find a service;

2 A *Service Agent (SA)* that represents the service offered;

3 A *Directory Agent (DA)* for intermediate storage of information about available services.

2.1.1 User Agent

The User Agent (UA) represents the client part in the process of discovering services. The UA may be a software component or an end-user that wants to locate a specific service. In most cases the UA will offer a low level functionality directed towards system components.

2.1.2 Service Agent

The Service Agent (SA) represents the service in the architecture. This can be the actual service, or some entity representing it. Such an entity is called a proxy-SA. An SA will advertise the services it offers by either broadcast or multicast service messages. If a Directory Agent exists, the SA will try to register with it.

2.1.3 Directory Agent

A Directory Agent (DA) acts as a cache and will merely collect information from the Service Agents and forward it on demand to User Agents. The UA and SA can use either passive or active DA discovery to find a DA. This is shown in Figure 2-1.

The service model can be divided into three architectures, the *distributed*, *centralized*, and *hybrid* architectures. In the distributed architecture, there are no DAs present. The UAs will issue a multicast message to find the SAs. The reply from the SAs can be unicast, or multicast if the UAs have the ability to use and manage a local cache of available SAs. In the centralized architecture, one or more DAs are present.

A variation of the centralized architecture is shown in Figure 2-2, where the UA is retrieving service discovery bindings from the DA according to the *pull* model. In this example, UAs are also proactively caching announced bindings according to the *push* model, as illustrated in the left part of the figure. (The

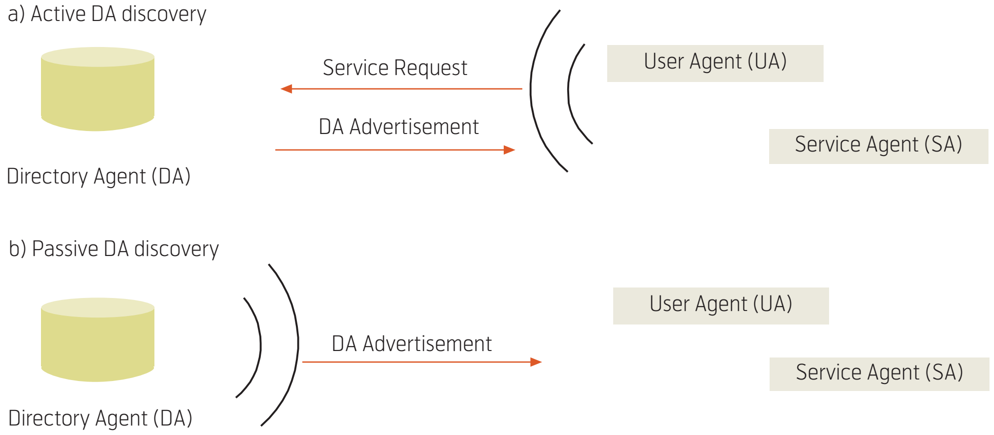


Figure 2-1 Active and passive Directory Agent (DA) discovery

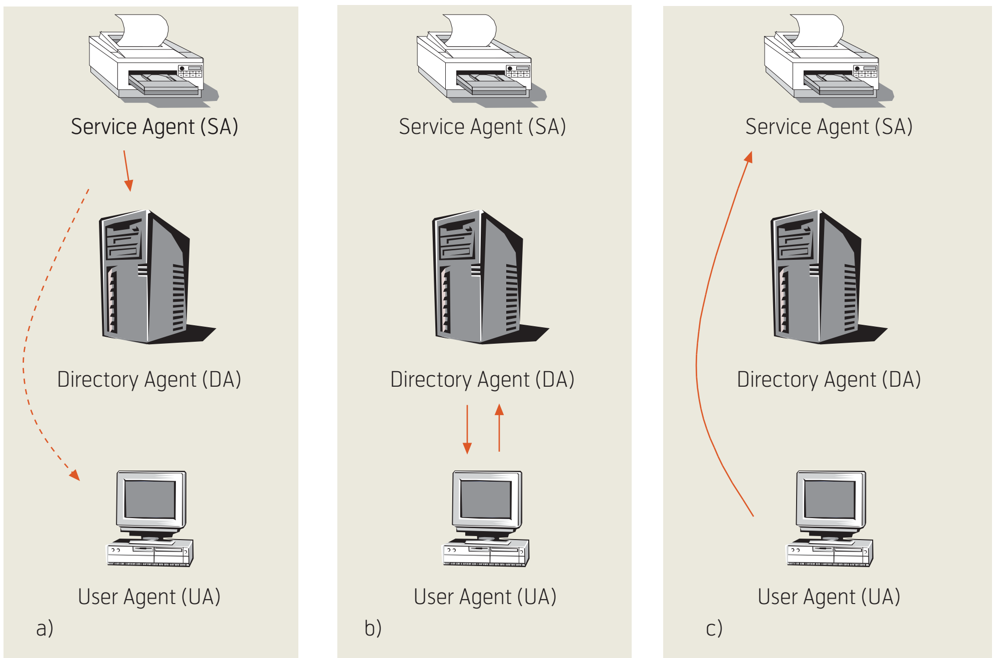


Figure 2-2 A Service Agent announces its services, and announcements can be cached in both the DA and the UA (a). If the UA is searching for a service it has not heard being announced, it can broadcast a service request or contact a DA directly. The DA will respond with information about where the service can be located (b). The UA can access a service it has learnt about via a DA, or from the SA directly (c)

ISSN 0085-7130 © Telenor ASA 2005

push and *pull* approaches were described in the previous chapter on name resolution.)

In the hybrid approach, a UA will first try to contact a DA according to the centralized architecture, and fall back to the distributed approach if no DA can be located.

2.2 Service discovery on the Internet

2.2.1 Current practice for service discovery

When an application on computer *A* wants to connect to an application on computer *B*, computer *A* requires the network address of computer *B* and the port number of the service. The network address is necessary to route *A*'s request to *B*. Since *B* may offer a multitude of services, a port number is used to distinguish between the different services offered at the same network address. Existing practice in IP networks is to go through a three-step process to obtain the network address and the port number of the services. The three steps consist of:

1 Mapping the service name to the name of the computer offering the service;

2 Mapping the port number of the service to a service name;

3 Resolving the computer name to a network address using DNS or a local name service.

There is no widely used mechanism to undertake the first step of this process in today's IP network. A common method of advertising services is to map the network address of the computer offering services and the service name with the DNS service. Examples of such services can be a public web server which is given the name www, or a public file transfer server which is given the name *ftp*.

Example:

www.some_domain.com
ftp.some_domain.com

The procedure for mapping port numbers to a service name is pretty simple, and a one-to-one relationship exists between the service name and the port number. The port number is assigned by the Internet Assigned Number Authority (IANA) and is normally maintained in a database on the local computer. The mapping between a service name and a port is typically:

```
<name> <port number>/<transport
protocol> <aliases>
```

An example of a port to service name mapping can be:

```
http  80/tcp  www  www-http
```

The mapping of a network address to the computer offering the service is normally done using address resolution through DNS or through a cache at the local computer, if there is no access to any name server.

As can be seen, the procedure for discovering services on the Internet is cumbersome and not very efficient. Other alternatives will be outlined in the following.

2.2.2 Service Location Protocol

The Service Location Protocol (SLP) is an emerging Internet standard provided by IETF for automatic service discovery on the Internet [16]. SLP provides a framework to allow networking applications to discover the existence, location, and configuration of networked services in enterprise networks. Traditionally, in order to locate services on the network, users of network applications have been required to supply the host name or network address of the machine that provides a desired service. Ensuring that users and applications are supplied with the correct information has, in many cases, become an administrative nightmare.

SLP was inspired by the AppleTalk ® protocol [15], which was a mechanism created by Apple and which proved to be a huge success much due to the simplicity and benefits of the solution. The only drawback was that it did not scale very well.

The main focus of SLP is to be a mechanism that acts as an enabler for Plug and Play functionality in IP networks with automatic and dynamic bindings between services and service users.

The SLP protocol introduces three major components in the network:

1 *User Agent (UA):* The SLP *User Agent* is a software entity that looks for the location of one or more services. This is usually implemented (at least partially), as a library to which client applications link, and it provides client applications with a simple interface for accessing SLP-registered service information.

2 *Service Agent (SA):* The SLP *Service Agent* is a software entity that advertises the location of one or more services. SLP advertisement is designed to be both scalable and effective, minimizing the use of network bandwidth through the use of targeted

multicast messages, and unicast responses to queries.

3 *Directory Agent (DA):* The SLP *Directory Agent* is a software entity that acts as a centralized repository for service location information. In a large network with many UAs and SAs, the amount of multicast traffic involved in service discovery can become so large that network performance is degraded. By deploying one or more DAs, both SAs and UAs make it a priority to discover available DAs, since the use of a DA minimizes the amount of multicast messages sent by the protocol on the network. The only SLP-registered multicast in a network with DAs is for active and passive DA discovery.

SLP introduces dynamic naming services without the need for any centralized name server or other agents. Since SLP uses IP multicast for this purpose, it requires the cooperation of IP routers that implement IP multicast. IP multicast is used for such features as IP-based audio and video broadcasting and video conferencing, but IP multicasting may not be completely implemented across some intranets. In the absence of IP multicasting, SLP name lookups will only work within the subnet on which they are performed, or within the groups of subnets over which IP multicast is supported.

The SLP protocol suffers from the lack of implementation support, since companies such as Apple and Microsoft, each on their own push a different service discovery technology. Generally, standardisation is a good thing, but not very useful if no one provides implementation support for the standards. The Service Location Protocol has proved to be useful, especially for UNIX® variants, and an open source project exists (an implementation of SLP can be downloaded from www.openslp.org) to support service discovery on operating systems such as Linux and BSD.

2.2.3 DNS Service Resource Records

An alternative to building up a new SLP infrastructure on the Internet is to reuse the existing DNS infrastructure and allow for service discovery as an extension to DNS. Extensions to DNS are enabled through the use of DNS Resource Records (DNS RRs). The most common resource record for service location is the DNS SRV resource record [17].

DNS SRV was originally designed to locate services on the global Internet. As an example, let us assume that *company_A* has implemented the use of SRV in its DNS server. Entries for the protocols *http* and *smtp* would look like this in the zone file for *company_A*:

```
$ORIGIN company_A.com.
@      SOA server.company_A.com.
root.company_A.com. (1995032001 3600
3600 604800 86400)
       NS server.company_A.com.
http.tcp.www SRV 0 0 80 server.com-
pany_A.com.
smtp.tcp     SRV 0 0 25 mail.com-
pany_A.com.
server  A  172.30.79.10
mail    A  172.30.79.11
```

For example, to locate a http server that supports TCP protocol and provides web service, it does a lookup for:

```
_http._tcp.www.company_A.com.
```

If the use of SRV had been widely deployed, a DNS server would have answered with a list of web servers that satisfied the searching criteria.

With no existing Directory Agent, this mechanism solely depends on the DNS system. Critics claim that the deployment of it puts an extra burden on an already overloaded DNS system. Furthermore, DNS SRV only allows for simple service name resolution and has little support for the type of service attribute negotiation that is accommodated by SLP.

2.2.4 XML Web Services / UDDI

An XML Web Service is a service that accommodates direct interaction using XML-based messaging (such as SOAP [18]) over Internet-based protocols, such as HTTP. The interfaces and bindings of the Web Service are defined, described and discovered by XML [19].

In addition to being able to describe and invoke a Web Service, publication of and discovery of Web services should also be accommodated. The Universal Description, Discovery and Integration (UDDI) specification [20] is commonly accepted to be the standard mechanism to handle this. UDDI registries provide a publishing interface to allow for creation and deletion of entries in the registry and an inquiring interface to search for entries in the registry by different search criteria. The interfaces are invoked by SOAP messages, and as such UDDI itself can be thought of as an XML Web Service.

Each entry in the UDDI registry contains three parts:

- The white pages contain business information;
- The yellow pages contain the service a business provides;

- The green pages contain the specific services provided and technical information sufficient for a programmer to write an application that makes use of the service.

Web services are described by XML using the WSDL specification [21].

2.2.5 Other service discovery protocols

The Salutation protocol [22] released by the Salutation Consortium in 1996 predates SLP. It introduces the same concept as a Directory Agent, referred to as the Salutation Manager (SLM). In the Salutation Protocol, User Agents and Service Agents are referred to as Clients and Servers, respectively. The Salutation Manager Protocol (SMP) and the Transport Manager (TM) are used for the actual communication, using Remote Procedure Calls (RPC). SLM will also assist in establishing a session pipe over which a service access between the client and the server can occur.

Jini® [23] is another technology for service discovery that runs on top of Java®. It allows clients to join a Jini lookup service, which maintains dynamic information about services in the network. The client can use it for simple service discovery by requesting information about a particular device. An attractive feature of Jini is that it allows clients also to download Java code from the lookup service, which is used to access the service. It requires however that the server has already uploaded the Java proxy that the client downloads from the lookup service. Jini also supports the concepts of federations, where groups of devices may register with each other to make their services available within the group.

2.3 Service discovery on link local networks

2.3.1 Simple Service Discovery Protocol

The Simple Service Discovery Protocol (SSDP) [24] is a part of Microsoft's *Universal Plug and Play* (UpnP™) [25] and provides a mechanism that network clients can use to discover network services. UPnP supports self configuration networks by enabling the ability to automatically acquire an IP address, announce a *name*, learn about the existence and capabilities of other elements in the network, and inform others about own capabilities.

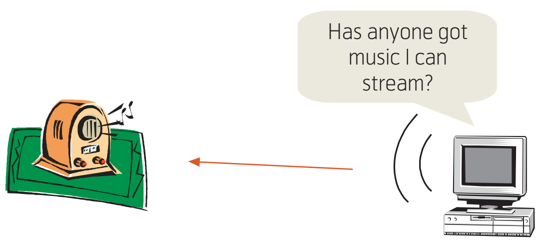


Figure 2-3 Illustrating the principle of multicast DNS

The UPnP protocols are based on open Internet-based communications standards. UPnP is based on IP, TCP/UDP, HTTP and XML. The SSDP protocol specifies the use of multicast of UDP/HTTP for announcements of services. The content of the service announcements are described using XML. HTTPU and HTTMU are used by SSDP to generate requests over unicast and multicast.

The SSDP architecture introduces three entities in the network:

- *SSDP Service:* The SSDP service is a Server Agent and represents the individual resources in an SSDP enabled network. The agent is defined in two versions, depending on whether an *SSDP Proxy* is available in the network or not. An SSDP Service without proxy support is a simple service where all messages are sent on an SSDP reserved multicast group.

- *SSDP Client:* The SSDP Client is a User Agent. When starting, the SSDP Client will search for a proxy, followed by the search for other relevant resources. If an SSDP Proxy is available, all requests are done using unicast. If not, the SSDP searches for services using multicast. The SSDP Client will cache all information about services and uses a time stamp to manage the accuracy of the cache.

- *SSDP Proxy:* The SSDP Proxy is a Directory Agent and gathers information about available resources in the network. The Proxy can be viewed as a regular resource with responsibility to cache and manage all service information. An SSDP Proxy is not a mandatory element in an SSDP enabled network, but improves the scalability when deployed in large networks.

2.3.2 Multicast DNS

DNS Service Discovery is a way of using standard DNS programming interfaces, servers, and packet formats to browse the network for services. As shown in the previous chapter, Multicast DNS (mDNS) [26] can be used to resolve names without the use of any DNS server. The same multicast mechanism can be used to search for services, by requesting a binding for the type of service wanted instead of requesting a binding for a name. This is illustrated in Figure 2-3.

When an mDNS query is sent out for a given service type and domain, any matching service replies with their names. The result is a list of available services to choose from.

For example, an application that is searching for a printer that supports TCP and is located in the *company_A* domain would issue a query for:

```
_lpr._tcp.company_A.com
```

Then every printer attached to the LAN will answer with information about its services, i.e. color, pages per minute etc.

Using a mechanism for automatic service discovery greatly simplifies the job of connecting PCs, terminals, wireless units and consumer electronics.

Caching of multicast packets can prevent hosts from requesting information that has already been requested. For example, when one host requests, say, a list of LPR print spoolers, the list of printers comes back via multicast, so that all local hosts can see it. The next time a host needs a list of print spoolers, it already has the list in its cache and does not need to reissue the query.

2.4 Service discovery in ad hoc networks

2.4.1 Service discovery mechanism for MANETs

Discovery of services and other named resources is an important feature for the usability of mobile ad hoc networks (MANETs). The characteristics of ad hoc networks were described earlier. We recall that a MANET is anticipated to be without any pre-existing infrastructure and that nodes may enter or leave the network at any time. This makes efficient and timely service discovery a challenging task.

In a MANET, any node may in principle operate as a server and provide its services to other MANET nodes. Any node may also operate as a client and use the service discovery protocol to detect available services in the network. The service attributes include service characteristics and service binding information, such as IP addresses, port-numbers and protocols, which enable the client to initiate the selected service on the appropriate server.

Existing service discovery mechanisms, described in previous sections, are designed with a fixed network in mind, and might not fit well to MANETs.

Before a service discovery mechanism for ad-hoc networks can be designed, we need to determine the necessary principles for service discovery in ad hoc networks and evaluate which service discovery architecture that is most suitable. The reader will observe that name resolution and service discovery have many similar features in terms of both discovery principles and possible architectures.

2.4.2 Service location architectures for service discovery on MANETs

The architectures available for name resolution, as described for name resolution above, are also available for service discovery. In the context of service discovery, the Caching Coordinator Entity is normally referred to as a Service Coordinator (SC), the User Entity is referred to as a User Agent (UA) and the Named Entity is referred to as a Service Agent (SA). Furthermore, the architectures are often referred to as the *Distributed*, the *Service-Coordinator-based* and the *Hybrid Service Location Architectures*. The Distributed architecture for MANETs is illustrated in Figure 2-4, and the Service-Coordinator-based architecture is shown in Figure 2-5. The hybrid architecture is a combination of the two: The UA tries

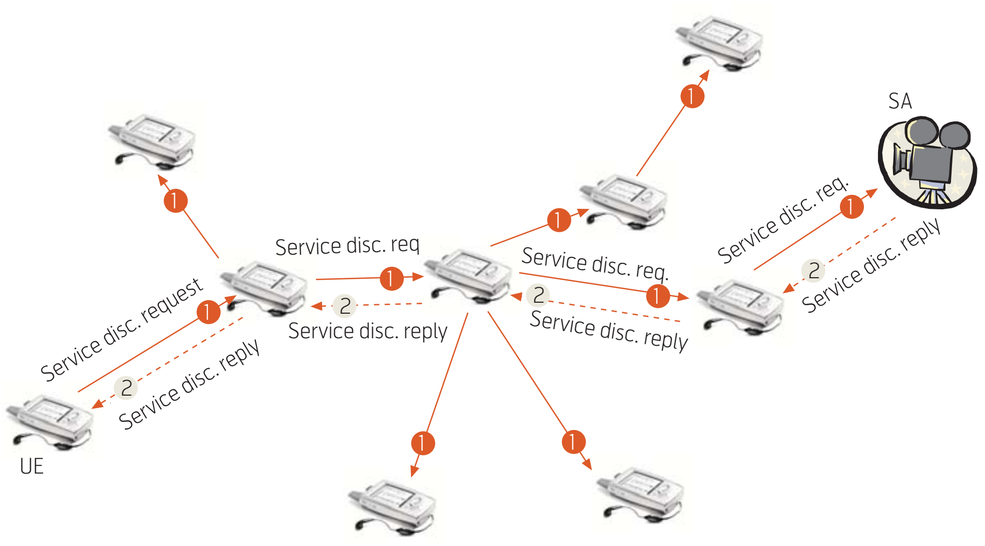


Figure 2-4 The Distributed architecture

ISSN 0085-7130 © Telenor ASA 2005

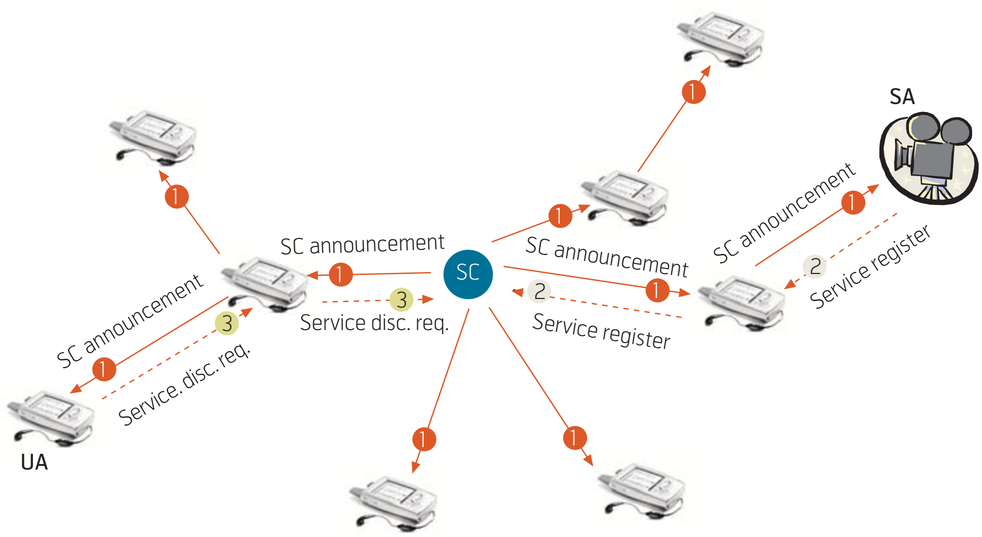


Figure 2-5 The Service-Coordinator-based architecture

to discover services according to the Service-Coordinator-based architecture, but falls back to the Distributed approach if the selected SC does not have the desired binding, or if no SC can be found.

2.4.3 Emerging principles for service discovery on reactively routed MANETs

The emerging principles for name resolution in reactive MANETs, in which resolution requests and replies are carried by routing messages (outlined in the previous chapter), are also useful for service discovery. In the context of service discovery, *Service Discovery Requests (SREQs)* and *Service Discovery Replies (SREPs)* are piggybacked on RREQ and RREP packets, respectively. This is illustrated in Figure 2-6.

The advantages of piggybacking service discovery on routing messages are:

1 Reverse routes to the User Agent (i.e. client) are established along with the SREQ so that no additional route discovery is necessary to relay the SREP back to the requestor.

2 Forward routes to the SC are established along with the SC Announcements so that SREQs and Service Registrations can be unicast to the SC.

3 A forward route is established along with the SREP so that no additional route discovery is necessary for further communication with the node issuing the reply.

Figure 2-6 shows how service discovery can be streamlined with the reactive routing protocol. Service Discovery Requests (SREQs) are piggybacked on Routing Request (RREQ) packets, and Service Discovery Replies (SREPs) are piggybacked on Routing Reply (RREP) packets. In addition, for the Hybrid architecture, the SC Announcements are piggybacked on RREQ packets, and Service Registrations are piggybacked on RREP packets. Thus, both the SC-based, Hybrid and Distributed architectures can take advantage of this.

Service Discovery Request (SREQ)	=	IP	RREQ	Service description
Service Discovery Reply (SREP)	=	IP	RREP	Service description and binding
Service Coordinator announcement	=	IP	RREQ	Service coordinator information
Service Registration	=	IP	RREP	Service description and binding

Figure 2-6 Routing packets carry Service discovery messages

2.4.4 Proposed solution for service discovery in reactive ad hoc networks

A solution for service discovery in reactive ad hoc networks is proposed in [27]. It basically uses the same mechanism as was presented for name resolution in the previous chapter. It is based on the distributed architecture without the use of any Service Coordinators, but here the intermediate nodes are allowed to cache service bindings and respond immediately if a valid binding is found. It also uses the same technique to carry the discovery messages by the routing packets to allow both services and the routes to the nodes providing these services to be discovered in one round-trip.

In [27] a *service binding* is defined as a mapping of a service name to an IP address. Different encoding schemes, such as Service Port Request or Service URL, can be used to request a binding for an IP address. An SREQ for a Service URL contains a service type string and a service request predicate of formats that are defined by SLP. The format of the URL and the authentication block contained in the corresponding SREP are also defined by SLP. Hence, not only are formats of SLP reused, but the authorization block also ensures that the service authorization features of SLP are maintained. The use of SREQs for a Service Port assumes that the User Agents know in advance the well-defined (TCP or UDP) port number associated with the requested service. Hence, the SREQ only needs to contain the port number associated with the service application requested.

The proposed service discovery protocol not only considers the case where the UA has neither a service binding nor an active route to a node providing the desired service. It also considers the case where the route is active, but the service binding has expired (or is absent), and the case where the service binding is active, but the route has expired.

2.4.5 Evaluation of service location architectures in ad hoc networks

As a slight simplification, one may say that all service discovery protocols presented above are based on two baseline mechanisms for the management of service discovery information:

1 Information about services offered on the network is stored on one or a few centralized nodes, referred to as Service Coordinators (SCs) in this article;

2 Information about each service is stored on each node that offers the service.

In previous sections we have defined the service discovery architectures according to the two mechanisms above. A solution that is only based on the first mechanism is referred to as a *Service Coordinator based architecture*, while a solution only based on the second mechanism is referred to as a *Distributed architecture*. Finally, a solution based on a mixture of both the first and the second mechanism is referred to as a Hybrid architecture.

In the next section we evaluate the performance of the Hybrid and Distributed architectures in a reactively routed MANET. The architectures were presented in detail in the previous chapter and summarized more briefly above in this article.

2.4.6 Architecture evaluation

The evaluation of the Distributed and Hybrid architectures is based on results from [28]. The architectures can be configured by different settings of the following two parameters:

1 *Flooding scope:* This parameter determines the maximum number of hops a flooded Service Discovery Request is allowed to traverse in the network. (For example, the flooding scope is of four hops in Figure 2-4 and of two hops in Figure 2-5.)

2 *SC announcement scope:* This parameter determines the maximum number of hops a flooded SC Announcement is allowed to traverse in the network. (For example, Figure 2-5 illustrates a situation with SC announcement scope of two hops.) This parameter is used only in the Hybrid architecture. Alternatively, the Distributed architecture can be considered as a special case of a Hybrid architecture where the SC announcement scope is set to zero.

The objective is to optimize the benefits of additional service availability provided by the use of Service Coordinators against the cost of additional overhead and possibly higher delay. We will consider the performance measured by delay, by the service availability and by the message overhead.

In [28] it was observed that the differences in delays between the two architectures are only in the order of a few milliseconds. Since service discovery is normally part of the service initiation, users would normally accept an initial delay (e.g. when retrieving a web page on the Internet or for setting up an IP Telephony call). Hence, it was concluded that the small observed differences in delay between the two architectures should be considered negligible in this context. With delay out of the picture, the key question is reduced to whether the increased service availability is worth the increase in message overhead.

The service availability can be defined as [29]:

$$Serv.Avail. = \frac{number\ of\ positive\ service\ replies}{total\ number\ of\ service\ requests\ generated}$$

A *positive Service Discovery Reply* means a successful contact to this server via the given access information, i.e. that a route to the resolved server can be found.

It is observed in Figure 2-7 that the service availability is indeed higher with the Hybrid approach. Figure 2-7 also shows how the presence of SCs (i.e. for the Hybrid architecture) influences the service availability. When comparing architectures that use the same flooding scope, we find that the Hybrid architecture improves the service availability as compared to the Distributed query-based architecture.

The main reason that SCs improve the service availability, is that in some cases the SC will be positioned

ISSN 0085-7130 © Telenor ASA 2005

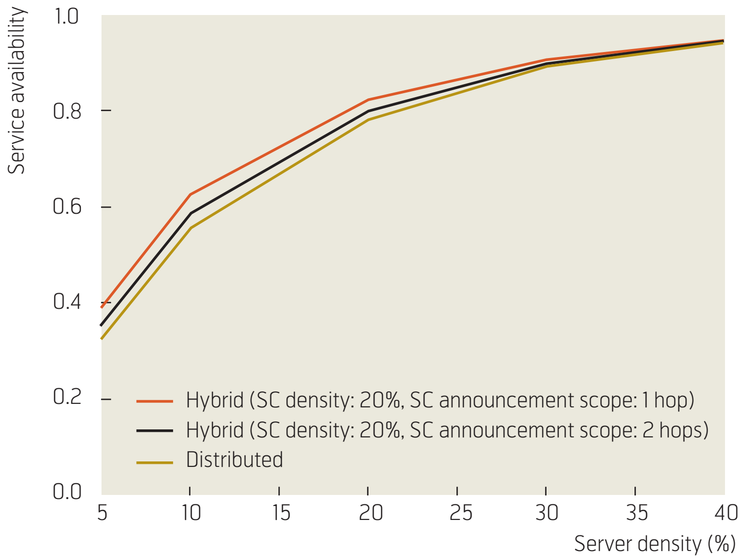


Figure 2-7 The introduction of SCs improves the service availability

in between the UA and SA. We may return to Figure 2-5 for an example of such a situation. Here the UA and SA are four hops apart, but both are only two hops away from the SC. The SC announcement and flooding scopes are both of two hops. Hence, the SA is able to register its services with the SC, and the UA is able to discover the server by means of the SC. However, since the UA is four hops away from the SA, the UA would not be able to discover the SA if the Distributed architecture with a flooding scope of two hops had been used.

From Figure 2-7 we observe that with SC announcement scopes of one or two hops, the service availability is improved by 8.7 % or 20.9 %, respectively, at a server density of 5 %. Since the introduction of SCs improves the service availability, it comes as no surprise that the service availability increases with an increasing SC announcement scope.

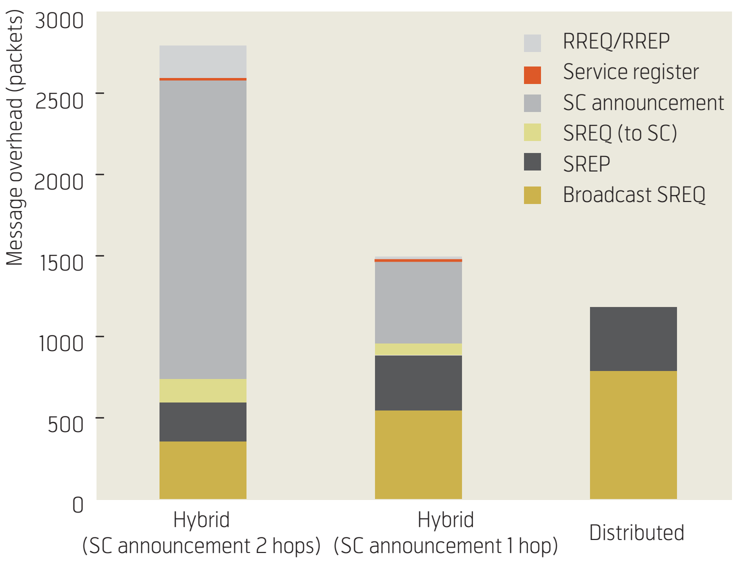


Figure 2-8 Detailed comparison of message overhead by message type. (Server density at 20 %, flooding scope of two hops)

Although the SCs introduced in the Hybrid architecture yield higher service availability, it also results in extra message overhead, as observed in Figure 2-8. The SCs introduce two proactive elements to the network, namely SC Announcements and Service Registrations. These messages will take up a fixed bandwidth regardless of whether or not there are clients doing service discoveries. In addition, these two types of messages will also trigger pure route discovery messages when a reactive routing protocol is being used.

By comparing Figure 2-7 and Figure 2-8 we observe that the additional cost of using SC in terms of percentage increase in message overhead is much higher than the additional benefits provided in terms of percentage increase in Service Availablity.

A more rigorous analysis that compares the two architectures are undertaken in [28]. It takes into consideration a large range of control parameters, such as server density, SC density, flooding scopes, SC announcement scopes, reasonable request frequencies, number of different types of services, level of mobility, and so forth. It is also argued that the conclusion is valid independent of the lengths of the service discovery messages.

In [28] it is generally observed that for any Hybrid configuration with a given SC announcement scope and flooding scope, it is always possible to find a distributed configuration (with some flooding scope) that outperforms the Hybrid configuration in terms of both higher service availability and lower messaging overhead. As the opposite is not the case, it is concluded that the Distributed architecture outperforms the Hybrid architecture. Hence, service discovery protocols that use Service Coordinators (or functionality similar to Directory Agents) do not work well in ad hoc networks with reactive routing. The main reason is that the increase in service availability by adding Service Coordinators is negligible compared to the extra message overhead it causes.

In addition to the analyses presented in [28], there are several other arguments that are in favor of not introducing Service Coordinators in MANETs at large. First, the Distributed architecture is considerably less complex than the Hybrid architecture. Furthermore, in a dynamic topology with network entries and departures, the Service Coordinators of the Hybrid architecture have the disadvantage of sometimes providing the User Agents with “false positives”, i.e. with outdated bindings of servers that have already left the network. Moreover, the Hybrid approach may call for a separate complicated mechanism for electing Service Coordinators, which might require a substantial amount of network resources.

References

1 Mockapetris, P. *Domain names – concepts and facilities*. STD 13, RFC 1034, Internet Engineering Task Force, November 1987.

2 Mockapetris, P. *Domain names – Implementation and Specification*. STD 13, RFC 1035, Internet Engineering Task Force, November 1987.

3 Microsoft TechNet. *Windows Internet Naming Service (WINS)*. http://www.microsoft.com/technet/archive/windows2000serv/evaluate/featfunc/nt5wins.mspx (Last visited: 7 April 2005).

4 Sun Microsystems. *The Network Information Service (NIS) / Yellow Pages*. http://www.sun.com (Last visited: 7 April 2005).

5 Wellington, B. *Secure Domain Name System (DNS) Dynamic Update*. RFC 3007, Internet Engineering Task Force, November 2000.

6 Chesire, S, Krochmal, M. *Multicast DNS*. IETF Internet Draft, draft-cheshire-dnsext-multicastdns-04.txt, February 2004. Work in Progress.

7 Aboba, B, Thaler, D, Esibov, L. *Linklocal Multicast Name Resolution (LLMNR)*. draft-ietf-dnsext-mdns-39.txt, March 2005. Work in Progress.

8 Macker, J, Corson, S. *Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations*. RFC 2501, Internet Engineering Task Force, January 1999.

9 Perkins, C E, Royer, E M, Das, S R. *Ad-hoc On Demand Distance Vector (AODV) Routing*. RFC 3561, Internet Engineering Task Force (IETF), July 2003.

10 Johnson, D B, Maltz, D A, Hu, J-C. *The Dynamic Source Routing Protocol*. draft-ietf-manet-dsr-10.txt, July 2004. Work in Progress.

11 Engelstad, P E, Do T V, Egeland, G. Name Resolution in On-Demand MANETs and over External IP Networks. *Proceedings of IEEE Int. conf. on Comm. 2003 (ICC 2003)*. http://www.unik.no/~paalee/publications/NR-paper-for-ICC2003.pdf

12 Engelstad, P E, Do T V, Jønvik, T E. Name Resolution in Mobile Ad-hoc Networks. *Proceedings of 10th Int. Conf. on Telecom. 2003 (ICT 2003)*. http://www.unik.no/~paalee/ publications/NR-paper-for-ICT2003.pdf

13 Clausen, T, Jacquet, P (eds.). *Optimized Link State Routing Protocol (OLSR)*. RFC 3626, Internet Engineering Task Force (IETF), October 2003.

14 Engelstad, P E, Egeland, G, Koodli, R, Perkins, C E. *Name Resolution in On Demand MANETs and over External IP Networks*. IETF Internet draft, draft-engelstad-manet-name-resolution-01.txt, February 2004. Work in Progress.

15 Apple Computers. *AppleTalk*. http://www.apple.com (Last visited: 7 April 2005).

16 Guttman, E, Perkins, C, Veizades, J, Day, M. *Service Location Protocol, version 2*. RFC 2608, Internet Engineering Task Force (IETF), June 1999.

17 Gulbrandsen, A, Vixie, P, Esibov, L. *A DNS RR for specifying the location of services (DNS SRV)*. RFC 2782, Internet Engineering Task Force (IETF), February 2000.

18 Gudgin et al. *SOAP Version 1.2 Part 1: Messaging Framework*. http://www.w3.org/TR/soap12-part1/, W3C Recommendation, June 2003.

19 Bray et al. *Extensible Markup Language (XML) 1.0 (Second Edition)*. W3C Recommendation, http://www.w3.org/TR/2000/REC-xml-20001006, October 2000.

20 OASIS UDDI. *Universal Description, Discovery, and Integration (UDDI) 2.0*. http://www.uddi.org (Last visited: 7 April 2005).

21 Christensen, E, Curbera, F, Meredith, G, Weerawarana, S. *Web Services Description Language (WSDL) 1.1*. W3C Note, http://www.w3.org/TR/2001/NOTE-wsdl-20010315, March 2001.

22 The Salutation Consortium. *Salutation Architecture Specification Version 2.1*. http://www.salutation.org, (Last visited: 7 April 2005).

 ISSN 0085-7130

23 Sun Microsystems. *Jini Network Technology*. http://www.jini.org (Last visited: 7 April 2005).

24 Goland et al. *Simple Service Discovery Protocol /1.0*. Internet Draft, draft-cai-ssdp-v1-03.txt. Work in Progress, Oct 1999.

25 Microsoft Corporation. *Universal Plug-and-Play (UPnP) Forum*. http://www.upnp.org (Last visited: 7 April 2005).

26 Chesire, S, Krochmal, M. *Multicast DNS*. IETF Internet Draft, draft-cheshire-dnsext-multicastdns-04.txt, February 2004, Work in Progress.

27 Koodli, R, Perkins, C E. *Service Discovery in On Demand Ad Hoc Networks*. IETF Internet draft, draft-koodli-manet-servicediscovery-00.txt, October 2002. Work in Progress.

28 Engelstad, P E, Zheng, Y, Koodli, R, Perkins, C E. Service Discovery Architectures for On-Demand Ad Hoc Networks. *International Journal of Ad Hoc and Sensor Networks*, Old City Publishing (OCP Science), 1 (3), 2005.

29 Engelstad, P E, Zheng, Y. Evaluation of Service Discovery Architectures for Mobile Ad Hoc Networks. *Proceedings of the 2nd annual conference on Wireless On-demand Networks and Services (WONS 2005)*, St. Moritz, Switzerland, Jan. 19–21, 2005.

Paal E. Engelstad completed his PhD on resource discovery in Mobile Ad hoc and Personal Area Networks in 2005. He has also a Bachelor and Masters degree (Honours with Distinction) in Applied Physics from NTNU, Norway, and a Bachelor degree in Computer Science from University of Oslo, Norway. After working five years in industry, he joined Telenor R&D where he focuses on IETF and IP technology (e.g. IP mobility, IPv6, QoS, MANET and AAA-issues) and IEEE wireless technologies (e.g. 802.11, 802.15 and 802.16). Paal Engelstad has published 28 refereed papers and holds three patents (two pending).

email: Paal.Engelstad@telenor.com

Geir Egeland holds a B.Eng (Hons) from the University of Bristol and has for the last ten years worked as a research scientist in the field of mobile networks. He is currently with Telenor R&D where his work is mainly focused on mobility for IP networks, with a particular emphasis on MANET and IPv6. Geir Egeland was formerly employed by the Norwegian Defence Research Establishment (NDRE) as a research scientist working on design and analysis of MAC and routing protocols for mobile ad hoc network.

email: geir.egeland@telenor.com

Terms and acronyms in Future Mobile Phones

ADSL	Asymmetric Digital Subscriber Line	A data communications technology that enables faster data transmission over copper telephone lines than a conventional modem can provide. The access utilises the 1.1 MHz band and has the possibility to offer, dependent of subscriber line length, downstream rates of up to 8 Mb/s. Upstream rates start at 64 kb/s and typically reach 256 kb/s but can go as high as 768 kb/s.
BT	Bluetooth	Bluetooth is an industrial specification for wireless personal area networks (PANs). Bluetooth provides a way to connect and exchange information between devices like personal digital assistants (PDAs), mobile phones, laptops, PCs, printers and digital cameras via a secure, low-cost, globally available short range radio frequency.
CDC	Connected Device Configuration	CDC is a larger subset of J2SE, containing almost all the libraries that are not GUI related. CDC is based on the cVM (compact Virtual Machine)
CLDC	Connected Limited Device Configuration	The CLDC contains a strict subset of the Java class libraries and is the minimal needed for a Java virtual machine to operate. CLDC is based on the kVM (kilo Virtual Machine). CLDC is basically used to classify the myriad of devices into a fixed configuration.
CIFS	Common Internet File System	CIFS defines a standard remote file system access protocol for use over the Internet, enabling groups of users to work together and share documents across the Internet or within corporate intranets. CIFS is an open, cross-platform technology based on the native file-sharing protocols built into Microsoft® Windows® and other popular PC operating systems, and supported on dozens of other platforms. With CIFS, millions of computer users can open and share remote files on the Internet without having to install new software or change the way they work.
DHCP	Dynamic Host Configuration Protocol	DHCP is a client-server networking protocol. A DHCP server provides configuration parameters specific to the DHCP client host requesting, generally, information required by the client host to participate on an IP network. DHCP also provides a mechanism for allocation of IP addresses to client hosts. DHCP appeared as a standard protocol in October 1993. RFC 2131 provides the latest (March 1997) DHCP definition. The latest standard on a protocol describing DHCPv6, DHCP in an IPv6 environment, was published in July 2003 as RFC 3315
DM	Device Management	Device Management is the generic term used for technology that allows third parties to carry out the difficult procedures of configuring mobile devices on behalf of the end user (customer). Third parties would typically be wireless operators, service providers or corporate information management departments. Through device management, an external party can remotely set parameters, conduct troubleshooting servicing of terminals, install or upgrade software.
GSM	Global System for Mobile Communications	GSM is the most popular standard for mobile phones in the world. GSM phones are used by over a billion people across more than 200 countries. The ubiquity of the GSM standard makes international roaming very common between mobile phone operators which enables phone users to access their services in many other parts of the world in addition to their own country. GSM differs significantly from its predecessors in that both signalling and speech channels are digital, which means that it is seen as a second generation (2G) mobile phone system. This fact has also meant that data communication was built into the system from very early on. GSM is an open standard which is currently developed by the 3GPP.
HLR	Home Location Register	The HLR is a central database that contains details of each mobile phone subscriber that is authorized to use the GSM core network. More precisely, the HLR stores details of every SIM card issued by the mobile phone operator. Each SIM has a unique identifier called an IMSI which is one of the primary keys to each HLR record. The next important items of data associated with the SIM are the telephone numbers used to make and receive calls to the mobile phone, known as MSISDNs. The main MSISDN is the number used for making and receiving voice calls and SMS, but it is possible for a SIM to have other secondary MSISDNs associated with it for fax and data calls. Each MSISDN is also a primary key to the HLR record.

ISSN 0085-7130 © Telenor ASA 2006

IEEE 802.11		Refers to a family of specifications developed by the IEEE for wireless local area networks. It also refers to the "Wireless LAN working group" of the IEEE 802 project. 802.11 specifies an over-the-air interface between a wireless client and a base station or between two wireless clients. The IEEE accepted the specification in 1997. There are several specifications in the 802.11 family, including i) 802.11 – provides 1 or 2 Mbit/s transmission in the 2.4 GHz band,; ii) 802.11a – an extension that provides up to 54 Mbit/s in the 5 GHz band. It uses an orthogonal frequency division multiplexing encoding scheme rather than FHSS or DSSS; iii) 802.11b provides 11 Mbit/s transmission in the 2.4 GHz band and was ratified in 1999 allowing wireless functionality comparable to Ethernet; iv) 802.11g provides 20+ Mbit/s in the 2.4 GHz band; v) 802.11z is a method for transporting an authentication protocol between the client and access point, and the Transport Layer Security (TLS) protocol. More variants are also under preparation, including support of 100 Mbit/s traffic flows. http://www.ieee802.org/11
IMAP	Internet Message Access Protocol	The Internet Message Access Protocol (commonly known as IMAP, and previously called Interactive Mail Access Protocol) is an application layer Internet protocol used for accessing email on a remote server from a local client. IMAP and POP3 (Post Office Protocol version 3) are the two most prevalent Internet standard protocols for email retrieval. Both are supported by virtually all modern email clients and servers, although in some cases in addition to vendor-specific, typically proprietary, interfaces. For example, while proprietary protocols are typically used between Microsoft's Outlook client and an Exchange server and between IBM's Notes client and a Domino server, all of these products also support IMAP and POP3 allowing interoperability with other servers and clients. The current version of IMAP, IMAP version 4 revision 1 (IMAP4rev1), is defined by RFC 3501.
IP	Internet Protocol	A protocol for communication between computers, used as a standard for transmitting data over networks and as the basis for standard Internet protocols. http://www.ietf.org
ISDN	Integrated Services Digital Network	A digital telecommunications network that provides end-to-end digital connectivity to support a wide range of services, including voice and non-voice services, to which users have access by a limited set of standard multi-purpose user-network interfaces. The user is offered one or more 64 kb/s channels. http://www.itu.int
ITU	International Telecommunication Union	On May 17, 1865, the first International Telegraph Convention was signed in Paris by the 20 founding members, and the International Telegraph Union (ITU) was established to facilitate subsequent amendments to this initial agreement. It changed its name to the International Telecommunications Union in 1934. From 1948 a UN body with approx. 200 member countries. It is the top forum for discussion and management of technical and administrative aspects of international telecommunications. http://www.itu.int
ITU-T	International Telecommunication Union – Standardization Sector	http://www.itu.int/ITU-T/
J2ME	Java 2 Platform, Micro Edition	J2ME is a collection of Java APIs targeting embedded consumer products such as PDAs, cell phones and other consumer appliances.
MIDlet		A MIDlet is a Java program for embedded devices, more specifically the J2ME virtual machine. Generally, these are games and applications that run on a cell phone.
MIDP	Mobile Information Device Profile	Designed for cell phones, MIDP boasts an LCD-oriented GUI API, and MIDP 2.0 includes a basic 2D gaming API. Applications written for this profile are called MIDlets. Almost all new cell phones come with a MIDP implementation, and it is now the de facto standard for downloadable cell phone games.
MIME	Multipurpose Internet Mail Extensions	MIME is an Internet Standard for the format of e-mail. Virtually all Internet e-mail is transmitted via SMTP in MIME format. Internet e-mail is so closely associated with the SMTP and MIME standards that it is sometimes called SMTP/MIME e-mail.
MSC	Mobile services Switching Centre	The MSC is a sophisticated telephone exchange which provides circuit-switched calling, mobility management and GSM services to the mobile phones roaming within the area that it serves. This means voice, data and fax services, as well as SMS and call divert. A Gateway MSC is the MSC that determines in which visited MSC the subscriber who is being called is curently located. It also interfaces with the Public Switched Telephone Network. All mobile to mobile calls and PSTN to mobile calls are routed through a GMSC. The term is only valid in the context of one call since any MSC may provide both the gateway function and the Visited MSC function; however, some manufacturers design dedicated high capacity MSCs which do not have any BSCs connected to them. These MSCs will then be the Gateway MSC for many of the calls they handle.

NAT	Network Address Translation	In computer networking, the process of network address translation (NAT, also known as network masquerading or IP-masquerading) involves re-writing the source and/or destination addresses of IP packets as they pass through a router or firewall. Most systems using NAT do so in order to enable multiple hosts on a private network to access the Internet using a single public IP address. According to specifications, routers should not act in this way, but many network administrators find NAT a convenient technique and use it widely. Nonetheless, NAT can introduce complications in communication between hosts. NAT first became popular as a way to deal with the IPv4 address shortage and to avoid the difficulty of reserving IP addresses. Use of NAT has proven particularly popular in countries other than the United States, which (for historical reasons) have fewer address-blocks allocated per capita. It has become a standard feature in routers for home and small-office Internet connections, where the price of extra IP addresses would often outweigh the benefits. In a typical configuration, a local network uses one of the designated "private" IP address subnets (such as 192.168.x.x or 10.x.x.x), and a router on that network has a private address (such as 192.168.0.1) in that address space. The router is also connected to the Internet with a single "public" address (known as "overloaded" NAT) or multiple "public" addresses assigned by an ISP. As traffic passes from the local network to the Internet, the source address on the packets are translated on the fly from the private addresses to the public address(es). The router tracks basic data about each active connection (particularly the destination address and port). When a reply returns to the router, it uses the connection tracking data it stored during the outbound phase to determine where on the internal network to forward the reply; the TCP or UDP client port numbers are used to demultiplex the packets in the case of overloaded NAT, or IP address and port number when multiple public addresses are available, on packet return. To a system on the Internet, the router itself appears to be the source/destination for this traffic.
NetBIOS	Network Basic Input/Output System	The NetBIOS API allows applications on separate computers to communicate over a local area network. It provides services related to the session layer of the OSI model but is not routable and must be encapsulated within a protocol in order to be of use in a wide area network. NetBIOS makes wide use of broadcast messages, which accounts for its reputation as a chatty interface. NetBIOS provides three distinct services: - Name service for name registration and resolution - Session service for connection-oriented communication - Datagram distribution service for connectionless communication
NFC	Near Field Communication Technology	NFC, jointly developed by Sony and Philips, was approved as an ISO/IEC standard on 8 December 2003. It was approved as an ECMA standard earlier on. On 18 March 2004 Nokia, Sony and Philips formed the NFC-forum to advance NFC development. NFC holds the promise of bringing true mobility to consumer electronics in an intuitive and psychologically comfortable way since the devices can hand-shake only when brought literally into touching distance
PAN	Personal Area Network	A PAN is a computer network used for communication between computer devices (including telephones and personal digital assistants) close to one person. The devices may or may not belong to the person in question. The reach of a PAN is typically a few metres. PANs can be used for communication between the personal devices themselves (intrapersonal communication), or for connecting to a higher level network and the Internet (an uplink). Personal area networks may be wired with computer buses such as USB and FireWire. A wireless personal area network (WPAN) can also be made possible with network technologies such as IrDA and Bluetooth.
POP3	Post Office Protocol version 3	POP3 is an application layer Internet standard protocol used to retrieve email from a remote server to a local client over a TCP/IP connection. Nearly all individual Internet service provider email accounts are accessed via POP3.
POTS	Plain Old Telephone Service	A very general term used to describe an ordinary voice telephone service. See also PSTN.
PSTN	Public Service Telephone Network	Common notation for the conventional analogue telephone network.

RFID	Radio Frequency Identification	RFID is an automatic identification method, relying on storing and remotely retrieving data using devices called RFID tags or transponders. An RFID tag is a small object that can be attached to or incorporated into a product, animal, or person. RFID tags contain antennas to enable them to receive and respond to radio frequency queries from an RFID transceiver. Passive tags require no internal power source, whereas active tags require a power source.
SAT	SIM Application Toolkit	The SIM Application Toolkit is a set of commands which defines how the card should interact with the outside world and extend the communication protocol between the card and the handset. With SIM Application Toolkit, the card has a proactive role in the handset (this means that the SIM initiates commands independently of the handset and the network).
SMB	Server Message Block	SMB was originally invented by IBM to turn DOS "Interrupt 33" local file access into a networked file system, but the most common version is heavily modified by Microsoft. At around the time when Sun Microsystems announced WebNFS, Microsoft coincidentally launched an initiative in 1998 to rename SMB to Common Internet File System (CIFS) and added more features, including support for symbolic links, hard links, larger file sizes and an attempt at supporting direct connection without all the NetBIOS trimmings – an effort that was largely experimental and required further refinement. SMB was originally designed to run on top of the NetBIOS protocol (which itself is typically run on NetBEUI, IPX/SPX or NBT), though SMB can also run on top of TCP/IP directly since Windows 2000. "SMB the protocol" is not to be confused with the SMB services that run on it, nor with NetBIOS, nor with the DCE/RPC services that utilise SMB as an authenticated Inter-process communication channel (over Named pipes), nor with the "Network Neighborhood" protocols which primarily but not exclusively run as datagram services direct on the NetBIOS transport. Because of the importance of the SMB protocol in interacting with the dominant Microsoft Windows platform, coupled with the heavily modified nature of the SMB implementation present in that platform, the Samba project was created to provide a free implementation of a compatible SMB client and server for use with non-Microsoft operating systems.
SIM	Subscriber Identity Module	A SIM is a logical application running on a UICC smartcard. Although the terms UICC and SIM are often interchanged, UICC refers to the physical card, whereas SIM refers to a single application residing in the UICC that collects GSM user subscription information. The SIM provides secure storing of the key identifying a mobile phone service subscriber but also subscription information, preferences and storage of text messages. The equivalent to a SIM in UMTS is a Universal Subscriber Identity Module (USIM).
SMTP	Simple Mail Transfer Protocol	SMTP is the *de facto* standard for email transmission across the Internet. SMTP is a relatively simple, text-based protocol, where one or more recipients of a message are specified (and in most cases verified to exist) and then the message text is transferred. It is quite easy to test a SMTP server using the telnet program. SMTP uses TCP port 25. To determine the SMTP server for a given domain name, the MX (Mail eXchange) DNS record is used.
SOAP	Simple Object Access Protocol	SOAP is a protocol for exchanging XML-based messages over a computer network, normally using HTTP. SOAP forms the foundation layer of the web services stack, providing a basic messaging framework that more abstract layers can build on. SOAP facilitates the Service-Oriented architectural pattern.
UMA	Unlicensed Mobile Access	UMA provides an alternative access to GSM and GPRS core network services via IP-based broadband connections. In order to deliver a seamless user experience, the specifications define a new network element and associated protocols that provide for the secure transport of GSM/GPRS signaling and user traffic over IP. The caller can automatically switch from GSM to IP-based networks and back again without interruptions. UMA also helps to make better use of the scarce GSM spectrum: as soon as a phone is in reach of an UMA access point, it will leave the GSM spectrum and continue its call over WiFi, leaving the GSM spectrum for those callers that really need it.
UMTS	Universal Mobile Telecommunication System	The European member of the IMT 2000 family of 3G wireless standards. UMTS supports data rates of 144 kb/s for vehicular traffic, 384 kb/s for pedestrian traffic and up to 2 Mb/s in support of in-building services. The standardisation work began in 1991 by ETSI but was transferred in 1998 to 3GPP as a corporation between Japanese, Chinese, Korean and American organisations. It is based on the use of WCDMA technology and is currently deployed in many European countries. The first European service opened in 2003. In Japan NTT DoCoMo opened its "pre-UMTS" service FOMA (Freedom Of Mobile multimedia Access) in 2000. The system operates in the 2.1 GHz band and is capable of carrying multimedia traffic. http://www.3gpp.org/

VLR	Visitors Location Register	The VLR is a temporary database of the subscribers who have roamed into the particular area which it serves. Each Base Station in the network is served by exactly one VLR, hence a subscriber cannot be present in more than one VLR at a time. The data stored in the VLR has either been received from the HLR, or collected from the MS. In practice, for performance reasons, most vendors integrate the VLR directly to the V-MSC and, where this is not done, the VLR is very tightly linked with the MSC via a proprietary interface.
VPN	Virtual Private Network	A VPN is a private communications network usually used within a company, or by several different companies or organizations, to communicate over a public network. VPN message traffic is carried on public networking infrastructure (e.g. the Internet) using standard (often insecure) protocols, or over a service provider's network providing VPN service guarded by well defined Service Level Agreement (SLA) between the VPN customer and the VPN service provider.
WAP	Wireless Application Protocol	WAP is an open international standard for applications that use wireless communication, for example Internet access from a mobile phone. WAP was designed to provide services equivalent to a Web browser with some mobile-specific additions, being specifically designed to address the limitations of very small portable devices. It is now the protocol used for the majority of the world's mobile internet sites, otherwise known as wap-sites. The Japanese i-mode system is the other major competing wireless data protocol.
WDMA	Wavelength Division Multiple Access	In optical communication different wavelengths are used to send packets from different sources. The wavelength constitutes independent communication channels and may carry different signal formats.
WSDL	Web Services Description Language	WSDL is an XML format published for describing Web services. Version V 1.1 has not been endorsed by the World Wide Web Consortium (W3C), however it released a draft for version 2.0 on 11 May 2005, that will be a recommendation (an official standard), and thus endorsed by the W3C. It is commonly abbreviated as WSDL in technical literature and often pronounced "Whiz'-Dull". WSDL describes the public interface to the web service. This is an XML-based service description on how to communicate using the web service; namely the protocol bindings and message formats required to interact with the web services listed in its directory. The supported operations and messages are described abstractly and then bound to a concrete network protocol and message format. WSDL is often used in combination with SOAP and XML Schema to provide web services over the internet. A client (program) connecting to a web service can read the WSDL to determine what functions are available on the server. Any special datatypes used are embedded in the WSDL file in the form of XML Schema. The client can then use SOAP to actually call one of the functions listed in the WSDL.
XML	Extensible Markup Language	XML is a W3C-recommended general-purpose markup language for creating special-purpose markup languages, capable of describing many different kinds of data. It is a simplified subset of SGML. Its primary purpose is to facilitate the sharing of data across different systems, particularly systems connected via the Internet. Languages based on XML (for example, RDF/XML, RSS, MathML, XHTML, SVG, and cXML) are defined in a formal way, allowing programs to modify and validate documents in these languages without prior knowledge of their form.
	XML Web Service	A Web Service is a software component that is described via WSDL and is capable of being accessed via standard network protocols such as, but not limited to SOAP over HTTP.