Telektronikk

1.2001

Wireless Future

Telektronikk

Volume 97 No. 1 – 2001
ISSN 0085-7130

Editor:
Ola Espvik
Tel: (+47) 63 84 88 83
email: ola.espvik@telenor.com

Status section editor:
Per Hjalmar Lehne
Tel: (+47) 63 84 88 26
email: per-hjalmar.lehne@telenor.com

Editorial assistant:
Gunhild Luke
Tel: (+47) 63 84 86 52
email: gunhild.luke@telenor.com

Editorial office:
Telenor Communication AS
Telenor R&D
PO Box 83
N-2027 Kjeller
Norway
Tel: (+47) 63 84 84 00
Fax: (+47) 63 81 00 76
email: telektronikk@telenor.com

Editorial board:
Ole P. Håkonsen,
Senior Executive Vice President.
Oddvar Hesjedal,
Vice President, R&D.
Bjørn Løken,
Director.

Graphic design:
Design Consult AS, Oslo

Layout and illustrations:
Gunhild Luke, Britt Kjus (Telenor R&D)

Prepress and printing:
Optimal as, Oslo

Circulation:
4,000

telenor

Contents

Feature: Wireless Future

Status

Guest Editorial

Per Hjalmar Lehne

2001 is a year filled with several events for mobile communications. In September we can celebrate the 20th Anniversary of the NMT 450 system in Norway. Another event is marking how mature mobile communications has become. On March 1, Telenor shut down its NMT 900 network. Later this year, UMTS will be launched.

The year 2001 has long been a symbol of the future. In 1968 Stanley Kubrick directed the movie “2001 – A Space Odyssey” based on Arthur C. Clarke’s novel. It is worth noting that most of the advanced technology described in this story is still not available. However, there are ideas and concepts which many engineers and scientists are much closer to now than 33 years ago. Even though we must all hope that computers will not develop the kind of behaviour of the mentally unstable “HAL 9000”, *talking* to computers has now become feasible. Speech based services are now being introduced in the telecom networks, with respect to recognition as well as synthesis. In his journey towards the space station “Hilton”, Dr. Heywood Floyd used an advanced video phone to wish his daughter a happy birthday. Video phones have not become widespread so far, maybe because people feel uncomfortable and unrelaxed by knowing that they can be viewed from the other end. The telephone, short message services (SMS), Internet chat, email etc. give the users some degree of “shield” against privacy intrusion. Clarke’s story gives an interesting view of what may happen when humans must relate to very advanced technology, but generally a positive attitude towards advanced computing and communications is brought forward.

A more pessimistic scenario is described by the Norwegian writer Tor Åge Bringsværd in the short story entitled “Codemus” (Norwegian: Kodemus) which was published in 1970. The story was also dramatized for the TV theatre in 1971 by NRK (the Norwegian Broadcasting Corporation). This story is a dystopian view of a future society where every person has his own “Little Brother” (women have “Little Sisters”). This pocket device functions as a personal terminal with telephone, diary, alarm clock and everything a person needs to function in this society. And it is completely speech controlled. It has continuous contact with a central database (“Big Brother”?) and one can not turn it off. The main character in this story, a young, single male named “Codemus”, suffers the problem that his Little Brother malfunctions. It gets disobedient to the central server (computer virus?) and directs Codemus to oversleep, not to go to work, approach a young lady (dating and marriage are strictly controlled and arranged by the central server) and even throw away his Little Brother. Little Brother insists that Codemus makes his own choices in life, however, he is not used to this and gives up and lies down crying in a park. Finally he is “saved” by automatic surveillance robots and taken care of. Codemus gets a new Little Brother and is happy again.

This is a science fiction story as well as a social satire. It points out the negative consequences of advanced computer and communications technologies if used by authorities as a means of social control. It also points to the danger of being too dependent and reliant on the technology, so that if it breaks down, we do not know how to handle simple everyday situations. Just think about our own reactions if the email system or the cellular phone break down during a normal working day.

This issue of *Telektronikk* is dedicated to what is happening at the frontier of mobile and personal communications. You will find papers from leading specialists in fields ranging from user and service aspects down to basic radio transmission problems. The content is organised in four thematic sections: *Basic technology, Heterogeneous access, Network enhancements* and, last, but not least, *User and service aspects*. Additionally, there is an introductory section giving a broader view on the visions and foreseen trends “beyond UMTS”, or towards 4th generation mobile communications.

Most of the papers are technology oriented but future mobile communications is also about providing advanced personal services on a pocket device to users, and the human and political sides are important to be aware of when developing the technology. Some of the questions are: Who is going to control the technology? Does anybody “own” the user? One of the subjects handled in this issue of *Telektronikk* is *user-centric communications*. It is time to put the user in focus, and let her/him control the situation.

Per Hjalmar Lehne

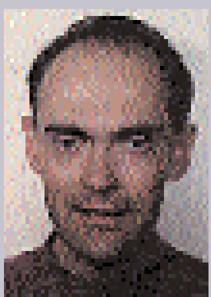
Towards Fourth Generation Mobile Communications

PER HJALMAR LEHNE, RAGNAR ECKHOFF, JOAR LØVSLETTEN, ANNE MARI NORDVIK AND STEIN SVAET

So far, the evolution in mobile communications has clearly been step-wise. The generation term stems from this fact. Now we are facing the introduction of the third generation systems – 3G, and the thinking about the next step, or "4G", has started. But will it be a "step" in the sense that we have seen before? There are a lot of signs pointing towards a smoother evolution, where keywords like heterogeneity, integration and diversity are central. This does not mean that technological development stops, rather the contrary.

Per Hjalmar Lehne (42) obtained his MSc from the Norwegian Institute of Science and Technology in 1988. He has since been with Telenor R&D working with different aspects of terrestrial mobile communications. 1988 – 1991 he was involved in standardisation of the ERMES paging system in ETSI as well as in studies and measurements on EMC. His work since 1993 has been in the area of radio propagation and access technology. He has participated in the RACE 2 Mobile Broadband Project (MBS), COST 231, and COST 259. From 1998 he was leader of Telenor R&D's smart antenna project. He is currently involved in work on 4th generation mobile systems and the use of MIMO technology in terrestrial mobile networks.

per-hjalmar.lehne@telenor.com



Ragnar Eckhoff (34) is Adviser at Telenor R&D, Kjeller. He is working in the area of mobile radio communication, particularly with radio propagation and smart antennas. Eckhoff graduated from the Norwegian University of Science and Technology (NTNU) in 1991 with a Master (siv.ing.) degree in radio communications. Before joining Telenor R&D he worked for Telenor Conax from 1992 to 1995 and NTNU from 1995 to 1999.

ragnar.eckhoff@telenor.com

Introduction

In modern cellular mobile communications, the *generation* term has been used to denote technological jumps. It all began with the first *automatic* cellular systems. The Scandinavian PTTs were pioneers when the NMT-system were specified (start 1969) and launched in 1981 (NMT-450). Together with other *analogue* systems, like the British TACS and the North American AMPS, they constitute the *1st Generation* or "1G". The 2nd Generation is of course the GSM generation, the first digital cellular systems. Other 2G systems are e.g. the North American D-AMPS (IS-54, IS-136), also called TDMA, and CDMA (IS-95). The Japanese standard PDC is also 2G.

We are now facing the introduction of the 3rd Generation systems (3G), in Europe called UMTS.

It follows that 4G denotes the *4th Generation*, currently a loose name for technologies and systems that may represent a new jump "beyond UMTS".

The time period between the introduction of new generations seems to follow a 10 year cycle. It has also taken approximately 10 years from idea to reality. The work on NMT started in 1969 and the system was launched in 1981 (12 years). GSM specification work started in 1982 and the system was launched in 1990 (8 years). The same year, work on the UMTS idea started, and the first commercial services will be available in 2001 (11 years).

Up to, and mostly including, 3G (UMTS), the generation steps have been quite distinct, but we now seem to face more diffuse transitions. The first automatic cellular systems represented something totally new, both from the user side, as well as containing significant, technological innovations.

The next step towards GSM and similar systems contained new jumps, however primarily on the technological side. The change from analogue to digital radio transmission led to the development of advanced speech coders (for example to save bandwidth) and equalizers (to handle the channel

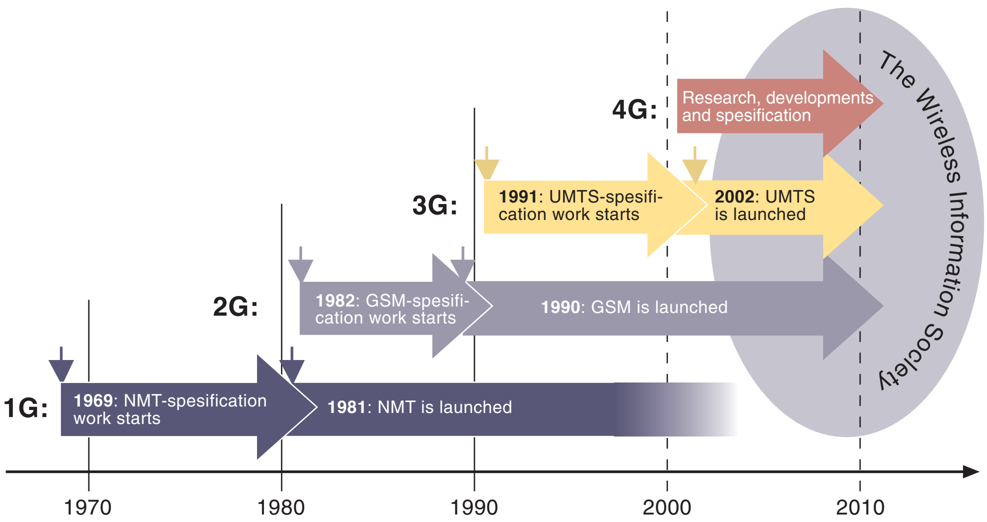


Figure 1 Specification and development time, as well as the generation cycle seem to be approximately 10 years

Joar Løvsletten (53) received his degree as Chartered Engineer in Telecommunications from the Norwegian University of Science and Technology in Trondheim in 1976. He has been with Telenor since his graduation. He has been working with radio relay and radio access systems. Since 1995 he has been with Telenor R&D working with DECT and future mobile communication systems.

joar.lovsletten@telenor.com

Anne Mari Nordvik (33) is Research Scientist at Telenor R&D, Kjeller, where she has been working in the mobile and personal communications group since 1997. Special interests include 4th generation mobile networks, UMTS and IP based cellular networks.

anne-mari.nordvik@telenor.com

dispersion introduced by multipath propagation). These two aspects had not represented major problems for the analogue narrowband systems. A good data bearer service was also introduced.

In UMTS, the introduction of a new multiple access method, CDMA, is a new technological jump. At the same time offered bandwidth increases. Additionally, a new world is introduced for the users, namely the possibility of "real" multimedia services on the mobile.

Important Events in the History of Wireless Communications

Radio communications all began when Guglielmo Marconi in 1895 demonstrated that electromagnetic radiation could be detected at a distance. Before that both he and Heinrich Hertz had performed fundamental experiments in the 1880s.

The cellular principle developed by Bell Laboratories in the 1970s is the basis for all the different 1G, 2G and 3G systems which are in use and being planned.

Another concept for mobile radio communications emerged from the early Internet. The original concepts underlying the Internet were developed in the mid-1960s at what is now the *Defense Advanced Research Projects Agency* (DARPA), then known as ARPA. The original application was the ARPANET, which was established in 1969 to provide survivable computer communications networks. The first ARPANET node was located at the University of California, Los Angeles. Additional nodes were soon established at Stanford Research Institute (now SRI International), the University of California at Santa Barbara, and the University of Utah.

At the same time, the *ALOHA Project* at the University of Hawaii was investigating packet-switched networks over fixed-site radio links. The ALOHANET began operating in 1970, providing the first demonstration of packet radio access in a data network [3].

The development of the ALOHA protocol for wireless packet transmission laid the foundation for today's wireless LANs. Better and more efficient protocols were developed taking into account some fundamental properties of the radio medium.

Background for UMTS

Research Activities

A lot of the early work towards continuously better, more efficient and flexible concepts was done in international research programmes. In Europe, the EU has played an important role as the driving force through the so-called research framework programmes: RACE, ACTS, and currently, IST. The project list also shows the increasing importance of mobile communications. In the first RACE programme (RACE 1), there was only a single mobile communications project, RACE 1043 Mobile (1988 – 1991). This project actually proposed ideas for both UMTS and MBS, a project in which Telenor participated.

In the RACE 2 programme several projects were dedicated to mobile communications. Most of them were directed towards studies and demonstrations towards UMTS, like access method studies in CODIT and ATDMA and network aspects in MONET.

The main focus of the ACTS programme was still UMTS and several projects developed important foundations. One of the most important to mention is the FRAMES project, which developed the WCDMA concept adopted by UMTS.

From Research to Standards

Standardisation towards what we today call UMTS started in the 1980s in ITU. The term used then was *Future Public Land Mobile Telecommunications System – FPLMTS*. Later the term used by ITU has been *IMT-2000*. In Europe, ETSI had started the preparations when the GSM work was at its highest, and in 1991 the SMG5 group was formed with the mandate to define UMTS. The first years of the work in

Some important events in early wireless history are [1]:

- **1901**: Marconi demonstrated *the first radio telegraph transmission* across the Atlantic Ocean.
- **1915**: *The first wireless voice transmission* between New York and San Francisco signalled the beginning of the convergence of radio and telephony.
- **1946**: *Public mobile telephone service* was introduced in 25 cities across the United States.
- **1947**: D.H. Ring at Bell Laboratories proposed *the first cellular concept* [2].
- **In the 1970s**: Researchers at Bell Laboratories developed the concept of the cellular telephone system, in which a geographical area is divided into adjacent, non-overlapping, hexagonal-shaped "cells".

Stein Svaet (41) is Senior Research Scientist at Telenor R&D. He has been with Telenor since 1986, and with Telenor R&D since 1988. He has been working within the field of mobile communications, in particular radio planning, system simulation and network design. His current interest is within network design and services for UMTS and future mobile communications systems. He is currently involved in the standardisation of UMTS in 3GPP.

stein-wegard.svaet@telenor.com

SMG5 were clearly a preparation phase. It started out with a similar approach as GSM, namely to define a complete, new system. The work went through a lot of important processes during the first years, but it was not until the work was reorganized that the progress evolved. When the GSM work declined, the UMTS responsibilities were split on the existing SMG sub-committees in 1995-96, and SMG5 was formally terminated in 1997.

In January 1998 ETSI decided on the radio interface for UMTS based on several candidates, of which two came from the above mentioned FRAMES project.

In the USA and Asia other groups were doing similar work towards a 3G system, and after some pressure from the ITU to harmonize the radio standards the *Third Generation Partnership Project – 3GPP* – [4] was established in 1999. 3GPP consists of the organizations ARIB, CWTS, ETSI, T1, TTA and TTC (see Table 1). 3GPP's mission is to produce standards for a 3G system based on evolved GSM core networks and radio access technologies they support, like UTRA.

ARIB	Association of Radio Industries and Businesses (Japan)
CWTS	China Wireless Telecommunication Standard Group (People's Republic of China)
ETSI	European Telecommunications Standards Institute (Europe)
T1	Standards Committee T1 Telecommunications (USA)
TTA	Telecommunications Technology Association (Korea)
TTC	Telecommunication Technology Committee (Japan)

Table 1 Organisations behind 3GPP

Early Activities beyond UMTS

One of the early projects addressing communications with a significantly higher capacity than UMTS was the *MBS* project (1992 – 1995) [5] [6]. The aim of this project was to perform a system study and develop specification of a cellular wireless ATM concept with high mobility. The system should offer an end-to-end bitrate of up to 155 Mbit/s. A demonstrator was also built (Figure 2).

At the time of the project, it was believed that the MBS concept should be realizable on the same time scale as UMTS. However, the technological level (60 GHz mm-wave technology), as well as the lack of a market for such high bitrate services, made this impossible. If we look at it today, 5 years after the project ended, we can say that the concept was ahead of its time. The stud-

Figure 2 The MBS demonstrator at Daimler-Benz Research in Ulm, Germany, 1995. The "base station" vehicle with the mm-wave antennas raised using a hoist

ies continued in the ACTS programme with the SAMBA project, and a new and better demonstrator was built and shown during the EXPO'98 in Lisbon [7]. We can conclude that MBS is a 4th generation concept, and today no one will doubt that the bandwidth will be needed and exploited.

There were also other projects in the ACTS programme which addressed what we could call 4G, like MEDIAN and WAND. In these projects high capacity WLANs were studied.

There are three important trends which show that even more efficient and flexible technologies must be developed:

1. The number of cellular subscribers world-wide is growing exponentially.
   - In 2000 there were more than 500 million subscribers world-wide. The number is expected to pass 1 billion in 2005. At the same time, the number of mobile subscribers is expected to outnumber the fixed connections.
2. The number of Internet connections is growing exponentially.
3. The sale of wireless LAN solutions is growing significantly.

The Drivers towards 4G

It is important to raise the question whether UMTS will provide the platform for the new mobile multimedia world. UMTS is undoubtedly an important step on the way to integrating different communication needs on a mobile platform.

What are the driving needs for a system beyond the upcoming 3G (UMTS)? Will the forces be capacity, meaning traffic per unit area, or will it be a question of throughput to handle multimedia services? Are new services coming up which put new demands on the network other than the mentioned ones?

The Capacity Question

This seems to be a never ending demand. The FMC trend will reach the less developed and very highly populated countries demanding an even higher capacity. In the Scandinavian countries all the previous systems have run out of capacity, perhaps with the exception of GSM 1800 (so far).

The trend is towards a huge future demand for mobile Internet access. Communications and broadcasting will merge, and what we get access to by wire, we also want on the "mobile".

When work towards 3G started, no-one was able to see how much the Internet would grow.

So far, the systems have almost only been carrying voice traffic, however, this will not be the situation in the 4G time frame. The 3G network will give a good indication of the evolution from voice focused services to first generation multimedia services. It is therefore tempting to look into the situation in the fixed network as well as the PC software industry. This can give us a basis for what is a possible development when we are looking into future demands. In both cases the applications are more demanding regarding CPU power and throughput. The future network, or networks, must aim to follow the general development, taking into account the natural properties and impairments of the radio channel.

Which new types of requirements must be translated into more general requirements to a 4G network? Two important key words can be *mobility* (services, VHE) and *security*. To optimize such a system towards a global market scalability will also become more important than today.

What is 4G?

Up to, and mostly including, 3G (UMTS), the generation steps have been quite distinct, but we now seem to face more diffuse transitions. It is clear that we no longer specify a complete system.

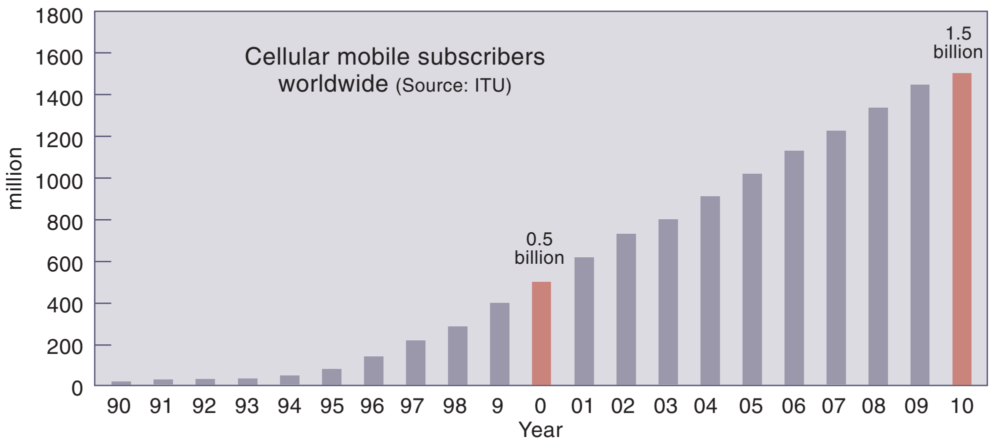


Figure 3 Projected cellular growth [8]

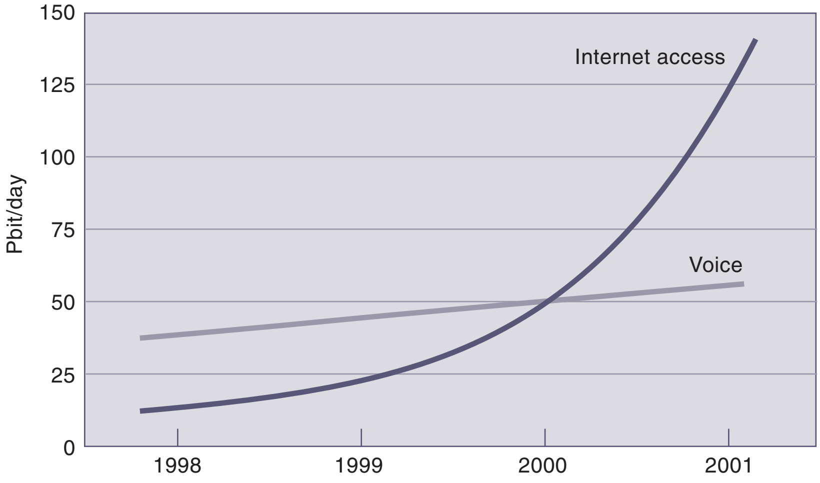


Figure 4 Internet growth (From OVUM, Global Telecoms & IP Markets, 1999) [9]

This is especially visible on the network side, where existing and alternative platforms are used. Great effort has also been put into widening the UMTS term to be more than a mobile access technology.

There are many opinions about what 4G represents. The EU is engaged in this development, primarily through the research framework programmes. As mentioned earlier, there have been projects with a distinct 4G viewpoint, at least viewed from a distance. In the 5th framework programme, IST, the term 4G is put forward, and there are several projects studying different aspects. The view of the European commission as well as one of the major projects in IST called BRAIN, is comprehensively described in other papers in this issue [10][11].

There is obviously a development towards closer interworking between different technologies, especially on the access side. An important factor is also the development of open, technology independent application platforms like e.g. OSA. The IP platform currently has great momentum, partly because it represents a possible common denominator for transport and access to services independent of the different access platforms.

In this paper, we go further into three main areas which we believe are important in order to understand what 4G is about and why the development does not stop with UMTS.

In the next chapter, we address the user aspects, the trends and possibilities. Further, we deal with the radio access technology aspects, spanning from antenna systems to the notion of "reconfigurable mobile communications". Finally, we look at the development towards realizing the so-called "All-IP" vision, which seeks to enable service mobility independent of access technology and a lot of new applications.

4G seen from the User

In 2G SMS has been a successful service that nobody did foresee, and it remains to be seen what will be the killer application in 3G. Trying at this point to foresee which kind of services and applications will be predominant in 4G is more like guesswork. However, there seems to be a common understanding of some trends.

The development of mobile communication from 3G onwards is expected to evolve in a smoother way than the generation steps so far where new heavy technical platforms have been introduced as the bases for new generations. New technologies emerging today and technologies still behind the horizon will, together with the 3G technology, form a heterogeneous network that step by step will increase capability.

Which direction or directions the development takes will to a greater extent than thus far be determined by the market. The introduction of 3G will open for a multitude of different types of services and applications which are new to the users. The response to all new possibilities of mobile communication presented will in addition to paving the way for new services and applications also indicate areas where enhanced capability is required and thereby influence the development of the technical platform.

Terminals

For the user the personal pocket terminal will be a central part of the 4G technology. In principle the terminal will be a powerful PC with a variety of communication capabilities, and a range of terminals having different capabilities will exist. The simplest ones will be able to handle a limited set of access technologies and a limited set of services and applications. As users get used to communicating using pictures and video in 3G, a video camera and a screen to view received videos will be a requirement for 4G terminals.

User Interfaces

The personal pocket terminal consists of a computing/communication part with a user interface. The user interface, in addition to being integrated in the computing/communication part of a 4G terminal, might also exist as a separate unit and communicate by using short range radio technology. In this case only the user interface will be a wearable for users. The computing/communication part may remain in the pocket or the briefcase during a communication session or when using an application of the computing part. One option for a remote user interface could be a screen built into a spectacle-like device which also includes a microphone for voice recognition and a video camera. Included in the user interface integrated in the computing/communication part will also be a built-in camera. The camera or cameras in the terminal will have the options to catch motives in the direction away from the person communicating, e.g. when transmission of video postcards take place, or to be used as part of a videophone turned towards the person communicating.

Large screens with high resolution will be available for use with personal terminals. Several types will exist. One type of screen will be an extremely thin foldaway, comparable to a sheet of paper, which accordingly is convenient to bring along. Even 3D pictures are expected to be available with 4G.

Input commands to the terminal will mainly be given in speech, but it will also be possible to use a keyboard. A combination of speech and pushing buttons or icons on a display will also exist.

Data Capabilities of the Terminals

In addition to being a communication device the personal pocket terminal will contain what a person usually carries with her and that can be made electronic.

Examples are keys, PDAs, business cards and contents of the wallet, e.g. money, ID-card, different kinds of magnetized cards, driving licence, passport, tickets, receipts and so on. The electronic money will be international and no change of currency will be necessary when travelling abroad.

Short Range Communication

An essential communication part of the 4G terminal will be a short range type similar to Bluetooth. In addition to the communication between the user interface and the main computing/communication part a short range link will be established when the terminal is used for example to unlock a door or to transfer money when shopping. The short range communication will also be used when a connection is transferred from the pocket terminal to another terminal which is more suitable for the type of communication taking place, for example a fixed terminal having a large screen for a video connection.

Terminal as Enabler

Keys for accessing private networks like intranets will be stored in the personal pocket terminal. The same goes for access to certain services or sites on the web. For example, when accessing a network behind a firewall from a foreign terminal, the pocket terminal will function as the enabler of this access.

Figure 5 What can be made electronic goes into the terminal

Personalisation of the Terminal

The pocket terminal will be personalised and contain the user's personal profile and sufficient information to offer personal context based services.

Such information may be based on initial configuration of the terminal and through a learning process of the user's ways of utilising the terminal. Necessary actions to establish a certain type of connection or perform some task may be initiated by a speech command or a keyboard, and the rest will happen automatically if the user does not want to do it manually.

User Scenarios

The "there and then" possibility introduced by mobile communication means that a call can be made there and then when a need occurs. This will in many cases make life easier. For example if somebody is picking you up at the station and the train is delayed, you can make a call or send an SMS from your mobile phone and inform about the delay. Without a mobile phone the person picking you up would either have to get information on the arrival time herself, or you could try to make a call from some station along the line.

Terminal mobility also offers benefits regarding incoming calls. When an important call is expected, you do not have to wait by the fixed phone, but can receive the call wherever you are.

Two generations of mobile phones have made subscribers used to making and receiving calls when needed more or less independent of circumstances. With 3G more new services will be available and used the same way. If a person suddenly remembers she has to pay the rent for her flat, she can make a transaction using her mobile terminal there and then wherever she is. With 4G the same development will continue with even more advanced services.

Video communication as a utility service:

- In a car accident where people are injured remote assistance from professional medical staff can be given when a video connection is established. People at the site of the accident can be instructed in how to handle an injured person and can be supervised while carrying out first aid.
- A video recording as documentation for an insurance company of what has really happened in an accident can be sent to a personal server and stored.
- If problems occur in the production process in a factory a remote specialist can get detailed information on the situation if a high quality video is transferred and thereby be able to solve the problem.
- Video postcards can be sent from more or less exotic places when people are travelling. Video connections can also be established when people are in a particular situation or place and they want to show it to others.

When trying to identify services and applications in 4G and how they are used one approach is to divide the utilisation into two categories based on usefulness. One category represents all the cases where 4G might be regarded as a useful tool making work more efficient and everyday life easier. The other is where 4G is used for entertainment. All services can then be placed under one of the two categories utility and entertainment.

A lot of examples can be found where future mobile communication can be used for utility purposes. To foresee how the capabilities of 4G can be used for entertainment is more difficult. However, the greatest potential for high revenue services is expected to be in the entertainment segment.

Images and Video

Transmission of pictures and video is expected to become an increasingly more important part of mobile communication after it is introduced in 3G. Enhanced capability of terminals and networks towards 4G will make live pictures become a normal part of a communication session. Pictures and video might be used both in communication between people and in man-machine communication.

When people communicate the terminals might be used as videophones sending pictures of the people talking along with the voice. However, it will probably become more and more common to show something from the surroundings or a certain object that is related to the theme of the conversation, i.e. the camera is turned away from the person speaking. Examples of video communication under the utility category are many (see frame).

A situation quite unpopular by most people is having to stay at home to wait for a workman that needs access to the flat during working hours. If you are away and the workman needs to enter, you are notified from the home network and can verify through a video transmission that he is the person he makes out to be. Then the door can be unlocked remotely by sending the right command to the home network.

The entertainment category is expected to contain many new applications. Different types of games used on PCs and TVs will continue to increase performance regarding animation and graphics. Types of games where more than one person participate will also be available on personal terminals. In families where one person is often away travelling, games where the whole family participates may be popular.

Different kinds of lotteries and gambling services are expected to be interesting businesses utilising 4G networks.

The sex industry was an early adopter of Internet and this industry may also want to utilise the video capability in the future mobile technology to offer services.

Short Range and Location Based Services

Short range technologies will play an increasingly important role in the development of information and communication technology. It will be essential for utilising different applications of the computing part of the personal pocket terminal.

The location capability of future mobile systems will be widely used and will form a platform for many new services (see frame on next page).

Context Based Services

Location based services are an essential part of context based services. An example of a context based service could be that the terminal knows what time dinner is and also that the user is away from home and therefore lets advertisements from nearby restaurants through.

Another example may be that the user is relaxing in his hotel room and his favourite soccer team are playing. The terminal then makes him aware of this and asks if he wants to watch. Should he be driving a car the terminal would know this because it would have been used for starting the engine, and consequently would not ask if he wanted to watch.

Advertising on personal pocket terminals is expected to be a large business area in 4G. However, most users will accept advertising and other push services only in certain situations, and advertising towards personal pocket terminals will therefore to a large extent be context based. Context based services, especially towards young people, are comprehensively treated in two other papers in this issue [12][13].

Access Technology Trends

Heterogeneous Radio Access

Traditionally, a mobile radio system – like GSM – has been synonymous with a specific radio access technology. With the introduction of 3G (UMTS) things have started to change. Even if UMTS has its own radio interface – UTRAN – it was of great importance during the standardisation to make the system as compatible with GSM/ GPRS as possible. The original UMTS concept included a satellite access part for large-area coverage, although a definite solution has yet to be standardised. At the moment, work is ongoing to integrate the upcoming HIPERLAN/2 technology into UMTS, the idea being to use HIPERLAN/2 as a supplement to UTRAN in hot-spots where additional capacity is needed.

The UMTS case points in the direction of the future, where a number of different access technologies are expected to play together in an integrated manner as shown in [10]. It is important to note that 4G does not mean a specific new radio technology, but rather a way to integrate a number of different technologies – with reconfigurable radio being a central element.

Observe that the 4G access portfolio covers much more than just traditional cellular technologies like GSM. 4G is not a new cellular system, but rather the convergence of a range of systems like cellular, WLAN, broadcasting and so forth.

A journey abroad by plane can illustrate how the short range communication and location services can be utilised in different situations:

- When booking a hotel room the booking registration together with the accurate position of the hotel is stored in the personal pocket terminal. Also a map of the foreign city is downloaded.
- The ticket for the transport to the airport is paid for and stored using the terminal. The receipt is also stored in the terminal. To pay for the ticket money is transferred electronically from the terminal by the short range communication.
- The electronic ticket for the flight is automatically controlled at the airport and information on gate number and updated departure time is conveyed to the terminal. If requested, a map showing the existing position of the passenger and the position of the gate can be downloaded to the terminal. Also estimated walking time to the gate can be given.
- At the gate the ticket is automatically validated by short range communication.
- On arriving at the destination the electronic passport is controlled in an automatic way.
- On the bus to the city centre the terminal communicates with the internal information system on the bus, the bus stop closest to the hotel is presented on the terminal and the position of the bus stop is indicated on the map. A walking route from the bus stop to the hotel is also presented on the map.
- Arriving at the hotel a short range communication takes place between the information system of the hotel and the terminal. As both the terminal and the hotel system have the room reservation the terminal is recognized as an enabler to access the room and an electronic key is loaded into the terminal.
- When it is time for dinner the terminal will provide information of nearby restaurants.

Interworking

With the multitude of access forms that are expected in future mobile communication it is important to clearly understand and find novel solutions for access network interworking. This is reflected in EU's IST programme [14], where several projects deal with these issues. The BRAIN project looks at interworking between a network based on an improved HIPERLAN type technology and cellular networks like GPRS and UTRAN, [9][11] while the WINEGLASS project covers interworking between UTRAN and more traditional WLANs. In the DRIVE [15] and MCP projects, the focus is on interworking between terrestrial broadcast systems and cellular systems to support multimedia applications in cars.

Interworking is also an important issue for Telenor. There are projects dealing with both hybrid access as well as a common IP based core network.

Software Defined and Reconfigurable Radio

From the end user perspective, a multi access type scenario contains two extreme cases:

a) One terminal for each access type, e.g. a TV set for digital terrestrial TV (DTV) and a cellular WAP phone for GPRS. This is essentially what we have today, but in the future it is expected that there will be a close interaction between the different terminal types. One solution is the virtual terminal concept as described in [16]. With this concept all the different terminals, or communication devices, will be considered as one big virtual terminal with multiple input and output capabilities. From the network perspective only one terminal is seen – the virtual terminal – to which all kinds of services are delivered. The virtual terminal then decides which communication device is most appropriate for interaction with the user at the given time.

The idea of reconfigurable mobile communications raises a number of problem areas, including:

- how to detect which radio networks are present in a given area;
- how to determine the most suitable network to connect to;
- how to perform handover between networks with different radio technologies, including download of necessary software;
- how to cope with different quality of service handling in the different networks;
- how to provide acceptable security solutions;
- how to handle "charging & billing" in such complex scenarios.

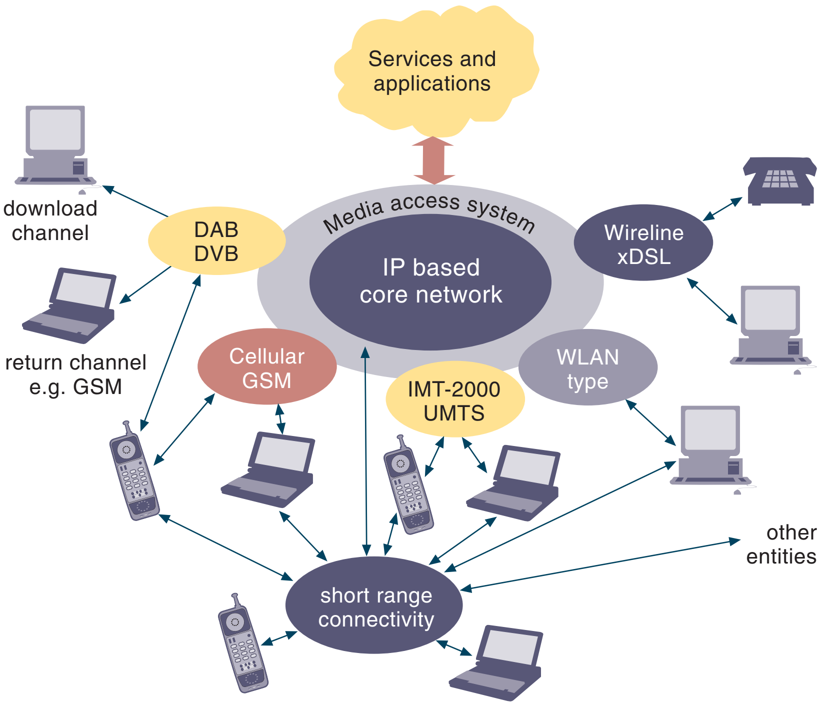


Figure 6 The future mobile network as seen by the IST TRUST project (from [19])

b) The same terminal for all access types, e.g. a PDA-type device for both GPRS and WLAN access. This will require multiple radio modules in the same device, intelligence to select the most appropriate access form at the given time and functionality for keeping the communication session while switching between access forms. The concept of software defined radio (SDR) is regarded as a key enabling technology in this context and it is further described in [17].

In practice we would likely end up somewhere in between the two cases outlined above, i.e. with a set of different terminals each supporting one or a limited set of access forms. In this case, both the virtual terminal concept (or something similar) and SDR can be expected to play important roles.

The SDR concept is not only applicable on the terminal side, but also on the infrastructure side. In Eurescom project P921 – UMTS Radio Access [18] – the use of SDR in implementation of a hybrid UTRAN and GSM/GPRS access network has been investigated. The project concludes that the use of SDR in combination with hybrid fibre radio (HFR) can be a cost-effective solution in certain scenarios. SDR as research area is surrounded with great interest internationally.

The introduction of SDR in the user terminal opens up interesting possibilities. By updating the software the terminal behaviour can be changed, e.g. to support another access technology. The update could be done via the air interface to enable the terminal to use an access technology not originally supported. In the ultimate case, this will eliminate the need for standardised radio access solutions. However, it is not trivial to make this work. While the SDR technology itself has started to mature, no solution exists today to do safe and effective software upgrade via the air.

The IST TRUST project [19] addresses the topic of reconfigurable radio and aims at finding solutions to these problems. The project is taking the user's perspective by seeking solutions that give the user what he/she really is after and in a manner that is user-friendly.

Figure 6 illustrates the TRUST project's vision for the future mobile network.

Emerging Radio Technologies

Radio is a complicated medium in the sense that the radio channel is time varying and unreliable and that radio spectrum is a scarce resource. As a result, the available data rates over radio systems have always been considerably lower than over comparable wired systems. The radio hop is the bottleneck of most wireless communication networks and there is a great demand for more capacity over radio.

One clear trend is that both applications as well as hardware requirements grow much faster than radio transmission capabilities as documented by Noll and Burrachini in another paper in this issue of *Telektronikk* [17].

Even though this comparison may be claimed to be unfair it illustrates that the increase in data rates for radio systems has shown a considerably slower progress than what has been observed in other areas of electronics.

In order to meet the ever-present demand for more capacity, basically two different approaches can be taken:

a) Take into use new, previously unused parts of the radio spectrum;

b) Improve the spectrum utilisation, i.e. increase the number of transferred bits/s per bandwidth unit (Hz).

Both approaches are being followed in current research activities. In the case of new spectrum, the bands around 40 and 60 GHz have been given great attention in recent years. Improvements in spectrum utilisation can be achieved by improving the physical and link layers of the radio systems and in some cases also the network layer. In the following sections we will look briefly into some promising emerging technologies.

Use of 40 and 60 GHz Bands

Use of frequency spectrum in the millimetre wave range is considered a promising way to realise high datarate wireless systems because of the large amount of available spectrum in these bands. The bands of greatest importance lie in the 40 and 60 GHz ranges.

In the 40 GHz range, CEPT has allocated the spectrum from 40.5 to 43.5 GHz, a total of 3 GHz bandwidth, to so-called multimedia wireless systems (MWS) [20].

In the frequency range around 60 GHz, oxygen-absorption of radio waves is significant resulting in a higher propagation attenuation. Realising that this opens up for a high frequency reuse factor in radio systems, CEPT has recommended that frequencies in the range 54.2 – 66.0 GHz are used for terrestrial fixed and mobile systems [21].

The 60 GHz range was selected by the MBS project [5] run under the European RACE 2 programme. The MBS system used the bands 62 – 63 GHz and 65 – 66 GHz, divided into 34 (outdoor) or 17 (indoor) subchannels (carriers), using 4-OQAM (offset quadrature amplitude modulation) or 16-OQAM. Data rates up to 155 Mbit/s could be provided using more than one carrier. The system was designed for small cells (max 500 m radius) and required generally line-of-sight between the base station and mobile terminal. To support fast moving mobiles (cars, trains), advanced handover algorithms were required.

In the follow up ACTS project SAMBA [22] frequencies in the 40 GHz band were used, with a target bit rate of 34 Mbit/s using OQPSK modulation. Again, the system was designed for fast moving mobile terminals.

Ideas from the MBS project were also adopted by the ACTS MEDIAN project [23]. This time, the focus was on WLAN type scenarios (i.e. low mobility) using the 60 GHz band. Using OFDM type modulation, the target was to be able to provide data rates up to 155 Mbit/s.

No standards or commercial products have emerged so far, most likely due to the immaturity of the technology and lack of a real market demand. The only known standardisation work going on in this area is by the Japanese MMAC project, where a "Ultra High Speed Wireless LAN" is being specified for the 60 GHz band with a target bit rate of 156 Mbit/s [24]. It is to be expected, however, that work in this area gains new momentum when the market for high-rate radio systems matures.

From a broadcast and fixed access viewpoint, there is still significant activity on systems for the 40 GHz band. The IST EMBRACE project aims at designing a low cost LMDS type fixed radio access system [25]. The system is designed with a downlink capable of supplying several Mbit/s for both broadcast channels as well as high speed communication data and an uplink for virtual no return (such as 64 kbit/s ISDN) up to several Mbit/s for high quality contribution and data transfer.

Advanced Coding and Modulation

While the 2G systems, like GSM, marked the introduction of digital radio technology into cellular telephony, 3G is the great breakthrough for CDMA technology, pioneered by the American 2G IS-95 system. CDMA has many desirable properties, one of the most prominent being the frequency reuse factor of 1, i.e. the possibility to use the same frequency in all cells.

At the edge of 4G, the arguably hottest technology is OFDM – orthogonal frequency division multiplexing. OFDM has already been introduced into digital broadcasting systems like DAB and DVB-T, and is selected for next generation WLAN systems like HIPERLAN/2, IEEE 802.11a and MMAC High Speed Wireless LAN. OFDM possesses excellent properties in dispersive radio channels, i.e. with high delay spreads, eliminating the need for complex equalisers.

Another important trend is the use of adaptive modulation. Radio communication is characterised by time-varying and complex radio channels. Traditionally, radio systems have been designed for worst-case conditions, the aim being to keep the outage probability below a certain, low value. By allowing the modulation to

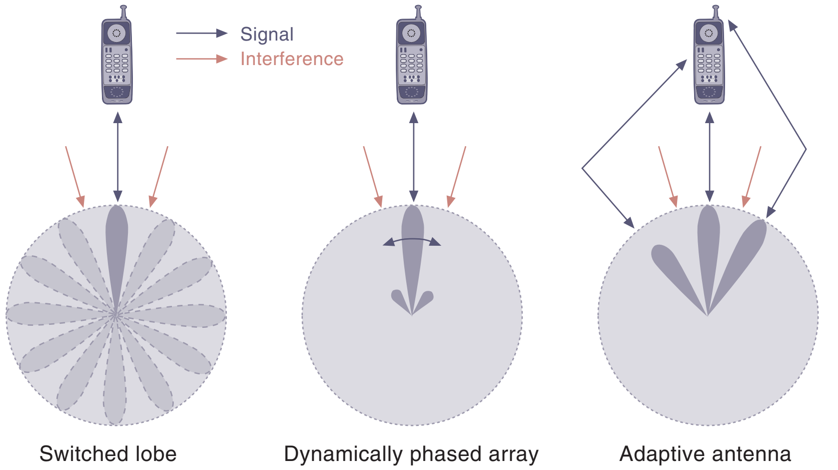


Figure 7 Smart antenna realisations (from [30])

adapt to the changing channel, using high-order modulation when the channel is good and low order modulation when the channel is bad, considerably higher average data rates are achievable. Adaptive modulation is already being introduced with systems like HIPERLAN/2, where the modulation can be changed from one OFDM symbol to another. In [26], it is claimed that even better performance is achievable if one also allows different modulation of each subcarrier in one OFDM symbol depending on the frequency response of the channel.

Developments are also done on the coding side, with turbo-code being shown great interest recently. This topic is comprehensively covered in [27] in this issue of *Telektronikk*.

Smart and Smarter Antenna Concepts

Smart antennas have great potential in increasing the spectrum utilisation in radio communication. In a traditional smart antenna, phase and amplitude of the elements of an antenna array are adjusted to form a beam towards the desired receiver/transmitter. This improves the link budget and at the same time reduces the interference transmitted to or received from the environment. Such smart antennas can be divided into three classes: switched beam, phased array and adaptive antenna (Figure 7).

In the switched lobe case, one of a predefined set of lobes is selected based on maximum received signal power. With a phased array, the lobe is adjusted to point in the direction of the strongest transmitter. The full adaptive antenna seeks to maximize received power and at the same time minimize interference, i.e. optimise C/I. This is achieved by directing lobes towards desired signal(s) and nulls towards interferers. Comprehensive tutorials on smart antennas can be found in [28] and [29].

The performances of the different concepts were evaluated in [30] based on real channel measurements at 2.1 GHz, assuming an 8-element linear array. The results were compared to a conventional sector antenna, and indicated a gain of around 6 dB in C/I for the switched lobe and phased array, and 12 – 16 dB for the adaptive antenna.

A new type of smart antenna technology has gained considerable attention the last couple of years. It is popularly termed MIMO – multiple input multiple output – and was first proposed by Bell Labs in the so-called BLAST concept [31] (Figure 8).

In traditional smart antennas, the same information symbol is sent on all antenna elements, with the phase and amplitude of each element being adjusted to direct the power in the desired direc-

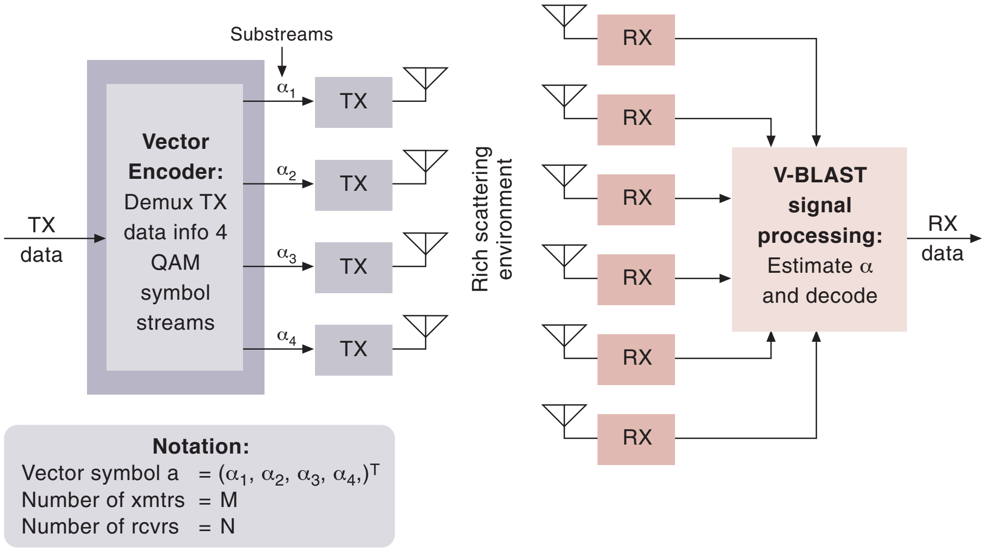


Figure 8 Bell Labs' BLAST concept (from [31])

tion. In MIMO systems, different symbols are sent on the different antenna elements. By using antenna arrays on both the transmit and receive sides and assuming a radio channel characterised by rich scattering, a number of uncorrelated radio channels can be realised. In the ideal case this number is equal to the number of elements of the "smallest" array (4 in Figure 8). Compared to the case with one antenna at both transmitter and receiver, the capacity of the link is increased by a factor equal to the number of uncorrelated channels. For MIMO to work, an accurate channel estimate has to be made by the receiver. In laboratory experiments, Bell Labs have demonstrated spectrum efficiency in the order of 20 – 40 bit/s/Hz.

The principles and possibilities related to communications with multiple antennas are treated in [32] in this issue.

Packet Based MAC Protocols

In the future, it is expected that most of the traffic on wireless systems will be packet based data, as opposed to circuit-switched voice in today's cellular systems. In order to optimise the use of the radio spectrum, the MAC protocols should be designed to reflect this change. In [33], AT&T proposes an OFDM based cellular system using 5 MHz channel raster, the same as being used by UMTS. By optimising the MAC protocol for packet based data transfer, a significantly higher capacity is achieved than in UMTS. AT&T claim that the system will be able to deliver 2 – 5 Mbit/s in macro-cells, while the goal for UMTS is 384 kbit/s.

Relay protocols

In most radio systems, there is only one radio hop. However, from a pure link-budget viewpoint, it is beneficial to split a single, long hop into several smaller hops. This is one of the main ideas behind radio relay protocols. In the ODMA concept, originally proposed as a candidate for UMTS, the idea is to use other mobile terminals as relays between a given terminal and the base station. By doing this, the amount of power used for transmission – and thereby the total interference level in the system – can be decreased, and the range of a base station increased. The concept is shown in Figure 9.

ODMA is adopted by 3GPP as an option in the TDD mode, but the standardisation is only in the initial stage and has shown little progress lately. Preliminary results documented in [34] indicate that ODMA has a great potential for increasing the capacity in UMTS, although there are still unresolved issues before the system can be implemented – not the least regarding routing protocols.

Trends on the Network Side

The trend is towards convergence between the telecom world and the Internet. IP technology is to a large degree introduced in UMTS and the UMTS vision is to standardise an "All IP" core network in the future. Organisations like the Internet Engineering Task Force (IETF) and the IST BRAIN project [9][11] develop IP based mobility algorithms to be used in wireless access networks and mobile networks.

Why Strive towards Integration?

The customers want seamless use of services across different access networks. Mobility is a major driver in the industry. With IP it is possible to make a generic mobility handling mechanism across different networks.

The All IP Vision

The "All IP" vision is a vision shared by many. This architecture (Figure 10) shall allow operators to deliver services, both real-time and non real-time services for speech, data and multimedia over a common service platform based on IP. The user shall be given full service and terminal portability. Personalised services shall be offered across different access networks, both wireless and wireline. Eventually all services shall be handled by the IP core network.

Evolution from 3rd Generation Systems

4th Generation mobile networks will most likely be based on evolution from 3G. For the core network and service platforms we see a migration from UMTS and IP/Internet towards a common core network/service platform handling all services across different access networks. Figure 11 describes the migration path foreseen in 3GPP.

Eventually other access networks than the ones specified in 3GPP will be connected to the All IP platform. ETSI BRAN works to specify inter-

Figure 9 ODMA – Opportunity driven multiple access (from [34])

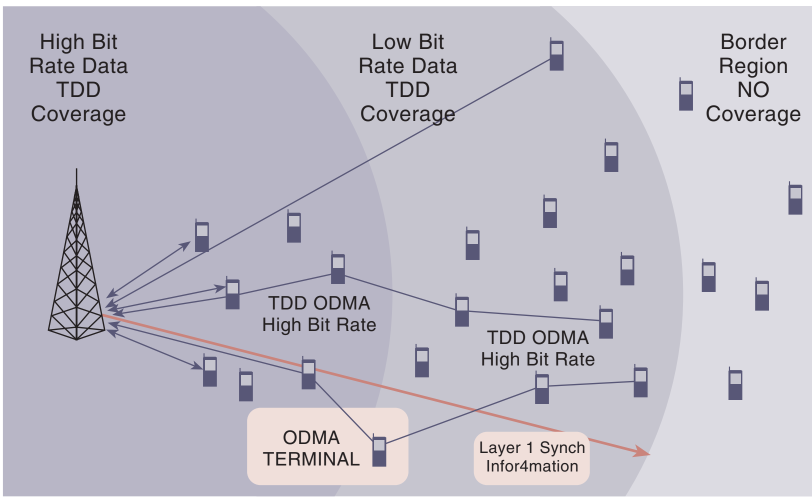


Figure 10 The All IP vision

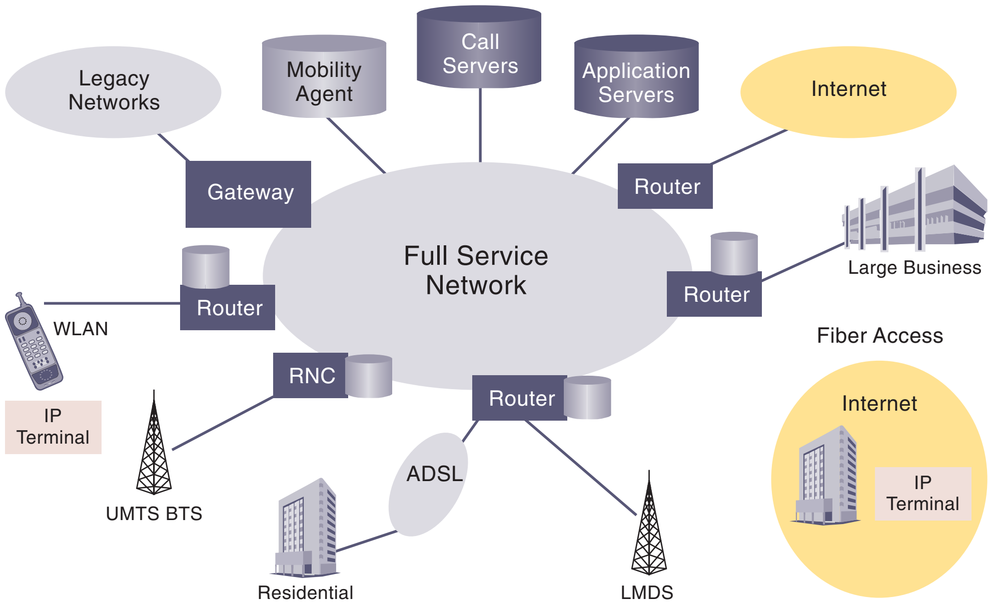
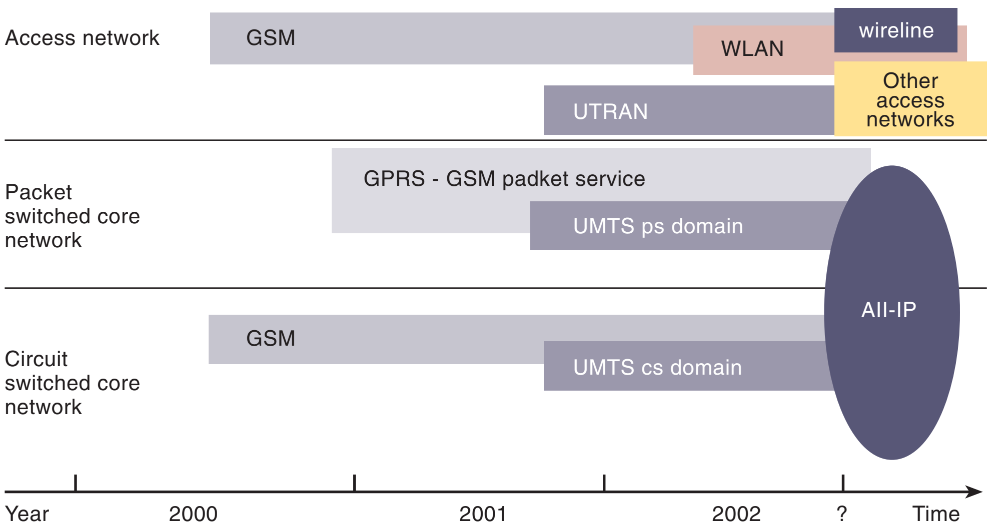


Figure 11 3GPP migration path



connection between WLAN and the UMTS core network. In UMTS the circuit switched core network will exist for some years in parallel with the packet switched core network. However, eventually the packet switched core network will offer services similar to those in the circuit switched core network, like for example voice and other real time services. When the packet switched technology can offer sufficient QoS also for real time services, the investments in circuit switched technology become superfluous, and in the end all services will be handled by packet transport. Figures 12 and 13 describe different steps along the migration path.

The Internet GPRS Support Node is a combination of the SGSN and GGSN node in UMTS, with some additional requirements. Here Foreign Agent (FA) functionality has been added to the IGSN. We find UMTS specific mobility handling internally in the IGSN, and Mobile IP+ will take care of mobility management higher in the hierarchy, between IGSNs and for intersystem mobility. This post 2000 scenario is described by 3GPP [35].

Figure 13 describes another scenario which has been evaluated by Telcordia which is a further integration of IP into the UMTS network. Here the traditional core network as we know it today in UMTS is totally removed and replaced by an IP network. The RNCs are connected to a gateway serving one domain. Mobility locally within one RNC is still handled by UMTS specific signalling. For mobility within the gateway, it is suggested to use Cellular IP, and for mobility outside the domain, Mobile IP+ will be deployed.

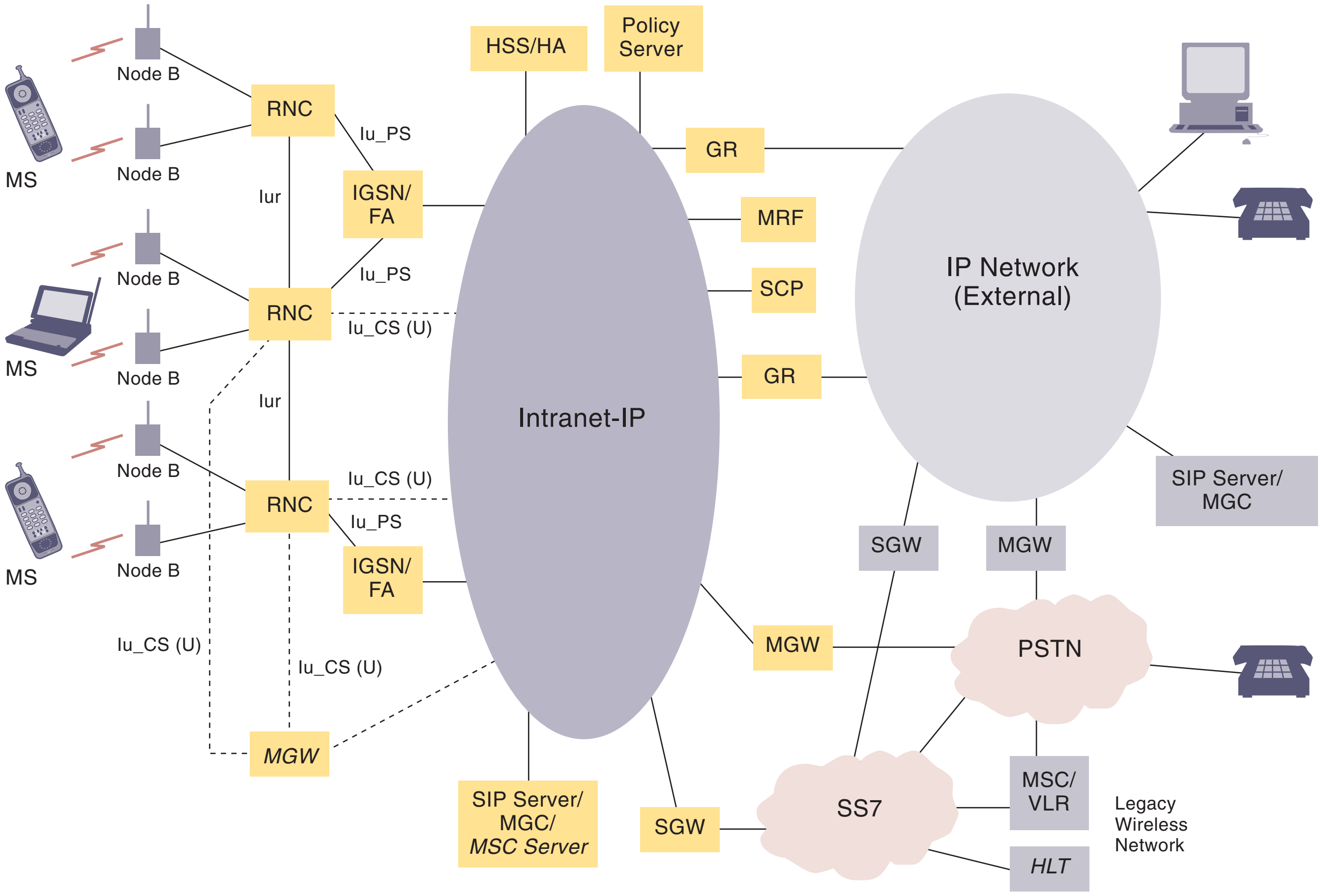


Figure 12 IGSN based All IP network

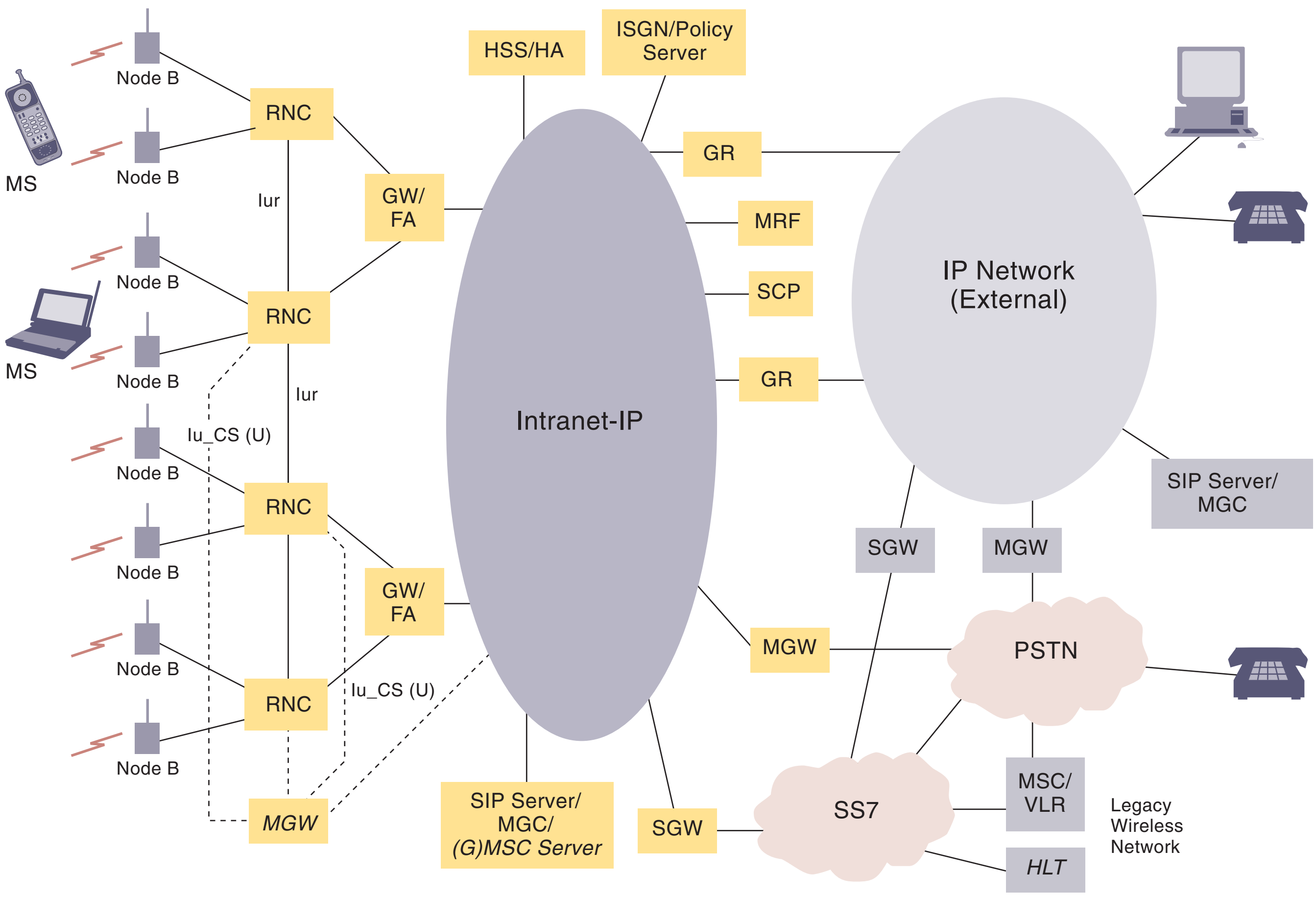


Figure 13 IP-to-RNC based All IP network

The IP world has some problems to solve before introducing new mobility algorithms in mobile networks. Some challenges are:

- Mobility management based on IP must be fast enough to handle real-time services.
- The overhead and length of headers need to be minimized for optimized use of radio resources.
- Differentiated service quality must be implemented (more generic IP problems).
- Security functions must be at least as good as in 3G networks.

IP Based Mobility Handling Mechanisms

Several mechanisms for mobility management based on IP are now being specified and looked at in different fora. These mechanisms are generic and will be independent of the access networks used. So the goal is to use these algorithms to create a seamless service environment across different access networks and replace the technology specific algorithms handling mobility with the new generic ones.

Mechanisms for both micro and macro mobility are being specified in IETF and evaluated by for example the IST project BRAIN [9] [11].

Many of the new IP based algorithms are two layer mechanisms to handle micro and macro mobility separately. The network is separated into administrative domains, where every domain includes a number of base stations. Local mobility within the domain (micro mobility) is handled by a specific mechanism, and mobility outside the domain is handled by Mobile IP in most algorithms.

One reason for this split is to reduce the global signalling load. This is possible since local mobility is handled locally, without notifying the home network.

Some alternative IP based algorithms for intra-domain mobility are:

- Cellular IP;
- HAWAII;
- Mobile IP Regional Tunnel Management + Hierarchical Mobile IP with Fast handoffs;
- TeleMIP;
- HMMP;
- HMIPv6.

Mobility in IP will not be handled further here, but is described comprehensively in another paper in this issue of *Telektronikk* [36].

Service Network Architecture

In the future standardised interfaces towards the service creation environment and the application servers will be implemented. Already in UMTS we see a migration in that direction. The goal is to achieve flexible introduction of new services. In UMTS three alternative technologies are introduced:

- OSA – Open Service Architecture which defines an open API for introducing and modifying services;

- JAIN Parlay – The Parlay group identifies a set of technology independent interfaces. Java APIs for Integrated Networks (JAIN) Community standardise Java for different network technologies;

- Internet proposal – Different proposals for programming SIP services are suggested in IETF. These include the use of Call Processing Language (CPL), Common Gateway Interface (CGI) for SIP, SIP Servlets APIs, JAVA enhanced SIP (JES).

Concluding Remarks

Some trends and topics of what can be called Fourth Generation Mobile (4G) have been elaborated in this paper. It can almost be said that any technology or system is old-fashioned when launched, so even for UMTS. While mobile communications up till now has evolved in steps, or generations, we now see that *integration*, *heterogeneity* and *diversity* are the key words for further evolution.

References

1 Committee on Evolution of Untethered Communications, Computer Science and Telecommunications Board, Commission on Physical Sciences, Mathematics, and Applications, National Research Council. *The Evolution of Untethered Communications*. National Academy Press, Washington D.C., USA. (2000, December 11) [online] – URL: http://www.nap.edu/readingroom/books/evolution/

2 Zysman, G I et al. Technology Evolution for Mobile and Personal Communications. In: *Bell Labs Technical Journal*, 5 (1), 107–129, 2000.

3 Abramson, N. The ALOHA System – Another alternative for computer communications. In: *1970 Fall Joint Comput. Conf., AFIPS Conference Proceedings*, vol 37, Montvale, NJ, 281–285, 1970.

4 *3GPP – Third Generation Partnership Project*. (2001, January 18) [online] – URL: http://www.3gpp.org/

5 Fernandes, L. *R 2067 Mobile Broadband System Final Project Report*. (CEC deliverable R2067/CPRM/1.1/MR/L/015.b1). Brussels, March, 1996.

6 Rokitansky, C-H (ed). *R 2067 Mobile Broadband System Updated version of SDD*. (CEC deliverable: R2067/UA/WP215/DS/P/68.b1). Brussels, December 1995.

7 Fernandes, J et al. The SAMBA Trial Platform: Initial Results. In: *Proceedings from 3rd ACTS Mobile Communications Summit*, Rhodes, Greece, June 8–11, 1998. Trochos Technical Editions, Athens, 1998, 461–466.

8 ITU. *IMT-2000 Press Releases and Related News. Slide Presentations and Video Clips. Mobile Market Growth: Cellular Market*. (2000, September 26) [online] – URL: http://www.itu.int/imt/4_news_arch/Slide_video/Slides/mobmkt.html

9 Wisely, D, Mohr, W, Urban, J. Broadband Radio Access for IP-Based Networks (BRAIN) – A key enabler for mobile internet access. In: *Proceedings from 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications – PIMRC 2000*, London, Sept 18–21, 2000. IEEE, 2000, 431–436.

10 Pereira, J M. A Personal Perspective of Fourth Generation. *Telektronikk*, 97 (1), 20–30, 2001 (this issue).

11 Wisely, D, Mohr, W, Urban, J. The BRAIN developments for a broadband mobile access network beyond third generation. *Telektronikk*, 97 (1), 58–64, 2001 (this issue).

12 Karlsen, M E, Sollund, A. Youngster: Focusing on future users in a mobile world. *Telektronikk*, 97 (1), 99–105, 2001 (this issue).

13 Pascotto, R, Neureiter, G. Giving mobile users access to context-aware services – Opportunities and challenges. *Telektronikk*, 97 (1), 127–133, 2001 (this issue).

14 *IST – Information Society Technologies*. (2000, December 14) [online] – URL: http://www.cordis.lu/ist/home.html

15 *The DRiVE project Homepage*. 2000, December 06. [online] – URL: http://www.comnets.rwth-aachen.de/~drive/

16 Do van Thanh. Towards user-centric communications with the virtual terminal. *Telektronikk*, 97 (1), 106–126, 2001 (this issue).

17 Noll, J, Burrachini, E. Software radio – a key technology for adaptive access. *Telektronikk*, 97 (1), 31–39, 2001 (this issue).

18 *P921 – D1 – SDR and HFR for low cost radio independent access*. (2000, December 14) [online] – URL: http://www.eurescom.de/~public-web-deliverables/P900-series/P921/D1/

19 *The TRUST Homepage*. (2000, December 14) [online] – URL: http://www.ist-trust.org/TRUST_Flash.html.

20 CEPT. *ERC Decision of 1 June 1999 on the designation of the harmonised frequency band 40.5 to 43.5 GHz for the introduction of Multimedia Wireless Systems (MWS), including Multipoint Video Distribution System (MVDS)*. 1999. (ERC decision ERC/DEC/(99)15.)

21 CEPT. *Provisional recommended use of the frequency range 54.25–66 GHz by terrestrial fixed and mobile systems*. 1991. (CEPT recommendation T/R 22-03.)

22 Prögler, M (ed). *AC204 System for Advanced Mobile Broadband Applications - Specification of the Air Interface*. ACTS programme, 1997. (CEC deliverable A0204/DB/F3F/DS/P004/a1.)

23 *MEDIAN – Wireless Broadband CPN/LAN for Professional and Residential Multimedia Applications*. (2000, December 14) [online] – URL: http://www.imst.de/median/median.html

24 *MMAC – Multimedia Mobile Access Communication Systems*. (2000, December 14) [online] – URL: http://www.arib.or.jp/mmac/e/index.htm

25 *EMBRACE – Efficient Millimetre Broadband Radio Access for Convergence and Evolution*. (2000, December 14) [online] – URL: http://www.telenor.no/fou/prosjekter/embrace/

26 Thoen, S et al. Predictive Adaptive Loading for HIPERLAN II. In: *Proceedings of the 51st Vehicular Technology Conference VTC 2000-Fall*. Boston, IEEE, 2000, 2166–2172.

27 Hole, K J, Øien, G. Adaptive coding and modulation: A key to bandwidth-efficient multimedia communications in future wire-

less systems. *Telektronikk*, 97 (1), 49–57, 2001 (this issue).

28 Pettersen, M, Lehne, P H. Smart antennas – The answer to the demand for higher spectrum efficiency in personal communications. *Telektronikk*, 94 (2), 54–64, 1998.

29 Lehne, P H, Pettersen, M. An overview of smart antenna technology for mobile communications systems. *IEEE Communications Surveys*. *Fourth Quarter 1999*, 2 (4), [online] – URL: http://www.comsoc.org/pubs/surveys/

30 Lehne, P H et al. *Smart antenna performance based on directional radio channel measurements*. Kjeller, Telenor Research and Development, 1998. (R&D report R 42/98.)

31 *BLAST: Bell Labs Layered Space-Time*. (2000, December 14) [online] – URL: http://www1.bell-labs.com/project/blast/.

32 Bach Andersen, J. Multiple antennas – the promise of high spectral efficiency. *Telektronikk*, 97 (1), 40–48, 2001 (this issue).

33 Chuang, J, Sollenberger, N. Beyond 3G: Wideband wireless data access based on OFDM and dynamic packet assignment. *IEEE Communications Magazine*, 38 (7), 78–87, 2000.

34 Haslestad, T et al. *An evaluation of ODMA*. Kjeller, Telenor Research and Development, 2000. (R&D Report 42/2000.)

35 *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects*. *Combined GSM and Mobile IP Mobility Handling in UMTS IP CN*. Sophia Antipolis, France, May 2000. (3GPP Technical Report 3G TR 23.923 version 3.0 0.)

36 Paint, F, Egeland, G. Seamless Mobility in IP networks. *Telektronikk*, 97 (1), 83–91, 2001 (this issue).

Abbreviations and acronyms

16-OQAM	16-state Offset Quadrature Amplitude Modulation
3GPP	Third Generation Partnership Project
ACTS	Advanced Communications Technologies & Services
ADSL	Asymmetric Digital Subscriber Line
AMPS	Advanced Mobile Phone System
ARIB	Association of Radio Industries and Businesses
ATDMA	Advanced TDMA Mobile Access (RACE project R 2084)
ATM	Asynchronous Transfer Mode
BLAST	Bell Labs Layered Space-Time Architecture
BRAIN	Broadband Radio Access for IP-Based Networks (IST project IST-1999-10050)
BRAN	Broadband Radio Access Network
CDMA	Code Division Multiple Access
CEPT	Conference Européenne de Post et des Télécommunications
CGI	Common Gateway Interface
CN	Correspondent Node
CODIT	Code Division Testbed (RACE project R 2020)
CPL	Call Processing Language
CWTS	China Wireless Telecommunication Standards Group
DAB	Digital Audio Broadcasting
D-AMPS	Digital AMPS
DRiVE	Dynamic Radio for IP-Services in Vehicular Environments (IST project IST-1999-12515)
DTV	Digital TV
DVB-T	Digital Video Broadcasting – Terrestrial
EMBRACE	Efficient millimetre broadband radio access for convergence and evolution (IST project 1999-11571)
ETSI	European Telecommunications Standards Institute
FMC	Fixed and Mobile Convergence

FPLMTS	Future Public Land Mobile Telecommunications System
FRAMES	Future Radio Wideband Multiple Access Systems (ACTS project AC090)
GGSN	Gateway GPRS Service/Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HAWAII	Handoff-Aware Wireless Access Infrastructure
HFR	Hybrid Fibre Radio
HIPERLAN	High Performance Radio Local Area Network
IETF	Internet Engineering Task Force
IGSN	Internet Gateway Support Node
IMT-2000	International Mobile Telecommunications 2000
IP	Internet Protocol
ISDN	Integrated Services Digital Networks
IST	Information Society Technologies
JAIN	Java APIs for Integrated Networks
JES	JAVA Enhanced SIP
LMDS	Local Multipoint Distribution System
MAC	Medium Access Control
MBS	Mobile Broadband Systems (RACE project R 2067)
MCP	Multimedia Car Platform (IST project IST-1999-13046)
MEDIAN	Wireless Broadband CPN/LAN for Professional and Residential Multimedia Applications (ACTS project AC006)
MIMO	Multiple Input Multiple Output
MMAC	Multimedia Mobile Access Communication
MONET	MObile NETworks (RACE project R 2066)
MWS	Multimedia Wireless Systems
NMT	Nordic Mobile Telephone
ODMA	Opportunity Driven Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OQPSK	Offset Quadrature Phase Shift Keying
OSA	Open Service Architecture
PDA	Personal Digital Assistant
PSTN	Public Switched Telephone Network
RACE	Research in Advanced Communications for Europe
SAMBA	System for Advanved Mobile Broadband Applications (ACTS project AC204)
SGSN	Serving GPRS Support Node
SIP	Session Initiation Protocol
SMS	Short Message Service
T1	Standards Committee T1 Telecommunications
TACS	Total Access Communications System
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TeleMIP	Telecommunications-enhanced MIP
TRUST	Transparent Re-configurabel UbiquituoS Terminal (IST project IST-1999-12070)
TTA	Telecommunications Technology Association
TTC	Telecommunication Technology Committee
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
VDSL	Very high data rate Digital Subscriber Line
WAND	Wireless ATM Network Demonstrator (ACTS project AC085)
WAP	Wireless Application Protocol
WCDMA	Wideband Code Division Multiple Access
WINEGLASS	Wireless IP Network as a Generic Platform for Location Aware Service Support (IST project IST-1999-10669)
WLAN	Wireless Local Area Network

A *Personal* Perspective of Fourth Generation

JORGE M. PEREIRA*

Jorge Pereira (40) obtained his Engineering and MSc degrees in Electrical and Computer Engineering from Instituto Superior Técnico (IST), Lisboa in 1983 and 1987, respectively, and his PhD in Electrical Engineering-Systems from the University of Southern California, Los Angeles in 1993. He is now on leave of absence from his position as Assistant Professor at IST, and has since 1996 been with the European Commission, DG Information Society, as Scientific Officer in the area of Mobile and Personal Communications. He is involved in Third Generation / UMTS and Broadband Wireless systems, Spectrum and Regulatory issues, Enabling Technologies, Reconfigurable Radio Systems and Networks, Terrestrial Positioning Technologies and Location-based Mobile Value Added Services, and Fourth Generation systems.

Jorge.Pereira@cec.eu.int

At a time when Second Generation (2G) enjoys a tremendous success, still growing exponentially in Europe, and Third Generation (3G, also known as IMT-2000) is nearing deployment, with some operators willing to pay billions for the 3G licenses, discussing Fourth Generation (4G) might seem premature, even far-fetched ...

In this contribution, we propose to discuss the distinct roles envisioned for 4G and evolving 3G systems by emphasising the *personal* perspective of 4G, thus dispelling fears of 4G constituting a "distraction" at a time when all efforts need to be focused on 3G, with the further risk of reducing the "value" of 3G licences.

In spite of timid steps towards accommodating some type of private (in fact, corporate) usage, past and current mobile generations – from First (1G) to Third (3G) – focus exclusively on public use. However, as 3G networks are (to be) deployed on top of existing Second Generation (2G) networks, with IP becoming ever more pervasive, and with the advent of reconfigurable radio [1], the opportunity is there, looking forward towards Fourth Generation (4G), to really start focusing on private, unlicensed use at par with the envisioned use of 4G systems as an extension of current and planned public systems.

> Fourth Generation will finally put the User in control, as they will be able to decide in every occasion (application), and for every environment (mobility, coverage), the right system, and even the right terminal. It will offer personalised service irrespective of the underlying network, and make the best use of the scarce spectrum by directing each data stream in a communication session to the most appropriate (i.e., efficient) network.

Indeed, our perspective of 4G is not one of *only* higher data rates – that was from the start the perspective for 3G-Phase 2, what some called IMT-2010 – still with a public service focus. What is at stake is a different concept of service, more than yet another technological revolution, and this constitutes a bigger challenge [2]: the focus is now squarely on the User, not on operators/service providers.

That said, some aspects of 4G, namely the integration of heterogeneous networks, will provide operators, already in the short term, with essential tools to manage their increasingly complex networks and provide seamless, efficient access to information – i.e. to remain competitive [3].

Adding to this, there is a clear need to start thinking *now* about Fourth Generation: the lapse between the initial studies, the assignment of spectrum, and the commercial deployment is always large, well over 10 years [4].

> In a recent speech, Mr. Tormod Hermansen, President and CEO of Telenor, called for a "Permanent Revolution" as the sole means for a telecommunications operator to survive in an ever-changing, increasingly competitive world.
>
> The consideration of a Fourth Generation will bring about a paradigm shift. But "Change is not Necessarily Disruption". In fact, on the way towards Fourth Generation, solutions, tools and products will be developed of immediate relevance, offering competitive advantage to first movers.
>
> "Innovation is the Key" to survival!

---

* *The opinions expressed herein are those of the author and do not engage the European Commission. The author is on leave of absence from Instituto Superior Técnico, and Instituto de Telecomunicações, Lisbon Technical University, Lisboa, Portugal.*

I Introduction

After spending huge amounts of money for Third Generation (3G, also known as International Mobile Telecommunications-2000, IMT-2000) licenses in Europe, and with a similar perspective for North America, operators (and manufacturers alike) are understandably not eager to consider yet another generation of mobile communication systems.

However, there is a clear need to start thinking now about Fourth Generation (4G): the lapse between the initial studies, the assignment of spectrum, and the commercial deployment is large, well over 10 years [4]. Moreover, looking ahead, one ends up addressing issues relevant to current generations, and allowing for easier, more fruitful integration with existing systems.

Three concurrent trends define the context for 4G to play a key role of integrator and gap filler [3]: the continued exponential growth of mobile/wireless communications, both cellular and W-LAN, together with a similar growth of the Internet; the deployment of 3G networks on top of existing 2G networks, requiring substantial efforts of integration; and the growing requirement, from corporations, SME/SoHo and even individuals, for integrated private/public telecommunication solutions.

The exponential growth of mobile/wireless communications is briefly analysed in Section II. Section III discusses the evolution from 1G to 2G to 3G, the integration of 3G networks with existing 2G networks, and in that context what 4G is not and what to expect from 4G, particularly the issues of integration of public and private mobile/wireless systems.

In Section IV, we briefly review some of the current 4G concepts, and in Section V we propose an all-encompassing approach to 4G, which besides integrating current and planned systems also fills a clearly identified gap in the areas of private, indoor wireless systems, personal area networks (PANs), body-LANs, embedded systems and machine-to-machine (M2M) communications. The advantages public operators can extract from the integration process are emphasised.

Section VI presents ongoing and planned activities in the area of 4G in Europe, focusing on the research performed in the scope of the European Union's Information Society Technologies (IST) Programme.

II The Success of Wireless/ Mobile Communications

This section summarises the analysis in [2] of the exponential growth of cellular, W-LANs and of the Internet.

A Cellular

The tremendous success of cellular is undeniable; projections are that world-wide cellular subscribers will exceed one billion by 2003 and two billion by 2005.

Such growth will inevitably lead, sooner in some countries than in others, first to saturation of 2G spectrum (different bands in different regions of the world) and eventually to saturation of the IMT-2000 terrestrial spectrum assigned at World Administrations Radiocommunication Conference (WARC) '92 (1885–1980, 2010–2025 and 2110– 2170 MHz – the so-called IMT-2000 core bands), particularly taking into account the expected growth of mobile multimedia.

This has led, at the recent WRC 2000, to the allocation of additional spectrum for IMT-2000: the additional bands identified for the terrestrial component of IMT-2000 are 806–960, 1710–1885 and 2500–2690 MHz (IMT-2000 extension bands), to be assigned by the national administrations at their discretion and based upon market demand. The USA, having missed the opportunity to use the IMT-2000 core band due to its assignment to PCS, is now actively looking into the IMT-2000 extension bands for 3G, with auctions looming in the near future.

B Internet Explosion and IP Pervasiveness

The number of Internet users is also growing exponentially. According to NUA Internet Surveys[1)], at the end of June 2000 there were approximately 333 million users world-wide. Some projections point to around 1 billion users by 2005.

Seen in the scope of the increasingly pervasive use of IP in the backbone and its migration to the *wireless* access network, this ensures IP will become the *lingua franca* of mobile communications, as it enables the mobile data explosion. In fact, by the end of 2004 it is expected that there will be more handhelds connected to the Internet than PCs, with well above half of the mobile subscribers being mobile data users.

C W-LANs

Simultaneously, the W-LAN market is said to be exploding, with reported yearly growth figures

1) *www.nua.ie*

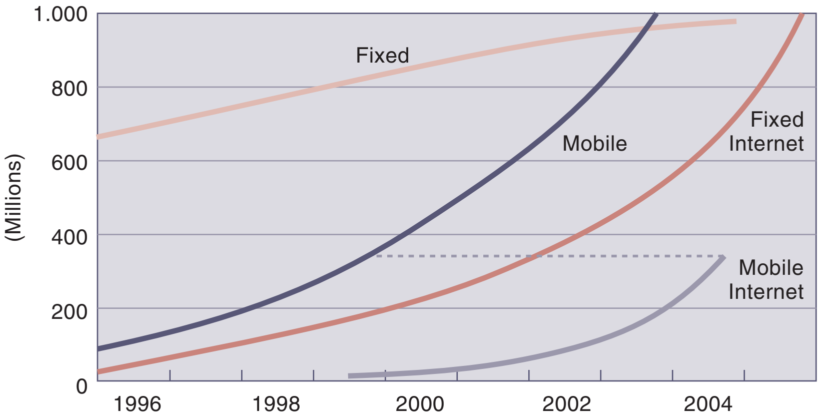


Figure 1 Exponential growth of Mobile and Internet

of 300 %, as finally the higher speeds, interoperability and lower prices that previously had been missing from this once niche market, bring it into the mainstream. As vendors embrace standardisation and interoperability, heightened competition will spur lower costs, new applications and flexible packaging.

Standardisation is in fact driving growth. In October 1996 the HIPERLAN Type 1 (20 Mbit/s) Functional Specification was ratified by ETSI as EN 300 652. The IEEE 802.11b standard (11 Mbit/s) followed in September 1999, and is currently being extended to reach 20 Mbit/s.

Two higher data rate standards are nearing completion: HIPERLAN Type 2 (54 Mbit/s) core technical specifications have already been published in March 2000, while IEEE 802.11a (25–50 Mbit/s) has not been approved yet.

IMT-2000 Core and Extension Bands

In the US, evolving to 2G meant to replace a single analogue system, AMPS, with a variety of incompatible systems at 800 MHz and 1900 MHz (PCS band), namely N-AMPS, TDMA IS-136, CDMA IS-95, GSM 1900, Omnipoint IS-661 and PACS.

By assigning to PCS part of the bands designated at WARC '92 for IMT-2000, the US is now in a situation where the IMT-2000 *core band* cannot be used for 3G.

As for the IMT-2000 *extension bands* decided at WRC 2000, the situation is not better. The whole 806–941 MHz band is used up or to be filled with AMPS, ISM, SMR, narrowband PCS and paging; the Government has exclusive use of the 1710-1850 MHz band; and the 2500–2690 MHz band is currently used for Satellite Broadcasting.

Facing this problem, the FCC plans for spectrum in the 700 MHz band (746–764 MHz and 776–794 MHz bands), currently used by UHF television channels 60-69, to be made available for "next generation high speed mobile services", although broadcasters have been given until 2006 to relocate.

III Mobile Generations: Not as Simple as 1,2,3 ...

We cannot approach the subject of 4G without taking into account the evolution from 1G to 2G to 3G, with its drivers and the requirement for coexistence of successive generations. Only then can we propose a meaningful approach to 4G that goes beyond a commercial *spiel*.

In the following, we will focus on what really changes and what stays the same from one generation to the next.

A From First to Second to Third

The evolution from 1G to 2G was a natural one, from analogue to digital, with all the advantages it brings. Moreover, the choice was made in Europe to go from a variety of incompatible analogue systems at 450 MHz and later at 900 MHz (NMT-450, NMT-900, TACS-900, RC2000, C-450) and a fragmented market, to a single digital system, GSM, to finally have *pan-European roaming* (today quasi-global) .

We are talking about a *new air interface*, but the *same service*, voice. And, as with 1G, 2G had an exclusive public service focus. However, the need for, and the advantages of, private networks using the same technology soon became evident (e.g. CERN's private GSM system [5]). In any case, efforts in that direction stopped short at the corporate/campus level. The integration of GSM with the corporate network has also been proposed, with micro-base stations capable of assuming the role of a PBX, and of being connected to the LAN thus allowing calls to be, when appropriate, directed to/made from the user's PC using voice over IP (VoIP, part of what is called today IP Telephony).

While the step from 1G to 2G was necessarily revolutionary from the technology point of view, the idea was from the start, at least in Europe, to have 2G as an *evolving platform*: thus GSM with its Phases 2 and 2^+, and the annual releases leading to the introduction of HSCSD (High Speed Circuit-Switched Data), GPRS (General Packet Radio Service) and EDGE (Enhanced Data rates for GSM Evolution).

The evolution from 2G to 3G (also known as IMT-2000) corresponds yet again to adopting a *new air interface* (UMTS and its W-CDMA and TD/CDMA modes), but most of all, to a change of focus from voice to *Mobile Multimedia*. Another first is the *simultaneous support of several QoS classes* in a single air interface.

Once more, 3G was conceived from the start in Europe as an *evolutionary platform*, with UMTS-Phase 2 planned to push data rates well

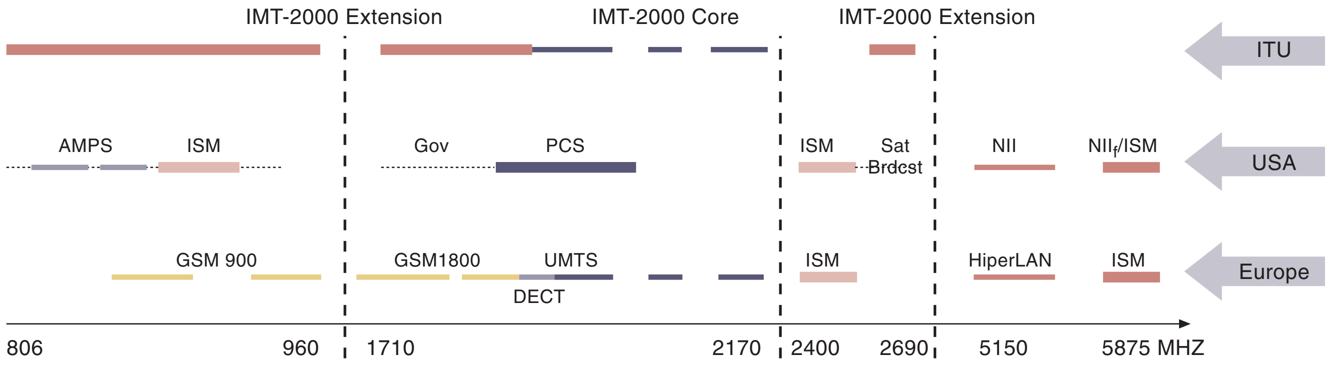


Figure 2 Spectrum allocation and usage in the 800 MHz to 5 GHz bands

above 2 Mbit/s into the 10 Mbit/s range. It was further conceived, as agreed upon by 3GPP, to be able to accommodate various (current, planned and future) air interfaces.

Even if the focus remains fully on *public service*, the above approach, going back to ETSI SMG's consideration of HIPERLAN Type 2 as an extension of UMTS, will at least "standardise" the interface with any such future system, increasingly of a personal nature, allowing full exploitation of VHE.

ITU's objective was to achieve *global roaming*, but that has become difficult with five different modes approved by ITU-R (CDMA Direct Sequence, CDMA Multi Carrier, CDMA TDD, TDMA Single Carrier, FDMA/TDMA), and the lack of common bands (see Figure 2).

B What about 4G?

We have just argued that 3G, at least as defined by 3GPP, has all the potential to evolve and to exploit other, enhanced air interfaces as extensions. So, in order for 4G to deserve that designation, it needs to constitute a clear step forward from 3G, and has to bring about a clear paradigm shift.

More than *new air interfaces* in some areas where gaps have been identified (personal area networks and body-LANs, low power sensors, networked appliances and self-configuring ad hoc networks), we contend that what will define 4G is the ability to *integrate all systems and offer access to all services, all the time and everywhere, irrespective of serving network*, allowing for the integrated provision of personalised, enhanced services over the most efficient/preferred networks, depending on the user profile, on the type of data stream under consideration, and on the traffic load in the available networks.

Furthermore, 4G will be designed to take into account multiple classes of terminals, adjusting content delivery to the terminal capabilities and the user profile.

Fourth Generation will put the User in control, allowing in every occasion (application) and environment (mobility, coverage) for the selection of the right system, and even of the right terminal. It will offer personalised service irrespective of the underlying network, and make best use of scarce spectrum by directing each data stream in a communication session through the most appropriate (i.e. efficient) network.

C Coexisting Generations

Mirroring closely what happens in human society, successive generations coexist in time, even if pursuing different agendas. The heavy investments made in 2G have not yet led to the demise of 1G, particularly in the US (but also in Italy, Germany and Spain), and likewise, the heavy investments foreseen for 3G (more than $200 billion in Europe alone, without considering the licence fees) will not lead to the quick disappearance of 2G systems.

In fact, in Europe, the result of recent 3G auctions and beauty contests, and some subsequent deals, was that until now all 2G incumbents obtained (access to) 3G licences. It is therefore in their best interest to make optimum use of their assets, both old and new, and this immediately suggests a much higher level of integration than has been possible, or even considered, until now.

The evolution towards 4G will crystallise the need to fully integrate all systems, as by then 2G will, from all projections, still be going strong. And, to this public service dimension one will have to add a private, unlicensed dimension, as

What 4G is NOT about ... [3]

- NOT just higher data rates – this would correspond to UMTS-Phase 2 (what some call IMT-2010), and has already been done by MBS and HiperLAN;
- NOT only public system extension, but taking also due account of private, unlicensed systems;
- NOT technology driven, whereby the label 4G is seen as just a way to sell a new air interface (differentiation).

Defining Generation based upon Data Rate

For argumentation sake we introduce a simplistic definition of generation, based on data rate alone, G_{DR}, and will check it against *public cellular* systems.

We begin the notion prevailing in certain quarters that a new generation implies around ten times higher data rates. This suggests the logarithmic definition below :

$$G_{DR} \triangleq \log_{10} [R(\text{kbit/s})]$$

Applying such a definition leads to Figure 3, where the horizontal axis only roughly represents time.

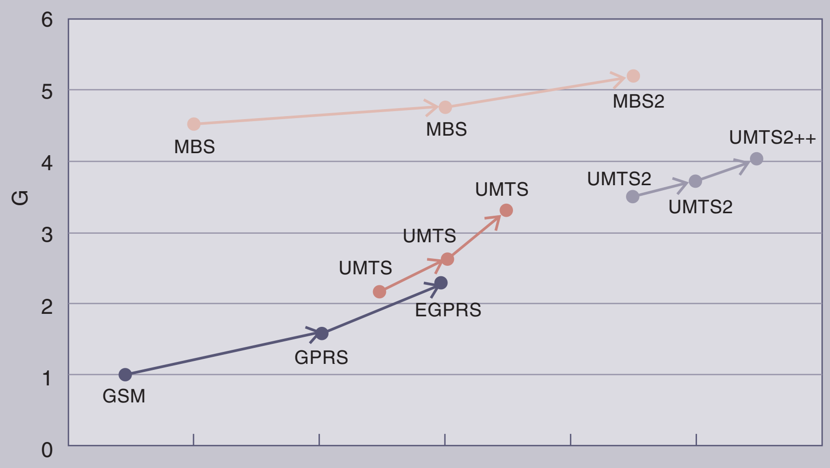


Figure 3 Simplistic Classification of Public Cellular Systems [3]

The first observation is that systems evolve, even when they are understood to be of a given generation (e.g. GSM being 2G). In fact, they start humbly enough before reaching its potential. In that sense, consecutive generations do compete very closely at the launch of the new generation – this makes more compelling the need for integration of successive generations.

The second observation is that generations do not necessarily follow each other in "proper" sequence: Mobile Broadband Systems (MBS), conceived from the start as a 3G system, was developed in parallel with UMTS and would deserve at least a 4 rating. MBS is expected, in Phase II, to reach 155 Mbit/s giving it a 5 rating.

Similarly, UMTS was from the start planned to be developed in two phases, without changing the premises of the system, with Phase II increasing data rate eventually up to 10 Mbit/s, bringing it close to a 4 rating. UMTS, through its planned evolution will reach, and through anticipated extensions even exceed by far, what one would expect from a 4-rated system.

It is then simple to conclude that such a simplistic definition as G_{DR} is not up to par: much more than higher data rates is needed to conjure a new generation.

W-LANs/Body-LANs and Home Networks explode into the mass market. 4G will then correspond to the implementation of the *Integrated Wireless World* [2] of Figure 4.

IV Perspectives of 4G

Let us now review some perspectives on 4G as the issue starts to be discussed.

After the agreement on IMT-2000 air interfaces, ITU-R working party 8F redefined its scope to also look into systems beyond IMT-2000, acting as a forum for user requirements and as a catalyst for translating those requirements into technical reality. Included in the work assigned to WP8F are issues such as spectrum requirements, higher data rate capabilities, and IP-based service needs. In that context, ITU-R organised in March 2000 a first workshop to discuss systems beyond IMT-2000.

At that workshop, and running the risk of oversimplifying, NTT DoCoMo effectively equated 4G with 2 Mbit/s basic rate and 20 Mbit/s best effort in the downlink [6], with a Fifth Generation already lurking, again only better in terms of data rate (see Figure 5).

In their 2010 vision [7], DoCoMo promises to deliver MAGIC: Mobile multimedia; Anytime, anywhere, anyone; Global Mobility support; Integrated wireless solution; and Customised personal service. However, the focus is (understandably) solely on *public systems*, as 4G is seen as an extension of 3G cellular service only (see Figure 6).

The main issues are clearly identified as Software Radio (a part of what we call Reconfigurable Radio Systems and Networks [1]), adaptive array antennas at the base station (although we contend that they are also to be deployed at the terminal) and the optical fibre backbone to support true broadband multimedia.

A different system perspective has always been proposed by the European Commission (EC), which introduced the concept of 4G in the IST Programme as early as 1998 [8]. Besides the need for accommodating the accelerated growth in the demand for broadband wireless connectivity, the focus is on ensuring seamless services provisioning across a multitude of wireless systems and networks, *from private to public, from indoor to wide area*, and providing for optimum delivery via the most appropriate (i.e. efficient) network available [2]. Particularly important is to cope with the expected growth in machine-to-machine (M2M) Internet-based communications: wireless low power sensors and actuators, Internet appliances, and a myriad of smart devices, capable of monitoring and interacting with the physical world.

Another initiative in this area comes from the German VDE, which produced a position paper covering also 4G [9]. Here the focus is not only on public systems, and namely MBS, building upon the ACTS (Advanced Communications Technologies and Services) Programme project SAMBA[2], but also on private W-LANs and W-CPN (customer premise networks) building upon the ACTS projects MEDIAN[3] and The Magic WAND[4]. Self-organising ad hoc networks are specifically identified as the "portable radio systems of the fourth generation", and embedded systems are expected to explode and with them the need for their networking: "Things that Think will Link".

The EC continued to actively promote discussion around 4G concepts. It organised, already in June 1999, a panel discussion on 4G in the scope of the ACTS Mobile Summit. The biggest motivation for 4G was identified as the need to improve the use of scarce resources through the optimisation of the combined use of communication options [10] given the growing demand discussed in Section II. The success, momentum and longevity of existing wireless solutions, together with the heavy investments, imply the need for an evolutionary, flexible, future proof solution, based upon open architecture concepts. Besides current systems, and their evolution, self-aware, self-organising as well as sensor-based and home area networks were identified.

The use of W-LAN for hot spot extension of 2G and 3G was also identified as promising, reflecting work in that direction in ETSI SMG.

4G's paradigm shift was identified as enhanced inter-operability through simultaneous support of several radio interfaces in a single terminal [11], allowing transparent use of the most suitable system. The biggest problem facing 4G is the fact that any new spectrum is likely to be very fragmented creating a challenge in terms of transceiver design [2]. Moreover, it will most likely vary in different regions of the world.

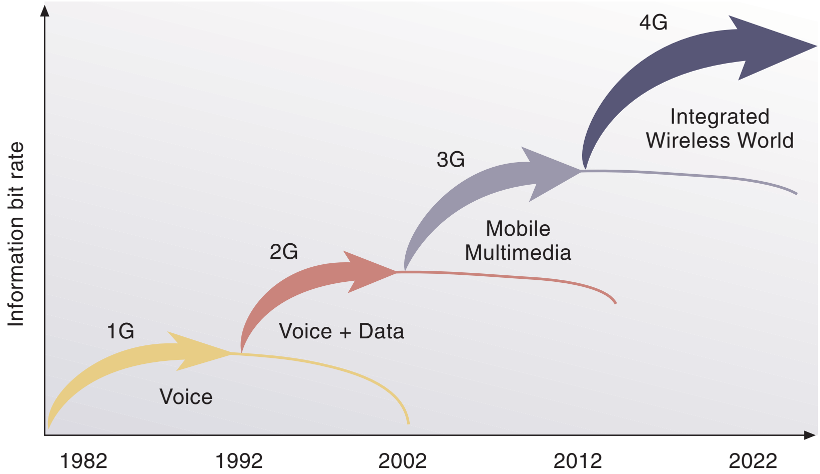


Figure 4 Coexistence of Successive Generations

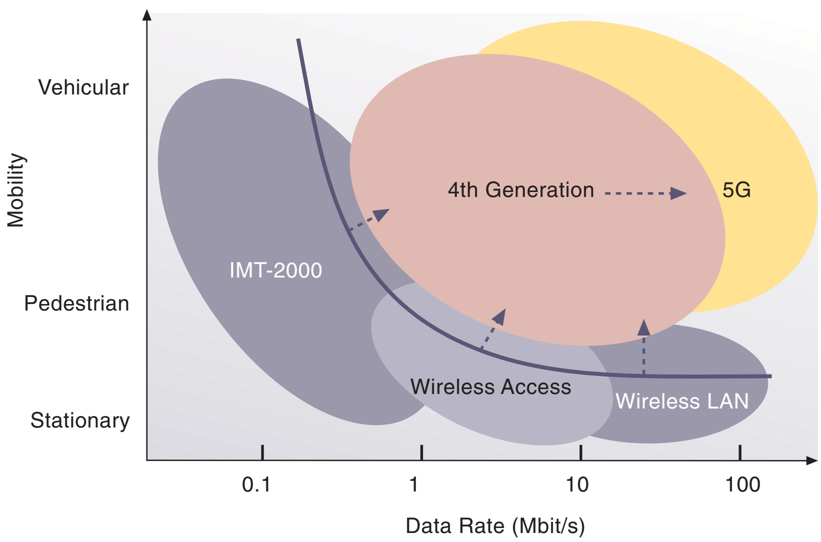


Figure 5 NTT DoCoMo's evolution perspective [6]

---

[2] *hostria.cet.pt/samba/index.htm*
[3] *www.imst.de/mobile/median/median.html*
[4] *www.tik.ee.ethz.ch/~wand/*

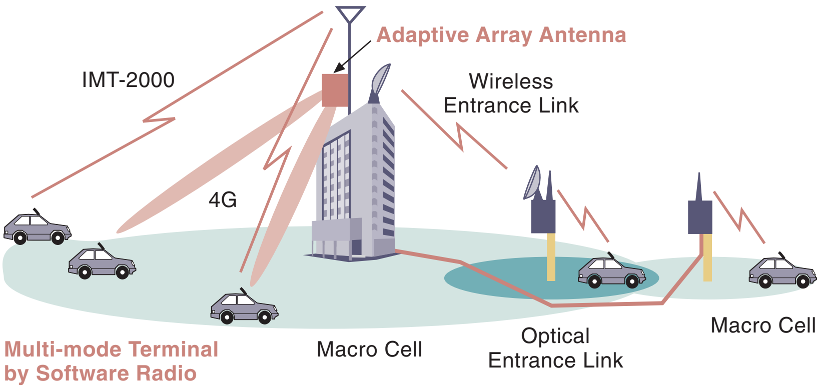


Figure 6 NTT DoCoMo's Public systems [7]

Later, in September 99, in the scope of the consultation towards the IST Workprogramme 2001, the Commission organised a one day workshop which covered, amongst other topics, 4G.

The fact that a cellular approach does not scale well under heavy multimedia traffic – the cellular breakdown [12] – and the expected increase of M2M communications, both low data rate and massive data exchange, suggest self-deployed, ad hoc networks as a way to break the infrastructure cost barrier, and of more dynamically sharing *unlicensed spectrum*. Furthermore, the necessarily partial, even spotty coverage, will require seamless public/private hand-over. Broadcasting is also targeted to fully exploit traffic asymmetry.

One of the issues that 4G is expected to resolve is that standards are not designed with interoperability in mind [13], and so the focus should be on integration with and interoperability across existing systems. Main motivations for 4G are cost savings, resource management for sustainability, automation, user-friendliness and *right-value services*.

More recently, in the scope of the concertation activities of the IST Mobile Domain[5)], the EC organised a mini-workshop on issues like Spontaneous Communications [14], ad hoc networking and Ultra Wideband Communications [15], as they are seen as some of the new technology elements of 4G. Further activities are planned that will keep the pressure on to flesh out the concept, and identify the research priorities in this area (see Section VI).

V An All-encompassing Perspective of 4G

After the short review above, it is time now to present our broader, all-encompassing perspective of 4G, focusing on the *User*, no longer on the operators or network providers [16], and paying due attention to private, unlicensed use.

A Based upon IP and Reconfigurable Radio

Challenging a fundamental premise of telecommunications as we know it, the user will no longer be "owned" by any operator [17]: the users, or their smart agents, will select at each instant, according to the user profile, the best system available capable of providing the performance required by the application. 4G will encompass *all systems*, from heterogeneous, hierarchical, competing but complementary, broadband networks (public and private, operator driven or ad-hoc) to personal area and ad hoc networks, and will inter-operate with 2G and 3G systems, as well as with digital (broadband) broadcasting systems. Finally, 4G will be *fully IP-based*, an expression of the wireless Internet.

This integrated perspective is illustrated in Figure 7. There we can see the broad range of systems that 4G has to integrate, from satellite broadband to High Altitude Platforms (HAP), to cellular 2G and 3G systems, to MBS, to Wireless Local Loop (WLL) and Fixed Wireless

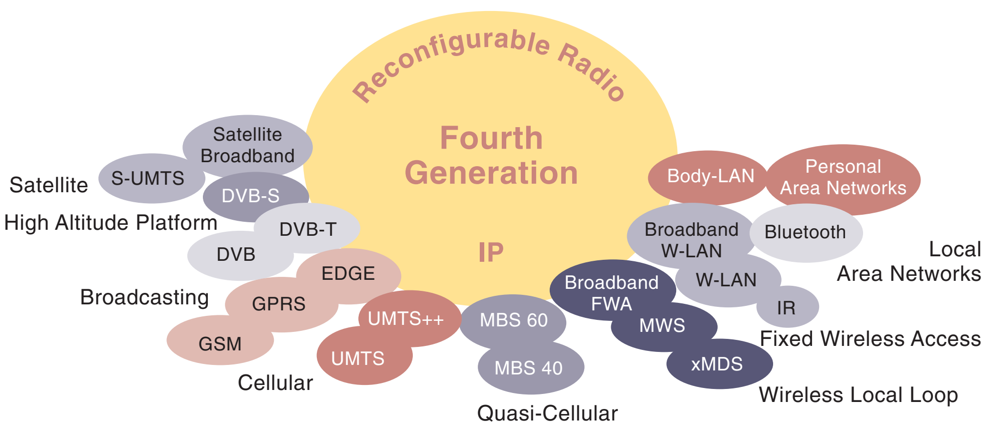


Figure 7 All encompassing, integrated perspective [2]

5) *www.cordis.lu/ist/ka4/mobile/2concerta.htm*

Access (FWA), to W-LAN, Personal Area Networks (PAN) and Body-LANs. *Reconfigurable Radio* is seen as the engine for such integration, and IP as the integrating mechanism, the *lingua franca* [2].

From the service point of view, 4G will implement adaptation to multiple standards and to delay sensitive and insensitive applications over channels of varying bandwidth, across multiple operators and service provider domains, with user controlled QoS levels and ensuring data privacy and information integrity, taking into account the user profile and the terminal characteristics.

What 4G is ... [3]

Fourth Generation is based upon

- Full-IP;
- Reconfigurable Radio; and
- Growing Role of Private, Unlicensed Systems,

and will build upon the integration of heterogeneous, hierarchical networks.

The leitmotiv of Fourth Generation is clearly the most efficient use of scarce Spectrum.

An open architecture which allows for differentiation of operators, service providers and application developers, thus promoting competition, is critical. Here, the contribution of Reconfigurable Radio is paramount [1].

B Efficient Use of Spectrum

Another critical element of 4G, one with immediate application, is *efficient use of spectrum*, allowing for optimal selection of delivery system according to the (different) data streams involved and the performance required by each of them. This will eventually lead to *full spectrum sharing*, at least in some bands.

Given the high prices paid for spectrum in recent auctions, and with 3G licenses covering the next 10–20 years, such lofty objectives might seem far-fetched, but many of the involved principles and associated technologies have immediate application.

C Growing Role of Private, Unlicensed Systems

We have already observed that the focus of current and planned telecommunication systems has been mainly on public service. However, the need for, and the advantages of, private networks using the same technology soon became evident (see for example CERN's private GSM system, and conversely the use of DECT for public service [5]).

Exploiting the full potential of private, mainly unlicensed broadband systems, fully integrated with public wireless (and wired) systems, it is our conviction that soon there will be a totally transparent public-private wireless broadband communication system that will extend the public network through an almost ubiquitous coverage consisting of multiple, mostly overlapping private systems.

Assuming the existence of such a distributed and inherently hierarchical system, it is easy to see that the user would then be able to have ubiquitous broadband access to multimedia content and applications in a most economical manner: provided transparent mobility management and the necessary security mechanisms are in place, the users would rely on the less congested and wider band private systems wherever and whenever possible, falling back to using the more crowded public systems in the remaining situations [18].

This scenario imposes a paradigm change in terms of the role of the public networks, as well as of the economics of telecommunications. As the use of private systems would "inherently" be free of charge (under reciprocity agreements, i.e. everyone would give access to everyone, and in turn be also given access[6)]), this would certainly reduce the *direct* revenue of the public wireless broadband systems.

If this perspective does not seem very interesting for the public operators, one has to consider that wireless broadband communications will allow access to a plethora of services and content, most of which will be paid for. The public operators will certainly extract a small commission out of the service/content providers, more than making up for the difference. In fact, it is in their interest to provide the best access possible, creating an incentive for the users to access those value-added services (VAS).

Provided the user can seamlessly transfer from one system to another, both amongst private systems and between public and private systems –

6) *The reciprocity principle would create the incentive for all private systems to expand to a accommodate any anticipated demands. Their reduced range will in any case limit the load to be accommodated.*

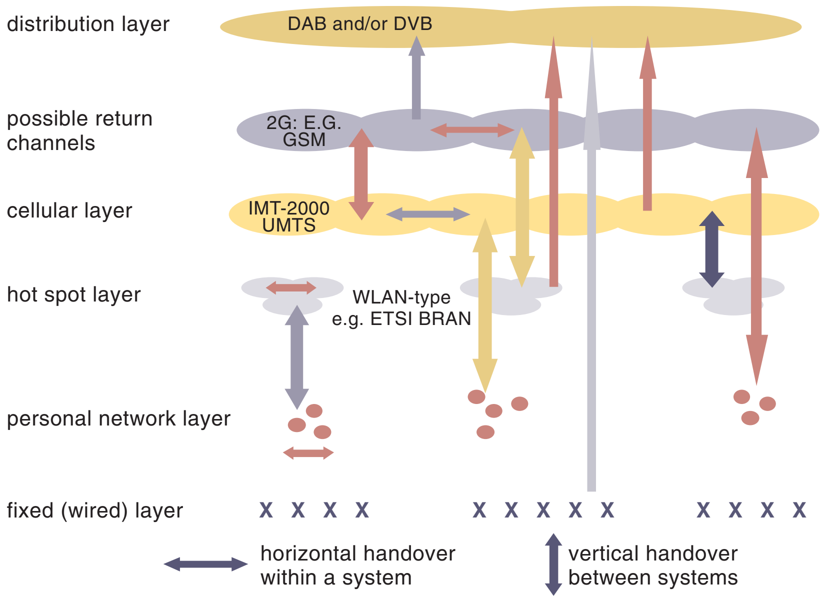


Figure 8 Hierarchical, heterogeneous integrated networks [Source: BRAIN Project]

through a combination of Reconfigurable Radio with network integration, at least at the level of IP delivery – we anticipate private, unlicensed systems will enjoy a tremendous success.

The added advantage of this approach is that it is rather "localised", and thus inherently modular, allowing for certain aspects of 4G to start being experimented with right now.

VI Ongoing and Planned Activities in Europe

Fourth generation was identified early in the IST Programme as an important area of research, initially more at a conceptual level than in terms of actual technology development.

A 4G Focus

Reflecting the perceived need to start exploring 4G concepts, an action line[7] was included in Workprogramme 2000 to prepare the ground for the likely technological and service evolution from current cellular and wireless systems and networks.

The focus was put on investigating and developing advanced and innovative concepts and technologies for self-aware, self-organising ad hoc wireless networks; on developing innovative air interface schemes allowing for scalable wireless connectivity; and on assessing potential spectrum requirements and co-existence issues, including the study of strategies and the development of appropriate tools allowing a distributed flexible management of the spectrum resources.

Further work will focus on inter-working with 2G and 3G systems and namely the extension of cellular networks into the indoors campus/office/home environment; on private systems (SOHO, home) and their integration and inter-working with public systems ensuring seamless service provision; on broadband W-LANs, PANs and Body-LANs; on roaming and VHE; on service to low-power sensors, particularly in a home-environment, in presence of (heavy) multimedia traffic; and on multi-hop/relaying systems for macro-cellular-type coverage, including issues of fall-back coverage, smart power control versus relaying availability and degree of service as a function of user density.

Of particular relevance is roaming in presence of heterogeneous systems (cellular/W-LAN/PAN/body-LAN – see Figure 8); the use of unlicensed spectrum in presence of guaranteed QoS requirement; and full spectrum sharing issues including "netiquette" and multi-streaming.

From the technological point of view, the need is perceived to develop a broad range of ultra-low power systems (systems on a chip) capable of transmitting a wide range of data rates (1 bit/s – 10 Mbit/s) over a wide range of average transmission output power levels (0.1 – 100 mW); to progress advanced microelectronics packaging; to develop high performance, low power, low cost RF devices; to explore spatial, time-diversity and multiple access schemes – all this for systems operating up to 300 GHz.

Additionally, new approaches are needed to *distributed and decentralised control* of network functionalities; to *dynamic resource allocation* to cope with varying traffic load, channel conditions and service environments; and to *network protocols* able to adapt dynamically to changing channel conditions, accommodating both low and high data rate users.

B Relevant IST Projects

As a result of the first Call for Proposals of the IST Programme, a few projects were launched which are starting to explore some of the issues relevant to 4G. Project BRAIN[8] focuses on a broadband extension of cellular systems such as GSM/GPRS/EDGE and UMTS, offering up to 20 Mbit/s for hot spot coverage. Its all-IP access network enables all real and non-real time ser-

[7] *www.cordis.lu/ist/ka4/mobile/areas.htm*
[8] *www.ist-brain.org*
[9] *www.comnets.rwth-aachen.de/~drive/*

vices in both public and private (home and corporate), licensed and unlicensed environments. The IP-based core network connects all radio access networks, allowing for vertical hand-over between complementary systems.

Project DRiVE[9)] aims at spectrum-efficient high-quality wireless IP provision in a heterogeneous multi-mode/multi-band/multi-system environment to deliver multimedia services, particularly to vehicles, ensuring ubiquitous access to information and support of "edutainment". To achieve this, DRiVE will build upon the complementarity of various communication and interactive broadcasting systems for the delivery of IP-based multimedia services.

DRiVE adds to the BRAIN approach a new Traffic Control (TC) layer enabling the selection of the most appropriate/efficient Radio Access Network (RAN) and eventually Dynamic Spectrum Allocation (DSA) – see Figure 9.

Finally, project WSI[10)] focuses on the initial concepts for future wireless systems beyond 3G. A think tank has been organised to promote discussion and ventilate ideas, taking in the input of the IST programme participants and industry at large, and a "Book of Visions" is currently available.

A MultiSphere concept centred on the User is proposed as a reference, spanning from PAN to Immediate Environment to Instant Partners to multiple Radio Access schemes to global Interconnectivity, through which one has access to the CyberWorld at large (see Figure 10).

The WSI partners (Alcatel, Ericsson, Nokia and Siemens) have just launched a new initiative, the Wireless World Research Forum (WWRF), open to everybody interested in contributing, where emerging concepts of 4G will be actively discussed.

At this point in time, the only thing certain is that a lot of thought needs to be put into fleshing out the concepts and gathering support from industry. And, most important of all, in bringing in expertise in the area of personal systems.

VII Conclusions

Fourth Generation is not a spoiler at a time when all the attention focuses on the deployment of 3G, but a challenge that will provide, already in the short term, some answers to the requirement of integration of heterogeneous, hierarchical networks for the purpose of optimally exploiting all available assets.

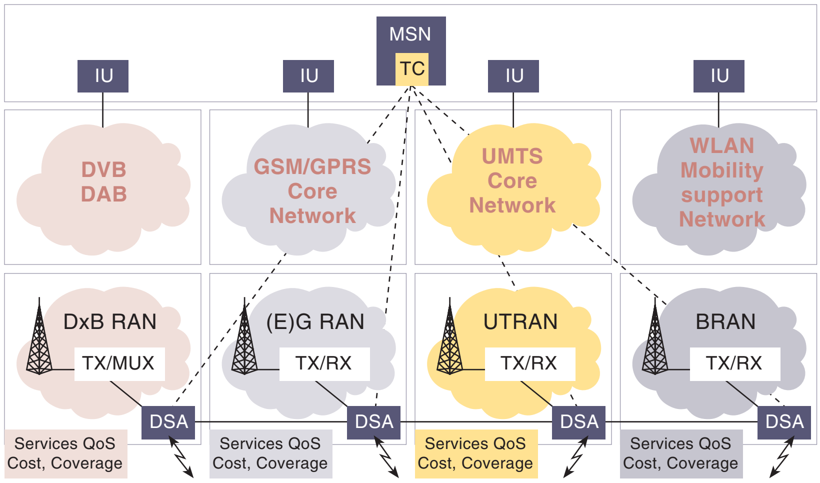


Figure 9 Architecture for Traffic Control and Dynamic Spectrum Allocation [Adapted from the DRiVE Project]

The result of such efforts of integration will be essential tools to help operators manage their increasingly complex networks and provide seamless, efficient access to information – i.e. to remain competitive.

Another strong motivation to start thinking now about 4G is the long time, well above 10 years, from the initial conceptual studies to the identification of spectrum requirements to the assignment of bands at WRC to commercial deployment, not forgetting where appropriate the need for regulation [2].

Finally, emphasising the role of private, mostly unlicensed usage systems will create new market opportunities, inclusive for public operators.

VIII Acknowledgements

Thanks are due to my colleagues at the European Commission, especially Dr. João da Silva, and many around the world that helped flesh out these ideas, through long discussions and despite

Figure 10 MultiSphere Concept [Source: WSI "Book of Visions"]

10) *www.ist-wsi.org*

A Co-Ordinated Approach to 4G

As with 3G, a co-ordinated approach [3] is necessary to ensure the success of 4G.

Combining R&D in the scope of the IST programme with Standardisation in the broadest sense, in ETSI, 3GPP, IEEE and IETF, with studies on Spectrum requirements by ERC focusing on mobile multimedia and taking into account traffic asymmetry, and a minimum of Regulation focusing mostly on issues of unlicensed use, access to/from public networks and privacy/security, will prepare the ground for a smooth introduction of 4G on top of existing systems, maximising the gains for the users (secure, personalised, user friendly, ubiquitous, low cost systems) and for the industry at large (open, modular architectures; spectrum efficient, self-planning/self-organising, fully integrated systems).

some initial scepticism. Particular thanks for the feedback from the many presentations on the topic, and for the comments received from many concerned operators, which led to the refinement of the positioning of 4G vis-à-vis existing and planned systems.

IX References

1 Pereira, J M. Redefining Software (Defined) Radio: Reconfigurable Radio Systems and Networks. In: *Special Issue on Software Defined Radio and Its Technologies, Transactions of IEICE*. June 2000.

2 Pereira, J M. Fourth Generation: Now, it is Personal! In: *Proceedings of PIMRC 2000*. London, September 2000.

3 Pereira, J M. *Balancing Public and Private in Fourth Generation*. Submitted to MobiCom 2001. Rome.

4 Verrue, R, Pereira, J M. Ensuring the Success of Third Generation. In: *Wireless Multimedia Network Technologies*. Ganesh, R, Pahlavan, K, Zovnar, Z (eds.). Kluwer Academic Press, 2000.

5 Pereira, J M. Efficient Management of Radio Resources – A Future Look. *European Microwave/Wireless '98*. Amsterdam, September 1998

6 Hata, M, Nakajima, N. Fourth Generation Mobile Communication Systems. *NTT DoCoMo, ITU-R WP8F Workshop*. Geneva, March 2000.

7 Murota, K. Mobile Communications Trends in Japan and NTT DoCoMo's Activities towards 21st Century. *NTT DoCoMo, ACTS Mobile Summit 99*. Sorrento, June 1999.

8 da Silva, J. Beyond IMT-2000. *ITU-R WP8F Workshop*. Geneva, March 2000.

9 *Everywhere and at any time: Wireless Multimedia Communications*. VDE-IGT Positionspapier Mobilkommunication 2005, Fall 1999.

10 Williams, F. Fourth Generation Mobile. *ACTS Mobile Summit 99*. Sorrento, June 1999.

11 Humo, H. Fourth Generation Mobile? *ACTS Mobile Summit 99*. Sorrento, June 1999.

12 Zander, J. Key Issues in 4th Generation Wireless Systems & Infrastructures. *Consultation Workshop*. European Commission, Brussels, May 2000.

13 Palomäki, M. Evolution of wireless systems, towards Fourth Generation. *Consultation Workshop*. European Commission, Brussels, May 2000.

14 Banâtre, M, Weis, F. Spontaneous Information Systems – A New Paradigm for Mobile Information Systems. *Consultation Workshop*. European Commission, Brussels, May 2000.

15 Scholtz, R (ed.). *Proceedings of the Ultra Wideband Radio Workshop*. Solvang, CA, May 1998.

16 Pereira, J M. What after Third Generation? *Symposium on Wireless Information Multimedia Communications*, Aalborg University. Aalborg, November 1999.

17 Pereira, J M. *Fourth Generation: Now, it is personal!* Presentation at the Communications Science Institute, University of Southern California, Los Angeles, April 2000.

18 Pereira, J M. Indoor Wireless Broadband Communications: R&D Perspectives in Europe. *Colloquium on Indoor Communications*. TU Delft, Delft, October 1997.

Software Radio – a Key Technology for Adaptive Access

JOSEF NOLL AND ENRICO BURACCHINI

Josef Noll (40) is Telenor R&D Fellow, responsible for Wireless Mobility. He holds a Dipl-Ing and PhD from the University of Bochum (D). His major working areas include broadband wireless access and mobility management in these systems. He is project leader of the EURESCOM Public Bluetooth Access and Local Provision of Next Generation Services projects. Prior to that he was head of the Mobility and Personal Communications Group and programme leader for Mobile Broadband Access at Telenor R&D. He has worked with SW radio issues since 1998, both at Telenor R&D and in the EURESCOM P921 project. He came to Telenor 1998 from the European Space Agency, where he was responsible for passive microwave remote sensing.
josef.noll@telenor.com

Enrico Buracchini (31) received his degree in Electronic Engineering from the University of Bologna in 1994, and was employed in the Mobile Services Divison of CSELT (R&D labs of Telecom Italia group) as a Research Engineer. His activity concerns the study of multiple access methods for mobile communications systems. He has been involved in several European research programs dedicated to 3G mobile systems, and from 1996 to 2000 he was part of the Italian delegation to ITU R TG8/1 standardisation group. Buracchini is now senior researcher and project manager in CSELT of the R&D project 'Radio Resource Management'.

enrico.buracchini@cselt.it

The paper provides a vision of an adaptive wireless access for next generation services. Depending on the requirements of the applications and the user situation, an optimum access method has to be selected. Software (SW) radio is the key enabling technology in this field. The functionality of a SW radio is described, and the challenges for future developments are identified. A major advantage of SW radio is the on-the-fly upgrade capabilities, both of terminals and of base stations. The paper concludes with a presentation of future trends; hybrid fibre distribution of radio signals and mechanisms for control of the various access networks.

1 Introduction

Already today we see a variety of different access networks: Internet via phone cable, via the Cable TV, via Satellite or even via the power line have become a reality. In addition to these wired access methods wireless access is more common in many scenarios. The most visible wireless access is the DECT based home phone and the GSM based public phone. Within a couple of years a variety of new wireless access networks will appear, ranging from local networks with a limited coverage of a few metres to public long distance distribution systems. The user will experience the need for different access methods depending on location and envisaged application. For an electronic bank transfer a secure connection of a few kbit/s might be sufficient, whereas the live transmission of a high resolution D-TV will require a radio link with more than 10 Mbit/s. A public network only cannot offer these varying demands. The wireless access has to be provided by an adaptive link, optimised for the application in mind. Europe has identified this vision as a key driver for the 4th generation network (4G) [DiBenetto2000]. Software radio is a key enabler for a 4G network, as it allows a seamless wireless access to a variety of access networks.

This paper will review a typical user scenario in chapter 2, explain the concept of SW radio in chapter 3, and provides in chapter 4 an overview of the challenges towards SW radio terminals and base stations. It concludes in chapter 5 with an outlook on the further developments in this field.

2 A Future Usage Scenario

A typical user scenario during the day includes activities at home, in a public place, at work and at hot-spots. A public place is understood as a typical travel environment, whereas the hot-spot scenario includes areas where many people meet, like a cinema or a city centre.

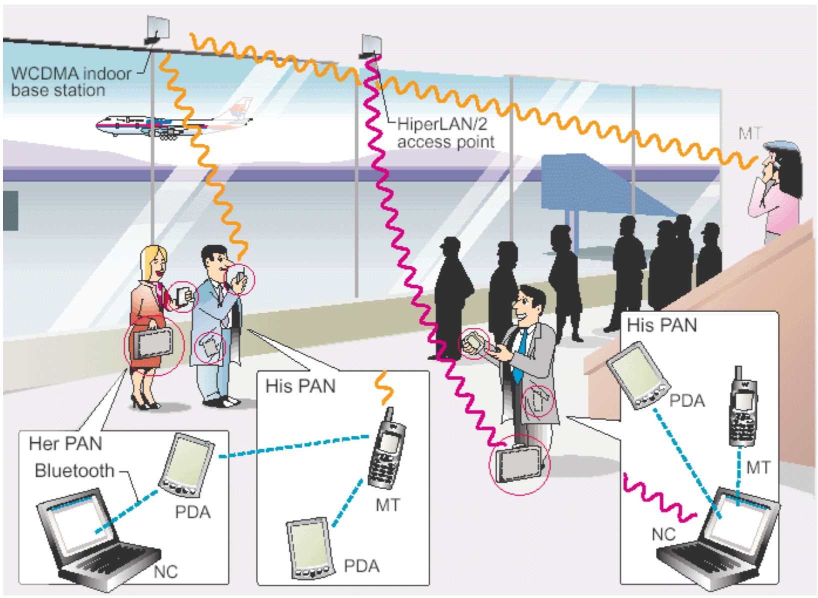
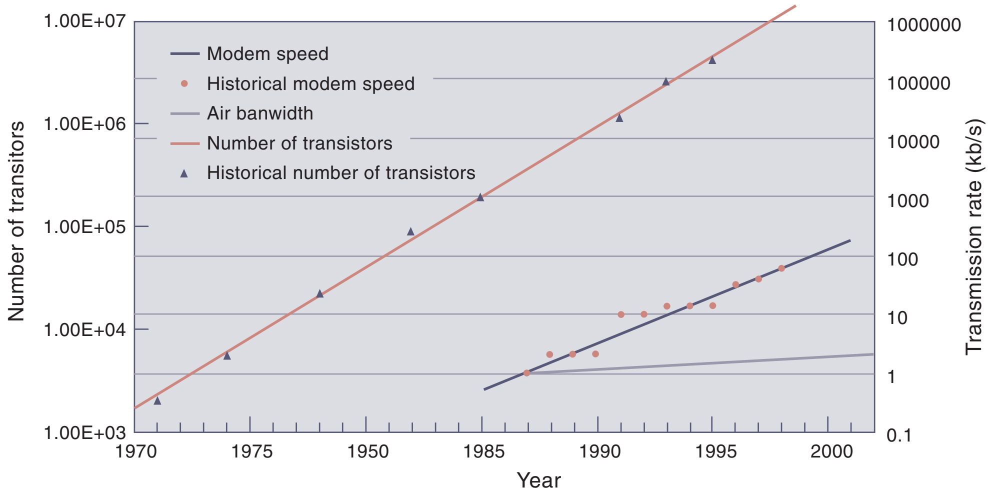


Figure 1 People on the move use a wireless connection to the nearest access point, (From [Ericsson2000])

Figure 2 Applications will grow faster than public available bandwidth



What is common in all these scenarios is the fact that the user wants to have wireless access to information (Figure 1). The user will find different access networks, e.g. DECT, WLAN or Bluetooth at home, HiperLAN/2 in the office, GSM/GPRS or UMTS FDD while he is travelling, and Bluetooth, UMTS TDD or HiperLAN/2 at the hot-spots. In order to access all these different technologies, his modem/terminal has to have the capability to adjust to different radio interfaces and operate in various frequency bands in the 1.8, 2.4 and 5 GHz region [Cuomo2000].

A comparison of number of transistors, available bandwidth on the fixed line and available spectrum efficiency on the radio link is shown in Figure 2. The number of transistors doubles approximately every two years, following Moore's law of information content in the information society. While modem speed over the fixed line followed the same trend, the capacity of the air interface, expressed in bits/s/Hz, shows only a slight increase in the last 10 years.

The trends and underlying system specifications can be summarised as follows:

- Applications and hardware requirements grow faster than modem capabilities.
- GSM and UMTS are developed for mobility up to a speed $v < 250$ km/h, thus sub-optimal for stationary use.
- The maximum network capacity is typical one Mbit/s in a UMTS network.
- High bit-rate data access is mainly requested from stationary users.

These elements suggest that the future access has to be adaptive to the user needs in a specific situation (Figure 3).

Figure 3 Each user scenario has an optimum access

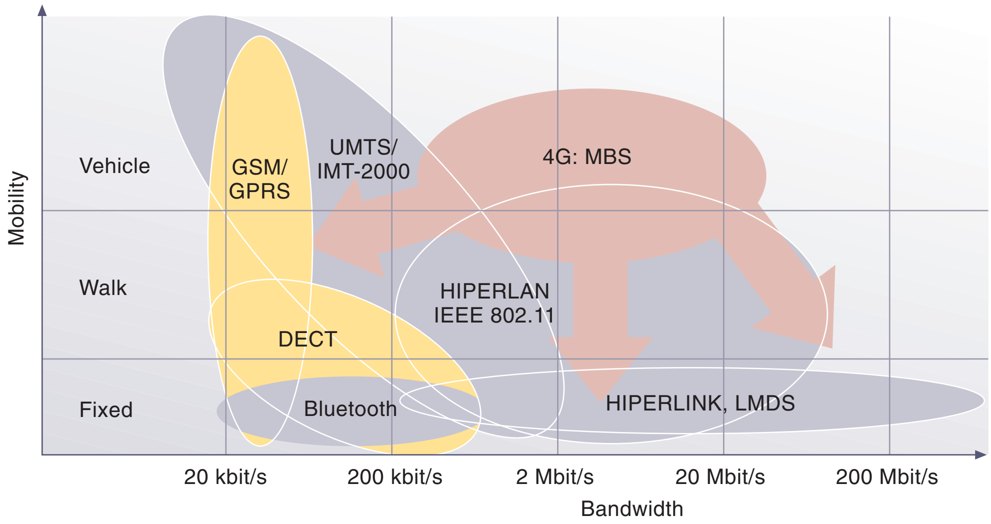


Figure 4 Mobility versus bandwidth in current and future wireless access systems

Software (SW) Radio has the capability of providing adaptive access, thus being the key enabling technology for future systems.

3 Software Radio

Chapter 2 has pointed out the need for a technology able to connect to various wireless access systems (Figure 4).

SW radio can be one of the key enabler for future systems, with a functionality including [Buracchini2000]:

- Flexible TX/RX architecture, controlled and programmable by SW;
- Signal processing to replace as much as possible of the radio functionalities;
- Ability to reconfigure the radio equipment by SW download;
- Multiple-mode operation by SW adaptation to the radio standard.

The SW radio shall further have programmable:

- Frequency band and radio channel bandwidth;
- Modulation and coding scheme;
- Radio resource and Mobility management.

As the whole functionality is operated by software, the network operator, the service provider, or the user could change the functionality at any time.

Previous European research projects have identified key elements of a SW radio architecture. Out of these the ACTS SORT (Software Radio Technology) and ACTS FIRST (Flexible Integrated Radio Systems Technology) projects are highlighted, as they have demonstrated the potential of SW radio for flexible air interface control very efficiently [FIRST, SORT].

3.1 SW Radio Functionality

Figure 5 shows the evolution from a multi-mode terminal to a fully adaptive, multi-mode, multi-frequency SW radio terminal. Currently multi-mode terminals are developed with specific ASICs for each mode. At a later stage the base-band (BB) processing is expected to be common for all modes, supported by mode specific RF modules. A fully and ideal SW radio uses advanced digital signal processing and digital to analogue conversion techniques to synthesise the radio signals as close as possible to the antenna (Figure 5). The incoming signal is bandpass filtered (BPF), amplified (LNA) and converted to a digital signal (ADC) allowing digital baseband processing. The baseband processing can be controlled and programmed using software upgrades alone.

Today, the capability of the RF components, of the baseband processors and the digital to analogue converters required for a full implementation of SW radio is not sufficient to do all of these functions using programmable hardware.

The quick development in this area, with a doubling of the capabilities every eighteen months, will progressively make SW radio an attractive solution, presumably already within the next two to three years.

3.2 Advantages of SW Radio

SW radio addresses issues involving all parts of the network and restructures its management. A necessary technological evolutionary step for a

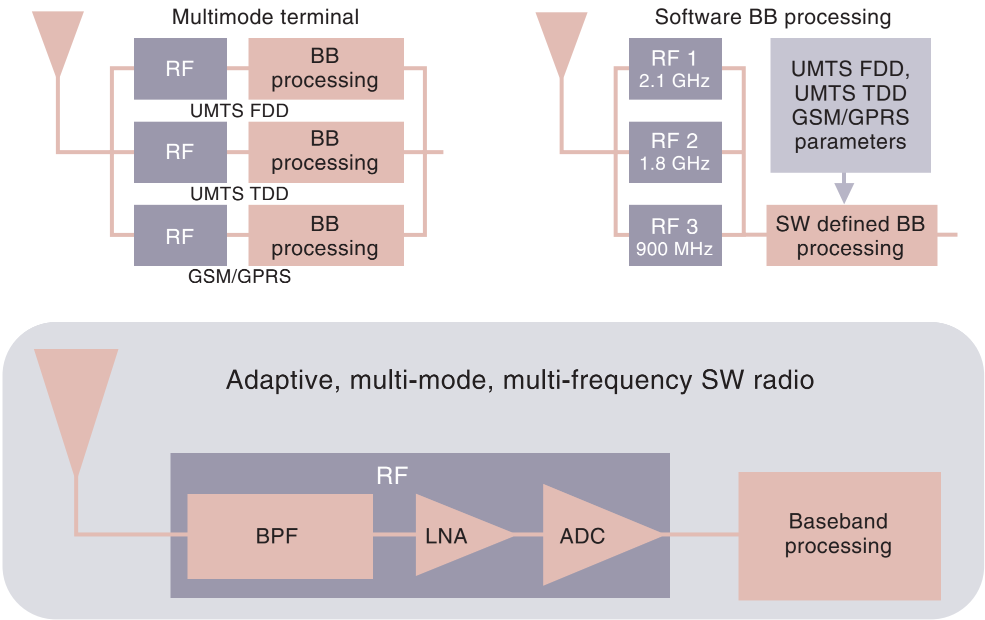


Figure 5 Evolution from multi-mode terminals to fully adaptive, multi-mode SW radio terminals

SW radio implementation is the introduction of intelligent base stations, which will incorporate part of the intelligence of the rest of the network. Such SW radio base stations (and terminals) will negotiate and select the mobile access method, based on best channel performance or best spectrum efficiency algorithm. Furthermore, a differentiated list of possible services (based on particular user needs or service provider offers) may be provided by the access network. Thus, SW radio deals with both operational issues (i.e. handover between heterogeneous base stations, networks and operators) and management aspects (i.e. the definition of flexible accounting procedures that enable different service providers to offer advanced services based on data transactions, location of the user, QoS assigned to a session, etc.).

Software radio has the potential to bring benefits to all actors in the telecommunications market:

- Manufacturers can concentrate on a reduced hardware set and expand the capabilities of the handsets by SW upgrades.
- Network operators can perform rapid roll-out of new services by SW upgrade of the base stations.
- Service providers can support a wide range of mobile multimedia applications using flexible transport services.

4 Challenges for SW Radio Realisation

The main challenges for a SW radio system are [Buracchini2000]:

- To move the border between analogue and digital worlds towards radio frequency (RF);
- To replace dedicated hardware by field programmable gate arrays (FPGA) or digital signal processing (DSP) for channelisation and baseband processing;
- To standardise a common interface for SW download.

4.1 Technical Issues

The main technical issues for the implementation of a SW radio are:

- Wideband RF and linear power amplifiers;
- Wideband, high-speed, and high-resolution A/D and D/A converters;
- High-performance signal processing devices (DSP, FPGA, μP);
- Standardised SW methods.

Supported by the European ACTS and IST research projects, developments are ongoing to establish technology for multi-standard, multi-frequency SW radio based solutions. A review of wide-band transceiver front-ends and their architecture is found in the literature [Hentschel1999, Tsurumi1999].

	Parameter	Current state	Future goal
A/D, D/A converters	Sampling rate, bandwidth, linearity, SNR, power dissipation SFDR, SNR > 70 dB	14 bit, > 80 Msps BW > 25 MHz SFDR, SNR > 100 dB	16-bit, 400 Msps BW ~100 MHz
Dynamic Up/Down Converter (DUC, DAC)	Bandwidth, I/O rate, tuning resolution, multi-carrier/channel support	16 bit, > 60 Msps BW 5 MHz SFDR > 90 dB	Multi-channel X x 5 MHz SFDR > 100 dB
DSP	Processing power, run time support, power dissipation	16-bit, 2 billion MACs 32-bit, 500 million MACs	Higher performance, efficient C-code execution
Power Amplifier (PA)	Linearity, power dissipation	BW > 60 MHz Gain up to 40 dB	BW > 100 MHz Gain up to 80 dB

Table 1 SW radio critical parameters [P921D1]

As sampling at RF is not possible with the current converter and signal processing technology, the radio signal has to be converted to an intermediate frequency. The programmable up- and down converters perform the most intensive processing tasks like filtering, modulation, and demodulation.

Signal-to-Noise Ratio (SNR), Signal-to-Noise and Distortion (SINAD) radio, Spurious Free Dynamic Range and Total Harmonic Distortion (THD), e.g. for multi-tone operation, are important radio performance parameters. The converters and signal processing part of the SW radio have to comply with the specifications given in the specific radio standard.

Table 1 provides an overview of SW radio critical parameters, the status of technology as of today and the expectations to be met for a SW radio terminal implementation.

4.2 SW Implementation and Download

The major challenge for the standardisation for SW radio is the definition of methods for SW download. This implies two aspects, the standardised interface for all terminals and the method for SW upgrade.

Manufacturers are likely to develop proprietary products. Therefore SW radio has to find a way to make the radio interface independent of the hardware platform. Like in computer systems, two potential solutions are seen:

- Develop a common HW platform and operating system. Even though this would allow an easy implementation of SW radio, the scenario is quite unlikely. As of today, a couple of operative systems are known for terminals and PDAs, as e.g. EPOC, PalmOS and PocketPC. It is not likely that manufacturers will join on one single platform and operative system.
- Resident compilers. The compiler generates an executable code specific to the hardware platform on which the code runs. This scenario allows differentiation on the supply side.

Java could be a suitable candidate for a resident compiler. Even though extensions for a real-time operative system are needed, most platforms have already today a virtual Java machine implemented. Further advances will come with DSP supporting Java directly [TI].

Figure 6 provides an overview of the hardware and software in a SW radio context: the virtual machine or virtual radio platform allows to abstract SW from the manufacturer specific hardware.

Most authors regard the reconfigurability of the air interface as the most important feature of a SW radio terminal. Reconfiguration of the service functionality is of the same importance [Daneshgaran2001]. Figure 7 provides an overview of capabilities of a future SW radio termi-

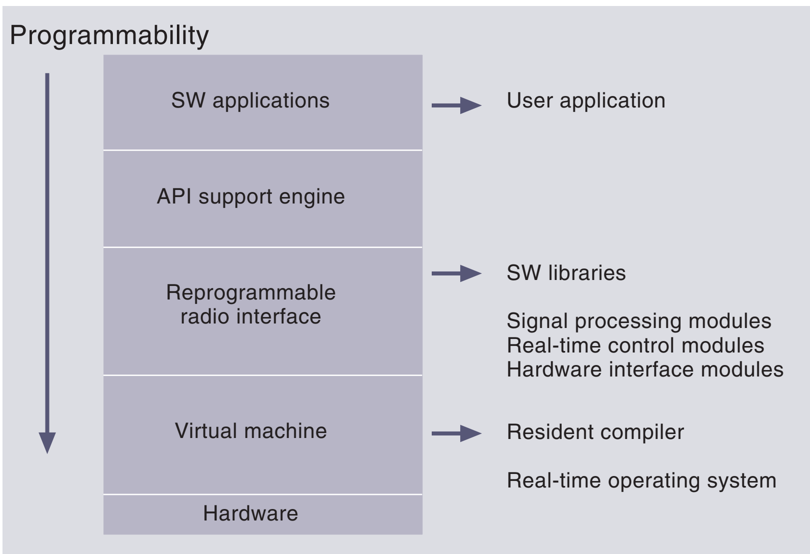


Figure 6 Relationship between HW and SW in an SW radio context

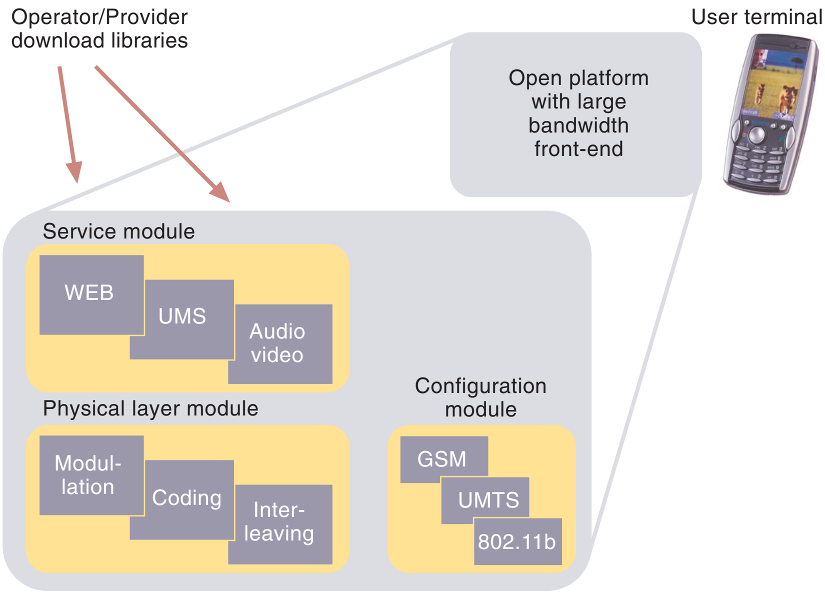


Figure 7 High level functionality of SW radio libraries

nal and the operator/provider specific libraries which need to be upgraded.

The implementation of fully programmable terminals and base stations requires the definition of methods for SW download. Concerning the base station, the SW download is normally executed when a new software version is released. In the base station the software updating does not have to be performed frequently. More critical is the software download at the mobile terminal side because this task may be performed more frequently due to the user's mobility, or if the user wants to change the air interface. Moreover, the software download in the mobile terminal must be as fast as possible and easy to perform. Different methods for software download concepts can be conceived:

- Smart card loading (e.g. SIM);
- OM/EPROM-based reconfiguration (different radio configurations are stored inside the SW radio);
- Service terminal (offline reconfiguration by connecting physically to a service terminal);
- Air interface download.

The air interface download is the most flexible one but needs a dedicated channel for the download. In the short term, the smart card loading seems to be the more practical one for implementation in software radio for mobiles, whereas for base stations the download can be managed over the network. The different download concepts for updating and reconfiguration of the system can be grouped in user-, vendor-, manufacturer-, and operator programmable.

4.3 Terminal Issues

The technological challenges for a SW radio terminal are quite stringent. Current handsets are optimised for low power supply and minimum size. Customers are used to change their mobile phone with a newer model quite often. If this trend holds, the market for SW radio terminals will be rather small.

It is however questionable that users are willing to change their terminal every time an advanced application or interface is available. This might open the market for SW radio terminals.

Quite some challenges are ahead prior to a product, including IF sampling, A/D and D/A design, and linear power amplifiers. Demonstration applications including multiband handsets, the SORT demonstrator, dual-use radios (Rhode & Schwartz) and car radio (DaimlerChrysler) are available and demonstrate the way further to commercial products [Mitola2000]. Realisations of a SW radio terminal are also available in the form of PC-based systems. Laddomada et al. have developed a prototype SW radio platform on a 700 MHz Pentium III system with an implementation of the front-end transceiver, the I/Q interface and the supporting programming environment [Laddomada2001].

The SDR Forum expects to see increased commercial, civil, and military use of SW radio in 2001, driven by the development and initial deployment of third generation commercial wireless systems and the need for multi-service capabilities in civil and military markets. By 2005, SW radio is expected to have been adopted by many manufacturers as their core platform [SDRForum].

4.4 Base Station Issues

Compared to terminals, the implementation of a SW radio base station appears simpler. Some vendors have already implemented solutions [SDRForum], but the technology is not yet mature enough for a global roll-out. The major challenges for the base stations are A/D, D/A converters and linear, broadband power amplifiers.

For mobile terminals where power consumption is quite strict, the use of highly oversampling sigma delta A/D-D/A converters is promising. For base stations, half-flash/weighting converters is an interesting option as fast 14-bit broadband converters are already available.

In Europe all mobile operators plan for the roll-out of UMTS, based on either an upgrade of the existing GSM 1800 network or a complete new installation. The benefits for the network operator in a SW radio based migrating to UMTS will be:

- The cost of the infrastructure is reduced;
- The air interface can be changed by reprogramming the DSP (Figure 8).

In the future evolution of the network, an integration of HiperLAN/2 (H/2) and UMTS TDD is foreseen. An SW radio base station will increase the hardware lifetime (both for the base station and the user terminal) and protect the investments. The system re-programmability allows for hardware reuse and the inclusion of future access methods such as UMTS TDD and HiperLAN/2. The same phenomena are likely to be experienced for mobile terminals.

4.5 Related Technologies

Technological developments seem to continue following Moore's law, with a doubling of processor speed about every two years (see Figure 2). Current developments in Quantum transistors expect terabit densities and processor frequencies as high as 700 GHz [Geppert2000].

On the radio side, Smart Antennas and multiple input, multiple output (MIMO) technologies have opened for increased bandwidth in the radio channel. These antenna technologies have to be seen together with advanced coding methods, as e.g. space-time coding or combined coding and modulation algorithms.

In the European vision of the "Wireless Information Society" these technologies are combined with the development and implementation of new access methods [Lu2000]:

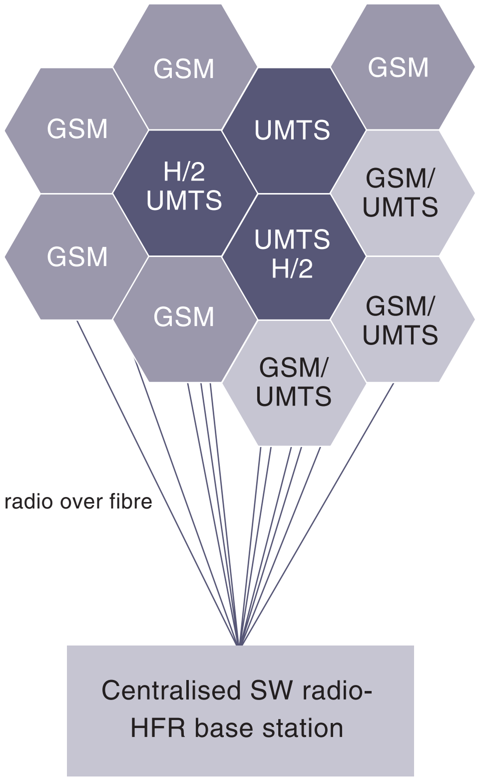


Figure 8 Migration from 2G to 3G networks

- DAB, DVB-S and DVB-T broadcasting;
- Mobile Broadband systems (MBS 40, MBS 60);
- Broadband wireless local loop;
- The proposal for new radio interfaces.

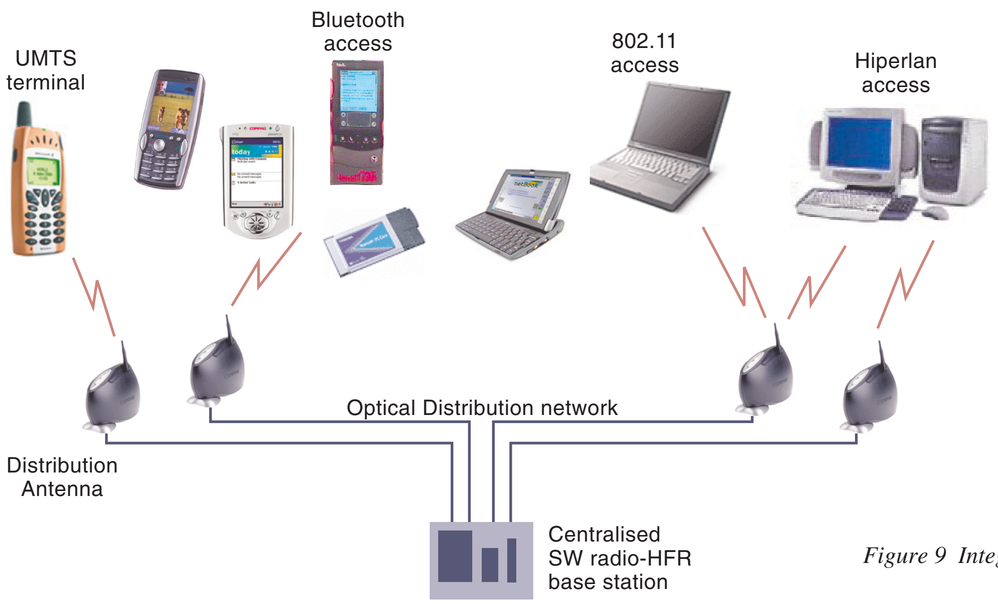


Figure 9 Integration of UMTS and LAN by SW radio-HFR base stations

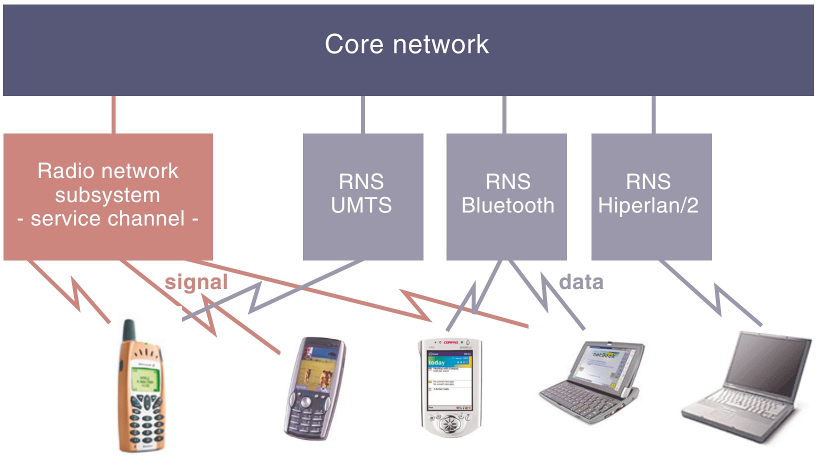


Figure 10 Pilot channel for access assignment

5 Outlook

The paper has described SW radio as a technology to support 4G mobility; that means the support of different radio access networks by one base station. Hybrid Fibre Radio (HFR) is a proven concept to distribute radio directly over fibre. In the EURESCOM P921 project a successful demonstration of UMTS over a 20 km long fibre was performed, suggesting the combined use of a SW radio base station with distribution of the radio signal over fibre to the remote antenna units [P921D1].

HFR is able to transmit all the common radio standards currently around like: DECT, GSM, UMTS, LMDS, MBS. However, the actual analogue optical feeder design may vary depending on the radio signal transmitted. Within the P921 project, HFR was identified as a very promising technology in the near future in order to give cheap, building independent wireless access by supporting small in-door radio cells (picocells) and outdoor GSM/UMTS microcells. SW radio in this sense would allow different radio standards to be transmitted in a flexible manner to these radio cells, thereby giving radio independent access (see Figure 9).

The integration of different access technologies has to be supported by an architecture for identification and assignment of the available access systems.

Figure 10 demonstrated the idea of using a pilot channel for the assignment of a data channel. Most European countries have an almost 100 % coverage with GSM. In the GSM system the BCCH channel is assigned for signalling. This channel could be used to transmit information to the core network about the availability of additional access networks. The data transport will then happen on the optimal access network, while the pilot channel (e.g. BCCH) carries the necessary information for handover.

Mohr and Kornhäuser suggest a layered structure for the seamless future [Mohr2000]. They suggest a service convergence protocol to support the different future services. In addition, they suggest a medium access control (MAC) layer to guarantee wireless quality of service and high spectrum utilisation by allocating dynamic bandwidth.

6 Conclusions

This paper described the evolution of mobile access beyond 3G. An optimised access has to be established for each scenario, depending on the requirements of the application and the location of the user.

SW radio is the evolving technology to support the optimised access. The ideal SW radio consists of a direct A/D conversion at the antenna, and the digital baseband processing in reprogrammable FPGAs or DSPs. As sampling at RF is not possible with the current converter and signal processing technology, prototypes work

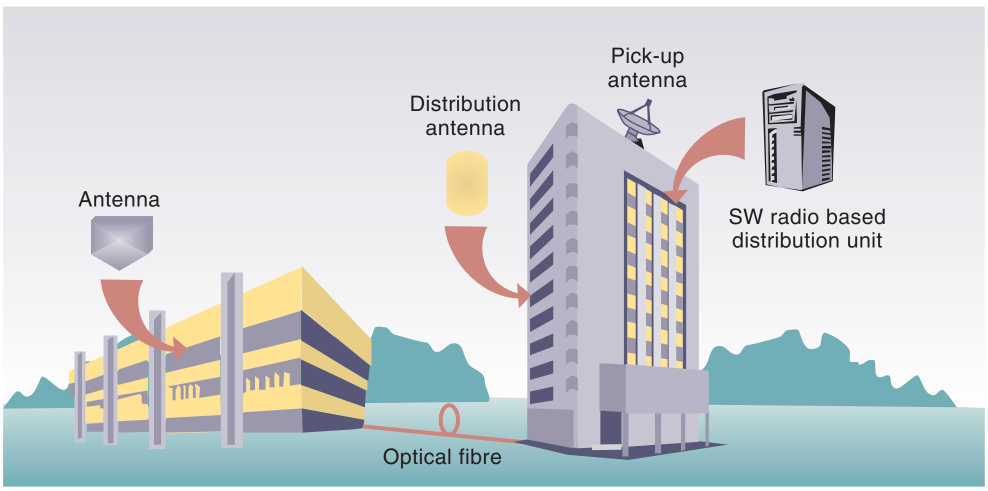


Figure 11 SW radio with a hybrid fibre distribution

with programmable up- and down converters, followed by base band processing.

Technology developments suggest the first prototypes of SW radio equipment within 2003, following the need for multi-standard 2G/3G terminals. SW radio technology is of even higher importance for base stations, as it increases the hardware lifetime, allows hardware reuse by reprogramming, and the inclusion of future access methods such as UMTS TDD and HiperLAN/2.

SW download and on-the-fly upgrade is a major advantage when using SW radio terminals. This implies standardised hardware solutions or resident compilers. The implementation of a real time Java virtual machine can be one of the most promising solutions in this field.

7 Literature

[DiBenetto2000] Di Benedetto, M-G, Tosco, F, Vatalaro, F. Mobile Radio Advances in Europe: Third Generation and Beyond. *IEEE Comm Mag*, 38 (9), 125–126, 2000.

[Ericsson2000] Frodigh, M, Johansson, P, Larsson, P. Wireless ad hoc networking – The art of networking without a network. *Ericsson Review*, 77 (4), 248–263, 2000.

[Cuomo2000] Cuomo, F, Baiocchi, A, Cautelier, R. A MAC Protocol for a Wireless LAN Based on OFDM-CDMA. *IEEE Comm Mag*, 38 (9), 152–159, 2000.

[Buracchini2000] Buracchini, E. The Software Radio Concept. *IEEE Comm Mag*, 38 (9), 138–143, 2000.

[FIRST] *ACTS FIRST, Flexible Integrated Radio Systems Technology*. (2001, March 16) [online] – URL: http://www.era.co.uk/first/first.htm

[SORT] *ACTS SORT, Software Radio Technology*. (2001, March 16) [online] – URL: http://www.ifn.et.tu-dresden.de/~sort/

[Hentschel1999] Hentschel, T, Henker, M, Fettweis G. The Digital Front-End of Software Radio Terminals. *IEEE Pers Comm*, 6 (4), 40–46, 1999.

[Tsurumi1999] Tsurumi, H, Suzuki, Y. Broadband RF Stage Architecture for Software-Defined Radio in Handheld Terminal Applications. *IEEE Comm Mag*, 37 (2), 90–95, 1999.

[TI] Texas Instruments. TI & JAVA: Bringing the Revolution in Network Intelligence to Wireless Systems. (2001, March 16) [online] – URL: http://www.ti.com/sc/docs/apps/wireless/techbrf.htm

[Daneshgaran2001] Daneshgaran, F, Laddomada, M, Mondin, M. Reconfiguration Issues of Future Mobile Software Radio Platforms. Submitted to *IEEE Pers Comm*, 2001.

[Mitola2000] Mitola III, J, Zvonar, Z. Software and DSP in Radio. *IEEE Comm Mag*, 38 (8), 140, 2000.

[Laddomada2001] Laddomada, M et al. A PC-Based Software Radio Transceiver for Third Generation Communication Systems. Submitted to *IEEE Comm Mag*, 2001.

[SDRForum] *Software Defined Radio Forum*. (2001, March 16) [online] – URL: http://www.sdrforum.org

[Geppert2000] Geppert, L. Quantum Transistors: towards nanoelectronics. *IEEE Spectrum*, 37 (9), 46–51, 2000.

[Lu2000] Lu, W W. Compact Multidimensional Broadband Wireless: The convergence of wireless mobile and access. *IEEE Comm Mag*, 38 (11), 119–123, 2000.

[P921D1] Sanmateu, A et al. SDR and HFR for low cost radio independent access. *EURESCOM P921 UMTS radio access project*. (2001, March 16) [online] – URL: http://www.eurescom.de/public/projects/P900-series/p921/

[Mohr2000] Mohr, W, Kornhäuser, W. Access Network Evolution Beyond Third Generation Mobile Communications. *IEEE Comm Mag*, 38 (12), 122–133, 2000.

Multiple Antennas – the Promise of High Spectral Efficiency

JØRGEN BACH ANDERSEN

Dr. Techn. Jørgen Bach Andersen (65) is Professor in Radio Communications at Aalborg University, Denmark, where he also heads Center for Personkommunikation (CPK), a research centre dealing with wireless communications. Jørgen Bach Andersen is a Fellow of IEEE.

jba@cpk.auc.dk

The combination of a scattering environment and multiple antennas at each end of a communication link leads to new solutions to the bandwidth dilemma. For a large number of incident paths from various directions it is possible in principle to change the usual, one-antenna fading channel into a set of parallel, non-fading channels, thus dramatically increasing the spectral efficiency in bits/s/Hz.

Introduction

Apparently there is a great need for wireless broadband related to the connection to the Internet and other sources of information. Compared with the wired networks, wireless networks have their own constraints and possibilities, arising from the special radio channel with the multipath fading, shadowing and other impediments. Present 2G systems are rather limited with respect to basic data rates, and while the planned 3G systems offer considerably higher rates, up to several hundreds of kb/s, it is still rather limited, and for the higher data rates with only limited range. The approach in this paper is to look at the more basic constraints arising from Shannon's and Friis' laws and use multiple element antennas at both ends of the link. The possibility of using several elements only at one end, typically at a base station, is of course also very important for interference reduction and gain enhancement, but this situation has been described in great detail elsewhere. Classical uses of antennas are already applied in Fixed Wireless Access (FWA) with high gain line-of-sight applications. We are here more concerned with the mobile situation. The use of adaptive antennas presumes a considerable amount of signal processing, either at RF or base band, but these problems are considered to lie outside the scope of the present paper.

Let us first discuss the free space situation in Figure 1, where there are two arrays each having two elements. The environment is simple, consisting of two isolated scatterers widely spaced angularly, and in the far field of the arrays.

The antenna elements are coupled to the input ports through a linear network supplying complex weights to the array elements making it possible to change the radiation patterns. There are two separate data channels, 1 and 2, indicated by the blue and red colour. In order to separate the two channels one easy solution is to send only the blue beam to the top scatterer having a null towards the lower scatterer, and vice versa for the red beam. Expressed analytically the radiation pattern for the array with two elements spaced d can be expressed as

$$F(\theta) = w_1 + w_2 e^{jkd \sin \theta} \tag{1}$$

By choosing the weights such that

$$F(\theta_0) = w_1 + w_2 e^{jkd \sin\theta_0} = 0 \tag{2}$$

where θ_0 is equal to the null direction, we are sure of the isolation between the two data signals. We have effectively doubled the capacity. It is also seen that all the degrees of freedom have been used, so we cannot have maximum gain in the wanted directions as noted in Figure 1. For that more antennas are needed. For M elements $M - 1$ null directions could be chosen, and the game could go on with more scatterers. There is a serious problem if the wanted direction is close to one of the nulls, so there are limitations to this technique.

Let us describe the situation in a little more general way and concentrate on the receive side only, where V_1 and V_2 are the signals at the two antenna elements and the combined signal after the weights

$$F = w_1 V_1 + w_2 V_2 \tag{3}$$

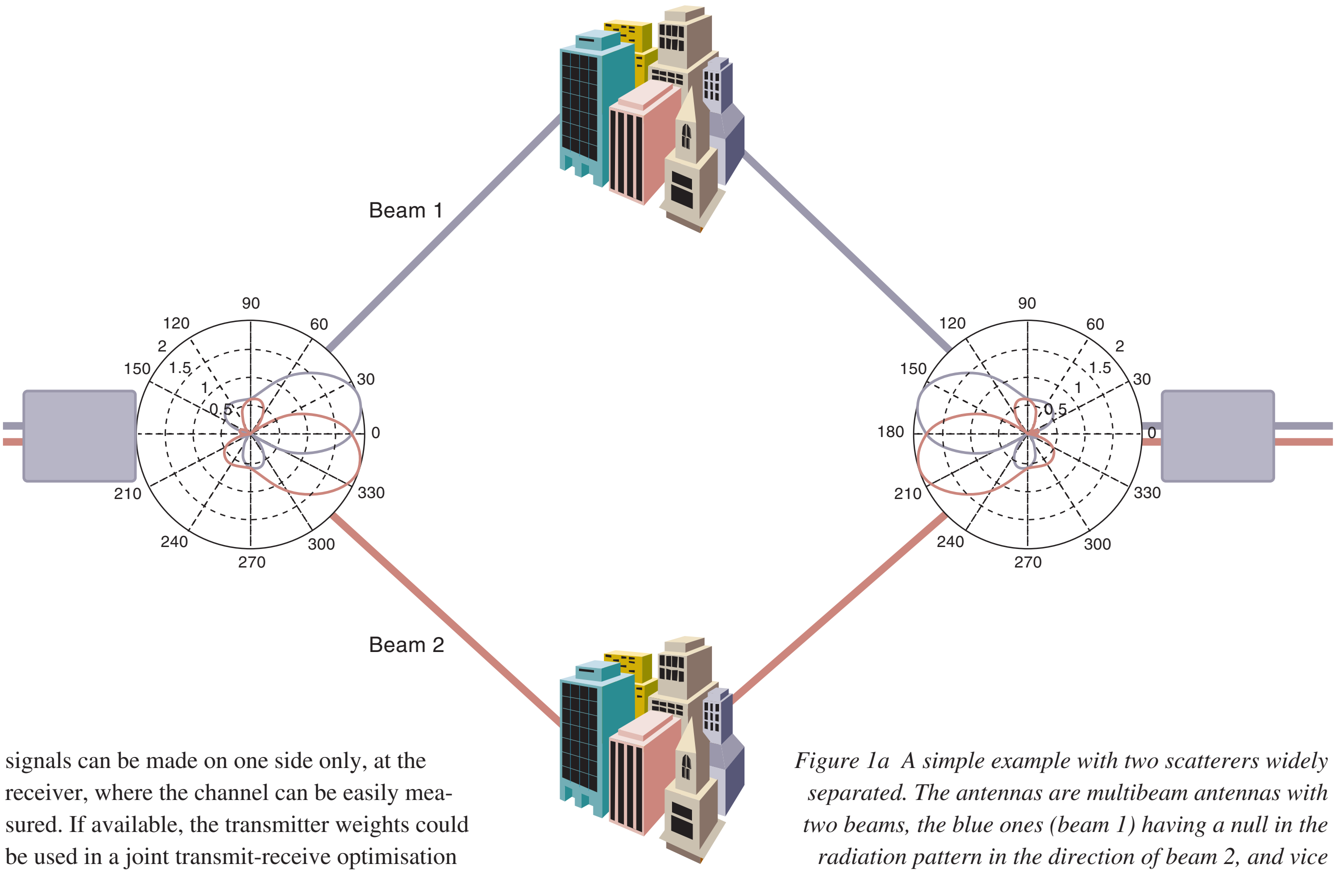
Each antenna signal is a combination of the two data streams

$$\begin{aligned} V_1 &= \alpha S_1 + \beta S_2 \\ V_2 &= \gamma S_1 + \delta S_2 \end{aligned} \tag{4}$$

so the output of the combiner is

$$F = (\alpha w_1 + \gamma w_2) S_1 + (\beta w_1 + \delta w_2) S_2 \tag{5}$$

and the output to port 1 can be made independent of the red signal S_2 by choosing $(\beta w_1 + \delta w_2) = 0$. In practice this is done by measuring the channel parameters α, β, γ, and δ. It is also seen that it is a condition that the factor to S_1 is different from zero, i.e. the signals must come from different directions. The separation of the



signals can be made on one side only, at the receiver, where the channel can be easily measured. If available, the transmitter weights could be used in a joint transmit-receive optimisation to maximise the wanted powers. The formulation in equations 4 and 5 is of course equivalent to the angular description above, but it leads to the more normal situation in Figure 2 where there is a multitude of rays, and there is no chance with a few elements to put many nulls in the radiation pattern. The key point is that it is not necessary to do so. The description in equations 4 and 5 is still valid so it is still possible to choose the weight factors appropriately to separate the two data streams and increase the capacity. Physically we can explain the situation as putting the unwanted signal in a deep fade instead of a null in a radiation pattern, the result is the same. Note that the scattering from widely separated scatterers is a necessity for the increased capacity; a line-of-sight situation where the two antenna arrays see each other directly would only lead to one channel, although with a larger gain. This follows directly from Shannon where the capacity in bits/s/Hz for independent parallel channels is given by

$$C = N \log_2 \left(1 + \frac{P}{N}\right) \qquad (6)$$

where P is the power over noise, or SNR, divided equally over the channels. The expression grows with N, the more so the larger P is, but the beauty of the array solution is that antenna gain will also grow with N, so we end up with the following

$$C = N \log_2 (1 + P) \qquad (7)$$

Figure 1a A simple example with two scatterers widely separated. The antennas are multibeam antennas with two beams, the blue ones (beam 1) having a null in the radiation pattern in the direction of beam 2, and vice versa. The antennas only have two elements, so they cannot also have a maximum in the wanted direction. The system has doubled the capacity, since the transmissions take place at the same frequency

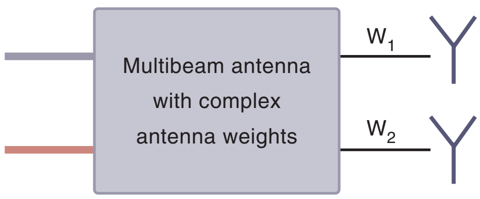


Figure 1b Two independent data streams are sent out and received over both antennas, and the weights are chosen such that there is isolation between the two beams

Thus parallelism is a very promising technique, it can only be realised in scattering environments with wide angular spreads, and multiple antennas at both ends are needed.

Although we have seen that only the receiver array has to know the channel, it is an advantage to let the transmitter array know the channel as well, which would increase the P in eq. (7). The information must in all cases be spread over the transmit elements, and considerable research has gone into finding optimum ways to do this using so-called space-time coding (Tarokh et al [8]).

Foschini [1] and Winters [2] realised that this situation could also be created with closely spaced elements in a scattering environment, where the scatterers would act as angularly widely spread parasitic elements of the array.

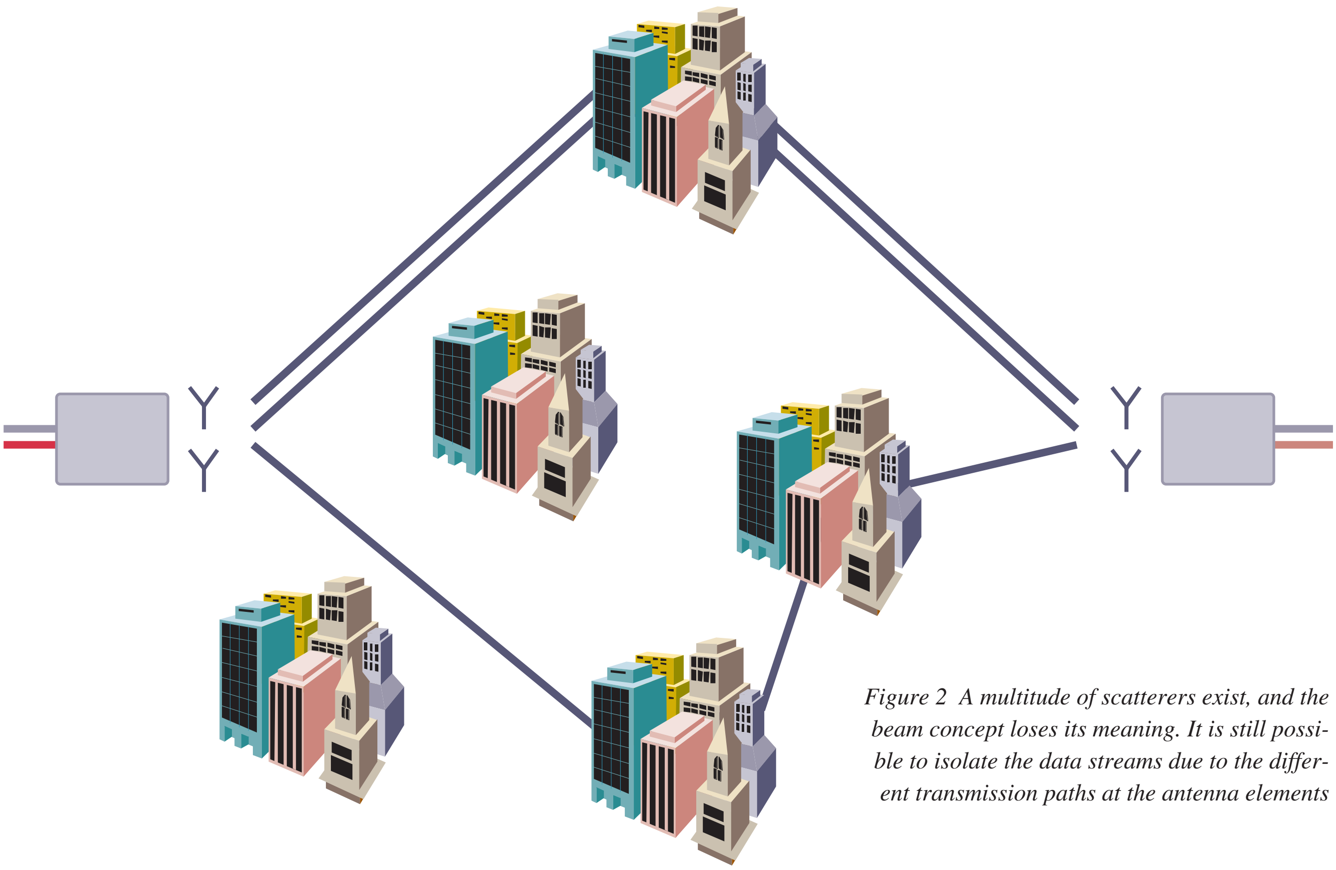
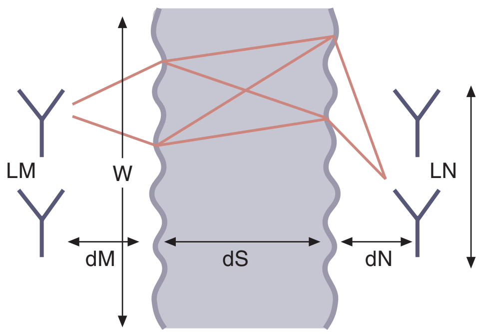


Figure 2 A multitude of scatterers exist, and the beam concept loses its meaning. It is still possible to isolate the data streams due to the different transmission paths at the antenna elements

Figure 3 Two arrays with M and N elements of total length LM and LN spaced a distance D illuminate a number of scatterers on two sheaths of width W. Distance from array to scatterer is dM and dN. A few ray paths are shown



Thus we have a situation where propagation and antenna aspects are interrelated in a new way. A tutorial overview may be found in Andersen [3] with more details in Andersen [4].

In the following we will leave the simplistic description and present some results using more advanced methods.

Theory

Some simplifying assumptions are made. The mutual interaction between array elements is neglected, and the spacing is fixed to half a wavelength.

The coupling between the two arrays may be described by a coupling matrix $\boldsymbol{H}(M,N)$ connecting each element to the left with each element on the right. Each element of the matrix is a sum of many transmission paths emanating from the element, scattered by the first layer, propagating to the scatterers in the second layer and to the element of the second array, see Figure 3. The reason for having two scattering layers is to be able to simulate many practical situations. Each scatterer is assumed to scatter isotropically, and the scattering cross-section is a random, complex number normally distributed. It will be assumed that the transmission matrix $\boldsymbol{H}$ is known both at the transmitter and at the receiver. When this is the case the complete link may conveniently be described through a singular value decomposition of the matrix, i.e.

$$\boldsymbol{H} = \boldsymbol{U} \cdot \boldsymbol{D} \cdot \boldsymbol{V}' \tag{8}$$

where $\boldsymbol{D}$ is a diagonal real matrix of singular values, $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, $\sqrt{\lambda_3}$, ... $\boldsymbol{U}$, $\boldsymbol{V}$ matrices of orthonormal excitation vectors for the receive and transmit side, respectively. There are in principle $min(M,N)$ different singular values, although in practice the number may be much smaller, as will be seen later. The singular values are ordered after decreasing magnitude, and if we concentrate for the moment at the largest we get

$$\boldsymbol{H} \cdot \boldsymbol{V}_1 = \sqrt{\lambda_1}\boldsymbol{U}_1 \tag{9}$$

from which we deduce that λ_1 is the power gain, or actually the maximum array gain, when the arrays are weighted with $\boldsymbol{V}_1$, $\boldsymbol{U}_1$. Since orthogonality is assured, the remaining singular values may be used simultaneously.

Maximum Gain and Diversity

Using just the maximum gain situation of eq. (9) it is of interest to study the behaviour of λ_1 as a function of the environments. In order to classify the various situations two dimensionless parameters are introduced

$$Fcor = \frac{LW}{d\lambda} \tag{10}$$

$$Fpin = \frac{W^2}{d_S\lambda} \tag{11}$$

where λ is the wavelength, $d = dM = dN$ the distance from array to nearest screen, d_S the distance between screens. We have here assumed symmetry around the middle, but it is clear that parameters may be defined separately for the two sides. The parameter *Fcor* is roughly the normalised phase difference over the array from the edge of the scatter screen, and similarly for *Fpin* between the two screens. When *Fcor* is small (<< 1) the screen looks like a point source seen from the array, and the antenna weights are uniform with no possibility for diversity. Contrary, for *Fcor* >> 1 the fields are spatially fading over the array, and diversity gain may be achieved. When *Fpin* is small the two screens are in each other's farfields, while for *Fpin* large, they are in the radiating near fields of each other. We can roughly define four different situations:

a) *Fcor* << 1 and *Fpin* << 1
This corresponds to a situation where everything looks like a point source. The total gain equals *MN*, and there is no diversity gain. There is only one eigenvalue, since there is effectively only one effective path between the two screens. It acts like a pinhole (Gesbert et al. [5], Chizhik et al. [6]), which explains the name *Fpin*. Since the link consists of a product of the two Rayleigh fading scatterers, the distribution is now a double-Rayleigh. Mathematically the probability density for the power is a modified Bessel function of zero order [6]. If there were many such screens we would end up with a lognormal distribution superimposed on the fast fading.

b) *Fcor* >> 1 and *Fpin* << 1
Locally around the arrays there is wide angular scattering, but the ensemble of scatterers still looks like a point from the other side, and there is only one eigenvalue. Each side may be optimized independently with full diversity gain, which also means a mean link array gain of *MN*, and a diversity order of *MN*.

c) *Fcor* << 1 and *Fpin* >> 1
The two sheets act like one scattering point in the far field of the arrays. Independent optimization on each side, no diversity, mean gain of *MN*, and the distribution is Rayleigh.

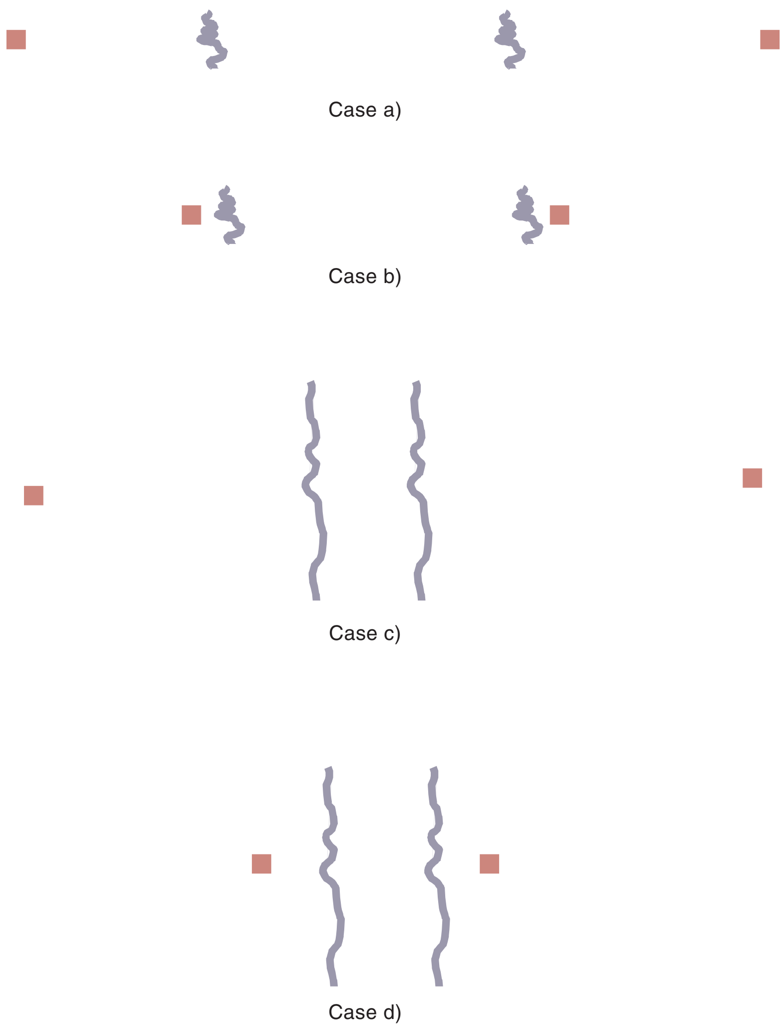
d) *Fcor* >> 1 and *Fpin* >> 1
Complete decorrelation at all points, the optimization at the receiver depends on the optimization at the transmitter, so it is a case of joint optimization. The mean gain of the best channel is now less than *MN*, and it may be shown (Andersen [3,4]) that asymptotically for large *M* and *N* the gain approaches

$$G = \left(\sqrt{M} + \sqrt{N}\right)^2 \tag{12}$$

The diversity order is high.

The four different cases are sketched in Figure 4.

Figure 4 The four different basic cases (see text)



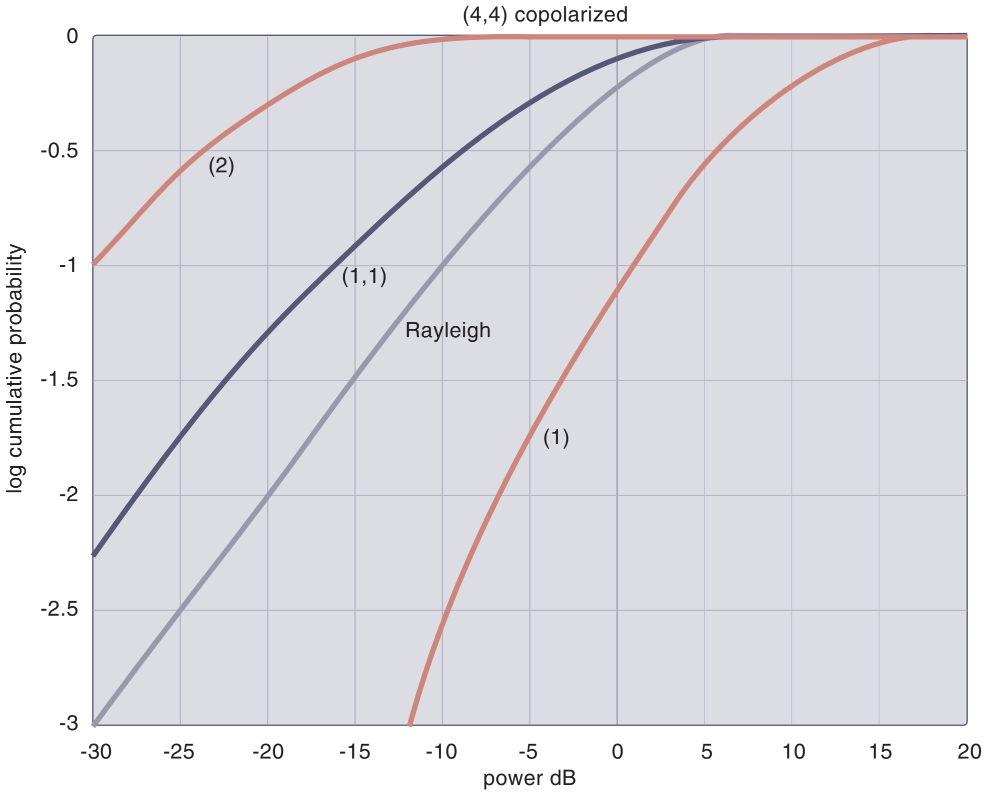


Figure 5 Cumulative probability distribution for two 4 element arrays. Fcor = 1 and Fpin = 0.1. The four curves from the right are the maximum eigenvalue channel, a reference Rayleigh distribution, a reference case of 1 antenna at each end, and the next largest eigenvalue channel. The two remaining channels have a gain too small to be shown

An illustrative example is shown in Figure 5 for partly correlated 4 element arrays and some pinhole effect (*Fcor* = 1, *Fpin* = 0.1). The (1,1) case is for one-element arrays, and it is seen that it is worse than Rayleigh. There should be four channels (eigenvalues), but only two have sufficient gain to appear in the graph. The mean gain for the maximum eigenvalue is close to 12 dB as it should be, and some diversity is noted, since the slope of the curve is higher than for the (1,1) case.

Another case with highly uncorrelated signals corresponding to case d) above is shown in Figure 6 (*Fcor* = 10, *Fpin* = 10). Now all four channels appear on the plot, and the diversity gain is clearly seen. It should be noted though that the

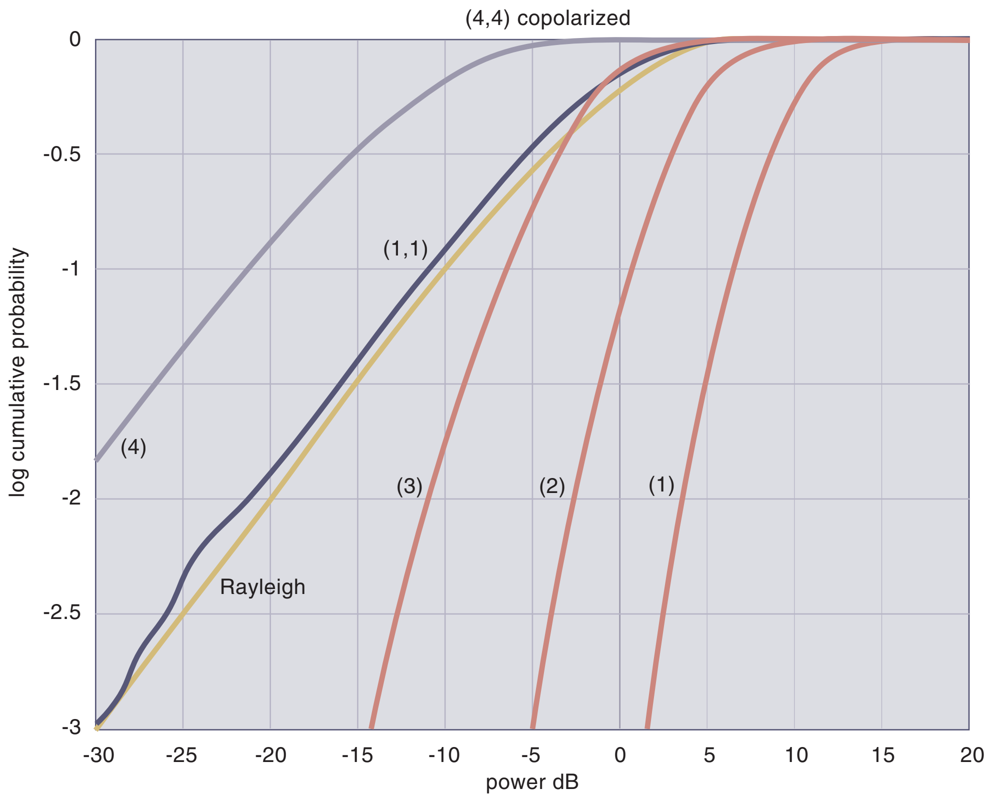


Figure 6 As in Figure 5, except Fcor = 10 and Fpin = 10. All four channels are now seen, the (1,1) case is close to Rayleigh, and the diversity gain is high, 25 dB at the 1 % level

slope of the gain curve corresponds to a diversity order of 4-6, less than one might expect.

It should be recalled that the gain values shown above rely on adaptive arrays, which all the time track variations in the channel and adjust the weighting factors on each side. This is a somewhat optimistic situation in practice, where often a fixed antenna beam will be used. If a fixed beam is used at one array, then the gain of that array will vary from M to 1, where the gain of M will correspond to *Fcor* small and 1 when *Fcor* is large (Andersen [3]). The latter case may also be understood by the effect of a narrow beam in a wide scattering situation, the antenna does not receive (or illuminate) all the scatterers, and the gain will decrease. There will of course be no diversity gain for a fixed set of weights.

Gain Impact on Data Rates, Range and Frequency

The data rate achievable over a link is a function of many factors like modulation, error distribution, coding et cetera, but in all cases a certain energy per bit is required, the well-known E_b/N_0. When the data rate increases the needed power increases with the data rate R. This is why the link antenna gain, the mean value and the diversity gain, are so important for wideband services, since the achievable data rate is proportional to the power gain.

The impact on the range depends on the decay of power with distance, so for a power law like

$$P = P_0\, d^{-n} \tag{13}$$

the range will vary with gain like

$$d = d_0\, G^{1/n} \tag{14}$$

As an example we can choose $n = 3.5$. If the required power is increased by a factor of 10 for a given data rate, the range is reduced by a factor 2, unless the gain is increased correspondingly by a factor 10. Four elements at each end will roughly give a mean gain of 10 dB.

As far as the carrier frequency is concerned the situation is more complicated, since it depends on how the path loss varies with frequency. The famous Hata law for urban propagation gives a frequency dependence of $f^{-2.6}$ for the received power, which corresponds approximately to the free space law and a shadow diffraction. This is valid for constant gain antennas, but if we instead consider constant area antennas filled with adaptive antenna elements, the situation changes dramatically. For all the situations discussed above where the gain was MN the received power now *increases* with the square of the frequency. The worst case is case d) above, where the gain is reduced, and in this situation the received power is *independent* of frequency.

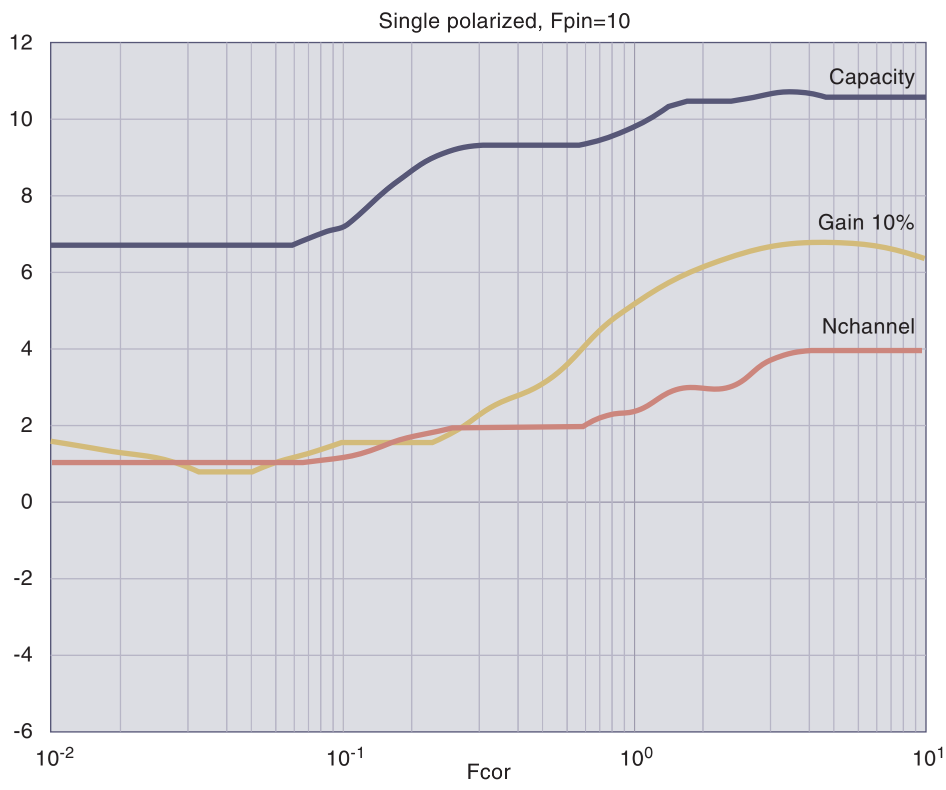


Figure 7 Two 4-element arrays with Fpin = 10 and SNR = 10dB. Nchannel is the mean number of active parallel channels, Gain10% is the gain in dB at the 10 % level, and Capacity is the capacity in b/s/Hz

Thus there seem to be possibilities of going up to 10–20 GHz for the mobile links if the adaptive arrays can be made economically feasible.

Parallel Channels

As already observed by Winters [2] the rich scattering environment may be utilised by having multiple channels. It is apparent from Figure 3 which shows 4 channels, 3 good ones with diversity gain and one bad one, worse than Rayleigh. It may be shown that the total capacity, or spectral efficiency in bits/s/Hz, may be given approximately by the following expression (Andersen [4])

$$C = N \log_2 \left(1 + \frac{MP}{N}\right) \tag{15}$$

with P being the signal-to-noise ratio for one antenna. It is assumed that M and N are large, and $M > N$. The formula may be simply interpreted as N parallel channels, each with power P/N and gain M. This is the ideal case of totally uncorrelated signals at the antennas, but it has this interesting aspect that there is in principle no upper limit to the data rate achievable over a given bandwidth for an infinite number of uncorrelated paths. The two-screen model is a good model for testing these assumptions. Reviewing the four cases again it is noted that cases a), b), and c) all have only one channel or one eigenvalue of any significant value, since either *Fcor* or *Fpin* or both are small. Only case d) offers the high capacity.

The two 4-element arrays will again be used as examples. Figure 7 shows the case of *Fpin* = 10 (highly interactive screens) as a function of *Fcor*. The *SNR* equals 10 dB, and the figure shows the mean number of active channels, $N_{channel}$, the gain in dB of the best channel at the 10 % level, and the mean capacity in b/s/Hz. At the left side of Figure 7 where *Fcor* is small, there is only one channel, no diversity gain and the basic capacity of 7 b/s/Hz for the one channel with 12 dB mean gain. In this region there is no reason to spread the signal over the antennas since there is only one channel. At the right side we get the four channels with some diversity gain, and the capacity increases to 11 b/s/Hz. The crossover point is between values of *Fcor* of 0.1 and 1.

Figure 8 represents the pinhole effect with *Fpin* = 0.01. A moderate increase in capacity with maximally 2 channels, but a fair diversity gain since the –4 dB at the left derives from the double-Rayleigh.

Effect of Polarization

In this section the case of dual polarized antennas is briefly discussed. This is especially pertinent for a small number of antenna elements, since the adaptive techniques ensure two inde-

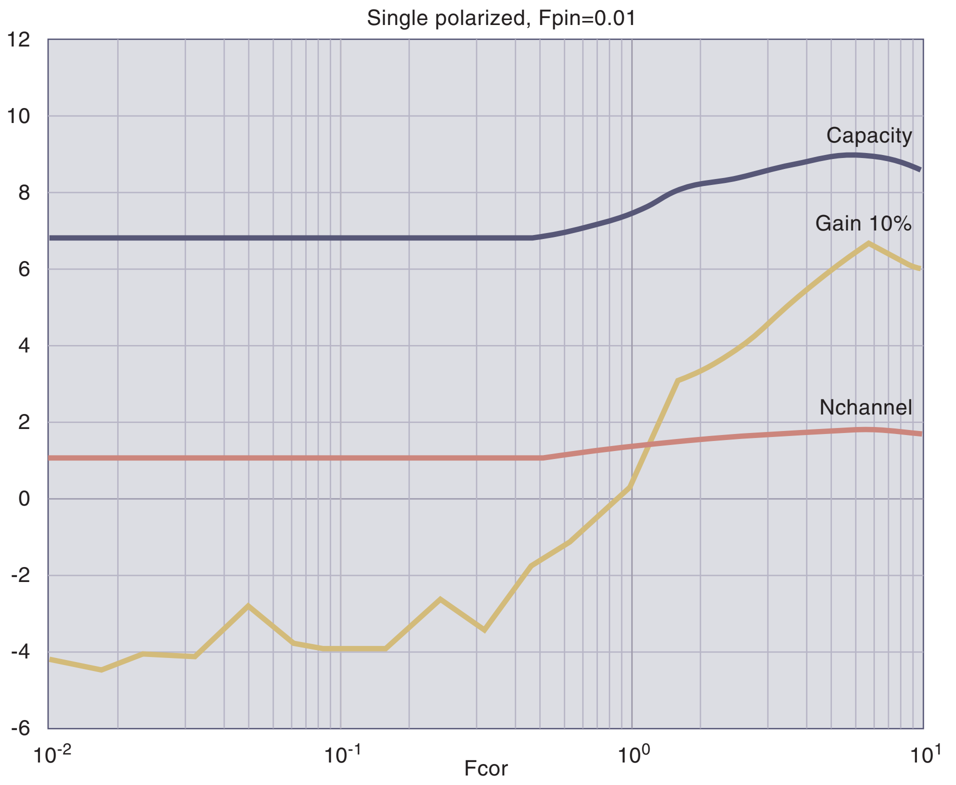


Figure 8 As in Figure 5, but with Fpin = 0.01

pendent channels even for low angular spread (low values of *Fcor*). We have here assumed an average cross-polarization level of –10 dB. In order to make a fair comparison there are two dual polarized elements. The result is shown in Figures 9 and 10 corresponding to Figures 7 and 8. The number of channels never drops below 2 as expected, and the capacity is now much less dependent on *Fcor* or the angular spread.

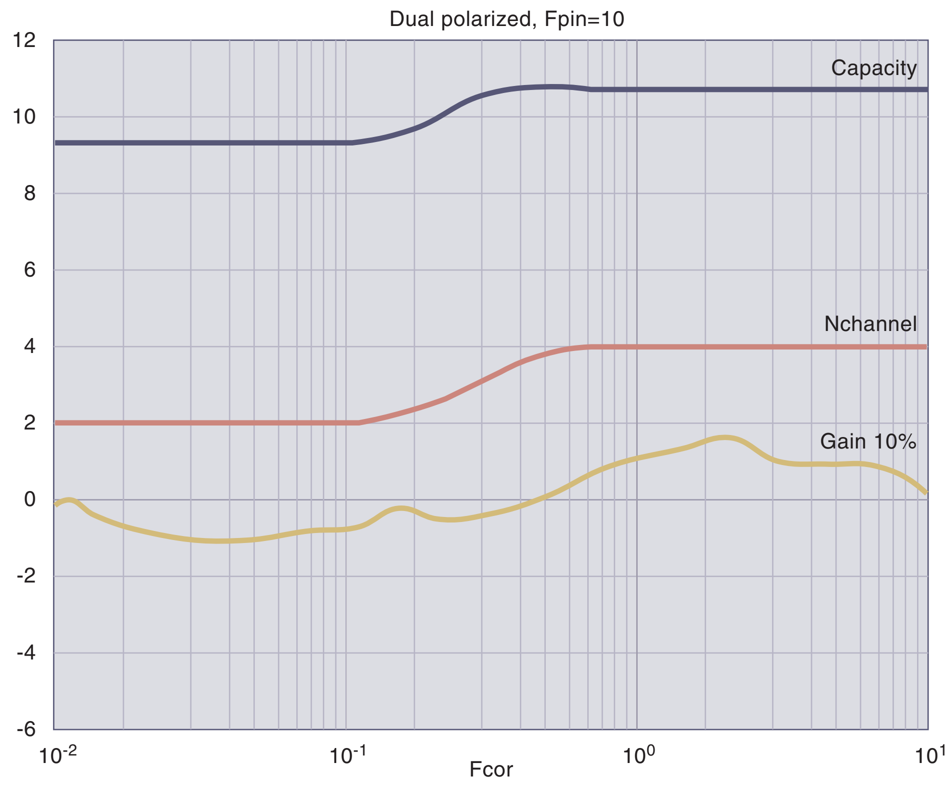


Figure 9 Two 2-element dual polarized arrays with Fpin = 10

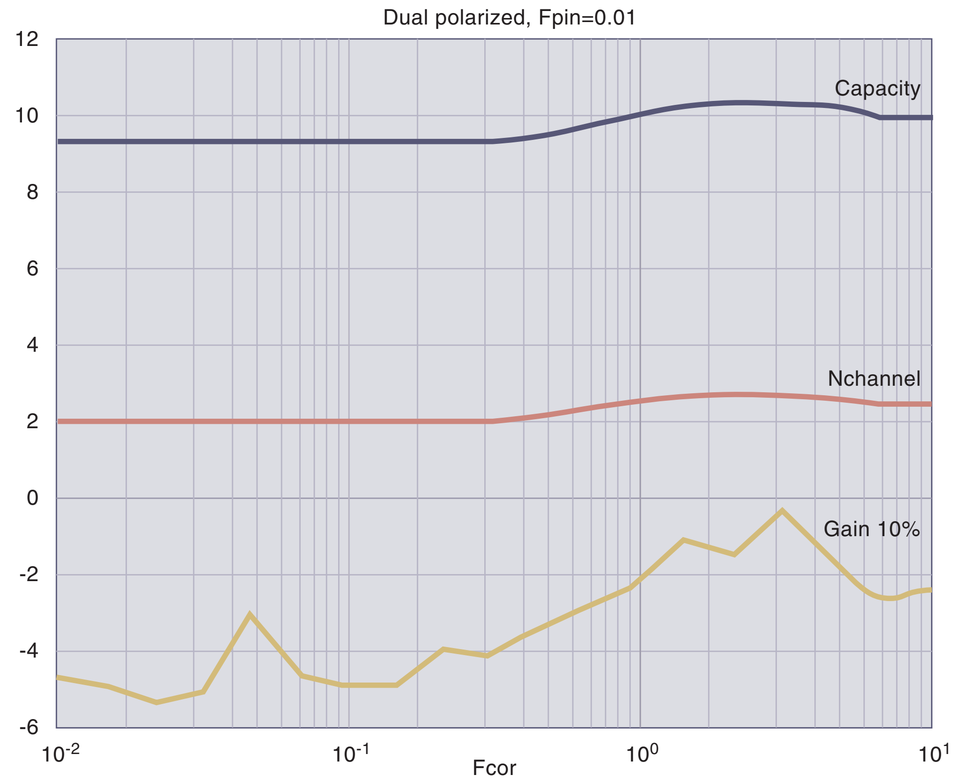


Figure 10 Two 2-element dual polarized arrays with Fpin = 0.01

Conclusion

The challenge of providing wireless broadband communications to the mobile users puts strong emphasis on the antennas. Antenna arrays at both ends of the link offer the gain needed for the higher data rates, where the gain is considered as the gain of the channel with the highest gain. It also offers the possibility of using higher carrier frequencies when the number of elements is increased accordingly. Also higher spectral efficiencies may be achieved if also the other channels are used. The multiple channels exist when the parameters *Fcor* and *Fpin* both are large. They express roughly speaking that the arrays and the extended scatterers are in the radiating nearfields of each other. If the screen of scatterers look as point sources seen from the arrays, or if the two screens appear as point sources, then the parameters are small and only one channel exists, although it may have maximum gain.

The use of polarization diversity is beneficial, since it guarantees two channels, even when the parameters *Fcor* and *Fpin* are small.

References

1 Foschini, G J. Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multi-Element Antennas. *Bell Labs Technical Journal*, 1 (2), 41–59, 1996.

2 Winters, J. On the Capacity of Radio Communications Systems with Diversity in a Rayleigh Fading Environment. *IEEE Journal on Selected Areas in Communications*, 5, 871–878, 1987.

3 Andersen, J B. Antenna Arrays in Mobile Communications : Gain, Diversity, and Channel Capacity. *IEEE Antennas & Propagation Magazine,* 42 (2), 12–16, 2000.

4 Andersen, J B. Array Gain and Capacity for Known Random Channels with Multiple Element Arrays at Both Ends. *IEEE Journal on Selected Areas in Communications*, 18, 11, 2000.

5 Gesbert, D et al. Outdoor MIMO Wireless Channels : Models and Performance Prediction. *Globecom 2000,* San Francisco.

6 Chizhik, D, Foschini, G J, Valenzuela, R A. Capacities of multi-element transmit and receive antennas : Correlations and Keyholes. *Electronics Letters,* 36 (13), 1099–1100, 2000.

7 Gallager, R G. *Information Theory and Reliable Communication.* New York, Wiley, 1968.

8 Tarokh, V, Seshadri, N, Calderbank, A R. Space-Time Codes for High Data Rate Wireless Communication : Performance Criterion and Code Construction. *IEEE Transact. Information Theory,* 44 (2), 744–765, 1998.

Adaptive Coding and Modulation:
A Key to Bandwidth-Efficient Multimedia Communications in Future Wireless Systems

KJELL JØRGEN HOLE AND GEIR E. ØIEN

Kjell Jørgen Hole (39) received his BSc, MSc and PhD degrees in computer science from the University of Bergen in 1984, 1987 and 1991, respectively. He is currently Senior Research Scientist at the Department of Telecommunications at the Norwegian University of Science and Technology (NTNU) in Trondheim. His research interests are in the areas of coding theory and wireless communications.

Kjell.Hole@ii.uib.no

Geir E. Øien (35) received his MSc and PhD degrees in electrical engineering from the Department of Telecommunications at the Norwegian Institute of Technology in 1989 and 1993, respectively. Øien is Professor in information theory at the Department of Telecommunications at the Norwegian University of Science and Technology (NTNU), Trondheim. His current research interests are in the areas of information theory and wireless communications.

oien@tele.ntnu.no

This paper gives an introduction to adaptive coding and modulation, and explains its potential in wireless communications over time-varying fading channels. If accurate channel modelling and channel state estimation can be performed, the average information rates (measured in transmitted information bits per second) might be significantly increased by the use of rate-adaptive transmission. Current theory predicts an increase in achievable average information rates of several hundred % over today's systems for the same number of users and the same overall bandwidth. We show results for a practical adaptive coding scheme applied to some well-known flat-fading channel models. An important performance measure discussed is average bandwidth efficiency (information rate per unit bandwidth) vs. bit-error-rate. Both single-user links and multi-user microcellular networks are used as examples of what might be achieved by a properly designed adaptive transmission scheme. The examples indicate that rate-adaptive transmission can provide the large information rates needed to support high-quality multimedia services in future wireless systems.

1 Introduction and Motivation

Wireless access, e.g. to the Internet, may in the future replace many fixed-wire connections in telecommunications. Wireless technology can be rolled out rapidly, thus bypassing the need for installation of new cables. The technology also encourages *mobility*, allowing people on the move to access interactive multimedia services such as video conferencing, e-commerce, location services, and (network) computer games.

An acceptable quality of service for future multimedia services (e.g. high audio and video quality, high reliability, strict real-time constraints) can only be achieved by realizing much *higher information rates* than those available in today's wireless systems. At the same time, *bandwidth* is becoming an ever-scarcer resource as the number of systems, users, and services increase. If real-time, high-quality multimedia services are to be supported in future wireless systems, there is thus a need for novel transmission schemes – i.e. compression, modulation, error control, and access techniques. The schemes must provide bandwidth-efficient, robust communication with low delay, supporting multiple users on wireless channels. Since time-varying channel conditions – and thus *time-varying capacity* – is an important feature of wireless and mobile communication systems, future systems should exhibit a high degree of *adaptivity* on many levels in order to reach these goals [1, 2]. Examples of such adaptivity are: information rate adaption, power control, code adaption, bandwidth adaption, antenna adaption, and protocol adaption.

In this paper we focus on the aspects of *information rate adaption*. The ideas presented are generic in the sense that they might be exploited both in fixed wireless access and satellite communications, as well as in fourth generation (4G) mobile cellular systems.

The paper is organized as follows: Section 2 models a single-user wireless fading channel, while Section 3 discusses the channel (Shannon) capacity and argues why the coding and modulation scheme should be adaptive. Section 4 introduces techniques for designing and analyzing rate-adaptive schemes. A multi-user cellular network utilizing adaptive coding and modulation on the wireless links is modelled in Section 5. The spectral efficiency of a particular adaptive scheme is presented in Section 6 for both the single-user and multi-user models. Section 7 concludes the paper and gives an overview of future research.

2 Wireless Channels

The signals transmitted on wireless channels are subjected to a number of impairments, notably reflections, attenuation, and scattering of power. Wireless channels are thus typically characterized by *multipath transmission* of the signal, i.e. the received signal results from summation of different replicas of the original signal. Each replica has its own particular amplitude attenuation and delay, which vary with time. This leads to time-varying signal strength and signal *fading*.

In the following we give some background on the modelling techniques used for wireless channels. We will focus on *narrowband* channels with *frequency-independent* or *flat* fading [3].

2.1 Narrowband Radio Channels

Narrowband radio channels are often represented by stochastic flat-fading channel models – such as Rayleigh, Rice, or Nakagami fading [3]. The models are described simply by a probability density function (PDF) for the fading amplitude, and the fading correlation properties.

We shall make use of the well-known *complex baseband* model. Denoting the transmitted complex baseband signal at time index k by $x(k)$, the received signal after transmission on a flat-fading channel can be written as $y(k) = \alpha(k) \cdot x(k) + w(k)$. Here, $\alpha(k)$ is the *fading envelope* and $w(k)$ is complex-valued additive white Gaussian noise (AWGN) with statistically independent real and imaginary components. The fading envelope ($\alpha(k)$) is a complex stochastic variable representing the resultant gain after summing all received signal components (line-of-sight, reflected and scattered) at the receiver. In this paper we shall assume that the receiver is able to perform perfect coherent detection, effectively performing perfect phase compensation and thus turning $\alpha(k)$ into a real-valued variable.

In general, the received *channel signal-to-noise ratio* (CSNR) at a given time k is defined as

$$\gamma(k) = \frac{P_r(k)}{P_n(k)}, \tag{1}$$

where P_r is the received signal power and P_n is the received additive noise power. The CSNR is often the most useful way of describing the channel's state. We shall assume that a constant average transmit power P [W] is used, and that the two-sided power spectral density of the AWGN is $N_0 / 2$ [W/Hz]. For a given channel bandwidth B [Hz], we then have

$$\gamma(k) = \frac{\alpha^2(k) \cdot P}{N_0 B} \tag{2}$$

with $E[\gamma(k)] = \bar{\gamma} = GP / (N_0 B)$ where $G = E[\alpha^2(k)]$ is the average received power attenuation.

2.2 Nakagami Multipath Fading: A General Flat-fading Channel Model

Of special interest is the so-called *Nakagami multipath fading* (NMF) model [3]. This is a very general statistical model covering or approximating a lot of interesting special cases, including the more well-known Rayleigh and Rice fading channels. The NMF model has been empirically shown to be able to provide a good fit to the fading experienced in a wide range of real-world wireless channels.

In the case of an NMF channel, the fading envelope[1)] a has a so-called Nakagami-m distribution, which causes the received CSNR to become *gamma* distributed with PDF

$$p_\gamma(\gamma) = \left(\frac{m}{\bar{\gamma}}\right)^m \frac{\gamma^{m-1}}{\Gamma(m)} \exp\left(-m\frac{\gamma}{\bar{\gamma}}\right), \; \gamma \geq 0. \tag{3}$$

Here, $m \geq 1/2$ is the *Nakagami fading parameter*, restricted in this paper to be an integer, and $\Gamma(m)$ is the Gamma function, equal to $\Gamma(m) = (m - 1)!$. The Nakagami fading parameter can be loosely interpreted as the ratio between received line-of-sight signal power and received signal power from scattering and reflections. In other words, the higher m is, the more line-of-sight power dominates, making for a better channel less prone to deep fades.

3 What to Learn from Information Theory

Information theory tells us that any communication channel is characterized by a maximal information rate, the *channel capacity*, which provides the ultimate upper limit for when reliable transmission is possible [4]. For any given channel with a certain average CSNR, this channel capacity is a constant number, regardless of how strict the demands on bit-error-rate (BER) may be. This number has come to represent the "holy grail" for designers of bandwidth-efficient transmission schemes. It may be stated as the maximal average number of information bits which may be carried per transmitted channel symbol, or – if multiplied by the maximum (Nyquist) channel symbol rate – as the maximal average number of information bits per second. For any information rate we may choose to specify below the channel capacity, theory tells us that there always exist a transmitter and a receiver which are able to communicate at this rate with arbitrarily low BER. For rates above the capacity, we can be sure that no such transmitter-receiver pairs exist.

Of course, finding the optimal transmitter-receiver pair may be a very complex task even if the specified information rate lies below capacity. Also, even if found, the capacity-achieving transmitter-receiver pair may be extremely complex to implement. Most current-day communication systems thus operate at rates significantly less than the theoretical capacity, although the advent of modern error-control coding schemes has decreased the gap between theory and practice considerably for some channels, and almost closed it in some special cases.

---

1) *We suppress the time dependence from now on for the sake of notational simplicity.*

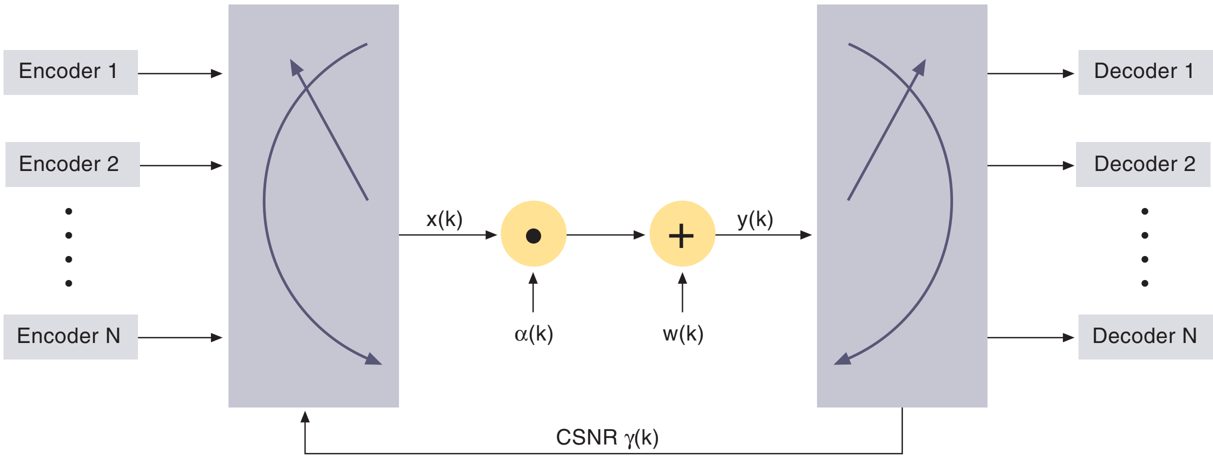


Figure 1 Rate-adaptive transmission

3.1 Information Theory for Wireless Channels

Claude Shannon developed the general theory behind channel capacity and derived famous formulas for some important channel models, most notably the AWGN channel [4]. However, for wireless channel models, closed-form capacity formulas, and the design of efficient transmission schemes approaching the capacity, have remained elusive – until the last 5 years or so.

In [5] Goldsmith and Varaiya derived the channel capacity of a single-user flat-fading channel with arbitrary fading distribution, when perfect *channel state information* (CSI) is available both at the transmitter and the receiver. The CSI needed by the transmitter is the CSNR γ. For a general flat-fading channel where the CSNR has PDF $p_\gamma(\gamma)$, the capacity is given by

$$C = B \int_{\text{all}\gamma} \log_2(1+\gamma) p_\gamma(\gamma) d\gamma \quad \text{[bits/s]} \quad (4)$$

when fixed average transmit power is assumed.

Alouini and Goldsmith [6] used this result to determine a closed-form expression for the capacity per unit bandwidth, or *maximum average spectral efficiency* (MASE), of an NMF channel

$$\frac{C}{B} = \frac{e^{m/\bar{\gamma}}}{\ln(2)} \sum_{k=0}^{m-1} \left(\frac{m}{\bar{\gamma}}\right)^k \Gamma\left(-k,\ \frac{m}{\bar{\gamma}}\right) \text{[bits/s/Hz]}, \quad (5)$$

where $\Gamma(\cdot,\cdot)$ is the *complementary incomplete Gamma function* [7]. This function is commonly available in numerical software.

To any channel capacity corresponds some specific transmission scheme, the so-called *capacity-achieving* transmission scheme for the channel under discussion. This is the scheme that must be used if we are to actually fully exploit the capacity. Goldsmith and Varaiya [5] showed that the channel capacity of an arbitrary flat-fading channel with constant average transmit power can be approached using a certain *adaptive* transmission scheme. This very fact constitutes the main motivation behind the study of rate-adaptive coding and modulation. To achieve capacity, the signal constellation size, and hence the information rate, must be *continuously* updated according to the channel quality, as measured by the CSNR γ. The transmission rate should be high when the CSNR is high, decreasing smoothly as the CSNR decreases. A block diagram depicting the main principles of a rate-adaptive scheme is given in Figure 1. Observe that the active encoder-decoder pair is determined by the CSNR at the receiver.

3.2 Practical Modifications

The theoretically optimal scheme described in the previous section must in practice be approximated by schemes using *discrete* updating of the information rate. One must then choose a particular set of discrete channel signal constellations, e.g. quadrature amplitude modulation (QAM) constellations of varying size [8]. For coded transmission, one must also choose a particular set of channel (error control) codes. The transmitter will then switch between signal constellations (and, in the case of coded transmission, channel codes) of varying size/rate at discrete time instances. At any given time, the transmitter chooses symbols from the largest constellation meeting the BER requirements with the available CSI, thus ensuring maximum spectral efficiency for the given acceptable BER.

There are different kinds of adaptive coding schemes [9] – [17]. Hereafter, we shall use the term *adaptive coded modulation* (ACM) as a collective term for all variants. As mentioned earlier, the CSNR determines the channel state at a given time. For any discrete ACM scheme, the CSNR must be periodically estimated at the receiver end. We assume that this estimation can be done without error. The set $[0, \infty)$ of possible CSNR values is divided into $N + 1$ non-overlapping intervals or "quantization bins". At any given time the estimated CSNR will fall in one of these bins, and the associated bin index $n \in \{0, ..., N\}$, is sent to the transmitter via a feedback channel, which is assumed error free (see Figure 1). The fading is assumed slow enough that the CSNR can be viewed as constant over the time used to estimate and transmit the CSI. The CSI used by the transmitter can therefore be assumed to be correct at all times.[2)]

When N is large, the CSNR is approximately constant within each bin, and the channel can be approximated by an ordinary *AWGN channel* for each n. Assume that $\gamma \in [0, \gamma_1)$ in bin 0, $\gamma \in [\gamma_1, \gamma_2)$ in bin 1, ..., $\gamma \in [\gamma_N, \infty)$ in bin N. Also assume that the BER must never exceed a target maximum BER_0. Hence, when $\gamma \in [\gamma_n, \gamma_{n+1})$ the ACM scheme may use a code designed to achieve a BER $\leq BER_0$ on an AWGN channel of CSNR $\geq \gamma_n$. We will show how to determine the bin boundaries $\{\gamma_n\}$ in the next section.

An ACM scheme stops transmitting when the CSNR γ falls in the bin $[0, \gamma_1)$ simply because the channel quality is too bad to successfully transmit any information with the available codes. When $\gamma < \gamma_1$, the ACM scheme experiences an *outage* during which information must be buffered at the transmitter end.

4 ACM Analysis and Design

In this section, we describe techniques for designing and analyzing ACM schemes. It is assumed that a set of N transmitter-receiver pairs, denoted as transmitter-receiver pair $n = 1, ..., N$, is available. Transmitter n has a rate of R_n information bits per second, such that $R_1 < R_2 < ... < R_N$. Transmitter-receiver pair n is to be used when γ falls in the CSNR interval $[\gamma_n, \gamma_{n+1})$.

4.1 Average Spectral Efficiency

The *average spectral efficiency* (ASE) of a transmission scheme is equal to $\bar{R} / B$ [bits/s/Hz] where $\bar{R}$ [bits/s] is the average information rate. For a system which adaptively switches between N channel codes, $\bar{R} / B$ is thus the expected value of the individual codes' spectral efficiency with respect to the probability distribution $\{P(\gamma_n \leq \gamma < \gamma_{n+1})\}_{n=1}^{N}$

$$\frac{\bar{R}}{B} = \sum_{n=1}^{N} \frac{R_n}{B} \cdot P(\gamma_n \leq \gamma < \gamma_{n+1}) \text{ [bits/s/Hz]}. \quad (6)$$

We see that we need to determine the bin boundaries $\{\gamma_n\}$ that ensure BER $\leq BER_0$ in order to find the probabilities and thus the ASE. To do this, we invoke the following fact:

When a given code, say code n, is operating on an AWGN channel of CSNR γ, the BER-CSNR relationship when the CSNR is varied turns out to be well modelled by an expression of the form [17]:

$$BER \approx a_n \cdot \exp\left(-\frac{b_n \gamma}{M_n}\right). \quad (7)$$

Here, M_n is the number of signal constellation points and a_n and b_n are constants depending on the code's weight distribution. These constants can be found for any given code by least-squares curve fitting of data from AWGN channel simulations to (7). The curve fitting must be done separately for each individual code in the set. Assuming that this has been done, and using (7) with equality for a given target BER_0, the bin boundaries are found as

$$\gamma_n = \frac{M_n K_n}{b_n}, \; n = 1, 2, ..., N \quad (8)$$
$$\gamma_{N+1} = \infty$$

where $K_n = -\ln(BER_0)\, a_n$. Manipulations then give

$$P(\gamma_n \leq \gamma < \gamma_{n+1}) = \frac{\Gamma\left(m, \frac{m\gamma_n}{\bar{\gamma}}\right) - \Gamma\left(m, \frac{m\gamma_{n+1}}{\bar{\gamma}}\right)}{(m-1)!} \quad (9)$$

Inserting this result in (6), we may explicitly compute the ASE of the overall scheme for any specific set of codes (see [17] for a detailed example).

4.2 The Choice of Codes

In general, it is necessary to implement N encoder-decoder pairs to realize an ACM scheme. However, when the ACM scheme utilizes QAM modulation in conjunction with a type of error control codes called trellis codes [18] – [22], a *single* encoder-decoder pair exists that can execute the encoding and decoding of all N codes [12, 17]. We therefore concentrate on ACM with QAM trellis codes in the following.

A trellis encoder consists of a binary encoder followed by a bit converter mapping the coded bits to a QAM constellation. Consider the encoding of code n in the ACM scheme. At every L^{th}

2) *In some situations this assumption breaks down, e.g. in the case of high terminal speeds (leading to high Doppler shifts and small fading correlation). The assumption is also less accurate the higher the carrier frequency used, a fact which will influence future wireless systems.*

channel time index, $k = L \cdot t$, $t = 0, 1, 2, ...$, the binary encoder takes as input $L \cdot i_n - 1$ information bits and outputs $L \cdot i_n$ coded bits, $L \in \{1, 2, 3, ...\}$. Every block of coded bits is subsequently mapped by the bit converter to a sequence of L channel symbols, each taken from a QAM constellation of $M_n = 2^{i_n}$ signal points. The maximum spectral efficiency of this code can be seen [17] to be $R_n / B = i_n - 1/L$ [bits/s/Hz].

The L two-dimensional QAM symbols generated at each time index k can be viewed as *one* $2L$-dimensional symbol, and for this reason the generated code is said to be a $2L$-dimensional trellis code. The ACM schemes described in [12, 13, 16] utilize sets of two-dimensional ($L = 1$) trellis codes, while the ACM scheme described by Hole and Øien [17] may also utilize sets of multidimensional ($L > 1$) trellis codes, i.e. in practice codes with dimensions 4, 6 and 8. Multidimensional trellis codes are of particular interest since some such codes offer a significantly better performance/complexity tradeoff than two-dimensional codes [21].

5 ACM in Microcellular Networks

Most researchers have considered ACM in single-user communications systems. However, wireless systems must support a large number of users. The authors [23] have analyzed the performance of ACM in an urban microcellular network with a large number of active users, where each individual wireless link is modelled as a narrowband fading channel.

Specifically, we have investigated the use of ACM in outdoor urban microcellular networks of the "Manhattan" type [3]. The principles are transferable to other cellular network topologies as well, with relatively minor modifications to the mathematical techniques used to model the network and to assess the resulting system performance. The situation described here thus serves as an illustrative example of the kind of modelling techniques that must be used and what performance to expect.

Our Manhattan network model consists of uniformly spaced quadratic cells, denoted *microcells* because their width, D [m], is assumed to be no more than 1000 m. We refer to Figure 2 for an illustration. Each cell has a base station (BS) located at its center. A vehicle-mounted or portable mobile station (MS) may be positioned anywhere in the streets, but not inside any of the buildings. Since urban centers typically contain a large number of active MSs, we assume that the cellular network model is fully loaded, i.e. the cells' communication links are all fully used.

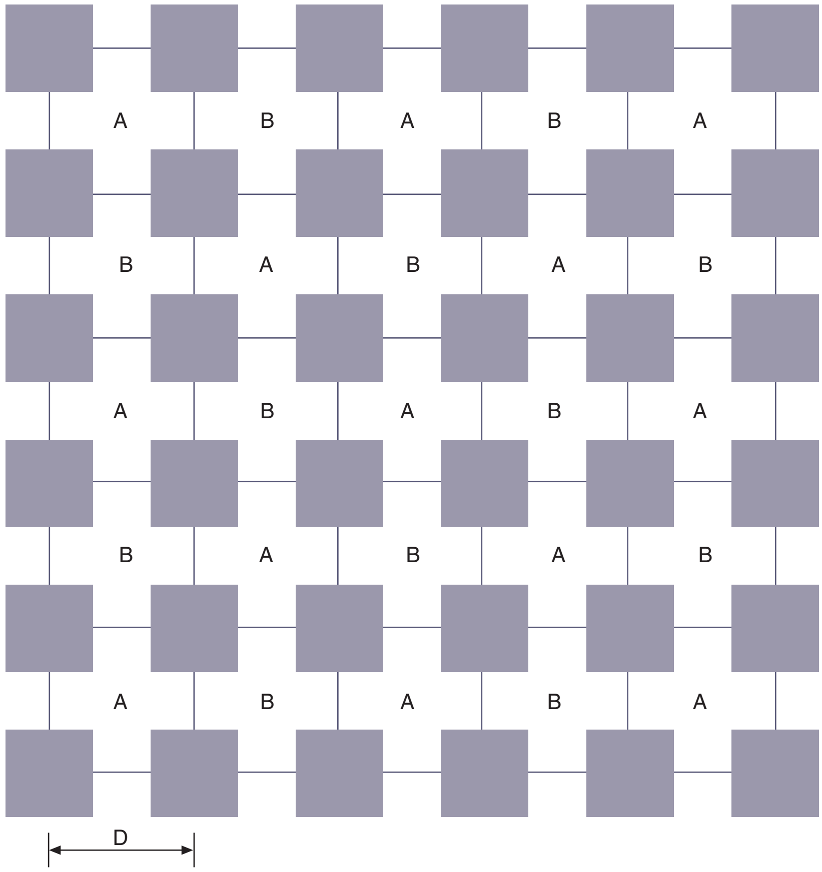


Figure 2 Idealized model of wireless Manhattan network consisting of square cells with length $D \leq 1$ km. The dark squares represent tall buildings and the space between the dark squares represent streets

The same set of carrier frequencies is used in each cell labeled A in Figure 2. A different set of carrier frequencies is used for all cells labeled B. When an MS in a cell accesses the network it is assigned two different carriers, one for the downlink (BS-to-MS) and one for the uplink (MS-to-BS). Since the battery power of the MS is severely limited compared to the available power at the stationary BS, the ability of the MS to transmit data over the uplink is much less than the ability of the BS to transmit over the downlink. We have therefore concentrated on the study of the uplinks in fully loaded networks.

The networks under study might use frequency-division multiple access (FDMA) or time-division multiple access (TDMA). The link signals are degraded by three random phenomena: a) multipath fading, modelled by a Nakagami-m PDF, b) shadowing, a slow variation of the received signal mean, and c) interference from other links.[3)] In addition there is a deterministic power loss, or *path loss*, due to the decay in the intensity of a radio wave propagating in space.

The impairments of the uplinks in a cell are modelled for the worst- and best-case configura-

[3)] *Note that AWGN is disregarded since it is typically less important than the other impairments.*

tions of interfering MSs in neighboring cells (see [23] for a detailed definition of these configurations of interfering MSs). Without going into further detail, it is possible to show that the instantaneous received *carrier-to-interference ratio* (CIR) at the BS is approximately log-normally distributed [23]. Given this model, adaptive coding schemes may be devised, optimized and analyzed.

As a performance measure, we initially use the *average link spectral efficiency* (ALSE), defined as follows

$$\text{ALSE} = \frac{\bar{R}}{B} \text{ [bits/s/Hz]} \tag{10}$$

where $\bar{R}$ [bits/s] is the information rate of the adaptive scheme, averaged over all CIRs, and B [Hz] is the effective bandwidth used per link in a cell. If there are U serviced MSs in a cell and the total available uplink bandwidth in the cell is W [Hz], we have $B = W/U$.

Because of the signal path loss, the expressions obtained for the ALSE in [23] is a function of the distance between the transmitting MS and the receiving BS. The available ALSE is higher the shorter this distance is, because the average CIR is then higher.

The careful network modelling carried out in [23] allows us to show that any spectrally efficient scheme for Manhattan networks must be able to operate over a large range of CIR values. This is the reason why it is possible to obtain a larger spectral efficiency with an adaptive coding scheme than with a non-adaptive coding scheme with the same overall transmission delay.

To determine the cellular layout of a Manhattan network, the available frequency spectrum is first divided over one *cluster* of adjacent cells such that the individual cells in the cluster utilize different sets of carrier frequencies. The complete network is then obtained by deploying many copies of the cell cluster. For the complete network, the *average area spectral efficiency* (AASE) provides a measure of the spectral efficiency. The AASE is measured in [bits/s] / [Hz · m^2] and is defined as

$$\text{AASE} = \frac{\text{ALSE}}{K \cdot A} \text{ [bits/s] / [Hz} \cdot \text{m}^2] \tag{11}$$

where K is the number of cells in the cluster, and A is the area of one single cell. As an example, the network in Figure 2 has $K = 2$. The cell area may be approximated by $A = D^2$.

We have approximated the AASE for ACM in Manhattan networks and shown how the AASE is influenced by the cellular layout. The reader is referred to [23] for a more in-depth discussion, and to the next section for some sample results.

6 A Practical ACM Scheme

This section presents results for a specific ACM scheme based on a set of 8 different 4-dimensional trellis codes utilizing QAM signal constellations with 4, 8, 16, 32, 64, 128, 256, and 512 signal points. The example scheme is based on the International Telecommunication Union's ITU-T V.34 modem standard (see [17] for more details). The target BER used is $\text{BER}_0 = 10^{-3}$.

6.1 Single-user Systems

Figure 3 shows the ASE (6) of the example scheme as a function of the average CSNR $\bar{\gamma}$ in dB for various values of the Nakagami parameter m (m = 1 corresponding to Rayleigh fading). As can be seen, the ASE for a given CSNR is much higher than in today's conventional, non-adaptive coding schemes for wireless links, e.g. GSM with constant spectral efficiency less than one regardless of the average CSNR.

Comparing our results to those of other researchers is not always straightforward. Goldsmith and Chua [12], for example, use adaptive transmit power as well as rate adaption, whereas the transmit power is constant in our system. However, it can be seen that in the special case of Rayleigh fading, our example scheme pro-

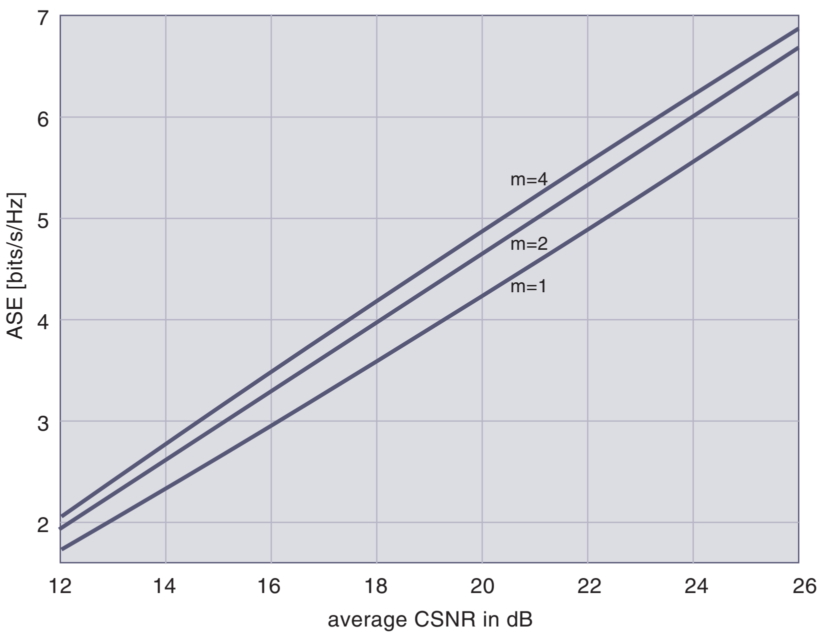


Figure 3 ASE as a function of the average CSNR [dB] for various Nakagami fading parameter values

vides an ASE increase of about 0.5 bits/s/Hz compared to the *uncoded*, constant-power, rate-adaptive QAM scheme of Alouini and Goldsmith [11]. Furthermore, comparing with the expression in (5) we find that the difference between the MASE and ASE of our example scheme is less than 1.9 bits/s/Hz [17].

6.2 Cellular Systems

Consider the link between an MS, called the *desired* MS, and its BS in the Manhattan network model. The other transmitting MSs in the fully-loaded network are the interfering MSs.

The ALSE (10) of the MS-to-BS link is plotted in Figure 4 for the worst-case interference configuration. Here, D [m] is the length of the square cells and r is the distance between the desired MS and its BS. As an example, consider the case when $D = 1000$ m and r is less than 100 m. We have that the ALSE > 7.75 bits/s/Hz, i.e. the ALSE is much larger than in today's systems.

In general, for a fixed length D of the cells, the ALSE decreases with growing distance r between the desired MS and its BS. When we keep r fixed, the ALSE decreases with shrinking D, i.e. smaller cell size. There are large differences in the ALSE values: When $r = 20$ m and $D =$ 1000 m, the ALSE has value 8.5 bits/s/Hz; for $r = 200$ m and $D = 400$ m the ALSE is 1.2 bits/s/Hz.

The *reuse distance*, $\mathcal{R}$ [m], is the minimum distance between two BSs that are within line-of-sight of each other and that use the same set of carrier frequencies for the MS-to-BS links. We define the *normalized reuse distance* as $\mathcal{R}_n = \mathcal{R}/D$. It can be seen that the normalized reuse distance is equal to the cluster size K of a Manhattan network. However, while K is a positive integer, we view the normalized reuse distance as a positive real number since this makes it easier to show how the AASE (11) varies with the cluster size.

The AASE is plotted as a function of the normalized reuse distance $\mathcal{R}_n$ in Figure 5. The distance between the desired MS and its BS is equal to the maximum possible value $r = D/2$. The plot contains four pairs of curves, one pair for each value of $D \in \{400,600,800,1000\}$. For a given pair, the upper curve is the AASE for the best-case interference configuration, while the lower curve is the AASE for the worst-case interference configuration. We observe that as $\mathcal{R}_n$ increases, the difference in AASE between the best- and worst-case interference configurations goes to zero.

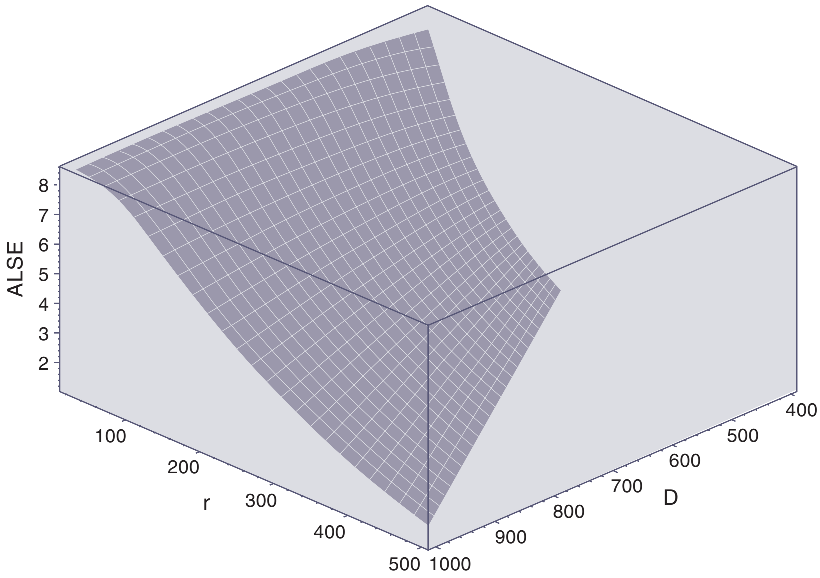


Figure 4 ALSE [bits/s/Hz] as a function of D [m] and r [m]

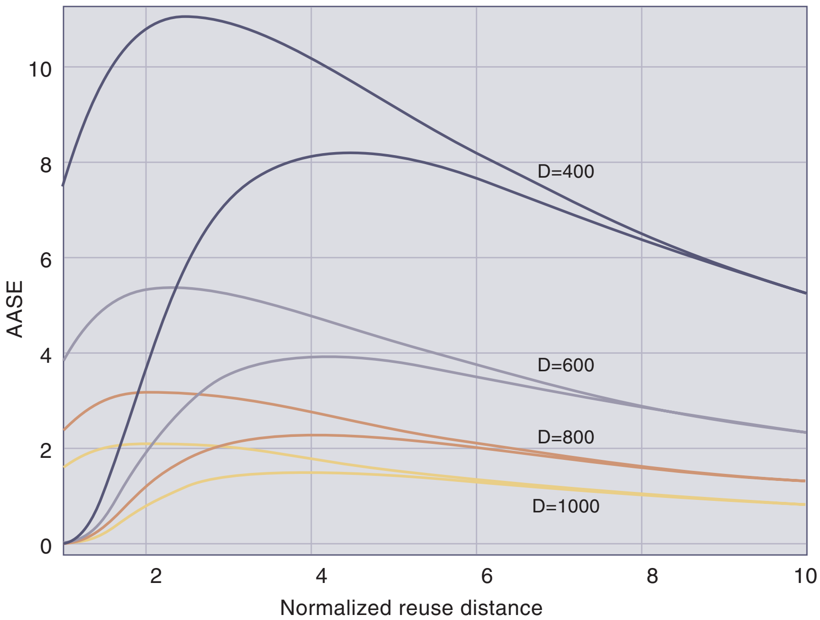


Figure 5 AASE [bits/s/Hz/km²] as a function of the normalized reuse distance $\mathcal{R}_n$ when $r = D/2$ and $D \in \{400, 600, 800, 1000\}$

For the best-case interference configuration, the AASE has maximum value for $\mathcal{R}_n = 2$ when $D \in \{600,800,1000\}$ and $\mathcal{R}_n = 3$ for $D = 400$. Unfortunately, the difference in AASE between the best- and worst-case interference configurations is very large for $\mathcal{R}_n = 2$. Hence, $\mathcal{R}_n = 3$ is a better choice. In other words, for the example ACM scheme, a cell cluster should consist of at least three cells using different sets of carrier frequencies.

7 Conclusion and Further Research

The results presented in this paper indicate that it is possible to obtain a much larger spectral efficiency in wireless systems than achieved in present-day systems. The large increase in spectral efficiency may be obtained by utilizing coding schemes which adapt the information rates according to the quality of the wireless links. This in turn may enable the introduction of novel bandwidth-intensive multimedia services.

7.1 Planned Research

While we have only considered narrowband channels with frequency-flat fading, adaptive coding techniques may also be used on broadband frequency-selective fading channels. *Orthogonal frequency division multiplexing* (OFDM) can be used to decompose a broadband frequency-selective fading channel into a set of parallel narrowband subchannels with flat fading [24]. Adaptive coding techniques may be utilized on each individual narrowband subchannel to achieve broadband OFDM transmission with large total average spectral efficiency.

The success of ACM/OFDM is ultimately dependent on reliable knowledge of the subchannel dynamics at the transmitter. Since the subchannel states have to be estimated at the receiver, there will always be a certain degradation of system performance compared to the idealized case of perfect subchannel knowledge. It is important to minimize this degradation by finding optimal subchannel estimation techniques for realistic channel models.

In November 2000, the authors started a new project with the title "Bandwidth-Efficient and Adaptive Transmission Schemes for Wireless Multimedia Communications" (BEATS). The BEATS project is a part of the IKT-2010 research programme funded by the Norwegian Research Council. The principal objective of this project is to develop bandwidth-efficient, adaptive OFDM schemes for broadband wireless multimedia communications. More information about the project may be found on the web-page http://www.tele.ntnu.no/beats/. As new results become available, they will be posted on this page.

The authors also participate in the joint NTNU-Telenor research project "Turbo Codes, Access, and Network Technology" (TURBAN). Our TURBAN project goal is to determine efficient channel codes for use in adaptive transmission.

References

1 Meyr, H. Algorithm design and system implementation for advanced wireless communications systems. In: *Proc. International Zurich Seminar on Broadband Communications (IZS'2000),* Zurich, 2000.

2 Bose, V, Wetherall, D, Guttag, J. Next century challenges : RadioActive Networks. In: *Proc. ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM'99),* Seattle, WA, 1999, 242–248.

3 Stüber, G L. *Principles of Mobile Communication*. Norwell, MA, Kluwer Academic Publishers, 1996.

4 Cover, T M, Thomas, J A. *Elements of Information Theory*. New York, Wiley, 1991.

5 Goldsmith, A J, Varaiya, P P. Capacity of fading channels with channel side information. *IEEE Trans. Inform. Theory,* 43 (6), 1986–1992, 1997.

6 Alouini, M-S, Goldsmith, A J. Capacity of Nakagami multipath fading channels. In: *Proc. 47th IEEE Vehicular Technology Conference (VTC'97),* Phoenix, AZ, 1997, 358–362.

7 Gradshteyn, I S, Ryzhik, I M. *Table of Integrals, Series, and Products*. San Diego, CA, Academic Press, fifth ed., 1994.

8 Webb, W T, Hanzo, L. *Modern Quadrature Amplitude Modulation*. Graham Lodge, London, Pentech Press, 1994.

9 Webb, W T, Steele, R. Variable rate QAM for mobile radio. *IEEE Trans. Commun.,* 43 (7), 2223–2230, 1995.

10 Goldsmith, A J, Chua, S-G. Variable-rate variable-power MQAM for fading channels. *IEEE Trans. Commun.,* 45 (10), 1218–1230, 1997.

11 Alouini, M-S, Goldsmith, A J. Adaptive M-QAM modulation over Nakagami fading channels. In: *Proc. 6th Communications Theory Mini-Conference (CTMC VI) in conjunction with IEEE Global Communications Conference (GLOBECOM'97),* Phoenix, Arizona, 1997, 218–223.

12 Goldsmith, A J, Chua, S-G. Adaptive coded modulation for fading channels. *IEEE Trans. Commun.,* 46 (5), 595–602, 1998.

13 Lau, V K N, Macleod, M D. Variable rate adaptive trellis coded QAM for high bandwidth efficiency applications in Rayleigh fading channels. In: *Proc. 48th IEEE Vehicular Technology Conference (VTC'98),* Ottawa, Canada, 1998, 348–351.

14 Ue, T et al. Symbol rate and modulation level-controlled adaptive modulation/ TDMA/TDD system for high-bit-rate wireless data transmission. *IEEE Trans. Veh. Technol.,* 47 (4), 1134–1147, 1998.

15 Goeckel, D L. Adaptive coding for time-varying channels using outdated fading estimates. *IEEE Trans. Commun.,* 47 (6), 844–855, 1999.

16 Kim, Y M, Lindsey, W C. Adaptive coded-modulation in slow fading channels. *J. Commun. and Networks,* 1 (2), 99–110, 1999.

17 Hole, K J, Holm, H, Øien, G E. Adaptive multidimensional coded modulation over flat fading channels. *IEEE J. Select. Areas Commun.,* 18 (7), 1153–1158, 2000.

18 Ungerboeck, G. Channel coding with multilevel/phase signals. *IEEE Trans. Inform. Theory,* IT-28 (1), 55–67, 1982.

19 Ungerboeck, G. Trellis-coded modulation with redundant signal sets – Part I: Introduction. *IEEE Commun. Mag.,* 25 (2), 5–11, 1987.

20 Wei, L-F. Trellis-coded modulation with multidimensional constellations. *IEEE Trans. Inform. Theory,* 33 (7), 483–501, 1987.

21 Chouly, A, Sari, H. Six-dimensional trellis-coding with QAM signal sets. *IEEE Trans. Commun.,* 40 (1), 24–33, 1992.

22 Pietrobon, S S, Costello, D J Jr. Trellis coding with multidimensional QAM signal sets. *IEEE Trans. Inform. Theory*, 39 (3), 325–336, 1993.

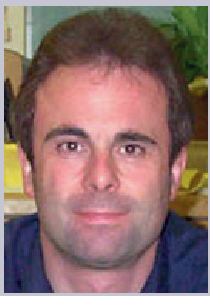
23 Hole, K J, Øien, G E. Spectral efficiency of adaptive coded modulation in urban microcellular networks. To appear in *IEEE Trans. Veh. Technol*. Paper may be obtained at http://www.tele.ntnu.no/beats/

24 Van Nee, R, Prasad, R. *OFDM for Wireless Multimedia Communications*. Norwood, MA, Artech House Publishers, 2000.

The BRAIN Developments for a Broadband Mobile Access Network beyond Third Generation

DAVE WISELY, WERNER MOHR AND JOSEF URBAN

The licenses for third generation mobile systems are currently allocated and auctioned across Europe. Some first UMTS networks will be operational in 2001. At the same time activities are starting to define and to develop the next generation of mobile networks. An important characteristic of these future systems is the user-friendly access to services across heterogeneous networks using different technologies including traditional cellular networks as well as other radio technologies such as Wireless LAN. The European research project BRAIN is contributing to these developments by defining an open architecture for wireless broadband Internet access. This article will provide an overview of major achievements of the BRAIN project during the first year of the project's lifetime.



Dave Wisely (39) graduated in 1982 with a first class degree in physics. He worked at the Atomic Weapons Research Establishment (AWRE) and at Marconi Space and Defence Systems. He obtained a higher degree in optics from Reading University, winning the Pilkington prize for best student, and joined BT Labs in 1988 doing research into micro-optic devices. He then took another MSc in telecommunications, winning a prize for his thesis. Since 1995 Dave Wisely has been working on various aspects of mobile communications and is now head of UMTS research at BT Labs leading a large team looking at IP for UMTS phases I and II, IP mobility and roaming, GPRS evolution, etc.

dave.wisely@bt.com

Werner Mohr (45) obtained his Masters degree in electrical engineering from the University of Hannover in 1981 and his PhD in 1987. From 1987 to 1990 he was senior engineer at the Institute of High-Frequency Technology at the University of Hannover. In 1991 he started work for Siemens AG, Mobile Network Division in Munich. He was active in the RACE-II project ATDMA, ETSI SMG5 for UMTS standardisation, and was manager of the ACTS FRAMES project. Since 1998 he has been Vice-President Pre-Engineering in the Chief Technical Office of the Communication on Air business area of Siemens ICM with responsibility for cross-functional research activities.

werner.mohr@icn.siemens.de

Introduction

The licensing process for the provision of third generation networks is well under way. It is anticipated that around 100 licences will be awarded over the next 12 months. Some first UMTS networks will be operational in 2001 [1].

While 3G networks are rolled out, the research community is already thinking about the next generation. For example, NTT DoCoMo and Hewlett Packard announced recently a joint research effort with the aim to improve multimedia delivery and network applications over fourth generation (4G) wireless broadband networks. Alcatel, Ericsson, Nokia and Siemens decided to create the Wireless World Research Forum (WWRF) [2]. The objective of this new forum is to develop visions on strategic research directions in the wireless field and to identify technical trends for mobile and wireless system technologies. The European research project BRAIN (Broadband Radio Access for IP based Networks) is contributing to this discussion about systems beyond 3G by defining an open architecture for wireless broadband Internet access. The BRAIN project is also actively contributing to standards bodies with contributions to the IETF, ETSI BRAN and the HiperLAN/2 Global Forum delivered or under preparation.

The BRAIN Project

The project is sponsored under the IST (Information Society Technologies) programme of the European Commission [3]. The project consortium consists of the following manufactures, operators, and research institutes: Ericsson, Nokia, Siemens, Sony, BT, France Télécom R&D, NTT DoCoMo, T-Nova, INRIA, King's College London, and the Spanish Internet start-up Agora Systems [4].

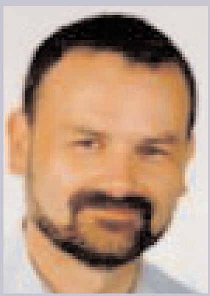
The general objective of the BRAIN project is to propose a system architecture which combines broadband radio access systems such as HIPERLAN/2 – in hot spot areas like railway stations – with UMTS and GSM to enable full coverage of seamless IP based services for users in hot spot areas and on the move. The design of the BRAIN system is optimised across application, network, and air interface layer with regard to QoS and mobility management. This enables the system to provide human-friendly services and applications to the mobile user of the future.

An overview of the most important aspects of the BRAIN work in the three areas of applications and services, the network layer and the air interface is given below.

The BRAIN Business Model and the Support for Adaptable Multimedia Services

The project has adopted a top-down approach to define the architecture and protocols needed for the seamless provision of IP services. In order to define the requirements on the terminal, network and air interface, the project described and analysed several scenarios which illustrate the usage of advanced mobile services [5]. For example, a scenario called 'leisure time' describes a situation in which a user is connected from the same terminal to a private HIPERLAN/2 network of a shopping mall as well as to public cellular networks (Figure 1).

The private WLAN network of the shopping mall offers to their visitors access to high bandwidth applications. Users are offered a range of services related to shopping in the big hall. These include: price and availability checks on goods, news of special offers, restaurant menus and booking and even transport timetables. Within the stores there are further customized services available, such as video-tailoring: based on your stored profile you can see an image of yourself wearing clothes for sale in a tailor's shop.



Josef Urban (42) holds a diploma in computer science from the Technical University of Munich and a Masters degree in the theory of science and logic from the University of Munich. He started his professional career in the central research labs of Siemens in Munich, gaining experience in applying object-oriented software techniques throughout the whole software development cycle. He was later involved in software development projects for the Siemens broadband switching systems and two European ACTS projects. In 1998 Josef Urban joined the European Commission in Brussels as project officer in the DG Information Society. He is now director in the chief technology office of the mobile communications division of Siemens and manages the IST project BRAIN.

josef.urban@icn.siemens.de

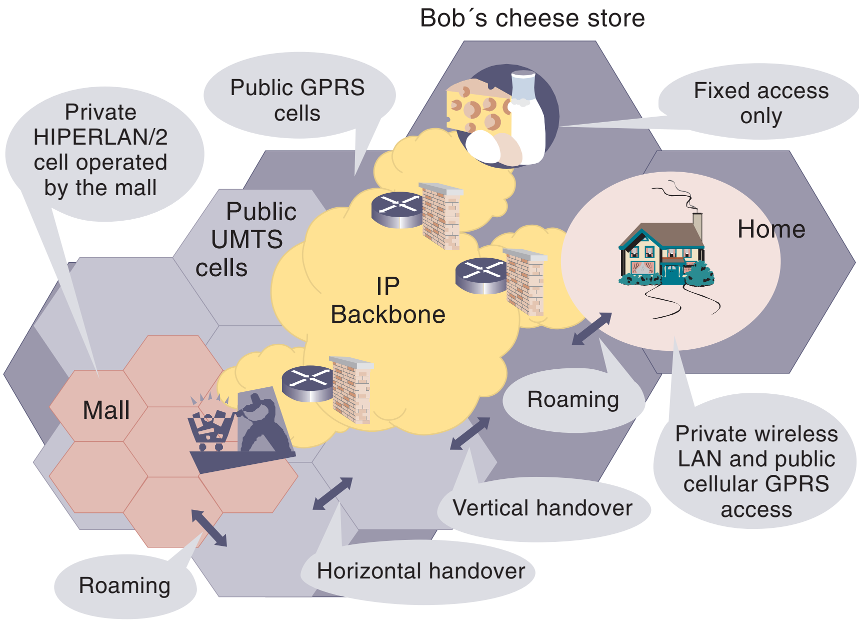


Figure 1 Usage scenario – leisure time

Simultaneously, people can be connected to the Internet or to their home via the public UMTS network. This will allow, for example, to set up a videoconference with a member of the family in order to discuss the purchase of some goods. The connection to the family home can continue to exist also on their way home and may include an automatic vertical handover to another access technology, such as GPRS, if required. The video conference will be adapted to the changed mobile environment and the user will perhaps notice that the video screen has changed from colour to black and white.

Based on this and other usage scenarios the project worked out a business model, which identifies the key actors in the delivery of future multimedia services. These are: Content service providers (provide music etc.); application service providers (host complete applications, e.g. video chat); proxy service providers (provide databases, e.g. user profiles); network service providers (own customers, e.g. Internet Service Providers); network infrastructure providers (access or core network providers); terminal suppliers (supply terminals).

The business model is illustrated in Figure 2 by means of the scenario 'leisure time'. The shopping mall instructed a company called StoreOperators to install and to operate a HIPERLAN/2 network. In this case StoreOperators is acting as an application service provider, infrastructure provider, and proxy service provider. The shopping mall does not want to bother with the technical details of running a network, but wants to be visible as a network service provider by offering a location based service to the customers. This location service is enriched with specific information about shops and restaurants which are close to the current position of a customer. By that, the shops play the role of a content service provider. The shopping mall does not charge the customers for its service, but gets its money from the stores.

Studying usage scenarios and business models, the project felt that the customer is the most uncertain factor in these models. He is already used to the free Internet services and will only be willing to pay higher fees if the services, their content, mobility, better quality of service (QoS) and the simple handling of the provided applications prove to add value. Based on these considerations the project defined a number of requirements on broadband mobile access networks. The requirements considered the most important are in brief:

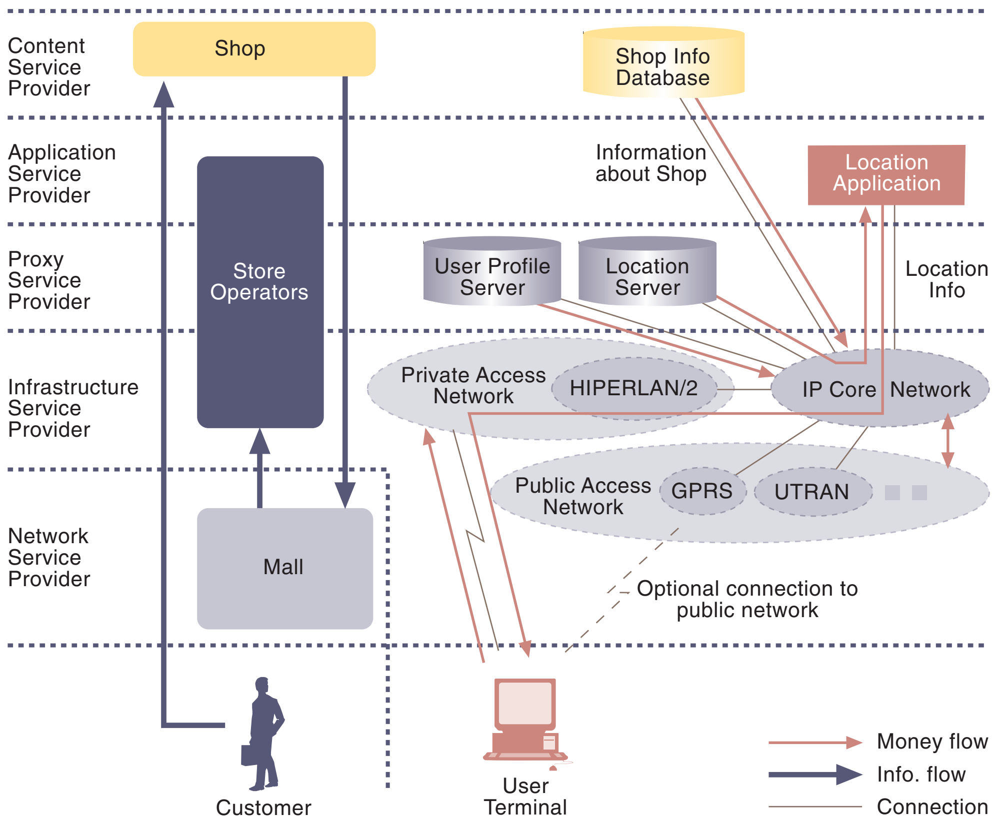
1 Personal mobility requirements
- The need for profiles;
- The need to adapt services to the access link and terminal being used.

2 Network requirements
- The need to support vertical handover (e.g. HIPERLAN/2 to UMTS);
- The need for flexible and modular QoS support;
- The need for autoconfiguration.

3 Billing and accounting requirements
- The need for a single bill.

4 Authentication and security required
- The need for systems to respect confidentiality;
- The need for unified authentication.

Figure 2 Business case study



Terminal Architecture

On the terminal side we have identified the requirement to support QoS in a generic way. Applications will need high functionality programming interfaces (APIs) to allow rapid application development. Some applications will want to be relieved of the need to perform adaptations to the variable QoS nature of wireless links. We have developed the BRAIN End Terminal Architecture (BRENTA) to support applications and to perform terminal QoS adaptation [6]. In the development of the BRENTA two approaches to the way applications will interface with the network were identified.

Firstly we have what has been termed the IETF protocol solution. In this approach the emphasis is on lightweight component protocols – i.e. there is no network API. Applications in the terminal interface directly with these protocols to set up sessions, negotiate QoS with the network and deal with QoS violations. An example application would be Microsoft NetMeeting with plug-ins in supporting SIP [7] (Session Initiation

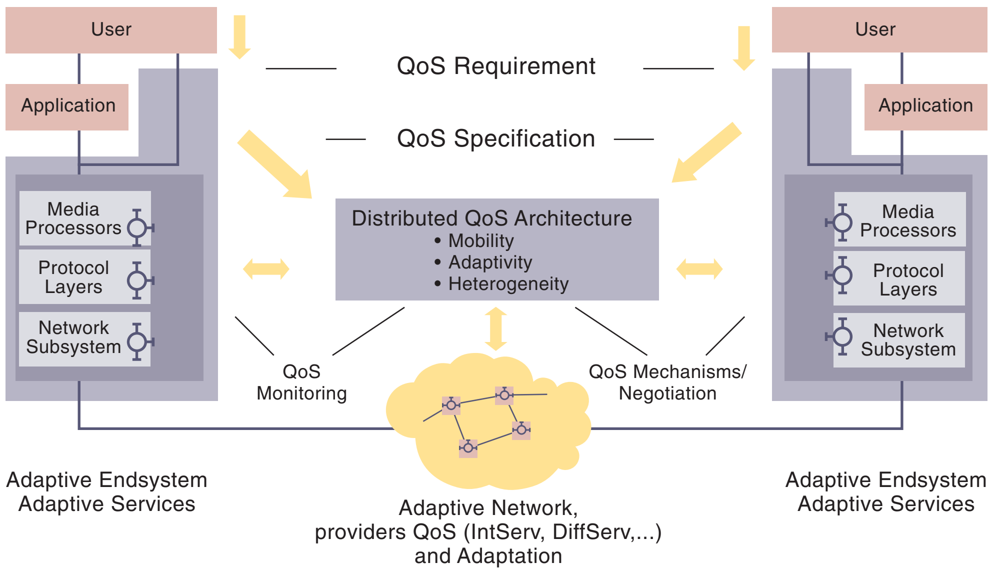


Figure 3 Adaptive QoS framework

Protocol) and RSVP [8] (Resource ReSerVation Protocol).

The other approach has been termed Adaptive QoS Middleware. This consists of both an enhanced end terminal stack as well as QoS brokers, mobility gateways and media filters located within the network – forming a complete distributed architecture for QoS management (Figure 3). Applications are presented with a standard API, rather than having to deal with session and QoS negotiation and violations themselves.

The BRAIN project has recognised the merits of both approaches and developed a modular, component-based architecture that encompasses both. We have designed enhancements to the terminal stack (Figure 4) that provide interfaces to a number of different application types: legacy (type A); those that utilise session protocols (type B); those that can make use of a component API (providing frame grabbers, packetizers ...) (type C); and those that can make use of a full blown QoS broker to deal with all connection issues (type D).

The QoS enabled transport interface, also called Service Interface (SI), allows a complete separation of the BRENTA from a specific network implementation which could be provided by over-provisioning or by more complex protocols such as RSVP. Well-known transport primitives at the SI interface will be enhanced by additional primitives to give applications the facility to use QoS, with a specific focus on the needs in the wireless and mobile domain.

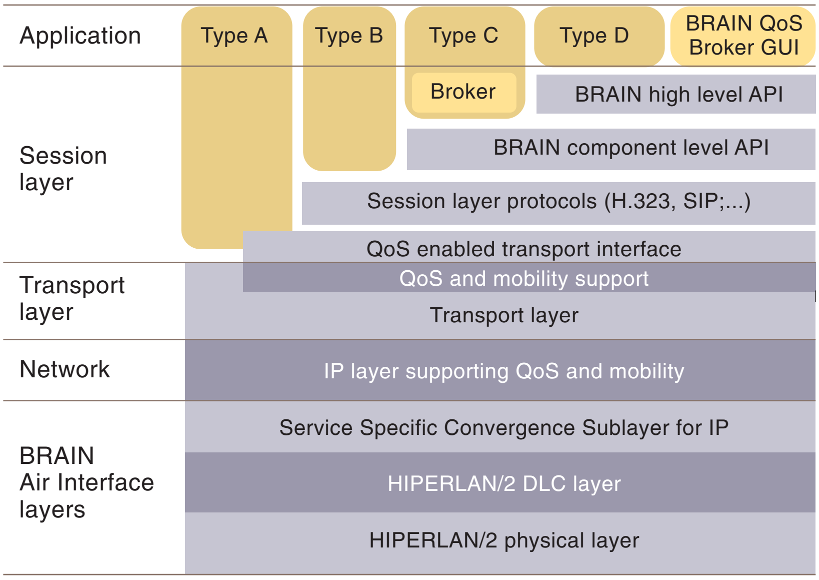


Figure 4 BRAIN End Terminal Stack

The Access Network

The BRAIN Access Network (BAN) (Figure 5) is intended to be deployed over a campus sized area – a city centre, Science Park or airport. Within this domain it will offer terminals Quality of Service (QoS), Terminal Mobility and Vertical Hand-over support. Personal mobility and service creation will largely reside outside of the BAN.

Terminal mobility will be provided by a modified micro-mobility protocol such as MER TORA [9], [10] or HAWAII [11]. A mobile host terminal will acquire an IP address for at least the duration of a session and the routers of the BAN will create per-host entries to allow packets to reach the mobile, wherever it moves within the domain. At present the BRAIN project is evaluating the currently proposed protocols for micro-mobility against an evaluation framework [12] derived from the perceived requirements from the usage scenarios outlined above.

The evaluation framework consists of three parts. Firstly a deconstruction of the protocol design issues involved: fast handover, paging, security, address management and so on. Secondly a classification of micro-mobility protocols against these design issues. Finally an evaluation against requirements such as scalability, robustness and efficiency.

Results present so far ([12], [13], [14]) show that there are only three fundamental approaches to IP micro-mobility and that a per-host routing approach is probably optimal – for BRAIN scenarios.

Another important area of BRAIN research is to produce an interface definition for the layer2/3

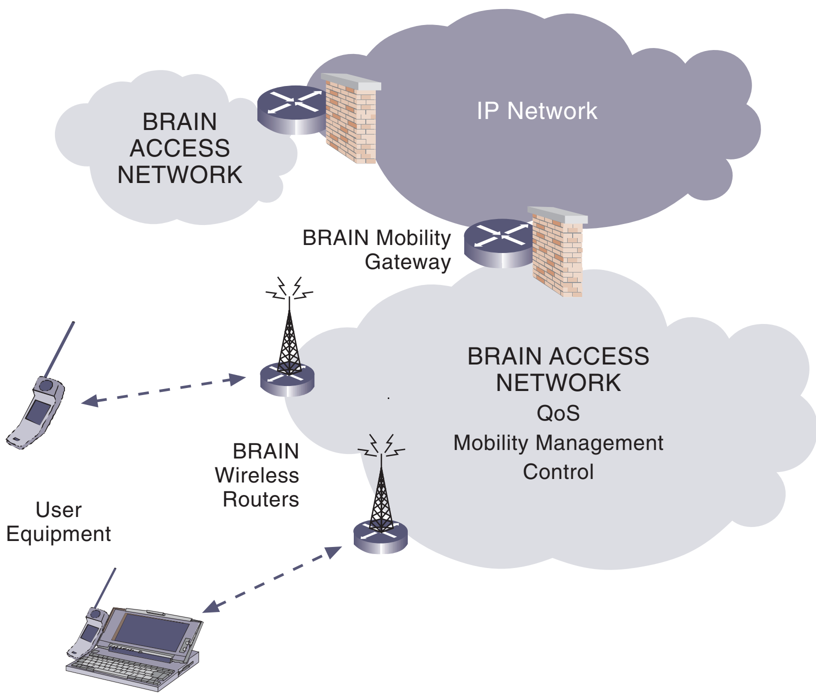


Figure 5 The BRAIN IP access network

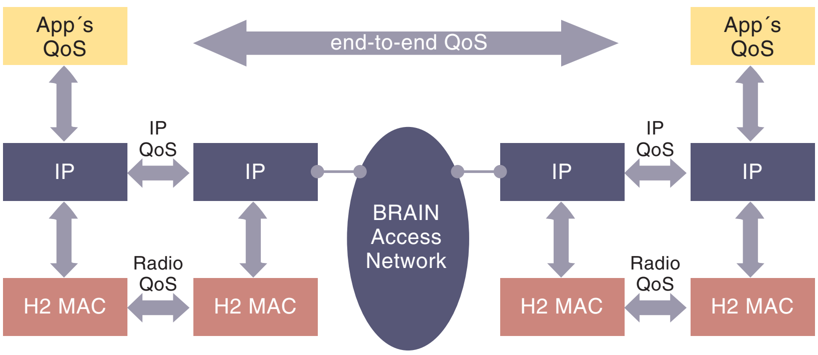
boundary ([15], [16]). This interface will attempt to standardise the way layer 2 access technologies interface to the IP layer – to support hand-over and QoS at both layers in a generic way. It is possible to perform all hand-over, QoS and security at the IP layer but this is inefficient when compared to a close co-operation between layers 2 and 3. However, only by standardising the IP2W (IP to Wireless) interface these efficiencies can really be exploited for a range of access technologies. In the case of HIPERLAN/2 a convergence layer will be written (see below) to implement the interface but, equally, Bluetooth or 802.11 link layers could be supported.

If the mobile host moves to another domain – either supporting another access technology or belonging to another authority – a macro-mobility protocol, such as Mobile IP, could be used to provide seamless handover. One of the research areas of the BRAIN is to look at service adaptation when, for example, a hand-over between HIPERLAN/2 and UMMTS/GPRS takes place. The adaptation will be achieved by co-operation between the terminal software and elements in the network.

In a similar way the BAN will support network QoS. In the same way that the BRAIN has produced an evaluation framework for mobility protocols so it has for QoS protocols. Current QoS schemes under evaluation include: IntServ, IntServ within the BAN only, DiffServ with bounded delay and IntServ over DiffServ. IntServ is useful at the edge of the network where the traffic becomes "lumpy" and a single large multimedia stream can dominate the local flows. Another area of research is how far the mobility management and QoS protocols should be coupled. Tightly coupled protocols tend to be inflexible, uncoupled protocols inefficient, and so we are looking at a loose coupling between the two.

The BAN is also responsible for the radio resource management and admission control parts of QoS.

Figure 6 BRAIN QoS architecture



The Air Interface

BRAIN has chosen a Wireless LAN standard – HIPERLAN/2 – as the basis for its broadband radio interface.

The BRAIN project will define enhancements to the Physical and Data Link Control (DLC) layers as well as define a BRAIN-specific convergence layer in order to make HIPERLAN/2 suitable for IP transport in the usage scenarios we have outlined and, thus, to implement the BRAIN QoS architecture depicted in Figure 6.

Since HIPERLAN/2 will operate in unlicensed spectrum and is intended to provide fixed network QoS, there is a need to optimize the system not only with respect to bandwidth efficiency – as it is usually done for systems designed for licensed deployment – but also with respect to transmission robustness, aiming to maintain the required QoS over the duration of an active link. The combination and joint optimisation of layer 2 scheduling (on DLC level), link adaptation and selected physical layer enhancements such as multiple antenna concepts offer the possibility to maintain the desired QoS for different classes of traffic and, thus, protect especially real-time applications from delays due to re-transmissions and due to unpredictable errors on the wireless link.

Specifically the BRAIN-enhanced HIPERLAN/2 system will support:

- Efficient transport of IP packets for all multimedia applications;
- A QoS service to the IP network layer;
- Network layer mobility management protocols – e.g. by providing paging;
- Hand-over of users to other BRAIN Wireless routers (horizontal hand-over) as well as non-BRAIN networks (vertical hand-over) – with minimum delay/loss of packets;
- Unicast, Multicast and Broadcast services;
- A transparent service to the IP layer.

At the physical layer the BRAIN will look at adaptive antennas, receiver diversity and smart antenna techniques to improve system robustness. In addition turbo codes and adaptive modulation will be analysed as further optimisation for HIPERLAN/2 aiming at an efficient use of spectrum ([17]). Both an increase in system robustness and spectrum efficiency will contribute to improving QoS for IP traffic.

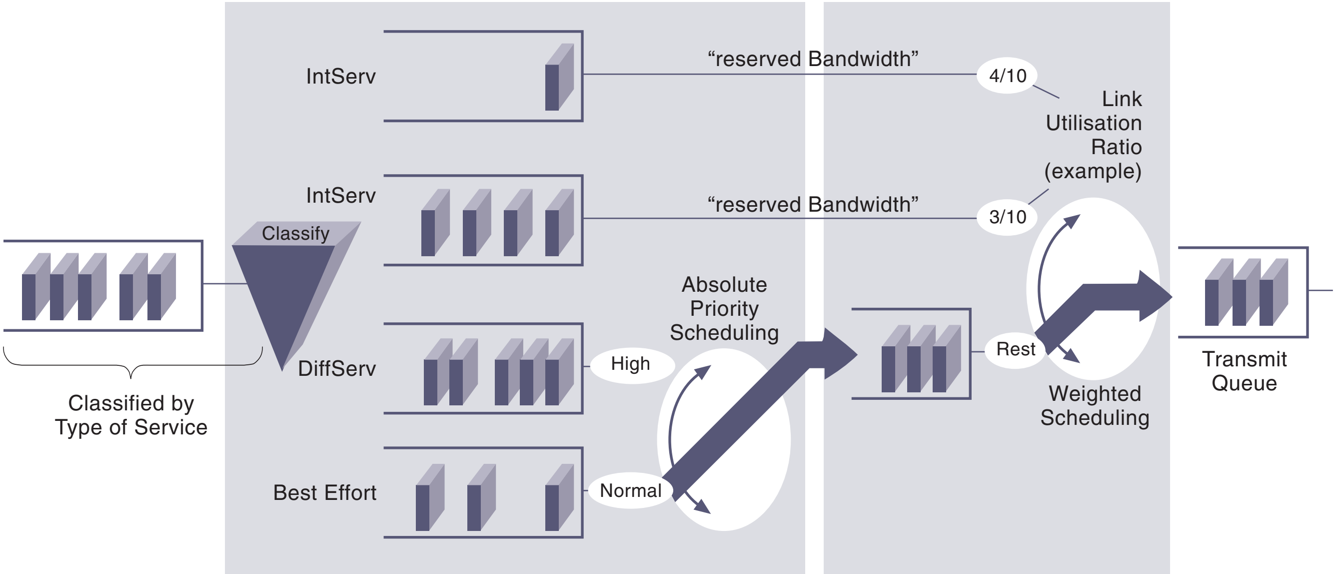


Figure 7 Example implementation of DLC packet scheduling

At the DLC layer dynamic channel allocation and link adaptation algorithms will be implemented – the modulation scheme being matched to the prevailing radio environment and the IP traffic requirements (e.g. QoS for instance in terms of low latency or low loss). The DLC will also feature a packet scheduler (Figure 7) to implement QoS over the radio link, this will take inputs from the interface between the convergence layer and the IP layer to interpret the IntServ and DiffServ messages and implement the appropriate QoS at the DLC layer.

Conclusion

This paper has provided an overview about some major achievements of the BRAIN project during the first year of its lifetime. The overall aim of this period was to design an access system which is optimised across application, network, and air interface layer with regard to QoS and mobility management. The following technical achievements have been highlighted: Usage scenarios and business models described at the beginning of the project have driven the top-down approach of BRAIN. A BRAIN ENd Terminal Architecture (BRENTA) has been designed which will supply an advanced, highly functional interface to applications. A service interface between the IP transport service and the terminal architecture has been specified; this is a clever way of de-coupling the terminal middleware from the underlying IP transport service. There is an explosion of IP mobility management protocols and possible approaches to IP QoS. The project has categorized these protocols, produced an evaluation framework and is currently proposing enhancements to the most promising ones. The IP to wireless interface solves the difficult problem of providing a generic way for IP (layer 3) to interact with a wireless layer 2. QoS and mobility support rely on co-operation between the two layers but this must be done without creating a stovepipe solution. A functional description of an IP service Specific Convergence Sub-Layer for HIPERLAN/2 has been drafted. The IP convergence layer is vital to the efficient QoS-supported transport of IP packets across the air interface. Finally the project has proposed enhancements to HIPERLAN/2 which cover improved spectral efficiency, support for handover and support of QoS.

Acknowledgement

This work has been performed in the framework of the IST project IST-1999-10050 BRAIN, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues from Siemens AG, British Telecommunications PLC, Agora Systems S.A., Ericsson Radio Systems AB, France Télécom – R&D, INRIA, King's College London, Nokia Corporation, NTT DoCoMo, Sony International (Europe) GmbH, and T-Nova Deutsche Telekom Innovationsgesellschaft mbH.

References

1 UMTS-Forum. *IMT-2000 Licensing Conditions & Status.* (2001, March 6) [online]. – URL: http://www.umts-forum.org/licensing.html

2 *Wireless World Research Forum.* (2001, March 6) [online]. – URL: http://www.ist-wsi.org

3 *IST programme.* (2001, March 6) [online]. – URL: http://www.cordis.lu/ist

4 *BRAIN homepage*. (2001, March 6) [online]. – URL: http://www.ist-brain.org

5 Robles, T, Umeda, N, Neureiter, G. Service scenarios and business model in BRAIN project. *BRAIN workshop in London*, Nov. 2000. (2001, March 6) [online]. – URL: http://www.ist-brain.org/

6 Kassler, A et al. An open end system architecture for adaptable multimedia services with QoS support. *BRAIN workshop in London*, Nov. 2000. (2001, March 6) [online]. – URL: http://www.ist-brain.org/

7 IETF. 1999. *Session Initiation Protocol*. (RFC 2543.)

8 IETF. 1997. *Resource ReSerVation Protocol (RSVP) – Version 1, Functional Specification.* (RFC 2205.)

9 IETF. 1998. *Temporally-Ordered Routing Algorithm (TORA) Version 1 – Functional Specification*. Work in Progress (draft-ietf-manet-tora-spec-01).

10 IETF. 2000. *Edge Mobility Architecture*. Work in Progress (draft-oneill-ema-02.txt).

11 IETF. 1999. *IP micro-mobility support using HAWAII*. Work in Progress (draft-ietf-mobileip-hawaii-01).

12 David, K et al. A first evaluation of IP based network architectures. In: *Proceedings of the IST Mobile Summit 2000,* October 2000.

13 Keszei, C et al. Mobility management and QoS in BRAIN access networks. *BRAIN workshop in London*, Nov. 2000. [online] – URL: http://www.ist-brain.org/

14 IETF. 2001. *Mobility Related Terminology*. Work in progress (draft-manner-seamoby-terms-00.txt).

15 Laukkanen, A et al. IP to Wireless Convergence Interface. *BRAIN workshop in London*, Nov. 2000. [online] – URL: http://www.ist-brain.org/

16 Bonjour, S et al. IP convergence layer for HIPERLAN/2. *BRAIN workshop in London*, Nov. 2000. [online] – URL: http://www.ist-brain.org/

17 Bolinth, E et al. BRAIN enhancements for the HIPERLAN/2 Air Interface to support QoS in Wireless Communications. In: *Proceedings of the IST Mobile Summit 2000,* October 2000.

Enabling Wireless Connections using Bluetooth

JOAKIM PERSSON AND JAAP HAARTSEN

A new wireless technology called *Bluetooth* has recently emerged as the *de facto* standard for wireless ad-hoc connectivity of devices within a relatively short distance from each other. The technology is supported by a large number of leading companies in the telecom, computer, and consumer areas. This paper discusses the basic functionality of this system, both from a technical standpoint as well as from a user perspective. Furthermore, we will also look ahead and see how some usage scenarios and their particular demands are likely to influence the technology, leading to the evolution of the system.

Joakim Persson (35) received his MSc in Computer Engineering and his PhD in Information Theory from Lund University, Sweden in 1990 and 1996, respectively. He joined the research department at Ericsson Mobile Communications AB in 1996, and from 1999 he is technical manager for the New Technology section within this department. He has been working with Bluetooth since 1997, and is one of the key contributors to the baseband specification. He is now playing an active role in the work within Bluetooth SIG that looks into the future evolution of the system.

Joakim.Persson@ecs.ericsson.se

Jaap C. Haartsen (38) joined Ericsson Mobile Communications in 1991 and has since worked at sites in RTP, USA and in Lund, Sweden in the area of wireless technology. In Sweden he laid the foundations for the Bluetooth radio concept. He is currently located in Emmen, the Netherlands, where he holds the position of Expert, Wireless Systems. He has recently been appointed adjunct professor at the Twente University of Technology, in the area of Mobile Radio Communications. He earned his MSc and PhD degrees in electrical engineering from the Delft University of Technology, the Netherlands. He has authored numerous papers and holds over 30 patents.

Jaap.Haartsen@eln.ericsson.se

1 Background

In 1994, a few people at Ericsson Mobile Communication sat down to discuss a particular problem regarding how to connect a laptop to a mobile phone. The outcome of the discussion was to opt for a wireless radio link. This simple application was the origin of what is now known as Bluetooth. Since then, the system has evolved considerably. The simple point-to-point connection that originally was the intended use for Bluetooth now also includes network functionality. Many interesting user applications have emerged which were not envisioned from the beginning, for instance background synchronisation (see Figure 1). The special needs of Bluetooth hosts running on batteries has been addressed with certain low power modes in the baseband and link manager protocols.

Certainly, one of the most important aspects of any communication system is the user experience and usability. History is full of examples where systems provide great technologies but have failed commercially due to insufficient usability. In the design of Bluetooth, a large effort has been put into such matters. For instance, to ensure interoperability between Bluetooth enabled devices from different manufacturers, a quite extensive set of rules cutting through the entire communication protocol stack, have been created for different usage scenarios. These rules are referred to as *profiles*. Furthermore, a certification procedure is being developed to ensure that Bluetooth equipment really complies to the specification, both for the radio interface and for the different profiles.

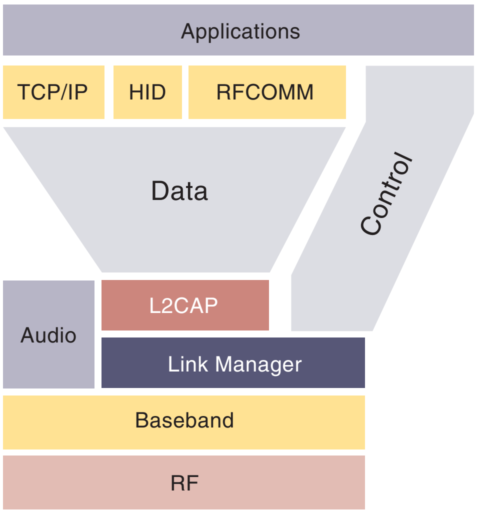
2 Technology

In this chapter we will give an overview of the technical aspects of the Bluetooth technology. We start with the physical layer and walk through the different layers defined above this. The Bluetooth protocol stack is visualized in Figure 2. The layers L2CAP and below form Bluetooth-specific protocols. L2CAP is a sort of adaptation layer between the Bluetooth radio and standard higher layers (like TCP/IP). However, for specific Bluetooth functions, additional higher layers have been defined like the Service Discovery Protocol, the IrDA interoperability (object exchange protocol over Bluetooth), RFCOMM (serial port emulation over L2CAP), Telephony control protocol, and interoperability requirements for Bluetooth as a WAP bearer. Figure 2 does not reveal all these protocols.

Clearly, we can only cover the most essential parts of the Bluetooth system in the given space. The complete specification [1] is obtainable from the URL http://www.bluetooth.com.

Figure 1 Synchronisation of laptop and PDA using Bluetooth and the cellular network

Figure 2 The Bluetooth protocol stack



2.1 Overview

Bluetooth is a radio technology to connect devices, without using wires, at a relatively short distance. The most common mode allows units to communicate within a distance of 10 m from each other. A high power mode is also defined for which the expected distance in free line of sight is 100 m. The radio works in the globally available 2.4 GHz *Industrial*, *Scientific*, and *Medical* (ISM) band.

Bluetooth devices connect in a star topology where the central unit is denoted master and the devices connected to this are denoted slaves. The master controls all traffic in the network. Both synchronous (voice) and asynchronous traffic (data) is supported. The data can be symmetric or asymmetric. The available user data rate is up to 433 + 433 kb/s (symmetric) and 723 + 57 kb/s (asymmetric), assuming no retransmissions are needed.

2.2 Radio

Bluetooth operates in an unlicensed band which is free for use all over the world. This ensures that Bluetooth equipped devices can be brought along and used when travelling abroad. Due to the unknown state of noise and interference in the band, it was not possible to design the Bluetooth radio interface for a known interference situation (in contrast to cellular systems such as GSM and UMTS, where the band is licensed and interference to a certain extent can be controlled). For instance, the IEEE 802.11 WLAN (and its derivatives) operates in the same band, and domestic microwave ovens heating food may leak a substantial amount of electromagnetic radiation here as well. The Bluetooth air interface has been specifically designed to accommodate for varying interference characteristics by incorporating *frequency hopping* (FH) spread spectrum techniques.

In the used ISM band there are 79 channels defined, each being 1 MHz wide. During communication, the devices spread with 1600 hops/s covering all channels according to a pseudo-random sequence. Binary GFSK modulation is used with a signalling rate of 1 Msymbols/s, which gives the radio signal an instantaneous –20 dB bandwidth of 1 MHz.

The allowable output power in the 2.4 GHz ISM band is restricted by different regulatory bodies in different countries. For Bluetooth transmission, the applicable paragraphs in ETSI 300 328 and FCC part 15 must be considered. Three different power classes (denoted class 1, 2, and 3 devices) have been defined for Bluetooth radio transmitters. They correspond to 100 mW (20 dBm), 2.5 mW (4 dBm), and 1 mW (0 dBm) output power, respectively. A power control functionality is required for class 1 Bluetooth devices transmitting over 0 dBm. This is to prevent unnecessary interference to other radio transceivers operating in the neighbourhood and in this ISM band.

In the design of the radio interface, much effort has been made to facilitate simple (low-cost) ASIC implementations of the radio hardware with none or only a few external components. For instance, the relaxed requirements on the noise figure (23 dB) allows a margin for substrate noise so that a low-current LNA can be used. The transmit/receive turn-around time (220 ms) is large enough to support a single synthesiser solution. The phase noise requirement (–89 dBc/ Hz at 500 kHz) is easy enough to allow for an integrated VCO. On the other hand, the requirements on the adjacent channel interference suppression are rather strict to ensure that system performance is satisfactory also in quite severely interfered environments. For further details on implementation issues, see [2].

In principle, the radio functionality is controlled by a state machine which usually is denoted baseband (or link) controller. The behaviour of the state machine is controlled by the link manager (described in Figure 3), which in its turn is controlled by the application running on the host.

2.3 Baseband

Bluetooth uses *time division duplex* (TDD) communication. The time axis is divided into 625 ms long slots. For each slot, a different hop carrier is used, resulting in a nominal hop rate of 1600 hops/s. The master starts transmissions in even slots, while a slave can only start transmitting in odd slots. A Bluetooth master unit and all slaves connected to it constitute a *piconet*. Only master/slave traffic is possible. For slave/slave traffic, either the master has to act as a relay between the slaves, or the slaves set up a piconet of their own (in which one of the slaves now becomes the master). The master decides which slave gets access to the piconet channel by addressing it. After having received something from the master, the addressed slave can return information in the following time slot. To give a slave a possibility to initiate communication, the master must poll the slave periodically with a maximal interval that was agreed upon when the slave joined the piconet.

When several piconets exist simultaneously in range of each other, they constitute a *scatternet*. From a system point of view, the aggregate throughput is increased by creating many piconets that occasionally interfere with each other, rather than having one or a few large piconets where interference between piconets is non-existent or very small. A unit (slave or master) can participate in more than one piconet at a time by utilising time sharing between the different piconets (note that by definition, a unit can only be master in one of the piconet(s) it belongs to). The process of sharing time between piconets is referred to as *interpiconet* scheduling. Similarly, interpiconet communication occurs when information flows between two piconets. Three sample topologies are depicted in Figure 4.

Each Bluetooth device obtains a unique, factory pre-set, 48 bit address, denoted BD_ADDR. This address is used when setting up a connection to a device. This procedure is denoted *paging*. If, however, a device does not know which other units are in range, it must find out this first through the *inquiry* procedure. Bluetooth devices that receive an inquiry message respond with their BD_ADDR and some timing information. Using this information, the inquirer can then page an arbitrary device that has responded. When the connection is made, the piconet master will give the slave a three bit active member address, AM_ADDR, that is used for subsequent addressing within the piconet. The three address bits restrict the number of active slaves to seven, since the all-zero AM_ADDR is reserved for broadcast messages.

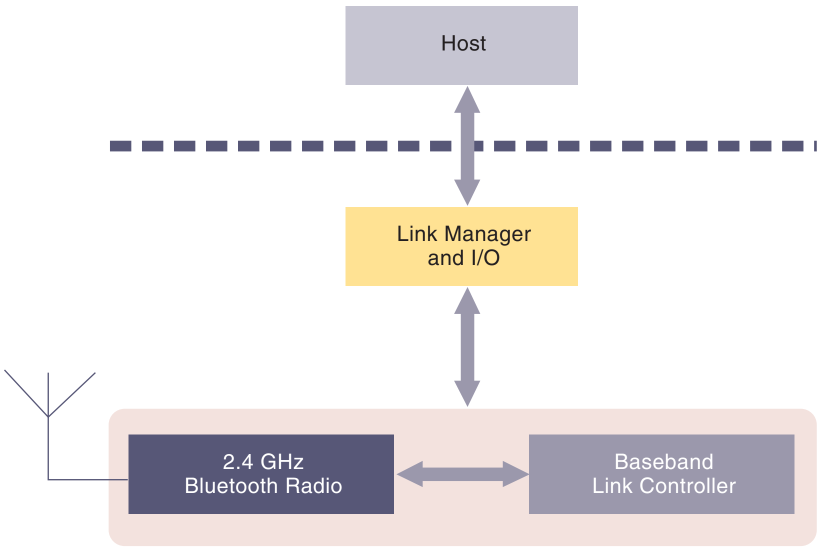


Figure 3 Functional blocks in the Bluetooth system

A Bluetooth packet consists of a 72 bit access code, a 54 bit packet header, followed by the payload data. The access code serves two purposes: it is used to fine-tune the receiver in order to remove DC-offset, and the receiver can time synchronise to align time slots with the transmitter. These operations are necessary due to relative drift between the crystals in the involved units. The packet header contains information about the recipient and what packet type the payload is. Information used by the ARQ protocol is also included there.

In the Bluetooth standard, different packet types have been defined. Packets that are typically used for data traffic contain cyclic redundancy check (CRC) bits to check the payload for errors. The payload length is typically 240 bits. If conditions are favourable (i.e. when the number of bit errors due to noise is relatively small) longer packets can be used with a payload length up to 2745 bits. The short packets can be sent within one time slot. Longer packets can occupy several (up to 5) time slots. Since the frequency is not changed during the transmission of a packet, the effective hop rate decreases when long packets are used. Packets typically used for

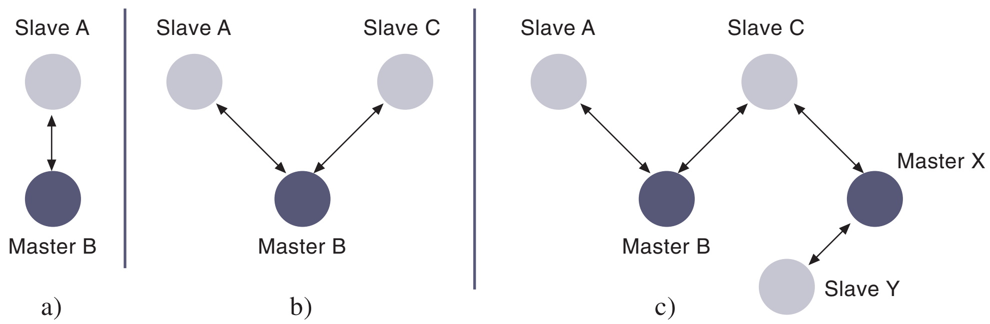


Figure 4 Different topologies for Bluetooth piconets; a) single slave operation, b) multi-slave operation, and c) scatternet with interpiconet communication

real-time traffic like voice do not contain a CRC field, since retransmission is not appropriate. A robust voice coding scheme is used which can tolerate many errors. This Continuous Variable Slope Delta modulation scheme follows the speech wave form by sampling the wave form and only transferring the difference between consecutive samples. Both data and voice packet may include coding bits for improved robustness.

To support both asynchronous traffic like file transfer or web browsing, and synchronous traffic like voice, two types of links with their corresponding packets have been defined. The asynchronous-connection-less (ACL) link is a point-to-multipoint link between the master and the slaves. ACL packets can be sent on any time slot. The slave address (AM_ADDR) is used to deliver the packet to the proper slave. The synchronous-connection-oriented (SCO) link is a point-to-point link between a master and a single slave (though the master can support several SCO links to several slaves). On the SCO link, SCO packets are transmitted at a fixed interval. Time slots are reserved at a fixed interval to provide quality of service for the real-time traffic.

To reduce power consumption in low-rate applications, several low-power modes are defined. In the *sniff* mode, the communication between master and slave only occurs during one TX and RX slot every N slots in which N is the sniff interval. In this case, the slave only has to scan (read sniff) the channel every N slots. In the *hold* mode, the entire communication between master and slave is suspended for a specified hold window. After the hold window expires, the slave wakes up again and communication can proceed. Finally, in the *park* mode, the master broadcasts a low-duty cycle beacon signal to which parked slaves are locked. The slave resides in an inactive state but is synchronised to the master. The beacon signal only contains broadcast information to all parked slaves, but can be used to move a parked slave from the inactive park state to the active state where direct communications between master and slave can be achieved.

For further details regarding the baseband functionality, see [3].

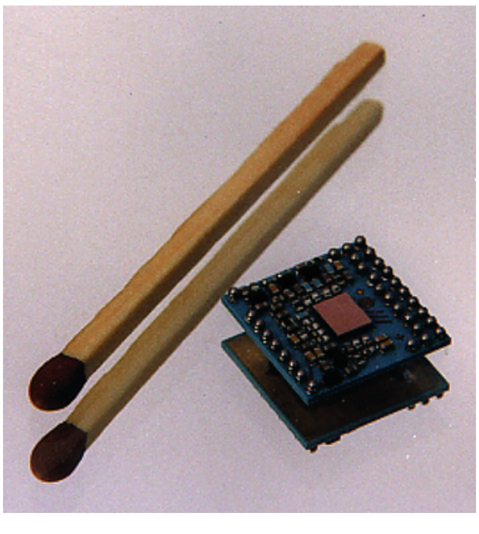


Figure 5 A Bluetooth radio module is fairly small

2.4 LMP

In general, the link manager (LM) is responsible for establishing and supervising the connections and logical links. For initial connection establishment (page and scan routines), only the baseband is involved. This keeps power consumption down in the standby mode. Once the master and slave are synchronized and indeed an FH channel has been established, the link manager protocols are activated. First, authentication procedures are carried out to prevent unauthorized usage of the connected devices. The link manager protocol also plays a role in the key distribution and encryption routine, see also section 3.3. After initial setup and authentication, the LMP takes care of the setup of SCO and/or ACL connections depending on the applications. During the connection, the LMP is involved in link supervision (i.e. selection of specific packets depending on the link quality and the required quality of service), but also takes care of low-power modes like *hold*, *sniff* and *park*. The link manager in the master is responsible for scheduling the traffic flows to and from the different slaves. Finally, the LMP is used to configure any (optional) parameters in the Bluetooth transceivers both at the start and during the connection.

2.5 L2CAP

The *Logical Link Control and Adaptation Protocol* (L2CAP) takes care of the multiplexing of different services, the segmentation and re-assembly of (datagram) packets, for example IP packets, and handles quality of service issues. The L2CAP forms an interface layer between Bluetooth-independent higher-layer protocols, and the Bluetooth-specific LMP layer. During segmentation, the higher-layer traffic is mapped onto baseband packets which are scheduled for transmission by the link manager. Received baseband packets are re-assembled by the L2CAP and relayed to the higher layers.

2.6 HCI

In many cases the Bluetooth functionality of the lower layers of the specification will be implemented into a module that can be fitted onto many different kinds of equipment. The *Host Control Interface* (HCI) forms the interface to the host (e.g. a PC or a cell phone). It provides higher layers with means for accessing some of the hardware status and control registers of the baseband controller and link manager of both the local and the remote Bluetooth devices. To accomplish this, an HCI driver residing in the host communicates over a physical bus (e.g. USB or PC card) with the HCI firmware residing in the local Bluetooth module (see Figure 5).

The HCI commands can be classified into three types. *Link Control* commands are used to directly control the links to other Bluetooth devices. For instance, these messages instruct the LM to establish contact with other devices or perform inquiries. It can also be used to require the other device to authenticate its identity. *Link Policy* commands affect the behaviour of the local and the remote LM. These commands determine the way piconets and scatternets are established and maintained (e.g. using the sniff and hold modes). Finally, the *Host Controller and Baseband* commands provide the host access to some registers of the local Bluetooth device. Examples of this are key maintenance, flushing the transmit buffers, and setting various timer values. The execution time of HCI commands are not defined, and it also varies between different commands. Therefore, the host controller sends specific *Command Complete* events back to the host upon completion of an HCI command.

2.7 Profiles

Interoperability is *the* key issue for Bluetooth. The profiles serve to define a selection of messages and procedures from the Bluetooth SIG specifications and gives an unambiguous description of the air interface for the specified service(s) and use case(s). Thus, devices conforming to a specific profile are able to work together regardless of manufacturer. The conformity to a profile is verified in the certification procedure that each vendor must pass before the Bluetooth brand name can be used for the product.

So far, 13 different profiles have been defined. The most important is the *Generic Access Profile* (GAP) that serves as the foundation for all other profiles. The GAP introduces definitions, recommendations and common requirements related to modes and access procedures that are to be used by the other profiles. It defines how Bluetooth devices in different baseband states (i.e. in idle modes or in connection modes) are to behave so that communication links and channels can be set up and maintained. It also defines procedures related to the use of different security levels and makes multi-profile operations possible. Furthermore, the common format requirements for parameters that are accessible on the user interface level are defined in this profile.

The *Service Discovery Access Profile* (SDAP) relates to how an application on one Bluetooth device can find out and locate a particular service that runs on or through a Bluetooth enabled device in the neighbourhood. There are also profiles defined for *Cordless Telephony*, *Intercom* (phone-to-phone), *Headset*, *Dial-up Networking*, *LAN Access* (via PPP), *Fax*, *Serial Port*, *Generic Object Exchange*, *Object Push*, *File Transfer*, and *Synchronisation*.

3 User Scenarios

There seems to be an endless amount of usage scenarios for which Bluetooth can serve as an enabling technology. A few of these are described below.

3.1 The Personal Area Network

When people communicate directly, the distance between them is fairly small. To some extent, Bluetooth communication can be said to mimic this behaviour since for the most common radio devices the range is limited to approximately 10 m; only devices within this distance can be reached. The relatively small physical size of any network that can be formed, justifies the common notion of it as a *Personal Area Network* (PAN). The PAN can be viewed as a sphere or "information bubble" around the user. Bluetooth devices within this bubble can be reached, others cannot be contacted.

The usefulness of a PAN can easily be illustrated with some examples. It should be noted that the trivial point-to-point connection in some sense also constitutes a PAN, as is described in some of the examples below.

- **Three-in-one phone**: At home this phone will connect via a Bluetooth link to an access point connected to the fixed line network. Alternatively, the phone may connect directly to some other Bluetooth phone using an intercom mode. When the user leaves home, the phone works as a conventional cellular phone (GSM or UMTS).

- **Universal remote control**: Your *personal digital assistant* (PDA) or cell phone may serve as a remote control not only in your home for residential audio/video equipment, but also as a programmable key to use for locks in your house and for your car. Moreover, it can be programmed for obtaining temporary access to office buildings and hotels you are visiting.

- **Cordless desktop computer**: Almost all wires connected to the back of a desktop computer can be replaced by Bluetooth links: keyboard, mouse, modem, microphone, webcam, speakers, *et cetera*.

- **Instant postcard**: A Bluetooth enabled digital camera can be used to immediately send a picture as an e-mail attachment to any recipient via a cellular phone. A text message can be supplied through the cellular phone keyboard or through a PDA.

- **Data access point**: Bluetooth equipped devices may serve as access points to an arbitrary backbone network. For instance, at an airport lounge such devices can be installed in a grid so that all seats are covered by at least one Bluetooth cell. These will give the users access to the airport *local area network* (LAN) that is connected to the Internet. Similarly, small office buildings and hotels can offer wireless access to their backbone LAN.

- **Hidden computing**: Ubiquitous computing means that small computers, dedicated to some specific task, are present almost everywhere without us even noticing them. Scattered all over a room, one of the largest problem is how to achieve the communication that is necessary for this scenario to work. Bluetooth is a very attractive candidate for this. A simpler variant of this theme is the so-called "briefcase trick". As a user disembarks an aeroplane and switches on his cellular phone, the laptop stored in his briefcase automatically synchronises its e-mail inbox through a Bluetooth connection to the cellular phone, which in turn connects to the corporate mail server over air. All this takes place without any specific user interaction.

3.2 Ad-hoc Connectivity

We define *ad-hoc* networking as forming and maintaining networks without any central administration. For many interesting user scenarios, this is a necessary functionality, and Bluetooth may serve as an enabling technology for these scenarios. Clearly, the piconet is by definition a sort of ad-hoc network. As already described, the Bluetooth specification provides means for creating sessions between devices without any *a priori* information about the other devices regarding addresses, clocks, and which services are run on devices in the vicinity. The specification also describes how units can enter and leave existing piconets, which facilitates dynamic network topologies.

The scatternets and interpiconet communication makes things even more interesting. By utilising relaying nodes, it is possible to perform multi-hop communication. A device can communicate with a node out of range using one or more intermediate nodes to reach the destination. However, for this to work it is necessary to develop routing in ad-hoc networks. This is a challenge, since the networks are not static. Users may enter and leave piconets at random time instants, and consequently the routes between source and destination may change frequently. Moreover, the quality of each link may change over time (people passing by, shadowing, fading), so for the entire route it is possible that certain services will encounter problems that are not seen on single-hop connections.

For a more thorough discussion on wireless ad-hoc networking, see [4].

3.3 Security Issues

The need for some security support in Bluetooth was recognised at an early stage of the system development. In general, radio communication is quite easy to listen in to without revealing this to the victim. Even though the pseudo-randomness of the FH channel gives some protection towards a casual eavesdropper (e.g. a person sitting close to you at an airport, following what websites you are reading while waiting in the departure hall), it provides no privacy in a cryptographic sense. For this purpose, a ciphering mechanism has been included in the specification. For some services, it is clear that means for access control is desirable. A Bluetooth headset connected to a mobile phone may serve as an example of this. Clearly, the phone owner would not like an arbitrary headset being able to connect to the mobile, since then someone could make calls using a personal headset while charging the phone owner without their knowledge. Only authorised headsets should be allowed access to the mobile.

To accommodate for security, all units must be able to prove their identity, the BD_ADDR. For this purpose, the *pairing* protocol is a prerequisite. Pairing is done when two units set up a connection for the first time. In this procedure, the units exchange a 128 bit secret key, denoted *link key*. This key is unique[1)] for each link. Thus, a device needs to remember one link key for each Bluetooth device it is paired to. The pairing involves some user interaction since a common pass key is necessary in both devices in order to generate the link key. The length of the pass key is defined to be from 1 to 16 bytes. In its simplest practical form, it can be an arbitrary 4-digit number entered on a keypad. Of course, the longer pass key used, the better security. For sensitive applications that need particular long pass keys, an alternative method more suitable for humans is to exchange it using some key agreement scheme (e.g. Diffie-Hellman).

[1)] *Exceptions to this do exist,* unit keys *and* master keys *are not unique. For details, see [1], chapter 14.*

The link key is used for *authentication*. This is a challenge-response scheme where the verifier sends a 128 bit challenge to which the claimant responds with a 32 bit number that is a hash function of the challenge and the link key. The verifier checks that the received claimant response is the correct one by computing what the true value is. All this is done automatically, and mutual authentication is possible by performing one challenge-response procedure in each direction. Once pairing has been done, the user(s) need not actively do anything on subsequent connections between the devices, since the authentication protocol will recall the link key from memory.

If link security is requested, both devices generate an encryption key after the identities of the units have been confirmed through the authentication procedure. This key (based on the link key, a device address, and a random number) is used for initialising the encryption hardware. The ciphering algorithm is a stream cipher sequence that is added bitwise modulo 2 to the data payload that is sent over air. To accommodate for different export regulations with respect to hardware containing encryption algorithms, the effective length of the encryption key can be restricted in hardware (1 to 16 bytes). Therefore, Bluetooth devices that want to set up a secure link need to negotiate what length to use before the keys are created. Figure 6 depicts the sequence of pairing, link key generation, and encryption key generation.

The security means defined in the Bluetooth specification relates only to the link. There is no end-to-end protocols defined. Data sent to a Bluetooth module is considered as plain-text strings, and delivered in the same format at the receiver module's host interface. The idea is only to protect the air interface from eavesdropping. Moreover, the identification process is based on device authentication, not message authentication. The receiver can be quite sure from which *radio* a particular message comes, but it cannot know which application (or user) that sent the message.

More details on security issues can be found in [5].

4 Future Evolution

The Bluetooth *Special Interest Group* (SIG) has formed working groups in different areas. The purpose of these is to increase the usability of the Bluetooth system by identifying new usage scenarios and create profiles for them. In some cases, the current specification may have certain characteristics that prevent or make implementation more difficult than desirable. To handle that, the specification will be revised regularly.

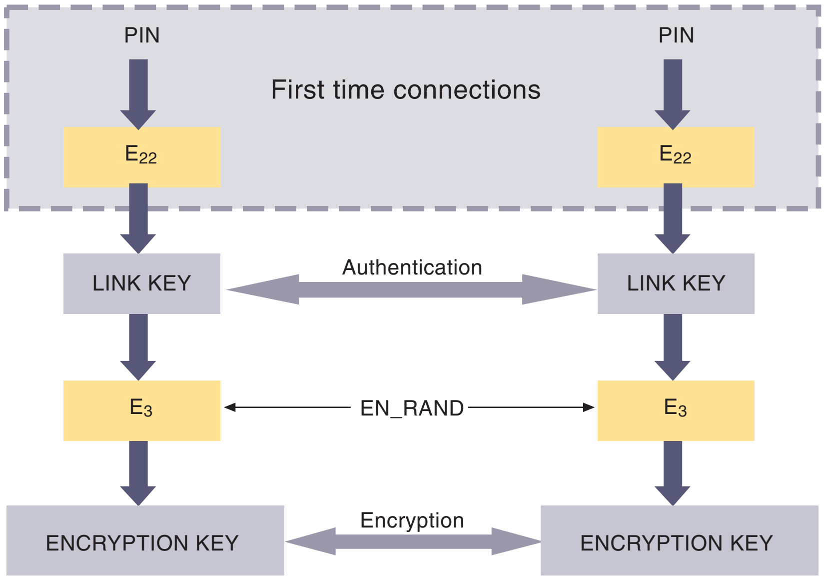


Figure 6 Order of key generation steps (top-to-bottom) for Bluetooth devices

The Bluetooth version 1.1 (successor of 1.0B) will be released relatively soon (early 2001). Below we will refer to versions 1.1, 1.2, ... as Bluetooth 1.x.

4.1 Improving the Current Specification

One of the most common wishes when it comes to improving the current specification, is to make connection set-up time faster. One scenario where this is desirable is walk-by toll gates, for instance at subway entrances. To guarantee that a pedestrian is registered, the Bluetooth link must be set up within 1–2 seconds at most. In the current version, the worst case for inquiry and paging is closer to 10 seconds. Even though this worst case truly is a rather unlikely value, for some applications it is not acceptable that there is a slight possibility for this to happen.

For multimedia services, it is easy to envision cases where both streaming data and conventional data are transferred between nodes. Mixing best effort and isochronous traffic is difficult within the current specification since no prioritisation functionality for messages controlled by the user exists. Because transmit queues cannot be pre-empted or scheduled, it is not possible to guarantee access to the channel at a certain time instant given that a retransmission sequence is ongoing. In the future, different means for controlling meaningful *Quality of Service* (QoS) levels are likely to be added to the specification.

The baseband specification is filled with hooks related to inter-piconet communication. However, to guarantee interoperability it is necessary to specify how to use these also on higher protocol layers. It is likely that profiles for this will be created. To make more advanced *ad-hoc* network visions come true, it is necessary to add support for this. Examples of what is needed is defining how to *efficiently* transfer a piconet with all slaves to a new master (master-slave switch), how to make roaming and hand-over between piconets in a scatternet scenario possible, and how to schedule time between several piconets and negotiating QoS requirements in these.

Many types of services will run over UMTS in the future. It seems advantageous to make sure that at least some of these can run over a Bluetooth link by defining some new profiles. Then, the service need not be terminated in the mobile phone, it can be directly forwarded to some other, more suitable terminal via Bluetooth without the need for costly transcoding. Furthermore, to achieve alignment with 2.5/3G cellular systems, it is desirable to introduce a 2 Mb/s data rate mode which will affect the air interface.

4.2 High Rate Mode

The available user data rate in the current specification is insufficient for some tasks that are likely to be popular in the future. In particular, applications requiring large files to be transferred in a short time frame or streaming high-quality video may put substantially higher requirements on bandwidth than exist in today's specification. For example, archiving pictures taken with a digital camera onto a PC would involve a relatively large file transfer from the camera's memory to the computer's hard disk. The same is true for quickly pushing a complex document file with pictures and/or graphics to a printer. Gaming is another interesting application that can benefit from higher data rates.

To meet the need of bandwidth demanding applications, the Bluetooth SIG has started a work group that will create a high-rate mode for Bluetooth. Some prerequisites for this mode is that it will be backward compatible with the current specification, it shall be compliant with improvements made to the Bluetooth 1.x specification, and high-rate enabled devices shall be world-wide usable according to the current radio specification. The goal is to provide the best throughput possible for cost-effective, small, battery-powered consumer devices, while still maintaining an acceptable interference resistance, a low cost target, and low power consumption.

In order to remain within the basic scope of Bluetooth, i.e. short-range, low-power, small-size and low-cost devices, the ambitions for higher rates are tempered. For the high-rate mode, an increase of a factor of 10 is envisioned providing a gross rate of about 10 Mb/s. It is important to notice that the high-rate mode is not to create an entirely new system, but is merely an extension to the existing Bluetooth 1.x specification. A high-rate enabled Bluetooth device must know how to communicate with old Bluetooth devices. Clearly, there will be an extra cost associated with the high-rate radio that some real low-cost applications can not tolerate, especially since the higher data rate is superfluous for many kinds of services. Thus, the high-rate mode will not be mandatory to implement in the future. Moreover, setting up a high-rate link is done using Bluetooth 1.x functionality. Switching to the high-rate mode is negotiated between the LM in the involved units.

References

1 *Specification of the Bluetooth System, version 1.1*. Bluetooth Special Interest Group.

2 Haartsen, J C, Mattisson, S. Bluetooth – A new low-power radio interface providing short-range connectivity. *IEEE Proceedings of the IEEE,* 88 (10), 1651–1661, 2000.

3 Haartsen, J C. The Bluetooth radio system. *IEEE Personal Communications Magazine*, 7 (1), 28–36, 2000.

4 Frodigh, M, Johansson, P, Larsson, P. Wireless ad hoc networking – The art of networking without a network. *Ericsson Review*, 77 (4), 248–263, 2000.

5 Persson, J, Smeets, B. Bluetooth Security – An Overview. *Information Security Technical Report*, 5 (3), 32–43, 2000. Elsevier Advanced Technology.

HiperLAN/2 – Overview and Evaluation of its MAC Protocol

OLAF RØSTBAKKEN

Olaf Røstbakken (32) received his PhD from Bristol University. At the time of writing this article he was a member of the Personal Communications group at Telenor R&D, Kjeller. Røstbakken is currently working for Inmarsat Ltd., London, and has previously been employed as an RF development engineer at NEC Technologies (UK). His research interests include antennas, propagation, MAC protocols and general radio access issues.

Olaf_Rostbakken@inmarsat.com

1 Introduction

The development in data and telecommunication during the last decades has by any standard been formidable, but it is only recently with the introduction of multimedia communication that these two "worlds" have started to "merge". This has also been the case in wireless communication, where the mobile systems have traditionally been concerned with voice communication and wireless network systems have mainly been used for non-real time data communication. However, this divide is soon to become less distinct with the introduction of UMTS, which has been particularly designed to handle multimedia applications/services. This trend is shown in Figure 1. UMTS – or more correctly UTRAN, as the access network in UMTS is called – will initially be designed to offer up to 2 Mbit/s indoors, 384 kbit/s outdoors (urban/suburban) and 144 kbit/s in rural environments. It is expected that the increase in data rates together with improved terminal display capabilities will pave the way for a stream of bandwidth hungry services/ applications like video conferencing, video clips, gaming, etc. The users will expect high quality of service, which in turn is related to delay, data rate and bit error rate requirements. As the demand increases more pressure will be put on the access network to transport the data to the user in an efficient manner (the frequency spectrum is a scarce resource and the air link operates in a harsh environment). The challenge is that it is difficult to implement an access technology that is efficient in all environments and for all services. Part of the answer to this challenge is software flexible radios, but unfortunately this technology is still immature and it will be some time before full-fledged software radio technology is common. In the meantime a possible solution may be to use several access technologies that are individually optimised (or as much as possible!) for the environment where they are deployed. For instance, with the introduction of bandwidth hungry applications (like video conferencing, browsing, etc.) hotspots (i.e. areas with large capacity requirements) are likely to appear in places like city centres, conference centres, airports, hotels, etc. It will be unreasonable for UTRAN alone to cover the whole communication need in these hotspots, and it will be desirable/beneficial to deploy alternative access technologies that are specifically designed to provide short-range wireless communication. For this purpose Wireless LANs (WLANs) are suitable, as they are designed to transport relatively high bit rates over a short distance. Most of today's WLANs are based on the IEEE 802.11 standard, which can deliver user data rates up to 5–6 Mbit/s and are mainly used in offices to provide a wireless data network extension (e.g. Ethernet). A limiting factor for IEEE 802.11 is that it is primarily designed for transport of non-real time data traffic. IEEE 802.11 equipment does also operate in the 2.4 GHz ISM frequency band, which means that it shares the limited frequency spectrum resources with other systems such as microwave ovens, garage openers and soon Bluetooth. The performance is therefore more susceptible to external disturbances (interference) than systems that use dedicated frequency bands.

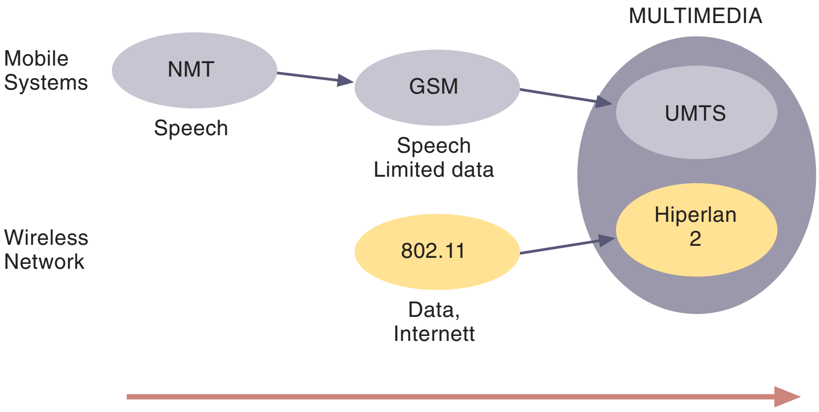


Figure 1 Developments in mobile wireless data networks

Table 1 Allocated HiperLAN bands (in Europe) [1]

Frequency band	RF Power limit	Comments
5.150 GHz – 5.250 GHz	200 mW mean EIRP	Indoor use only
5.250 GHz – 5.350 GHz	200 mW mean EIRP	Indoor use only and implementation of DFS[1] and TPC[2]
5.470 GHz – 5.725 GHz	1 W mean EIRP	Indoor and outdoor use and implementation of DFS and TPC

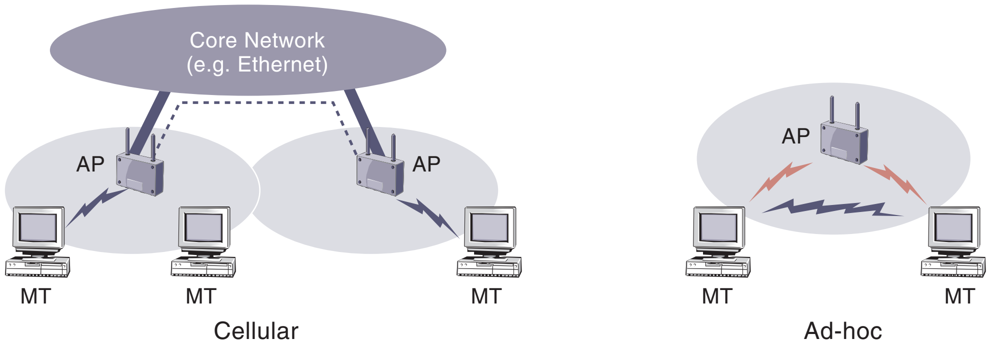


Figure 2 Cellular and ad-hoc network topologies for HiperLAN/2

A new WLAN standard, HiperLAN type2, which does not suffer the shortcomings of the 802.11 in regard to QoS, has recently been published. The remainder of this article will focus on this technology and aims to give an overview of HiperLAN/2 as well as explore its capabilities with particular emphasis on its MAC protocol.

2 HIPERLAN/2 Overview

HIPERLAN/2 is a wireless LAN standard being developed by ETSI, where the core components have been finalised in 2000. The standard will give high data rates (up to 54 Mbit/s on the physical layer) for short range (<150 m) communication and unlike today's WLANs (that are more or less exclusively based on the IEEE 802.11 standard), it will be able to handle multimedia applications. Some of its key features are listed below:

- High user bit rates, typically 20–25 Mbit/s;
- Mechanisms for QoS;
- Link adaptation; i.e. variable data rate in response to radio link quality and QoS requirements (through an adaptive modulation and coding scheme);
- Dynamic time division duplex for support of asymmetrical services;
- Dynamic frequency allocation to minimise interference, can operate in a plug-n-play manner;
- 455 MHz of spectrum allocated in the 5 GHz band (licence exempt, see also Table 1);
- Short range 30 m (non-line-of-sight), 150 m (line-of-sight).

These features make HiperLAN/2 a very flexible access technology that can be used in a variety of environments, including:

- Offices where it will primarily be used as a wireless extension of data networks;
- Home environments where it can be used to distribute data, TV, video and audio signals to various appliances;
- Public access, where HIPERLAN/2 is used as an access technology, for instance, by ISPs and /or UMTS operators.

The HiperLAN/2 standard allows for two types of network topologies (see Figure 2). One is a cellular network topology (or centralised mode) where the access point(s) (AP) is connected to a core network. All traffic (both data and control)

---

1) *Dynamic Frequency Selection*
2) *Transmit Power Control*

has to pass via the AP, even for communication between two MTs associated to the AP. It is assumed that the majority of traffic for the centralised mode is between entities where only one is associated to the AP. The other network topology is called ad-hoc or direct mode, and is used for communication between two MTs associated with the same AP. Here, the user data is transmitted directly between the two MTs and not via the AP. The control information, however, is transmitted via the AP, and the AP still controls the medium access.

The HiperLAN/2 standard covers the two lowest layers of the OSI model (physical and data link control layers) and in addition several core network specific convergence layers (CL) are defined to allow HiperLAN/2 to be used as an access technology for UMTS, Ethernet, IEEE 1394. The protocol stack with the three basic layers; Physical layer (PHY), Data Link Control layer (DLC), and the Network Convergence layer (CL) is shown in Figure 3. Next, descriptions of the main features associated with the different layers are attempted. Other overview articles on HiperLAN/2 can be found in [2], [3] and [4].

3 Description of the Different Layers

3.1 Physical Layer

In broad terms the physical layer receives (or delivers) data (called protocol data unit, PDU) from the MAC layer, then performs scrambling, coding (FEC) and interleaving on the data before multiplexing the data on to a number of orthogonal sub-carriers. This multiplexing process is also called Orthogonal Frequency Division Multiplexing (OFDM) [5], and is a special form of multicarrier modulation. By splitting the high rate data stream into several lower rate parallel data streams the signals are less sensitive to inter-symbol interference (ISI), which can be a problem in multipath environments. Another important feature of the physical layer is that seven different modulation and coding modes can be used on the sub-carriers and are displayed in Table 2. This makes link adaptation possible where the most appropriate PHY mode is used according to the current radio channel and application requirements. Mode 1 (BPSK) represents the most robust mode (lowest error rate for a given C/I) and gives a bit rate of 6 Mbit/s, whereas mode 7 gives the highest bit rate of 54 Mbit/s but with the disadvantage of being more sensitive to C/I. More detailed discussions on the performance of the physical layer can be found in [6] and [7].

The basic parameters of the physical layer are shown in Table 3.

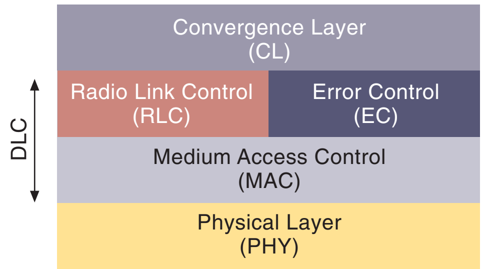


Figure 3 Layer model of the HIPERLAN/2 system

Mode	Modulation	Coding rate R	Nominal bit rate [Mbit/s]
1	BPSK	1/2	6
2	BPSK	3/4	9
3	QPSK	1/2	12
4	QPSK	3/4	18
5	16QAM	9/16	27
6	16QAM	3/4	36
7*	64QAM	3/4	54

Table 2 PHY modes defined for HIPERLAN/2 (optional)*

Parameter	Value
Channel spacing (and system clock)	20 MHz
FFT length	64
Number of used subcarriers	52
Number of data carriers	48
Number of pilot carriers	4
Modulation schemes on subcarriers	BPSK, QPSK, 16 QAM (64 QAM optional)
Demodulation	Coherent
Guard Interval length	800 ns (400 ns optional)
Channel coding	Convolutional code, constraint length 7

Table 3 Basic PHY layer parameters

3.2 DLC Layer

The Data Link Control (DLC) layer is divided into three major entities [4]:

- Medium Access Control (MAC) protocol [8];
- Error Control (EC) protocol [8];
- Radio Link Control (RLC) protocol [9].

3.2.1 MAC Protocol

One of the prime tasks of the MAC layer is to map the data and control information from the higher layers on to transport channels (transport channels are discussed in more detail in section 4.1). These transport channels are then used together with a preamble to construct PDU trains (PHY bursts) that are delivered to and received from the physical layer. This process is based on a centrally controlled TDMA/TDD scheme [8], where the AP (or central controller) controls all the resource allocations inside its cell. The MTs request resources to the AP which in turn decides which MTs are allowed to transmit. The MAC frame has a fixed duration of 2 ms and consists of five different phases (see Figure 4); Broadcast (BC), Down Link (DL), Direct Link (DiL), Up Link (UL) and Random Access (RA). The length of the individual phases can vary inside the MAC frame and is set by the AP depending on the traffic situation.

- **Broadcast (BC) phase**
  The BC phase consists of three transport channels, broadcast (BCH), frame control (FCH) and access feedback (ACH) channels (see Figure 6). The broadcast channel (BCH, downlink only) contains control information that is sent in every MAC frame and reaches all the MTs. The BCH provides information about transmission power levels, starting point and length of the FCH and the RCH, wake-up indicator, and identifiers for identifying both the HIPERLAN/2 network and the AP. The frame control channel (FCH, downlink only) contains an exact description of how resources have been allocated (and thus granted) within the current MAC frame in the DL- and UL-phase and for the RCH. The access feedback channel (ACH, downlink only) conveys information on previous access attempts made in the RCH.

- **Downlink (DL) and Uplink (UL) phases**
  The DL and UL phases carry user specific control information and user data. The user data uses the Long transport CHannels (LCHs, 54 bytes long) and the control information uses Short transport CHannels (SCHs, 9 bytes long) but can also use LCHs. The MTs can use UL phase to request resources for following frames.

- **Direct Link (DiL) phase**
  The DiL phase carries user data traffic (LSCs and SCHs) between MTs without direct involvement of the AP. The AP is indirectly involved by receiving Resource Requests from MTs for these connections and transmitting Resource Grants.

- **Random Access phase (RA)**
  The RA phase carries a number of Random Access Channels (RCHs), which are used by MTs who have no SCHs in the current MAC frame to request resources. The RCH is also used by non-associated MTs to initiate contact with the access point.

3.2.2 Error Control

The EC is responsible for detection and recovery from transmission errors on the radio link and three types of modes are defined:

1. Acknowledged mode uses a selective (SR) ARQ to provide a reliable transmission and a discard mechanism is used for low latency.

2. Unacknowledged mode provides an unreliable, low latency transmission. No feedback channel is generated.

3. Repetition mode is used to repeat LCH PDUs in order to enhance reception. The transmitter can arbitrarily retransmit PDUs, and this mode is typically used to broadcast data.

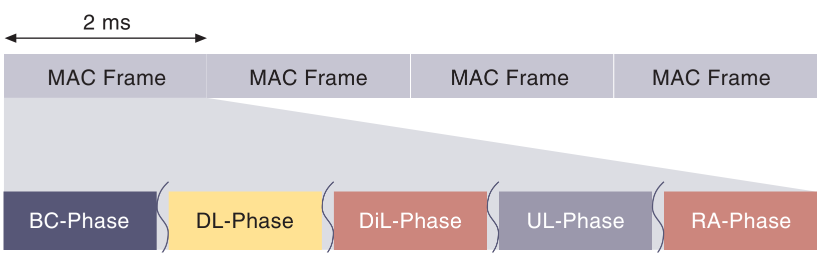
3.2.3 Radio Link Control Protocol

The Radio Link Control (RLC) sublayer is made up of three main control functions. These three entities comprise the DLC control plane for the exchange of signalling messages between the AP and MT [9]:

1. The DLC Connection Control (DCC) is responsible for setup and release of connections.

2. The Association Control Function (ACF) is responsible for the association procedure of an MT to an AP. The tasks of the association control are: association, encryption, authentication, and key exchange. The default encryption scheme is DES (56 bits), optionally triple DES or no encryption can be used. The key

Figure 4 Basic MAC frame structure



exchange is based on the Diffie Hellman procedure.

3 The Radio Resource Control (RRC) is responsible for the surveillance and efficient use of available frequency resources. Important functions of RRC are dynamic frequency selection (DFS), handover and power control. The DFS ensures that HIPERLAN/2 will operate in a plug-and-play manner and frequency planning will not be required. The decision to initiate a handover is primarily done by the MT (forced handover initiated by the AP is also possible) and three types of handover are supported. Sector handover if a sector antenna is deployed (Inter sector/Intra AP), radio handover if the AP consists of several transceivers (InterAPT/Intra AP) and network handover (Inter AP/Intra Network). The network handover involves higher layers, and signalling via the backbone may be needed.

3.3 Convergence Layer

The convergence layer acts as a bridge between higher layer protocols and the DLC Layer. This involves adapting the service requests from the higher layers to that used by the DLC layer, and segmentation/re-assembly of packets to fit the fixed length of DLC packets. Currently, convergence layers for ATM, IEEE1394 and Ethernet have been standardised, and work on a UMTS CL is underway. The convergence layers can be split into two different types, cell based (ATM) [10] and packet based (Ethernet, IEEE1394) [11]. The structure of the packet based convergence layer is displayed in Figure 5, and consists of a common part (with the segmentation/re-assembly) and a service specific part.

4 Performance of HIPERLAN/2 MAC protocol

As mentioned in section 2 the HIPERLAN/2 system is intended to provide multimedia communication in a variety of deployment scenarios. It is therefore of major interest to investigate the overall performance in terms of throughput, delay and QoS support. In order to predict the overall performance, the characteristics of the individual layers need to be evaluated. Already, there is published work considering the performance of the HIPERLAN/2 system [12], [2].
In this section the performance of the MAC protocol for different settings will be simulated.
In essence this is done by calculating the ratio between overhead and user data for the concerned MAC setting. The number of OFDM symbols in each phase of the MAC frame can be found from the following steps. Using the values from Table 3 the OFDM symbol interval t_{OFDM}, can be calculated

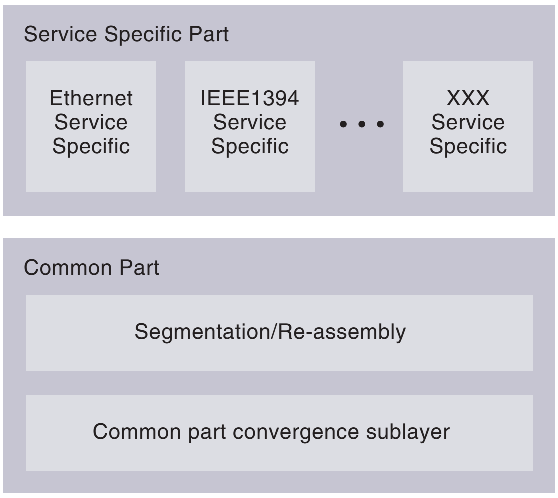


Figure 5 Structure of packet based convergence layer

$$t_{OFDM} = 64 \cdot T_S + T_G = 4\ \mu s \qquad (1)$$

where

T_S = sampling rate = 20 MHz or 50 ns
T_G = guard time = 800 ns

The number of OFDM per frame is then:

$$N_{OFDM} = \frac{t_{frame}}{t_{OFDM}} = \frac{200\ \mu s}{4\ \mu s} = 500 \qquad (2)$$

4.1 PDU Train Structure

It was mentioned in section 3.3 that the phases of the MAC frame were constructed using transport channels. These transport channels are used together with preambles to form Protocol Data Unit trains, or PDU trains, and represents the interface between the DLC protocol and the PHY layer. Three types of preambles are defined [8]:

- Broadcast control channel preamble, P_{BCH}, enables frame synchronisation, automatic gain control, frequency synchronisation and channel estimation. The length of P_{BCH} is 16 µs or 4 OFDM symbols.

- Downlink traffic preamble, P_{DL}, is used for channel estimation only, and is 8 µs long or 2 OFDM symbols.

- Uplink traffic preamble enables channel and frequency estimation. Two types of uplink preambles are defined $P_{UL\text{-}S}$ (short) = 12 µs or 3 OFDM symbols and $P_{UL\text{-}L}$ (long) = 16 µs or 4 OFDM symbols.

The standard specifies 6 types of PDU trains [8]:

1 Broadcast PDU train;
2 FCH and ACH PDU train (multiple antenna elements only);

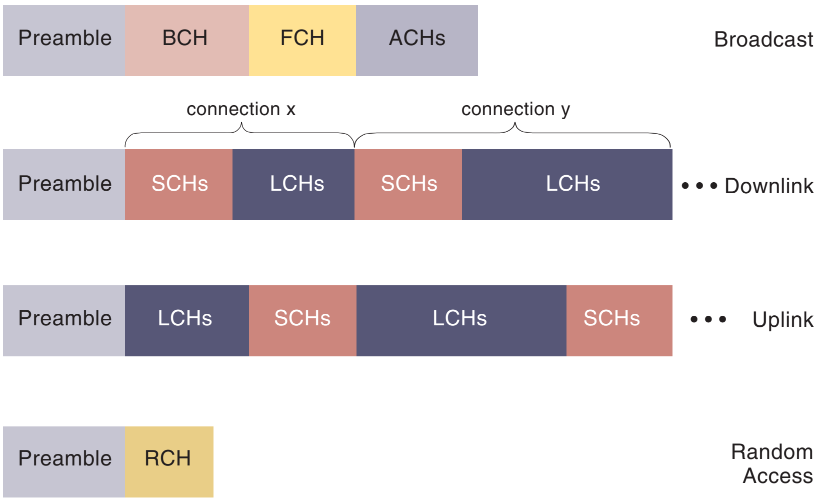


Figure 6 Structure of the PDU trains

3 Downlink PDU train;
4 Uplink PDU train with short preamble;
5 Uplink PDU train with long preamble;
6 Direct Link PDU train.

In this exercise non-sector antennas and centralised mode are considered only (i.e. PDU train 2 and 6 are not considered). The structures of the different PDU trains are shown in Figure 6.

4.1.1 Broadcast PDU Train

The preamble of the BCH is 4 OFDM symbols (16 µs). $P_{BCH} = 4$.

The BCH has a fixed length of 15 bytes, and uses always PHY mode 1, i.e. 5 OFDM symbols:

$$L_{BCH} = \frac{15}{3} = 5 \tag{3}$$

The FCH is built on fixed size IE (Information Elements) blocks. The AP determines the number of blocks. Each IE block contains three information elements with a length of 8 octets and a CRC of length 24 bits. The FCH length equals:

$$L_{FCH} = \left\lceil \left\lceil \frac{2 \cdot n \cdot 8}{24} \right\rceil \cdot \frac{216}{N_{DBPS(FCH)}} \right\rceil \tag{4}$$

where

N_{DBPS} is data bits per OFDM symbol
n is the number of active connections (duplex).

The minimum length of the L_{FCH} for non-sector antennas is 9 OFDM symbols (36 µs). The ACH has a fixed length of 9 bytes (uses always PHY 1 mode)

$$L_{ACH} = \frac{9}{3} = 3 \tag{5}$$

The overall size of the broadcast train is:

$$L_{BPT} = P_{BCH} + L_{BCH} + L_{FCH} + L_{ACH} \tag{6}$$

4.1.2 Downlink PDU Train

The downlink and uplink PDU trains consist of Short Channel (SCH) and Long Channel (LCH) PDUs. The SCH PDU is 9 bytes long and carries DLC control information such as ARQ acknowledgements and resource request for next frame (uplink only). The LCH PDU is 54 bytes long, where 48 bytes are allocated to the payload and the rest is used for DLC header.

The preamble of the Downlink PDU train is 2 OFDM symbols. MTs might have a number of SCHs and LCHs per frame. The size of the downlink PDU train excluding the LCH PDUs is

$$L_{DPT} = k \cdot \left(P_{DL} + \sum_{m=1}^{k} \left\lceil \frac{j_m \cdot 72}{N_{DBPS(SCH)}} \right\rceil \right) \tag{7}$$

where

k = number of active users
J_m = number of SCH PDUs for user m.

In the simulations the SCH (i.e. ARQ protocols) in the downlink is omitted. The size of the downlink PDU train excluding the LCH PDUs will then be:

$$L_{DPT} = k \cdot P_{DL} \tag{8}$$

4.1.3 Uplink PDU Train

The preamble for an uplink PDU train is 3 (short) or 4 (long) OFDM symbols. Just as for the downlink train MTs multiple of SCHs and LCHs per frame. Notice that the order of the LCH and SCH has been exchanged compared to the downlink train.

The number of RCH slots r determines the RCH size. Each slot has the length SCH (PHY 1) with uplink preamble.

The size of uplink PDU train excluding LCH PDUs is

$$L_{UPT} = k \cdot \left(P_{UL} + \sum_{m=1}^{k} \left\lceil \frac{j_m \cdot 72}{N_{DBPS(SCH)}} \right\rceil \right) + r \cdot (3 + P_{UL}) \tag{9}$$

In the simulations, one SCH is used per active connections for resource request purposes. The corresponding size of uplink PDU train excluding LCH PDUs will be:

$$L_{UPT} = k \cdot \left(P_{UL} + \left\lceil \frac{72 \cdot n}{N_{DBPS(SCH)}} \right\rceil \right) + r \cdot (3 + P_{UL}) \quad (10)$$

The total size of the overhead per MAC frame, L_{OH}, will equal the sum overheads of the PDU trains.

$$L_{OH} = L_{BPT} + L_{DPT} + L_{UPT} \quad (11)$$

The number of OFDM symbols available for LCH, L_{LCH}, will be the total number of OFDM symbols, N_{OFDM}, minus the overhead, L_{OH}.

$$L_{LCH} = N_{OFDM} - L_{BPT} \quad (12)$$

The data throughput in bits/s will be calculated as the number of LCH PDUs in a MAC frame multiplied by the number of data bits per LCH PDU and divided by the frame length.

$$Throughput = \left\lfloor \frac{L_{LCH}}{\left\lceil \frac{432}{N_{DBPS(LCH)}} \right\rceil} \right\rfloor \cdot \frac{N_{data}}{t_{frame}} \quad (13)$$

where

N_{data} = number of data bits in LCH = 384 bits (48 bytes).

4.2 Simulation

In this exercise the throughput of a HIPERLAN/2 system for three different PHY modes is evaluated using equation (13). These simulations will show the attainable performance of MAC protocol for a varying number of connections and terminals per MAC frame. The properties of the three modes are displayed in Table 4.

	mode 1	mode 2	mode 3
$N_{DBPS(FCH)}$	144	24	24
$N_{DBPS(SCH)}$	144	24	24
$N_{DBPS(LCH)}$	144	144	24

Table 4 PHY modes considered in simulation

In the first scenario the number of active terminals is varied while assuming one active connection per terminal (n = 1). Figure 7(a) shows the throughput of the three modes. It is apparent that the throughput depends strongly on the PHY mode. Another factor influencing the throughput is the size of the overhead. The size of the overheads L_{BPT}, L_{DPT} and L_{UPT} are shown in (b). It can be seen that the L_{UPT} is the predominant overhead for all three modes in this scenario, with a larger increase for mode 2/3 (7 OFDM symbols per MT) than for mode 1 (5 OFDM symbols per MT). The larger size of the L_{UPT} for mode 2 than for mode 1 is therefore responsible for a steeper decline in throughput for mode 1.

In the second scenario the number of MTs is set to one (k = 1), while the number of active connections, n, is varied. The throughputs of the three modes are shown in Figure 8(a). Comparing Figure 8(a) and Figure 7(a), it can be seen that the throughput for modes 2 and 3 are similar in both scenarios, whereas for mode 1 the throughput was considerably higher in the second scenario. The higher throughput in the second scenario for mode 1 can be explained by comparing the overhead characteristics in Figure 7(b) and Figure 8(b). In the first scenario the changes in overhead for mode 1 were dictated by L_{UPT} and L_{DPT}. In the second scenario (for mode

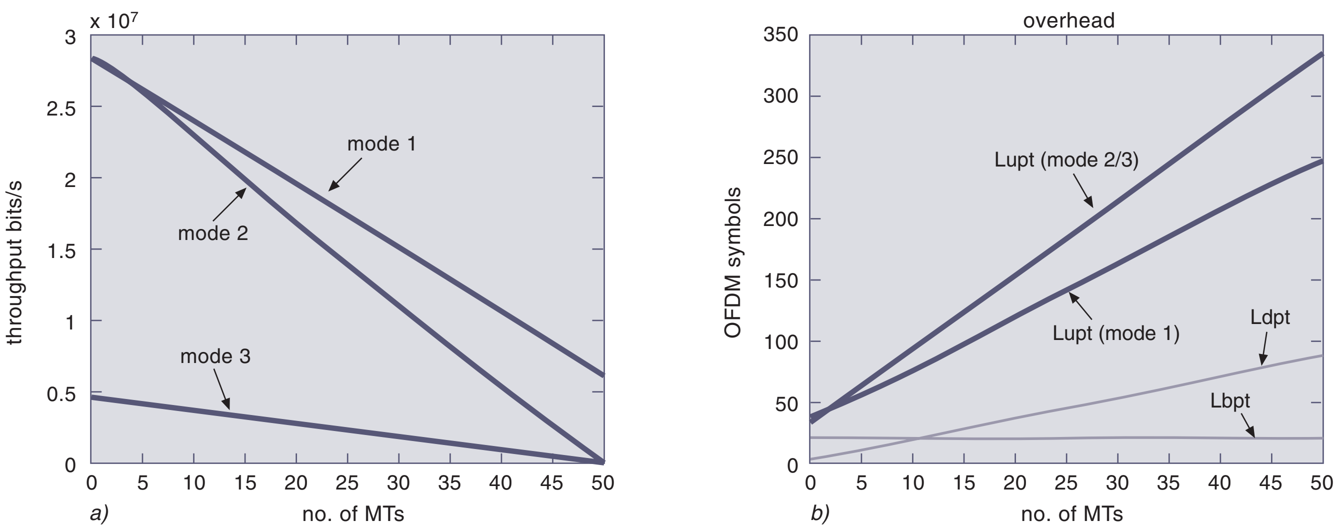


Figure 7 a) Throughput for different PHY modes versus number of MTs, one connection per MT; b) corresponding overhead characteristics

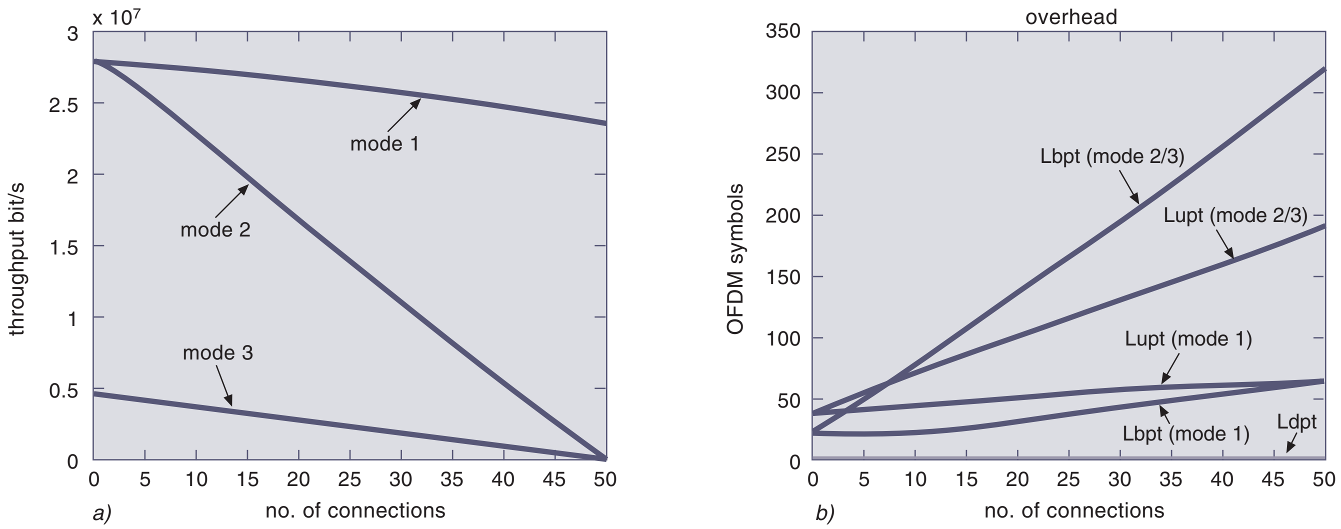


Figure 8 a) Throughput for different PHY modes versus number of connections for one terminal; b) corresponding overhead characteristics

Figure 9 Throughput for mode 1

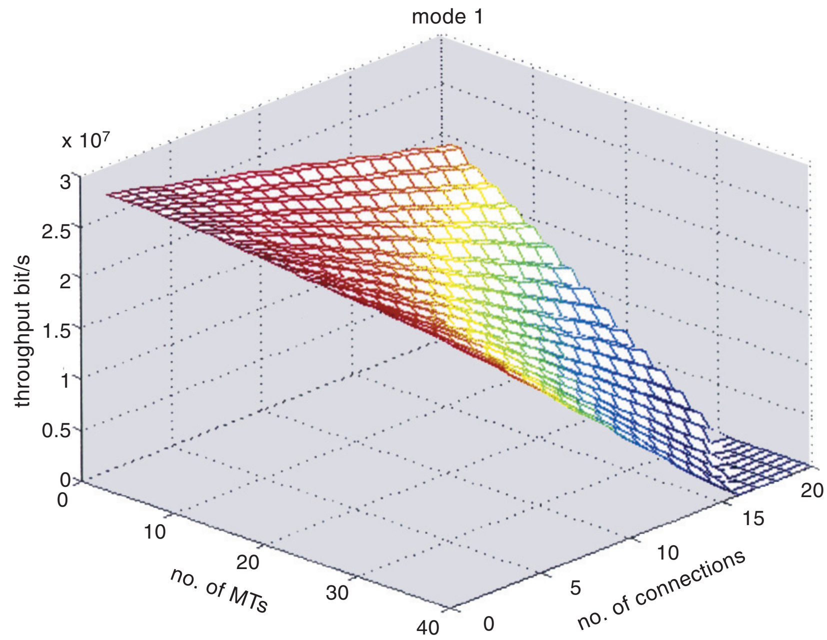
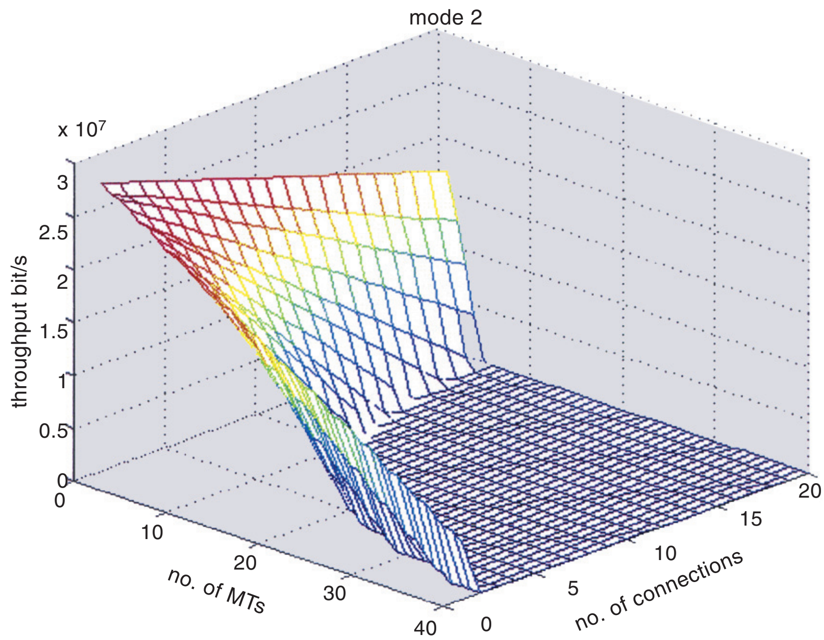


Figure 10 Throughput for mode 2



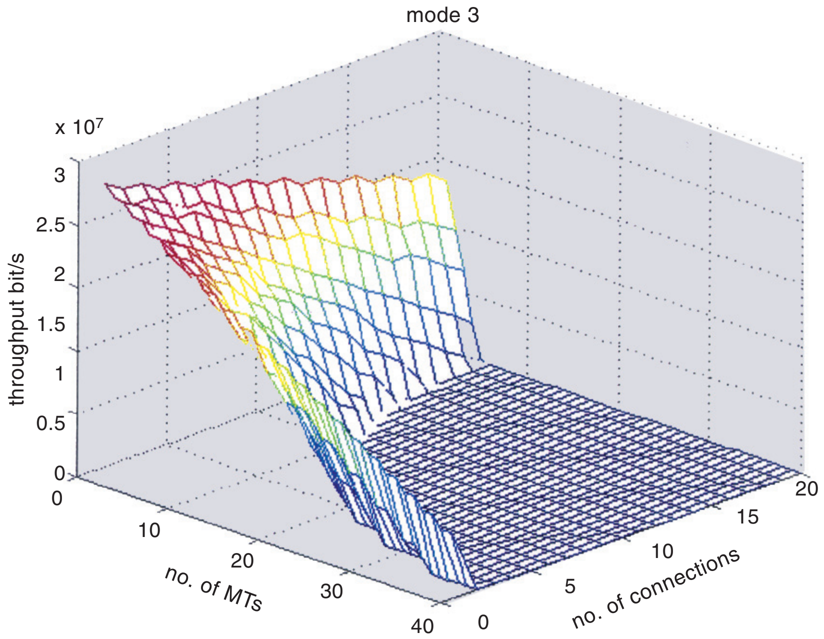


Figure 11 Throughput for mode 3

1) the gradient of the L_{UPT} is $^1/_{10}$ compared to the first scenario (5/*k* and 0.5/*n*, respectively). In addition the gradient of L_{BPT} (in the second scenario) is considerably smaller than the L_{DPT} in the first scenario. The overall result is that the throughput for mode 1 decreases more slowly with the number of connections than with the number of MTs. For mode 2 and 3 the gradient of L_{UPT} and L_{DPT} is smaller in the second scenario, but is offset by a larger increase in the gradient for the L_{BPT}. The result is that the overall throughput of the first and second scenario is similar.

The throughputs of mode 1, 2 and 3 for different numbers of MTs and active connections are shown in Figure 9, Figure 10 and Figure 11, respectively. Mode 1 gives a maximum throughput of nearly 28 Mbit/s ($k = 1$, MT = 1), and the throughput is inverse proportional with the number of MTs and the number of connections. From Figure 9 it can be seen that when the number of MTs exceeds 30 ($k > 30$) and the number of connections exceeds 16 ($n > 16$), the size of the overhead exceeds or equals the number of available OFDM symbols, resulting in zero throughput. For mode 2 (see Figure 10) the maximum throughput is similar to mode 1 (~28 Mbit/s), but the number of terminals and connections that can be supported is considerably smaller than for mode 1. For instance, the overhead exceeds or equals the number of available OFDM symbols at $k = 40$, $n = 2$ or $k = 5$, $n = 20$. For mode 3 (see Figure 11), the region for zero throughput is similar to mode 2. The maximum throughput (~4.6 Mbit/s) for mode 3 is however significantly lower than that of mode 2. These examples clearly illustrate that $N_{DBPS(FCH)}$, $N_{DBPS(SCH)}$ together with *k* and *n* determine the overhead size (and thereby the number of terminals and connections that can be supported), whereas $N_{DBPS(LCH)}$ determines the throughput for a given overhead size.

Consider an AP serving 10 terminals with two active connections each. For PHY mode 6 (Figure 9) the MAC throughput will be 26.9 Mbit/s, or a MAC efficiency of 75 %. In a worst case scenario, for PHY mode 1 (Figure 11) the MAC throughput will be 3.2 Mbit/s or a MAC efficiency of 53 %. This illustrates that the throughput of the HIPERLAN/2 will be high, even for poor radio conditions.

In this section the throughput (or the overhead) of the MAC protocol has been investigated for a number of settings. It has been shown that the number of MTs and active connections, as well as the PHY mode used on the different PDU trains influence the throughput. Using robust PHY modes will reduce the throughput but will be essential for operation in harsh radio conditions. It has been shown that the PHY modes used in FSC and SCH determine the number of terminals and connections that can be supported.

5 References

1 CEPT. *ERC Decision of 29 November 1999 on the harmonised frequency bands to be designated for the introduction of High Performance Radio Local Area Networks (HIPERLANs)*. (CEPT/ERC Decision (99)23.)

2 Khun-Jush, J et al. HIPERLAN type 2 for broadband wireless communication. *Ericsson Review*, 2, 108–119, 2000.

3 Johnsson, M. HiperLAN/2 – The Broadband Radio Transmission Technology Operating in the 5 GHz frequency Band. *HiperLAN/2 Global Forum, Version 1.0*, 1999.

4 ETSI. *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; System Overview.* Sophia Antipolis, 2000. (ETSI TR 101 683.)

5 ETSI. *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer*. Sophia Antipolis, 2000. (ETSI TS 101 475.)

6 Khun-Jush, J et al. Link performance of the HIPERLAN/2 Physical Layer in fading Environments. In: *Proceedings of the European Wireless '99,* Munich, 145–149, 1999.

7 Kluge, K. Effect of Link adaption on Overall Capacity in HIPERLAN Systems. In: *Proceedings of the European Wireless '99*, Munich, 163–167, 1999.

8 ETSI. *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) layer; Part 1: Basic Data Transport Functions*. Sophia Antipolis, 2000. (ETSI TS 101 761-1.)

9 ETSI. *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) layer; Part 2: Radio Link Control (RLC) Sublayer.* Sophia Antipolis, 2000. (ETSI TS 101 761-2.)

10 ETSI. *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Packet based Convergence Layer; Part 1: Common Part*. Sophia Antipolis, 2000. (ETSI TS 101 493-1.)

11 ETSI. *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Cell based Convergence Layer Part 1: Common Part.* Sophia Antipolis, 2000. (ETSI TS 101 763-1.)

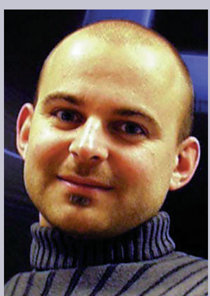
12 Kadelka, A, Hettich, A, Dick, S. Performance Evaluation of the MAC Protocol of the ETSI BRAN HIPERLAN/2 Standard. In: *Proceedings of the European Wireless '99*, Munich, 157–162, 1999.

Seamless Mobility in IP Networks

FREDERIC PAINT AND GEIR EGELAND

Frederic Paint (27) has been working for Telenor R&D since his graduation from ENST Paris (telecom engineering school) in 1998. His work has focused on 3G core networks and their evolution. This effort included participation in research projects (e.g. Eurescom P920 and Eurescom P1013) and standardisation activities (3GPP). More recently he has been involved in the field of mobility in IP networks specifically on micro-mobility support and inter-access mobility.

frederic.paint@telenor.com



Geir Egeland (34) is Research Scientist at Telenor R&D, where he has been employed since 1997. He has been working with Mobile IPv4 and Mobile IPv6, Service Discovery Protocols, TCP performance in wireless networks and the Next Generation Inernet Protocol (IPv6). His current research interests are mobile ad-hoc networks, Peer-to-Peer Computing and IPv6 networks.

geir.egeland@telenor.com

Mobile computing has gained considerable interest in the past few years. Maintaining secure IP connectivity as well as having transparent access to computing devices while moving across networks is a challenging problem being addressed in the Computing Industry. The operators are also looking at the IP mobility problem as they see the "all IP" as the enabler of true convergence. In particular the Cellular Industry is now considering architectures whereby base stations become IP enabled. In that context, mobility support in IP becomes even more challenging.

The purpose of this paper is to review the solutions being designed in the IETF to provide IP mobility support. Specifically we give some insight into the Mobile IP protocol and IP micro-mobility schemes as well as discussing the future developments of IP mobility support in the Internet Community.

1 Introduction

Today there is a considerable contrast between the mobile services provided in cellular networks and mobile services for Internet access. Although the Internet offers access to information sources world-wide, typically we do not expect to benefit from that access until we arrive at some familiar point such as the home, office or school. Being able to do mobile computing in the same way as in cellular networks offers many advantages. Access to the Internet anytime, anywhere will free us from being tied to our desktop. Having the Internet available to us as we move will give us the tools to build computing environments wherever we go, thus changing the very way we work.

Mobile computing and networking should not be confused with portable computing and networking that we have today. In mobile networking, computing activities are not disrupted when the user changes the computer's point of attachment to the Internet. Instead, all the needed reconnection occurs automatically and non-interactively.

Already solutions exist to provide ISP users the possibility to roam to other countries and use the local ISP's Internet access services. Such solutions provide support for portability. In other scenarios such as that of Wireless LAN access to IP networks, movement across IP subnets should be transparent to a user having ongoing communications. The Mobile IP standard provides that capability by making the mobility between subnets transparent to upper layers such as TCP, maintaining existing transport-layer connections. Furthermore users can continuously use a single IP address independently of their point of attachment, thus providing simple means for a mobile node to be reachable for other corresponding hosts.

The cellular industry is now considering employing IP up to the base stations for the sake of cost reduction. In that context questions have been raised as to whether Mobile IP scales well for micro-mobility, and in particular whether it can efficiently support fast handoffs. Several IP micro-mobility schemes extending or interworking with Mobile IP have been proposed in the *mobileip* working group of the IETF to provide better support of micro-mobility. These schemes focus primarily on solving the signalling latency introduced by mobile IP when roaming between subnets.

The purpose of this paper is to give some insights into Mobile IP and IP micro-mobility schemes as well as discussing the future developments of IP mobility in the Internet Community. The paper is organised as follows: First we present the Mobile IP protocol for IPv4 and IPv6 and its recent enhancements. Then we describe schemes proposed in the IETF for solving the micro-mobility and discuss their deficiencies. Finally we discuss the future development of IP mobility.

1.1 Acronyms

AAA	Authentication, Authorisation, Accounting
ACOA	Alternative Care-of Address
BS	Base Station
BU	Binding Update
CH	Correspondent Host
CIP	Cellular IP
COA	Care-of Address
DNS	Domain Name Service
DRR	Domain Root Router
FA	Foreign Agent
GFA	General Foreign Agent
GW	Gateway
HA	Home Agent

IETF	Internet Engineering Task Force
IP	Internet Protocol
ICMP	Internet Control Message Protocol
LCOA	Local Care-of Address
LCA	Least Ancient Ancestor
MAP	Mobility Anchor Point
MIP	Mobile IP
MH	Mobile Host
QoS	Quality of Service
RSVP	Resource Reservation Protocol
TCP	Transport Control Protocol
UDP	User Datagram Protocol

2 Mobile IP

There are some technical obstacles that must be overcome before mobility in the Internet can be a reality. The most fundamental obstacle is the way the Internet Protocol routes packets to their destinations according to IP addresses. These IP addresses are associated with a fixed network location much as a non-mobile phone number is associated with a physical jack in the wall. When a mobile node moves from one point of attachment to another, the mobile node must change its IP address so it is associated with the new network number, thus making transparent mobility impossible.

Mobile IP [1] is a standard proposed by a working group in the Internet Engineering Task Force (IETF) which is designed to solve the mobility problem by allowing the mobile node to use two IP addresses; a fixed home address and a care-of address that changes at each point of attachment. The proposal introduces a Home Agent that resides at the home network and tunnels packets destined to mobile nodes to their new point of attachment, and a Foreign Agent that provides mobile services to mobile nodes. Mobile IP is completely transparent for all layers above IP, e.g. for TCP, UDP and of course for all applications. Therefore, DNS entries for a mobile node refer to its home address and do not change if the mobile node changes its Internet access point.

Mobile IP consists of the co-operation of three mechanisms:

- Discovering Agents and obtaining a care-of address;
- Registering the care-of address;
- Tunnelling to the care-of address.

2.1 Discovering Care-of Addresses

The protocol used by mobile nodes to discover home and foreign agents is based on the existing protocol *Router Advertisement* [8]. Mobile IP does not modify any of the existing fields of the protocol, but expands it so that mobility functionality can be associated with it. This way a Router Advertisement can carry information about default routers, just as before, but in addition also contain information about one or more care-of addresses. Carrying this additional information, these router advertisements are called *Agent Advertisements*.

Home and foreign agents will broadcast Agent Advertisements at regular intervals, typically once every one to ten seconds. If a mobile node needs a care-of address and does not wish to wait for an agent advertisement, the mobile node can broadcast (or multicast) a solicitation that will be answered by any home/foreign agent that receives it. Mobile nodes use router solicitation to detect any change in the set of mobility agents available at the current point of attachment. If advertisements are no longer detectable from a foreign agent that previously had offered a care-of address to the mobile node, the mobile node should presume that the foreign agent no longer is within range of the mobile node's network interface. In this situation, the mobile node should start looking for a new care-of address. The mobile node may choose to wait for an agent advertisement, or it may send an agent solicitation.

2.2 Registering the Care-of Address

When a mobile node has a care-of address, it must inform its Home Agent about it. The mobile node will send a Registration Request message (carried in a UDP packet) with the information about its new care-of address. If the Home Agent approves this request, it will update its routing tables according to the new information and send a registration reply back to the mobile node.

The registration request message is protected by an authentication mechanism. This is because when a Home Agent accepts a registration request, it will associate the home address of the mobile node with the care-of address until the registration lifetime expires. A registration request is a form of *remote redirect*, because it is sent remotely to the Home Agent and affects the home agent's routing table. Without some sort of authentication for these remote redirect, a malicious node could cause the Home Agent to alter its routing table with erroneous care-of address information.

2.3 Tunnelling to the Care-of Address

Figure 2-1 shows the main principle of the mobile IP tunnelling operation. The default encapsulation mechanism that must be supported by all mobility agents using Mobile IP is IP-within-IP [2]. The Home Agent intercepts datagrams addressed to the mobile nodes and inserts an extra IP header, or tunnel header, in front of the IP header of the datagrams. The new tunnel

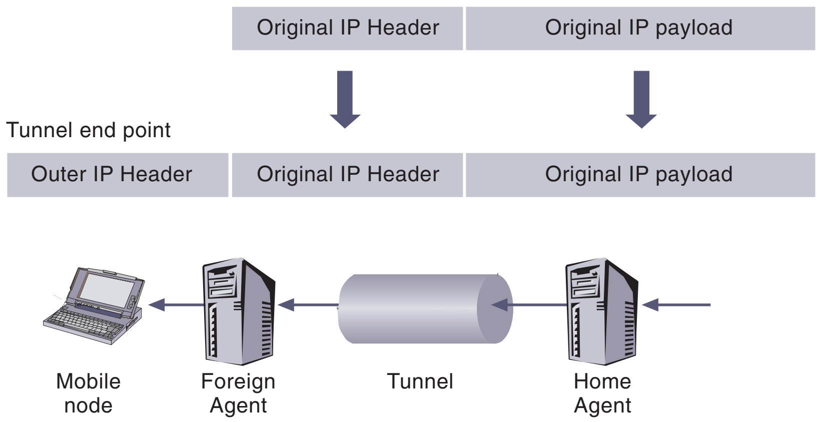


Figure 2-1 Tunnelling of IP datagram

header uses the mobile node's care-of address as the destination IP address, or tunnel destination. The tunnel source is the home agent.

2.4 Triangle Routing

All the datagrams that are sent to the mobile node must be routed through the Home network, being encapsulated and tunneled by the Home Agent using IP-in-IP encapsulation. Datagrams sent from a mobile node to a corresponding node on the other hand can be sent directly using standard IP routing. This is shown in Figure 2-2.

This asymmetric routing is called *Triangle Routing* and is in general far from optimal, especially when the mobile node is close to the corresponding node. An extension to Mobile IPv4 introduces the possibility for a Home Agent to inform corresponding nodes of the mobile node location. When a node tries to send a datagram to a mobile node, the Home Agent can provide information about the care-of address by sending what is called a *Binding Update* (BU). The BU is carried in a UDP packet with the mobile node's care-of address in addition to other parameters such as the lifetime of the care-of address. The corresponding node can now send datagram directly to the mobile node by setting up its own tunnel, thus getting a more optimal routing.

2.5 Mobile IPv6

Mobile IPv6 [3] includes many features for streamlining mobility support that are missing in Mobile IPv4. The design of Mobile IP support in IPv6 (Mobile IPv6) represents a natural combination of the experiences gained from the development of Mobile IP support in IPv4 together with the opportunities provided by the design and deployment of a new version of IP itself (IPv6) and the new protocol features offered by IPv6. Mobile IPv6 thus shares many features with Mobile IPv4, but the protocol is now fully integrated into IP and provides many improvements over Mobile IPv4. Mobile IPv6 requires the exchange of additional information. All new messages used in Mobile IPv6 are defined as IPv6 Destination Options.

The most important improvements are listed below.

Route Optimisation

Route Optimisation is now built in as a fundamental part of the protocol, rather than being added on as an optional set of extensions that may not be supported by all nodes as in Mobile IPv4. This integration of Route Optimisation functionality allows direct routing from any correspondent node to any mobile node, without needing to pass through the mobile node's home network and be forwarded by its Home Agent, and thus eliminating the problem of triangle routing present in the base Mobile IPv4 protocol.

Ingress Filtering

Support is also integrated into Mobile IPv6 and into IPv6 itself for allowing mobile nodes and Mobile IP to coexist efficiently with routers that perform Ingress Filtering. A mobile node now uses its care-of address as the Source Address in the IP header of packets it sends, allowing the packets to pass normally through ingress filtering routers. The home address of the mobile node is carried in the packet in a Home Address destination option, allowing the use of the care-of address in the packet to be transparent above the IP layer. The ability to correctly process a Home Address option in a received packet is required in all IPv6 nodes, whether mobile or stationary, whether host or router.

Foreign Agents

There is no longer any need to deploy foreign agents as used in Mobile IPv4. In Mobile IPv6, mobile nodes make use of IPv6 features, such as Neighbour Discovery and Address Autoconfiguration, to operate in any location away from home without any special support required from its local router.

Figure 2-2 Triangle routing in Mobile IPv4

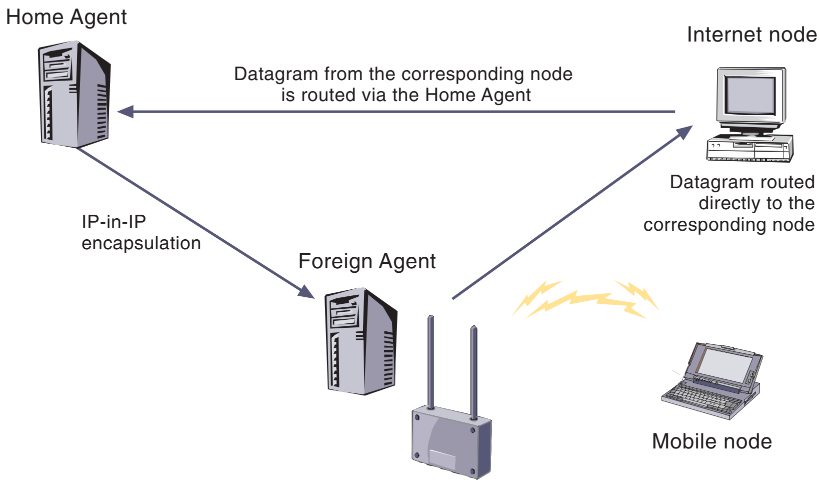
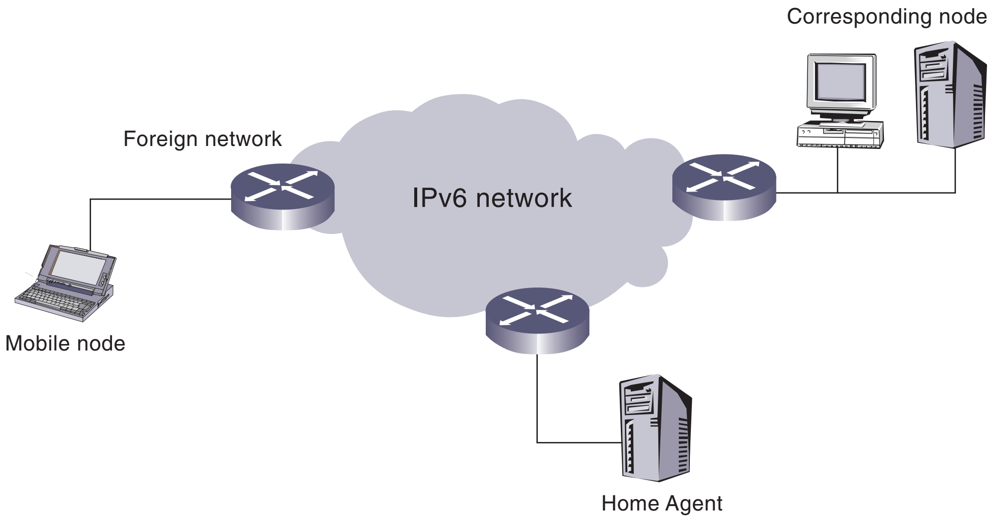


Figure 2-3 Mobile IPv6 architecture



Security
Unlike Mobile IPv4, Mobile IPv6 utilizes IP Security (IPsec) for all security requirements (sender authentication, data integrity protection, and replay protection) for Binding Updates (which serve the role of both registration and Route Optimisation in Mobile IPv4). Mobile IPv4 relies on its own security mechanisms for these functions, based on statically configured mobility security associations.

IP-tunnelling
Most packets sent to a mobile node while away from home in Mobile IPv6 are sent using an IPv6 Routing header rather than IP encapsulation, whereas Mobile IPv4 must use encapsulation for all packets. The use of a Routing header requires less additional header bytes to be added to the packet, reducing the overhead of Mobile IP packet delivery. To avoid modifying the packet in flight, however, packets intercepted and tunnelled by a mobile node's Home Agent in Mobile IPv6 must still use encapsulation for delivery to the mobile node.

Agent Advertisements
Mobile IPv6 defines an Advertisement Interval option on Router Advertisements (equivalent to Agent Advertisements in Mobile IPv4), allowing a mobile node to decide for itself how many Router Advertisements (Agent Advertisements) it is willing to miss before declaring its current router unreachable.

2.5.1 Movement Detection
As soon as a mobile node detects that it has moved from one link to another and it has discovered a new default router, it registers its new care-of address with its Home Agent on the home link using a BU. The primary movement detection mechanism for Mobile IPv6 uses the facilities of IPv6 *Neighbour Discovery* [9], including *Router Discovery* and *Neighbour Unreachability* Detection. In IPv6 there is a limit for how often a Router Advertisement can be sent. This limitation, however, is not suitable for providing timely movement detection for mobile nodes. Mobile nodes detect their own movement by learning the presence of new routers as the mobile node moves into wireless transmission range of them (or physically connects to a new wired network), and by learning that previous routers are no longer reachable. Mobile nodes must be able to quickly detect when they move to a link served by a new router, so that they can acquire a new care-of address and send Binding Updates to register this care-of address with their Home Agent and to notify correspondent nodes as needed. Thus, to provide good support for mobile nodes, Mobile IPv6 relaxes this limit such that routers may send unsolicited multicast Router Advertisements more frequently; in particular, on network interfaces where the router is expecting to provide service to visiting mobile nodes (e.g. wireless network interfaces), or on which it is serving as a Home Agent to one or more mobile nodes.

A mobile node may use any combination of mechanisms available to it to detect when it has moved from one link to another, and the mobile node can supplement the movement detection mechanism with other information available to the mobile node (e.g. from lower protocol layers).

2.5.2 Further Work
One can think of the Mobile IPv6 protocol as solving the network-layer mobility management problem. Some mobility management applications, for example handoff among wireless transceivers, each of which covers only a very small geographic area, have been solved using link-layer techniques. For example, in many current wireless LAN products, link-layer mobility

mechanisms allow a *handoff* of a mobile node from one cell to another, re-establishing link-layer connectivity to the node in each new location. Within the natural limitations imposed by link-management solutions, and as long as such handoff occurs only within cells of the mobile node's home link, such link-layer mobility mechanisms may offer faster convergence and lower overhead than Mobile IPv6. Extensions to the Mobile IPv6 protocol have been proposed to support a more local, hierarchical form of mobility management.

3 Micro-mobility and Fast Handoff

3.1 Background

The design of Mobile IP is based on the assumption that the rate of movement between subnets is low. This leads to the fact that the protocol is not suited for micro-mobility support for which the change of IP subnet is assumed to be frequent. More specifically, the registration process requires that a new registration is sent to the Home Agent for every change of subnet. This registration process induces unnecessary high signalling load on the global Internet and high processing needs in the Home Agents. It also leads to unnecessary latency during handoff from one subnet to another.

Several schemes have been proposed to solve the issue of micro-mobility management. All of these proposals assume a hierarchy in order to localise the scope of the mobility management to a domain (collection of subnet administrated by a single provider). Doing so reduces the global signalling traffic and enables a lower latency period during a change of subnet. These proposals assume Mobile IP for inter domain mobility and mainly focus on the intra domain mobility. There are two main concepts for managing the intra-domain mobility:

The **Host-based routing** approach: Mobility management is distributed among all the nodes within the domain. Host-specific routing is employed to keep track of the current location of the Mobile Host (MH) within the domain. Special routers are thus needed. The address of the user is the care-of address assigned/authorised by the domain.

The **multiple Care-of Address** approach: The Mobile Host is assigned multiple care-of addresses (COA), with each address identifying a specific agent/node in the hierarchy. Packets are routed through each node in the hierarchy using the corresponding care-of address. The Mobile Host only changes the care-of address of the lowest hierarchical level at every change of subnet. This approach uses conventional IP routing in the domain and distributes the mobility management over only a subset of nodes in the mobility domain.

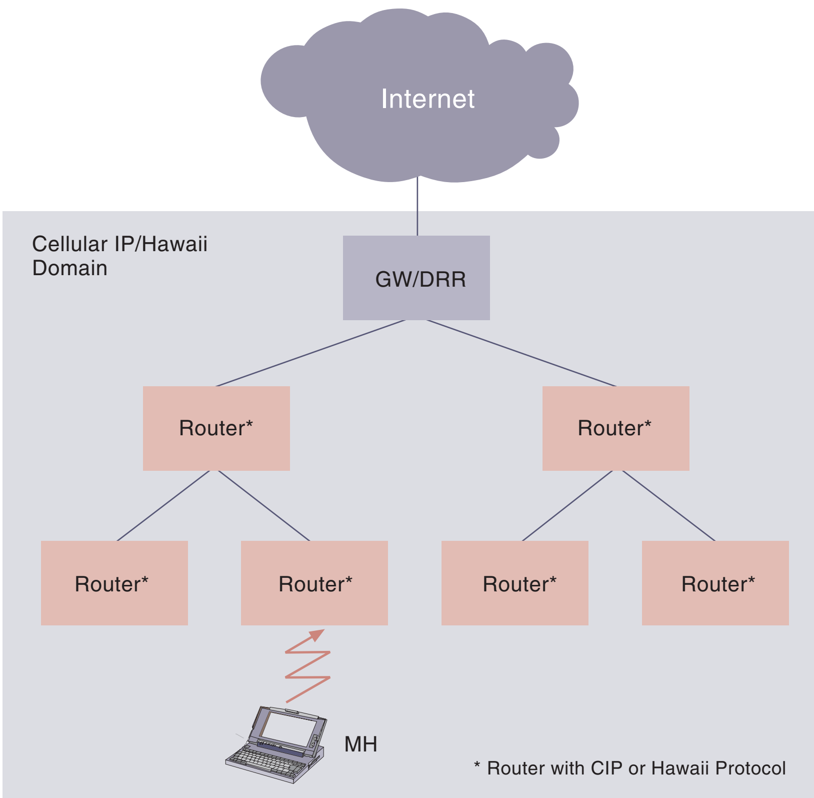
In the following we present these two concepts in more detail. We also review some of the proposed schemes and analyse their strong points and deficiencies. We note that at this point in time the schemes presented are in the form of Internet drafts and are thus under discussion in the IETF.

3.2 Host based Routing Schemes

3.2.1 Principles

Many schemes belonging to the host based routing category have been proposed. The Cellular IP and Hawaii schemes have received most of the attention thus far. Figure 3-1 illustrates the basic functional architecture and the implied tree like network topology of the Hawaii and Cellular IP solutions. Both approaches create host-specific entries through a specific path in the network topology. The ingress point into the domain is the root of the tree and is called the Gateway (GW) for Cellular IP and Domain Root Router (DRR) for the Hawaii scheme. Packets destined to the MH are forwarded through the sequence of nodes, the first one being the ingress point and the last one being the MH. It should be noticed that the edge device may not be a base station so that the schemes are applicable to solving mobility management in fix access networks as well. As an MH moves inside this

Figure 3-1 Cellular IP and Hawaii architecture



domain, it attaches to a new leaf in the hierarchy. Establishment of the correct path then essentially consists of propagating a new route establishment message up the tree towards the least common ancestor (LCA) of the current and old leaf nodes. Thus, the latency involved in the intra-domain update process is defined by the latency of communication between the leaf and the LCA (In the worst case it is as high as the latency between the leaf node and the ingress point). Cellular IP and Hawaii differ in the way the route update and refreshes are provided to the nodes. We further describe these protocols in the following sections. Additionally they differ in the way they interwork with Mobile IP. We note that these schemes were designed for IPv4 but can be adapted for IPv6. Here we only describe the version for IPv4.

3.2.2 Cellular IP

Cellular IP [4] provides local mobility and fast handoff support, relying on Mobile IP for macro-mobility. The Cellular IP network consists of a set of Cellular IP routers, several base stations which are also Cellular IP enabled, and a Cellular IP gateway (see Figure 3-1). Every base station is IP addressable and the Gateway acts as a Foreign Agent. The main idea behind Cellular IP (CIP) is that the IP address (Home Address) of the Terminal equipment is used as the user identity within the Cellular IP domain (one domain per gateway), and uplink user traffic packets (from the MS to the Gateway) update the path to the user. The following provides a summary of the main procedures specific to Cellular IP.

Mobility States: The MH is either in idle state or in active state. When an idle MH receives an IP packet (paged from the network) it moves from idle mode to active mode and sends a route update packet. When a subsequent packet is received, the active state timer is reset. When this timer times out, the MS goes back to idle. In active state, any Packet Data Unit (PDU) sent uplink will update the route from the MS to the gateway. A PDU can be a traffic PDU or a control PDU (route update message) if the route update timer has timed out. Whenever a traffic-PDU is sent this timer is reset. Thus in active state at least one IP packet is sent during a time equal to the value of the route update timer. In idle state paging updates are sent instead of route updates. The paging update is meant to update the paging cache of the nodes on the path. Route updates or paging updates uses ICMP.

Cellular IP routing: When a node receives an uplink packet, it reads the source address and updates the corresponding cache and then routes the packet towards the gateway. When a downlink packet is received, a node checks the cache. If the routing cache has not been updated the node is paged. Otherwise it sends to the next node that is indicated in the binding table.

Handovers: Inter-gateway handoff is like a new registration at a new gateway. The location information at the old gateway expires and is not explicitly removed. There are two options for intra domain handoffs, hard and semi-soft.

- In hard handoff, when the MH decides to handoff it sends a hard handoff route update message (for security reasons, an uplink packet is not sufficient) to the new base station. This route update heads towards the gateway. Each Cellular IP node along the way creates a new entry in its routing cache for the mobile node (or updates the existing entry). When the message reaches the LCA, the old entry is replaced by the new entry, pointing towards the new base station. At this point, the handoff is complete, and the route cache entries along the path from the LCA to BS 1 are left to expire when they time-out. Meanwhile packets are sent to the old base station and are thus lost.

- Semi-soft handoff addresses the problem of packet loss by requiring the mobile to switch back to the old base station for a short period, the semi-soft period, while waiting for the semi-soft route update to reach the crossover node. When the LCA is reached, a new entry (second entry) for the mobile node is created in the route cache. Hence, during the semi-soft period, packets are forwarded to both base stations. When the semi-soft period is over, the mobile node switches back to the new base station and sends a hard handoff route update which eventually removes the entry towards the old base station in the LCA.

Security

Cellular IP ensures the integrity and confidentiality of mobility related control messages. To ensure that, only authenticated packets can establish or change cache mappings in a Cellular IP network and all control packets, i.e. paging and routing update messages, must be authenticated. Data packets can only refresh existing mappings. This of course reduces the signalling efficiency of the original scheme in which the main benefit was to have the data packets acting as control packets to update the route.

3.2.3 Hawaii

In Hawaii [5] the addressing is different from CIP. The mobile host is assigned a co-located care-of address from its foreign domain and packets are tunnelled to the care-of address by a Home Agent in its home domain. When moving within the foreign domain, the mobile host keeps

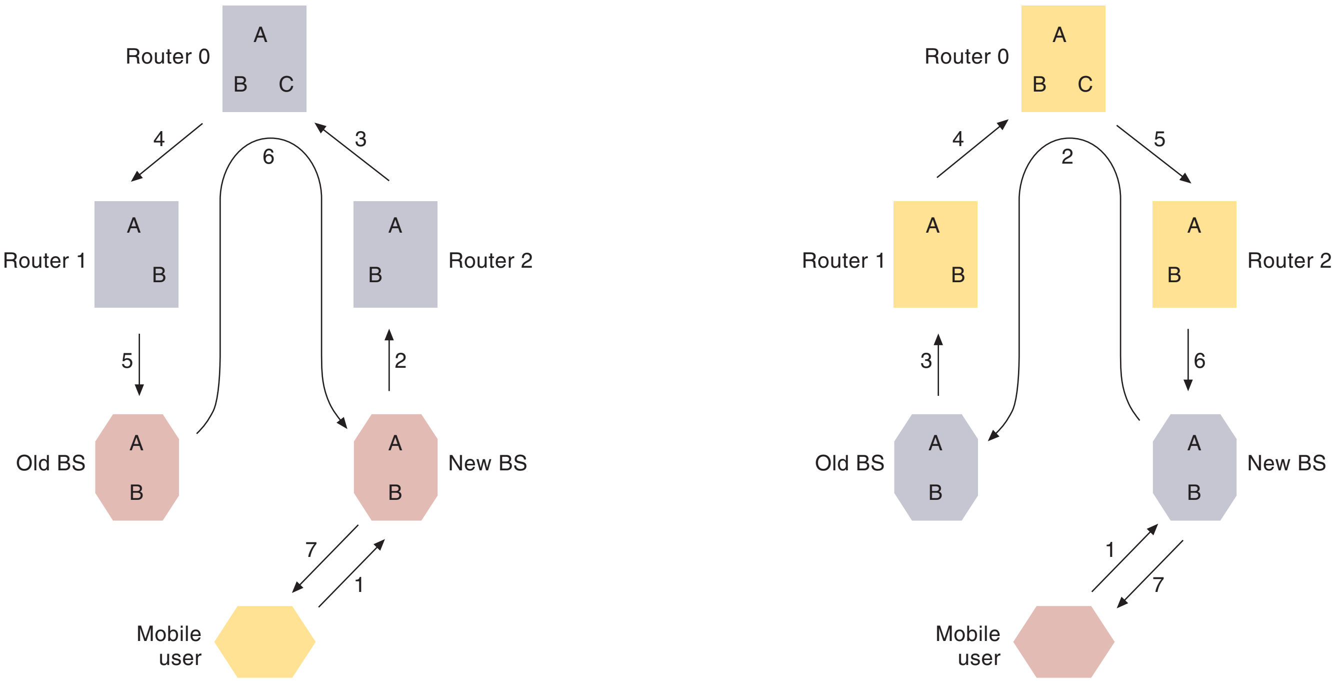


Figure 3-2 The two handover mechanisms of Hawaii

its care-of address. Hawaii maintains the paths in the domain through the specific path set-up mechanism that is described below.

Registration Phase

When the Mobile powers up it sends a Mobile IP registration message to the BS covering the area the MS is camping on. The BS triggers a registration procedure towards the HA which is located within the home domain of the MH. After completion of that procedure (assumed here valid) the BS sends a path set up message (Hawaii specific message) up to the gateway via a default route. Every router on that path will set its host based forwarding entries for the MH co-located care-of address. Then the BS sends an acknowledgement to the MH. The forwarding entries are kept as soft state meaning they expire after a certain time. They are kept alive by hop-by-hop refresh messages.

Handover

When the MS moves to another cell the new path has to be established and the old path has to disappear in the network. Two schemes are defined (see Figure 3-2). The first one, the forwarding path set up scheme, is optimised for networks where the mobile host cannot receive/transmit from/to different base stations (macro diversity), e.g. TDMA based systems. The second one, the non-forwarding path set-up scheme, is optimised for networks where macro diversity is possible (e.g. CDMA).

- **Forwarding scheme**. The packets are forwarded from the old BS to the new BS while the new path is not set-up. When the MH enters a new cell it sends a Mobile IP registration message that includes the old base station's address. When receiving this message the new BS sends a Hawaii message to the old BS. The old BS changes the forwarding entry for the MH to the uplink router (router 1), which is specified in its routing table, and forwards the message to that router. This router will do the same and forward the message further along the way towards the new BS. The message will eventually reach the new BS. The new BS sends a reply to the MS and the handover procedure is completed.

- **Non-forwarding scheme**. The MH sends a Mobile IP registration to the new BS which sends a Hawaii message to the old BS. This message is transmitted hop-by-hop and every router on the path between the new and the old BS will create/update its entry for the MH accordingly. When the message has reached the LCA the path is diverted. Meanwhile the packets are received by the MH on the old air interface which the MS is able to receive from if macro-diversity functionality is implemented. No packet loss is therefore experienced. The old BS will send an acknowledgement to the new BS after receiving the message. The handover procedure is then completed.

3.2.4 Analysis

The strong point of these techniques is that tunnelling within the domain is not necessary. Tunnelling is detrimental to the transport efficiency in the domain. Within the cellular IP network, routing is host-based. Hence Cellular IP or Hawaii nodes must be special routers that have a special routing function. In particular the nodes

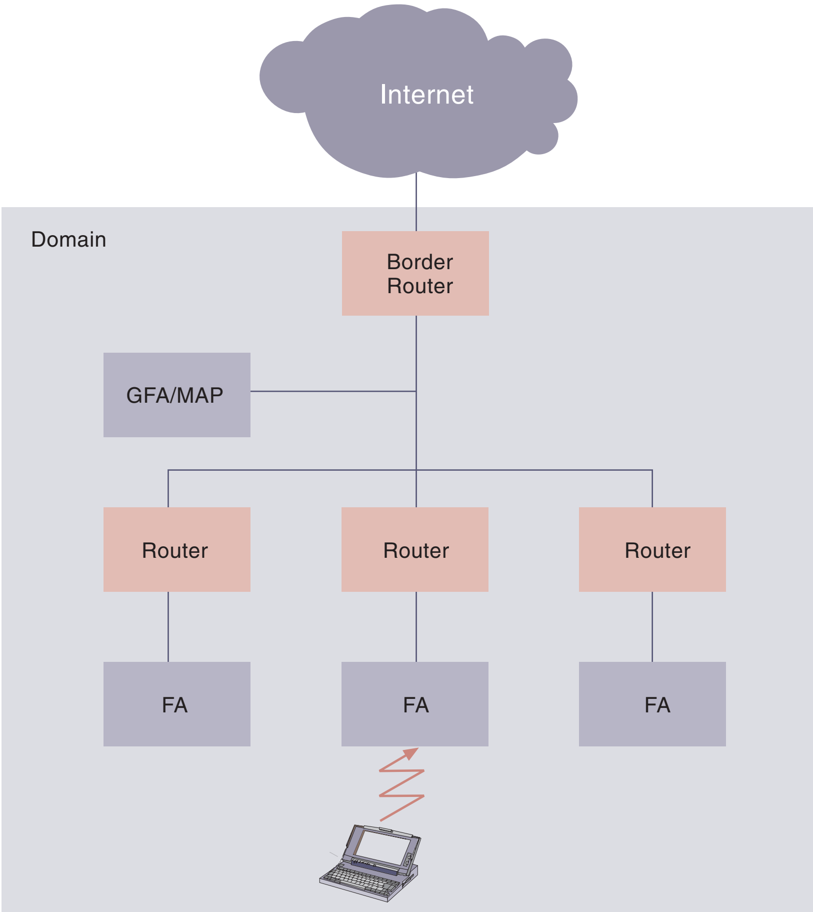


Figure 3-3 Architecture of a multiple care of address scheme

need to maintain a forwarding entry for all active nodes. It also means that table look-up becomes a demanding task in terms of processing. These schemes may thus present scalability problems as the network grows and the number of active nodes increases. Furthermore operation assumes a tree-like structure since the gateway is a single entry point into the cellular IP domain. Reliability and performance are thus also a concern.

3.3 Multiple Care-of Addresses

3.3.1 Principles

There are several proposals that follow the multiple care-of address approach. Among others, Mobile Regional Registration and Hierarchical Mobile IP (for IPv6) have gained considerable attention in IETF. All these proposals allow the domain to have any arbitrary topology. The architecture (Figure 3-3) typically consists of subnet-level agents, which essentially provide an MH with a local COA (LCOA). This COA identifies the MH's current subnet of attachment. A separate agent, the GFA (Global Foreign Agent) or MAP (Mobility Anchor Point) resides at a higher level in the domain hierarchy and provides the MH with a stable point of attachment. This point of attachment remains the same while the MH is in the same domain. The global care-of address either refers to an address associated with the top-level agent or an address for which packets are intercepted by the agent.

Since only the global COA is available to a CH located outside the domain, all packets destined to the MH use this COA. The MA/GFA intercepts these packets and then forwards them to the MH's local COA.

Having previously focused on IPv4, we thought it adequate to present a scheme designed primarily for IPv6, so we only present Hierarchical Mobile IPv6 in this chapter.

3.3.2 Hierarchical Mobile IPv6

In this scheme [6] the MH discovers the address of the MAP through router advertisements. The MH registers with the MAP using its local care-of address and permanent home address. Then it sends binding updates to the CH and the HA indicating the MAP's address as the alternative care-of address (ACOA). The CH can then transmit packets using the routing header option directly to this ACOA. The MAP will then forward the packet towards the Local CoA (LCOA).

On subsequent movement within the domain, the MH obtains a new Local COA (based on the stateless autoconfiguration mechanism) and then updates its MAP of this new LCOA. The MAP can then tunnel packets to this new LCOA.

The domain of a MAP is defined as the area where the router advertisement is advertising the identity of the MAP. Thus movement across domains is detected when the MAP address in the router advertisement has changed. When the MH has moved into a new domain, it sends the address of the new MAP (ACOA) in separate BUs to each of its CHs and its HA.

The MH must also check for the existence of a routing header in the inner packet; if such a header exists, the MH can assume that the CH is aware of its ACOA and does not need to receive a new BU.

3.3.3 Analysis

The main benefit of a scheme like HMIPv6 is of course the reduced global signalling compared to Mobile IPv6. It also enables the use of standard IP routers in each domain thus avoiding the cost of manufacturing special routers. The main drawback is that tunnelling is required inside the domain (between the MAP and the MH) which results in

- Poor transport efficiency;
- High Header processing complexity in the MH and Intermediate foreign agents.

4 Discussion

Mobile IP has been studied in a number of wireless communication research projects and is now starting to appear in commercial products. The protocol has a great potential and will be of great importance in future wireless and mobile computing. Micro-mobility schemes have in the last year received much attention within the IETF, especially from vendors of cellular equipment.

All of these schemes resolve partially the problem of micro-mobility as they first and foremost address the signalling latency. After a change of subnet it can be necessary to transfer the MH context such as that of Header Compression and policy information. Furthermore, Security has not been considered at all for some of the schemes. For other schemes (e.g. Cellular IP) it has been integrated afterwards as an add-on thus leading to sub-optimal solutions. Additionally, paging in IP, though addressed by some proposals (Cellular IP and Hawaii), is to be looked at more closely. These issues are particularly challenging in the context of an “all-IP” cellular infrastructure where scalability is a main concern.

The IETF has therefore set up a new working group called Seamoby (as in seamless mobility) [7] that has the responsibility of designing the necessary protocols for efficiently providing support for micro-mobility. It is to be noted that Mobile IP is assumed as the macro-mobility protocol. The main focus of this group will be:

- Further work on fast handoff mechanisms;
- Looking at the benefits of a Context transfer mechanism; specifying one if it is seen beneficial;
- Develop a paging mechanism at the IP layer.

Whether this work will result in a valuable extension to Mobile IP remains to be seen, since many problems are not yet defined and substantial work needs to be done.

References

1 IETF. *IP Mobility Support*. October 1996. (RFC 2002.)

2 IETF. *IP Encapsulation within IP*. October 1996. (RFC 2003.)

3 IETF. *Mobility Support in IPv6*. February 2000. (draft-ietf-mobileip-ipv6-12.txt.)

4 IETF. *Cellular IP*. Internet Draft, Work in Progress. November 1998.

5 IETF. *IP micro-mobility support using HAWAII*. Internet Draft, Work in Progress, June 1999.

6 INRIA. *An Hierarchical Mobile IPv6 Proposal.* November 1998. (TR-0226.)

7 IETF. *Seamoby WG charter*. (2000, March 6) [online] – URL: http://www.ietf.org/html.charters/seamoby-charter.html

8 Deering, S E. *ICMP Router Discovery Messages.* September 1991. (IETF RFC 1256.)

9 Thomson, S, Narten, T. *IPv6 Stateless Address Autoconfiguration.* December 1998. (IETF RFC 2462.)

Providing Open Application Interfaces to Support Third-Party Service Providers and Developers

GEIR GYLTERUD AND GAUTE NYGREEN

This paper deals with the emerging standards for the provisioning of open application interfaces to provide services to subscribers in traditionally closed networks. It provides an overview of the emerging standards in this area and discusses the opportunities and consequences this will have for the next generation of wireless networks (UMTS and beyond). Traditionally, the provisioning of services and applications in the network has been the domain of the network operator, the network being wired or wireless. This way, the network operators have been able to generate substantial revenue and at the same time provide customers with valued services, increasing the customer's loyalty to a specific operator. So far this privilege of providing services to the network has been well protected, certainly to protect a good source of revenue, but also for the reason that there has not been any secure and standardised way to preserve the integrity and security of the network while opening up for others to provide services. This is now about to change. Deregulation in the telecom market along with heavier competition, and the tremendous growth of the Internet with the IT-world's richness in applications and services, are the driving forces behind the work of opening up network interfaces to other service providers and competitors. In standardisation bodies like ITU-T and 3GPP standards are already being prepared which will give the network operator the ability to give other service providers access to providing services while also preserving the security and integrity of the network. The provisioning of open interfaces will force the implementation of new business models for telecom operators and service providers. The success of this model will of course depend on the ability to create revenue for each of the participating entities, which will be a great challenge.

Geir Gylterud (35) is Research Scientist at Telenor R&D, Trondheim, where he has worked in the Open Platforms and Service Innovation group since February 1997. He has been working in the field of Value Added Services (IN) with focus on platforms and the integration of such to Internet applications. The focus has over the recent years turned to Open Service Platforms featuring middleware technology for the distribution of the service execution environment and using Open APIs as Parlay and OSA for service provisioning to fixed and mobile subscribers.

geir.gylterud@telenor.com

Gaute Nygreen (26) is Research Scientist at Telenor R&D, Trondheim, where he has worked in the Open Platforms and Service Innovation group since March 2000. His work consists mainly of prototyping using open APIs like the OSA and the Parlay API. Special interests include open application interfaces for telecom.

gaute.nygreen@telenor.com

1 Introduction – a Marketplace in Change

The market for mobile and wireless communication services is changing rapidly and is becoming more and more competitive. This means that the network operator has to offer a diversity of differentiated services to be able to attract new customers but also to keep the existing customers and gain their loyalty. Open application interfaces against external service providers will be one of the tools the network operator will use to offer these differentiated services.

Not that many years ago governmentally owned telecom operators had a monopoly to deliver telecom services to customers within their country.

In the last few years we have seen a dramatic change in the telecom market with the opening and liberalisation of the market. This is especially visible when it comes to the mobile or wireless arena. If we look at the Norwegian mobile communication market, we see that the competition among mobile operators is getting tougher. For the GSM network we have two major operators in Telenor AS and NetCom AS, plus some virtual operators renting the unused capacity from these network operators. With the introduction of UMTS, four operators have been allotted a licence to build and operate a UMTS network, and the competition to attract customers is steadily increasing. This increasing competition will necessarily force the incumbent operators to change the way they look at and serve their customers. Getting new customers, keeping them satisfied and creating a loyalty among them that makes them willing to stay with the same operator for a longer period of time is one of the most important tasks from the telecom operator's point of view. It is getting more and more important to provide customers with unique "can't do without" services that make you as an operator attractive to the customers and make it harder for the customers to move to another operator.

As the market is changing, so is the way of providing services. Figure 1 presents a view of the changes in the way services are provided in future mobile networks as presented by UMTS Forum [6].

The figure shows a relative change in the value-chain of telecommunication service provisioning. As the terminal gets more intelligent and provides a service runtime environment and sufficient storage capacity, the amount of terminal based services will increase substantially. The amount of services provided by the network operator might decrease and only a slight increase in network applications is here assumed. The reason for this is the large fore-

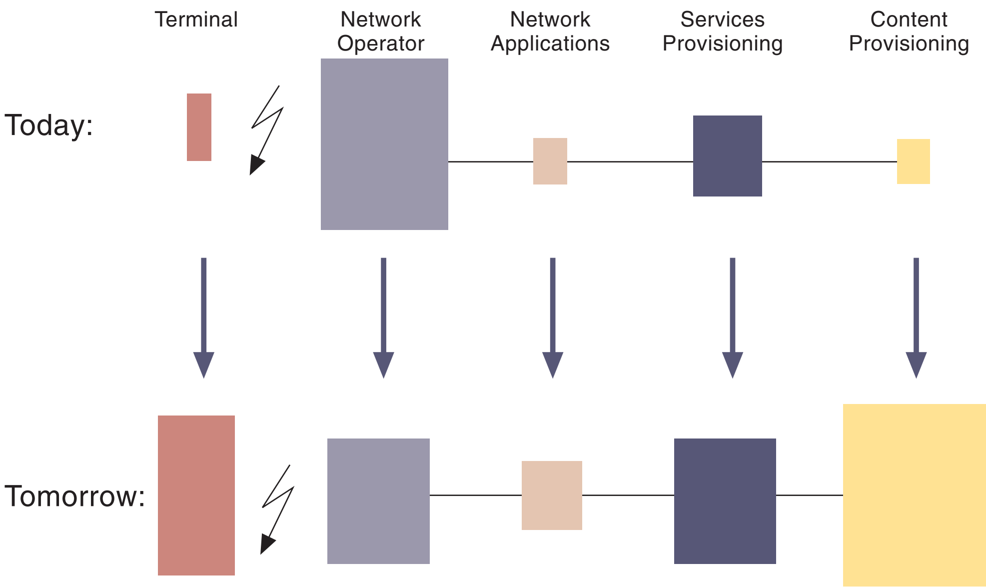


Figure 1 Relative changes in the value-chain of telecommunication service provisioning

seen increase in services provided by service and content providers. In UMTS and future mobile networks it is likely that we will see this shift in the way services and content are provided. One reason for the expected shift in service provisioning is the increasing focus given to the provisioning of open and standardised interfaces for service support. These open interfaces will give service providers both internal and external to the telecom operators organisational access to network functionality that they can use to build new and creative services. This also paves the way for independent software vendors delivering applications to the telecom industry. In the UMTS standardisation work in 3GPP, a lot of effort has been put into standardising the Open Service Access (OSA) interfaces which will give external service providers access to network functionality in a secure way. These external service providers will be able to deliver services more rapidly and much cheaper than the traditional telephone exchange suppliers. In the following chapters we will discuss the emerging standards in the area and what changes this will bring to the business model for service provisioning.

2 Emerging Standards, an Overview of Parlay and the Open Service Access (OSA) Interfaces in UMTS

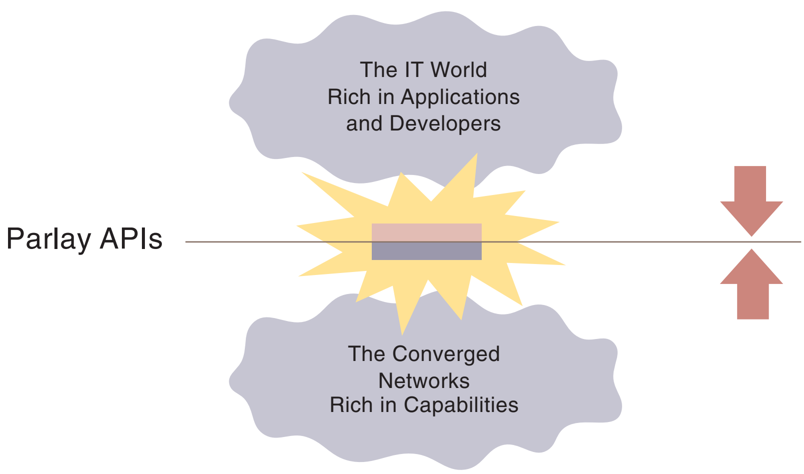
The demand for rapid development of new communication applications and integration of telecommunication and IT has seen many initiatives lately. One of these was the creation of the Parlay Group in 1998. In UMTS the work of the Parlay group has been adopted and specifications here already exist as the Open Service Access (OSA) interfaces.

2.1 The Parlay Group and Parlay API

The initial focus of the Parlay Group [5,7,8,9] was the development of an open Application Programming Interface (API) to link the rapid development of IT applications with the capabilities of the telecommunication world. Figure 2 shows the motivation behind the creation of Parlay API.

BT, Microsoft, Nortel Networks, Siemens and Ulticom – formerly DGM&S Telecom, funded the Parlay Group in 1998. AT&T, Cegetel, Cisco, Ericsson, IBM and Lucent joined the Parlay Group in May 1999. Since then the Parlay group has been reconstituted as a non-profit organisation where anyone who pays the entrance fee can join, and the group has grown considerably since this reorganisation late 2000.

Figure 2 Parlay convergence, merging IT and Telecom. From [7]



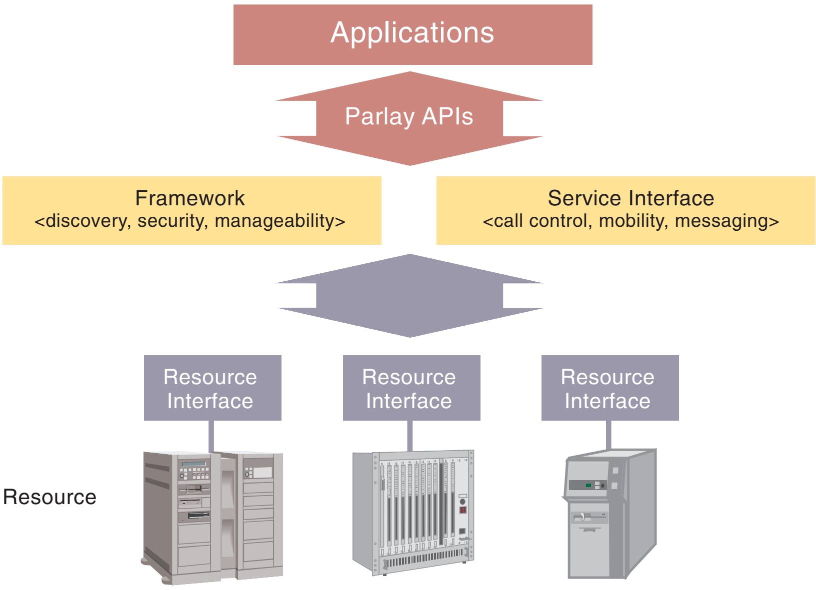


Figure 3 The Parlay API. From [7]

The Parlay API is shown in Figure 3. It consists of three parts:

- Applications, utilising the Framework and Service Interfaces and providing the actual service.

- Framework Interfaces, providing the mechanisms needed to let the Applications access the Service Interfaces. The Framework is the application's initial contact with the Parlay API. The application must first authenticate, and then it can discover different Service Interfaces. The Framework also supports manageability functionality like a heartbeat mechanism to keep track of applications or service interfaces going down.

- The Service Interfaces, providing the applications with service capabilities. The functionality of these capabilities is abstracted from the network resources. The different Service Interfaces are: Call Processing service interface consisting of Call Control and User Interaction; Generic Messaging service interface; Mobility service interface; and Connectivity Manager service interface. Before the Service Interfaces can be discovered by the applications they must register with the framework through the framework's registration service.

2.2 The Open Service Access (OSA) Interfaces in UMTS

The OSA interfaces are specified for UMTS by 3GPP [1,2,3,4]. The main focus of the OSA interfaces is to open the UMTS network to external service providers and to make rapid application development possible for the network operators. The work of the Parlay Group is firstly based on the fixed network and convergence of different networks. In contrast to this the OSA interfaces are based on the wireless network. In spite of this difference in starting point the two specifications are growing closer to each other; 3GPP and the Parlay Group are partially working together on the standards. The next releases of Parlay and OSA will probably have interchangeable Service Capability Servers.

The Open Service Access (OSA) specification has the same three parts as Parlay:

- *Application*
- *Framework*
- *Service Capability Servers*

The OSA infrastructure and API is depicted in Figure 4 showing an example of where these interfaces could fit in.

The Applications can be anything from common IN applications like number translation and Virtual Private Network to advanced location based or context-based applications. These applications reside on one or more application servers, which may be placed inside or outside the domain of the network operator.

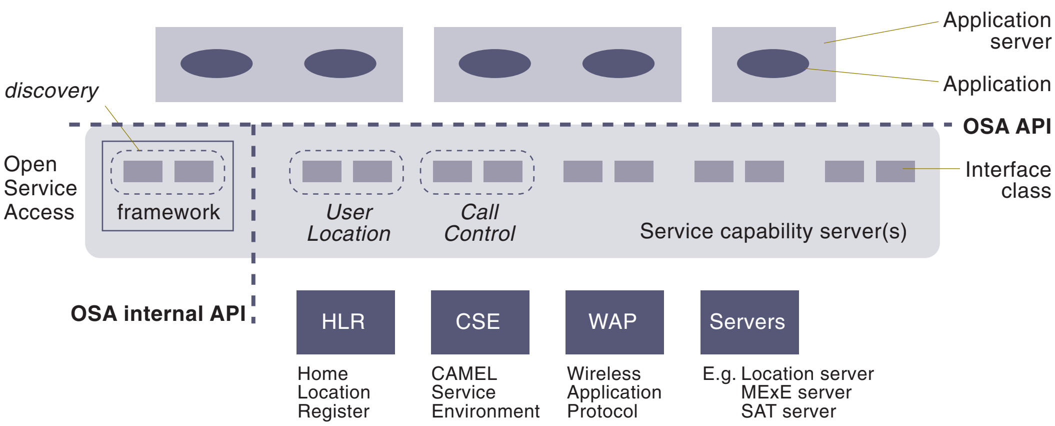


Figure 4 Overview of Open Service Access. From [2]

The Framework offers the essential capabilities for OSA applications to make use of the service capabilities available. The four main parts of the framework are:

- *Trust and Security Manager,* which contains the initial access point for OSA applications. Once connected to the initial access point two-way authentication precedes all other actions. When authenticated, the application can be authorised to access certain service capability features. (Authorisation is distinguished from authentication in that it is the act of determining what something/someone already authenticated is allowed to do.) The Trust and Security Manager is also responsible for signing of service level agreements; the agreements may consist of one off-line and one on-line part.

- *Service Registration* allows Service Capability Servers to register their service capability features for later discovery by OSA applications. The framework service capability features need not be registered because they are available by default. OSA applications may only discover service capability features after the features are registered and the applications are authorized.

- *Service Discovery* allows OSA applications to discover non-framework service capability features like Call Control and User Status.

- *Integrity Management* provides the framework with the means to keep track of its own integrity, the service capability features and the OSA application. The Integrity Management also allows the OSA application to query the integrity of the framework and service capability servers, and report about its own integrity.

The Service Capability Server (SCS) provides the applications with one or more Service Capability Feature (SCF). SCFs can be anything from Call Control to User Location functionality and are abstracted from the underlying network functionality. It is also possible to have more SCSs providing the same SCFs, either on top of different network types or on top of similar networks in different regions.

The network service capability features specified for OSA are:

- *Call Control* consists of two interfaces; the Call Manager, used to manage call related issues and letting the application enable and disable call-related event notification; and the Call interface to control ongoing calls and letting the application enable and disable call-events considering specific calls.

- *Data Session Control* consists of two interfaces; the Data Session manager used to manage data session related issues; and the Data Session interface which provides basic functionality for applications to control data sessions.

- *Network User Location* consists of a single interface providing terminal location information. The accuracy of the reported location is subject to the capabilities of the underlying network. Local legislation can hinder accurate location information being supported.

- *User Status* consists of a single interface that provides methods that allow the applications to obtain the status of the user's terminals.

- *Terminal Capabilities* will make it possible for applications to request terminal capabilities.

- *Message Transfer* consists of the Generic User Interaction SCF and the Call User Interaction SCF. The Generic User Interaction SCF is used by OSA applications to interact with the user. It consists of two interfaces; the User Interaction Manager, containing management functionality for User Interaction related issues, and the Generic User Interaction interface, containing methods to interact with the user. The Call User Interaction is used by the application to interact with users participating in a call. It consists of two interfaces; the User Interaction Manager, which is the same as used for the Generic User Interaction, and the Call User Interaction interface, which supplies call-specific user interaction.

2.3 Adapting OSA in the Network

Before OSA can be provided and used to provide services it needs to be implemented in the network. First the different ways of adding OSA functionality are shown, and then a short discussion follows on where to get different SCSs.

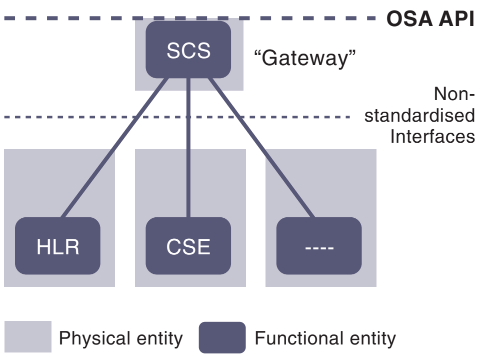


Figure 5 SCSs and network functional entities implemented in separate physical entities. From [2]

Figure 6 SCSs and network functional entities implemented in the same physical entities. From [2]

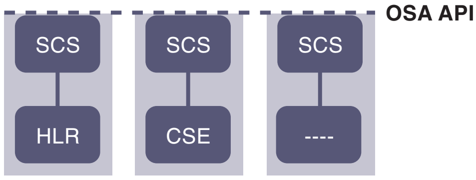
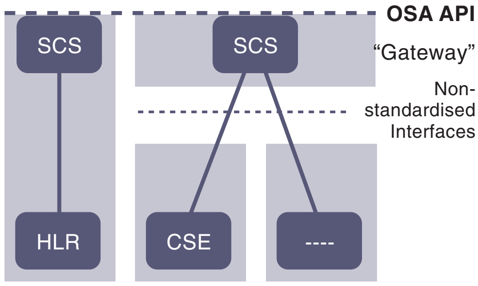


Figure 7 Hybrid implementation. From [2]



Different SCSs may be implemented on one or more physical entities, separate from the physical network entities. Figure 5 shows one SCS connecting different non-standardised network interfaces. Different SCSs can also use one 'Gateway' for each physical network entity.

The OSA SCSs can also be implemented in the physical network entity itself, as shown in Figure 6.

The possibility to have more than one SCS naturally renders the option of a hybrid solution with some SCSs implemented in the physical network entity and some SCSs in separate physical entities. This is shown in Figure 7.

Network operators wanting to offer OSA SCSs will probably implement the last solution shown. The reason for this is the need to use existing non-standardised interfaces already in the network and the need to maintain old services deployed in the network. They can then make or buy an SCS that can use the old interface or buy an upgrade from the infrastructure provider if or when upgrades are available. When expanding the network with new components network operators will in the future be able to buy components including SCSs. Some infrastructure providers will provide network entities containing SCSs, others will provide the SCSs as separate entities which open the market for third part SCSs. Third part SCSs will also be available against today's non-standardised interfaces.

2.4 Use of Open APIs

The provisioning of OSA, Parlay and possibly other open APIs will help network operators to rapidly develop and deploy new services that integrate functionality from several network resources, and this alone will motivate the introduction of OSA. Once the network operator implements OSA or equivalent APIs, the use of the SCSs can be sold to third party operators.

Before an external service provider is allowed to access a network operator's SCSs some sort of agreement is needed. The standard is a Service Level Agreement (SLA), normally a comprehensive contract. The SLA gives a detailed description of all aspects of the deal, such as the extent of the contract, the responsibilities of the network operator and the service provider and actions to be taken if one of the parties does not keep their part of the deal. Once the SLA is signed the service provider can start using the SCSs agreed on. Although the SLA is signed, both on-line authentication with digital signatures and on-line authorisation for use of the SCSs is needed.

There are predictions of rapid growth in the number of applications once the 3G mobile networks are available. Network operators will not be able to provide all these applications themselves. The applications provided may fall into four groups:

- Applications the network operators provide themselves. The reasons for providing some applications themselves range from being the most cost-efficient, to applications being so critical they have to provide them themselves.
- Applications the network operator needs to offer in their network, but chooses to let an external service provider run. These applications range from compulsory services the network operators are required to provide, to applications that are complementary to their own and able to generate extra traffic.
- Applications that a third party wants to run, with no other benefit for the network operator than the generated traffic.
- Applications that use SCSs from more than one network operator.

The largest growth will probably be in context-based services, and then especially in services using location information.

2.5 The Future of OSA

First some extensions and improvements already agreed on by the groups in 3GPP that are standardising OSA are described and then possible future enhancements are discussed.

The charging functionality of OSA will in the future allow the application to charge end-users' account for services provided. Both soft (e.g. download of music, software, video) and hard (e.g. CDs, books, DVDs) goods may be invoiced. The responsibility for the subscriber accounts can be assigned to the network operator or elsewhere. The network operator may also choose to be partially responsible for the account and send out invoices for traffic and subscription costs and leave the rest to a third party. Regardless of who handles the billing the means to divide the charging of a call controlled by an application between the network operator and the application provider is required.

Multi-Media Channel Control will be added. This will allow applications to control individual channels in IP Multimedia calls. The Multi-Media Channel Control functionality will be equivalent to the Call Control interfaces.

Future enhancements of the OSA API could include the following:

- *Logging functionality* is needed by the applications for charging and statistics. It can also be used for fraud detection.
- *Speech synthesis* support would be helpful in many applications such as reading of faxes and e-mails.
- *Speech recognition* support would be helpful in many applications such as automated directory inquiries.
- *E-mail and fax support*, with speech generation and recognition.
- *Unified Messaging* has been a hot topic for some time. Support for this would be useful.

3 Changes in the Business Model

So far we have seen little discussion around the topic of business models for UMTS but in opening up for external service providers this will also require a new model for how business is done.

Figure 8 shows a simplified view of the old business model. In both cases shown the telecom operator is responsible for delivering the service to the customer. In the model to the right a service provider is involved in provisioning the service, but the operator is still the one who delivers the service to the customer.

In the next figure a possible new business model that fits the new way of providing services is shown. Figure 9 shows an ideal layered model where each layer only has relations to the nearest layer above or beneath itself.

This is of course an idealistic model and the reality is far from that simple. An example of this is that it is likely that Service Providers also will provide services directly to the Consumer without the involvement of a Service Distributor. Other cross relations and other roles not mentioned here are of course also possible.

3.1 Network Operator

The Network Operator in this model provides the transport and access networks and is responsible for the operation, administration and maintenance of these networks. The Network Operator is also responsible for the different network resources; this includes the operation of the open interfaces, used by both internal and external service providers for the support of their services. This provisioning of open interfaces is the service provided by the Network Operator to the Service provider in the model in Figure 9.

3.2 Service Provider

The Service Provider is responsible for the operation, administration and maintenance of services and to provide these services to a Service Distributor. He is also responsible for maintaining the customer data used to configure the service for a particular user.

3.3 Service Creator

The Service Creator is the one who implements a service. This role can be filled by the Service Provider but can also be done by a company external to the Service Provider that in this case hosts or buys the service from the Service Creator. The Service Creator would have no other relations than to the Service Provider

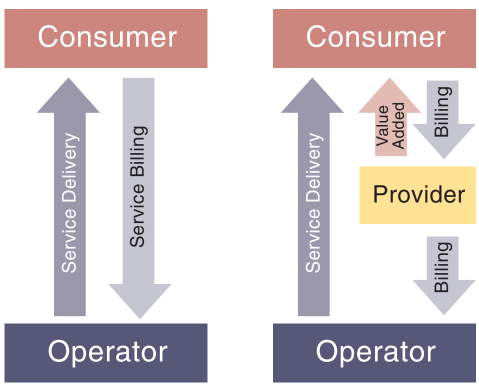


Figure 8 The old business model. From [10]

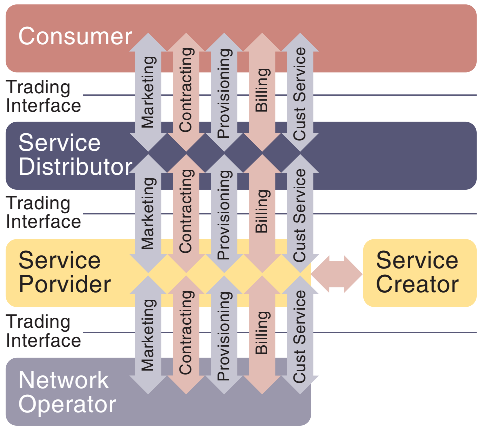


Figure 9 The new business model. From [10]

3.4 Service Distributor

The Service Distributor plays the role of being the one who handles the customers and is responsible for billing the actual customer for the use of a service. The Service Distributor has the freedom to bundle and combine services from multiple Service Providers before these services are offered to a Service Consumer. The Service Distributor is also responsible for the important task of marketing the services he offers to potential customers.

3.5 Service Consumers

The Service Consumer is the party who subscribes to and pays for the use of a service. The Service Consumer role can be split into Customer and User roles where the Customer can be a company and the User is the actual user of the service.

4 Summary and Conclusions

What opportunities then lie within the standardisation of APIs like OSA and Parlay for a telecom operator?

These interfaces will be able to support a wide range of new communication services to be built and deployed both within the network and outside of the network. This will give opportunities for service providers both external and internal to the telecom operator to create a diverse set of new "value added" services and make these services available to the mass market. The great increase in service providers and the relative ease of service development will also open the market for niche services directed at specific trades. These services have been too expensive for the end user up to now.

Both nationally and internationally the operator will have the opportunity of entering partnerships and alliances with external service providers giving the operator the ability to offer a broader and more varied spectre of personalised and integrated services. This will again lead to a greater loyalty among existing customers and also attract new customers. This is also likely to generate increased revenue as traffic increases and also generate a new source of income through the sale of service provisioning access to external service providers.

On the international arena the opportunity lies in creating international alliances with external service providers and distributors. As an operator is moving into new markets, the effort can then be concentrated on building and maintaining the core and access networks, and at the same time he can be able to offer a broad selection of services. It will also be easier for an operator to take already existing services for the national market and introduce those services internationally when moving into foreign markets.

As a service provider with long experience in providing services, the telecom operators will also have the opportunity to offer their services also for customers in the networks of other network operators nationally and internationally.

5 References

1 3G. 2000-11. *Stage 1 Service Requirements for the Open Service Access*. (TS 22.127 V1.1.1.)

2 3GPP. 2000-12. *Virtual Home Environment/Open Service Architecture Release 1999.* (TS 23.127 V3.3.0.)

3 3GPP. 2000-12. *OSA – Application Programming Interface – Part 1 Release 1999*. (TS 29.198 V3.2.0.)

4 3GPP. 2000-06. *The Virtual Home Environment Release 1999.* (TS 22.121 V 3.3.0.)

5 *The Parlay Group*. (2001, March 6) [online]. – URL: http://www.parlay.org/

6 *The UMTS Forum.* (2001, March 6) [online]. – URL:http://www.umts-forum.org/

7 Davis, S. *Parlay Concepts and Overview*. Boston, June 27–28, 2000.

8 *Parlay Group FAQs.* (2001, March 6) [online]. – URL: http://www.parlay.org/about/faqs.asp

9 *Parlay Group Past Events.* (2001, March 6) [online]. – URL: http://www.parlay.org/news/events/pastevents.asp

10 Parlay Group. *Parlay API Business Benefits White Paper 2.0.* January 12, 2000.

Youngster: Focusing on Future Users in a Mobile World

MARGIT KARLSEN AND ALF SOLLUND

Margit Karlsen (33) is Research Scientist at Telenor Research and Development in Grimstad. She holds a degree in Political Science and works in the field of Personal Communications, her work focusing on users and user issues.

margit-elise.karlsen@telenor.com

Alf Sollund (38) is a Research Scientist at Telenor Research and Development in Grimstad. He holds a degree in Opto-electronics and works in the field of Mobile Communications.

alf-martin.sollund@telenor.com

Introduction

During the past few years, research in mobile communication has been directed towards better, faster and cheaper access to traditional communications services. Now that GPRS (General Packet Radio Service) is up and running and UMTS (Universal Mobile Telecommunications System) is right around the corner, it is expected that Mobile Internet and Multimedia applications will take off. Young people between the age of 15 and 25 are an important target group in this regard.

This article will present an overview of the Youngster project. Since the Youngster project is only in its preliminary phase, we will not go into great detail on the technical aspects. We do however consider location services to be such an important technical building block that we will present a forecast and description of forthcoming technologies in this area.

The IST (Information Society Technologies) project Youngster under the 5th framework program will test new services for young mobile users while also looking at future concepts and needs. Our intention is to develop a more dynamic interface and new modes of use for young users.

IST Research Policy and Youngster

The IST 5th framework programme [1] is the European Community's (EC) research body. Unlike other joint technology driven initiatives such as EURESCOM, the IST is highly focused on improving the European community socially and economically rather than just implementing new technologies. The cornerstones are convergence of information processing, communications and media technologies. The program consists of four inter-related specific objectives:

- For the private individual, the objective is to meet the need and expectation of high-quality affordable general interest services.

- For Europe's enterprises, workers and consumers, the objective is to enable individuals and organisations to innovate and be more effective and efficient in their work, while also improving the quality of working life.

- In the sector of multimedia content, the key objective is to confirm Europe as a leading force, realising its full potential.

- For the enabling technologies, which are the foundations of the information society, the program objective is to drive their development, enhance their applicability and accelerate their take-up in Europe.

In order to be accepted, IST project proposals have to meet the socio-economic needs of European citizens and harmonize with EU policies contributing to European community added value. The utilization of project results must be visualized in an Exploitation Plan.

How will the Youngster Project meet the Ambitious IST Goals?

Quoting from the program strategy *"Our surrounding is the interface, to a universe of integrated services. This will enable citizens to access IST services wherever they are, whenever they want, and in the form that is most 'natural' for them"*.

Youngster fits this statement perfectly through its project goals that involve the creation of an interactive user interface in which support is provided for the easy creation and usage of the offered service features by a set of *tools* and new *content formats*. This means that the services the user can access through his/her mobile terminal is dependent upon the user's context, interests and choices as well as the interaction between these. Services can be chosen by the user, automatically or discarded depending upon the user's own interests and context (including location).

Youngster intends to make these new technologies *attractive to young people*, and to *use them as a user reference group*, hence the project name. Youths were chosen as a target group for this project based on the fact that they tend to adopt new services and technologies more readily than older groups, their lifestyle is more mobile than ever, and they are interested in personalising their user interfaces.

Youngster Team

The Youngster project is broken up into four distinct parts called "work packages". Work package 1 deals with the development of new business models, the mobile service platform is developed in work package 2, and this platform is then tested on a selection of users in work package 3. The fourth work package involves project management and dissemination of results.

A strong consortium backs up the Youngster project with a basis in professional experience and economic power. Each partner has specialized experience in different spheres of the project such that the project can be completed with maximum effect.

Deutsche Telekom represented by *T-Nova*, based in Berlin, has a great deal of experience in managing large multinational research and development projects. They are leading work packages aimed at project management and definition of new business models and application scenarios.

Siemens AG will manage work package 2, dealing with platform development. They bring to the project expertise in mobile middleware, architectures, context-aware mobile applications and UMTS services.

Sony International (Europe) is a leading manufacturer of consumer and professional electronic equipment. They contribute with their expertise in the areas of audio, video and telecommunications to the Youngster project. Both the Sony and Siemens teams are almost exclusively dedicated to the development of the mobile service platform on which applications and services will be developed.

Telenor, Norway's largest telecommunications company, brings to Youngster a strong multidisciplinary group, which includes both technical and non-technical specialists. Telenor will manage work package 3, which deals with field trials. Field trials will be administered in Norway among a selection of Norwegian youths.

The French company *Tecsi* will contribute to the definition of application scenarios addressing the needs and expectations of the users, contribute to the identification of requirements for the Mobile Service Platform services, design and implement generic application frameworks addressing the implementation of services in the areas of mobile eBusiness and eCommerce, and will specify, design and implement one or more demonstration services.

The *Heriott-Watt University* in Scotland has experience in different aspects of personalization and will contribute to the research and technical developments in this area – whether this is personalized presentation by the information/service provider or tools for the personalization of presentations for use by youngsters themselves. They will also contribute to the definition of the Mobile Service Framework and development of the Platform.

The *Norwegian Broadcasting Corporation* (NRK) is onboard as content provider. NRK brings its experiences in producing and providing content across different platforms. As a market leader in Norway on TV, radio and electronic media (digital TV, digital radio, Internet and telecom) NRK will contribute to the field trials by providing results from ongoing youth-research projects in NRK. NRK will also contribute in the definition of services, business models and by providing content for the application prototypes.

Youngster Objectives

The main goals of the Youngster project is to develop a new open, active, mobile multimedia environment which includes the following:

1 Development of an open, active Mobile Service Platform. The main characteristics of this platform will include:

- Components for the personalization of services;
- Customer specific information filtering and "pull" services;
- Context- and location-aware features;
- Seamless adaptation of services to characteristics of different end systems and networks;
- Content management and content interpretation tools;
- Mechanisms and components to ensure security and privacy.

2 The development of a new generation of enhanced mobile services using the enhanced Mobile Service Platform. These services will include the following features:

- Detection of, and smart interaction with other devices in the user's environment;
- Personalization and adaptation of services, user awareness;
- Integration of location aware services.

3 Verification of service and platform components in trials with a suitable test community. The trials will focus on young people and will be conducted in Norway by Telenor and NRK. The results of the trials will be evaluated and success will be assessed in terms of the response from the young people who participate in user testing.

4 Definition of new business models for providing the services to various groups of users.

Keeping in mind our focus on a young user group, we anticipate that the current business models (from the Internet or the mobile service area) will not be ideal for the target services. Thus new business models will be examined to ensure that developments arising from this project can be marketed to different user groups.

Development of a Mobile Service Platform

In order to realize our vision of an open active mobile multimedia environment, Youngster plans to develop an open Mobile Service Platform (MSP). This MSP will enhance existing platforms and incorporate open technologies such as Web servers, WAP gateways, Java engines and other technologies such as Lightweight Directory Access Protocol (LDAP).

The open MSP will be a unique platform for mobile services incorporating a wide array of context and personalization features. These include location information, time/date, social context and various other parameters relating to the current context combined with a high degree of personalization. The way this works is that MSP will consider details related to the user such as pre-selected interests, history of services used or used by others with the same interests and the current context. From this information, the MSP can "pull" services of potential interest to the user.

A central development on which the entire MSP concept depends is the "user retrieval system". This will hold individual user information specific to the user (age, gender, interests, preferences, etc.) as well as a wide range of context information, including location, and information directly related to the user's context (e.g. weather). Information in relation to this will be both static and dynamic. Some information will be held by the user, some by the service provider and some by the underlying support services. Security is an important issue in relation to this and is discussed in more detail later in this article.

Active space in the MSP will be used to store user preferences and is an additional module that assesses content and either pulls services to the user's menu or filters them out.

NRK will provide generic content of different media types that can be accessed as-is or combined with third-party content such as user comments. Community members or groups of friends can receive information or content as the user desires. This can either be original content or content manipulated by the user.

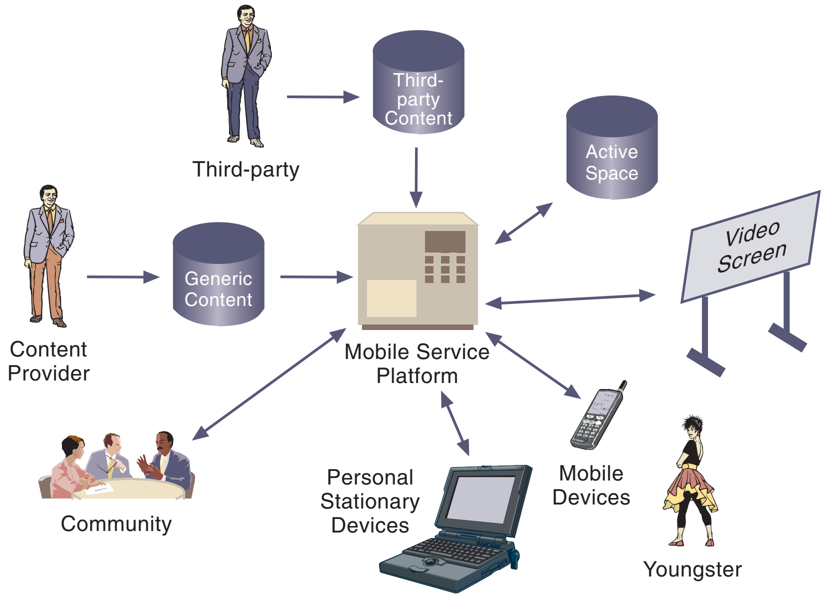


Figure 1 Youngster base system

Content can be accessed via a variety of devices (both mobile and stationary) and networks. This will allow users to access information on either their own or other devices in the immediate area (e.g. video screen or headphones in the user's environment). It also allows for the manipulation of content that they have downloaded for future circulation or storage.

Accessibility via a Wide Variety of Devices and Networks

The MSP will enable the user to use their device as a controller for other local devices. In order to accomplish this, one needs the ability to detect compatible devices that are both in the area and available (i.e. not in use by another user). The user can then select the device they wish to use. In addition, local devices can also deliver context information. An example of this would be if a user's mobile device was not equipped with a positioning device, but another device in the local area was, then the user might download information this way as an alternative.

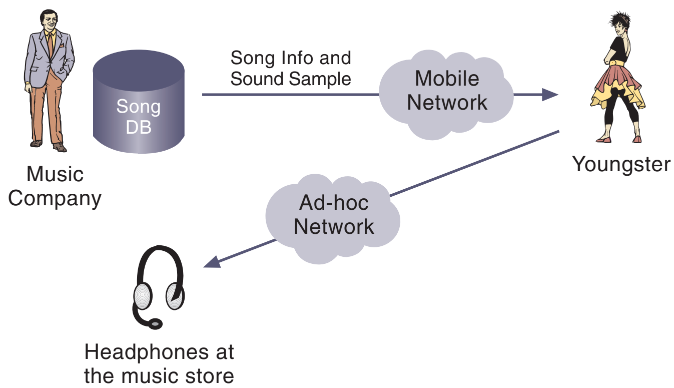


Figure 2 Interaction between devices

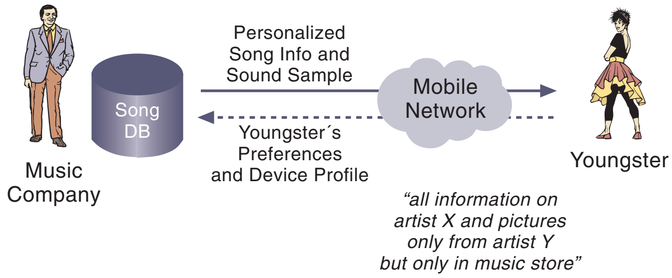


Figure 3 Service personalization

Location-awareness

Location is a very important part of the user's context. Location as a service will be enabled by the use of satellite-based systems or upcoming network-based positioning systems. The technical challenges are discussed in more detail later in this article. Map servers and Point-of-interest servers (PoI) may be combined with other services developed by the project to offer more value-added services.

User-awareness and Personalization

The personalization feature will allow the user to control how information is presented to them as well as to other users and devices and to filter information selectively depending on the user's interests and current context. The filtration of information is aimed at reducing the amount of irrelevant information sent to the user's device. This is connected to the user profile that helps to pull or discard information selectively.

Figure 3 demonstrates how the user can receive services/information adjusted to their needs. In this case, the youngster has stored information about his music preferences in his user profile and how he wants to receive this information. This means the user can decide the destination of information output. It could be his mobile phone, a video screen or speakers nearby.

Working Methodology

User Groups

In order to validate the goals in this project, our aim is to use a young user group, hence the title "Youngster". Youths between the age of 15 and 25 will be targeted and asked to participate in field trials that will enable us to develop services directed especially towards this group.

Our target group of young people is an especially important group. Young people are often the group most interested in what the latest technology has to offer. They are an extremely mobile group with active social lives and the desire to demonstrate this outwardly such that it is obvious to others. The Youngster system will allow them personalized services that are adaptive. This means that each youngster can create their own services simply in response to their own individual needs. The tools for creating these services will lie within the Youngster system and it is expected that youths will be eager to acquire the ability to put them into use. This in turn stimulates innovativeness.

Although the primary focus is on young people, the Youngster project is also aimed towards business people. This includes services for eCommerce and eWork. These services are not quite as differentiated, and requirements for these services are therefore easier to meet than the varied services youngsters will demand or create themselves. Older users demand more advanced services (e.g. banking and time schedules) that will help them co-ordinate their daily lives.

Non-technical Challenges

Privacy

The "user information retrieval system" is based upon a dynamic user profile. This raises concerns as to where the user information will be stored and how access will be controlled. Mechanisms such as logging, automatic negotiation of privacy agreements and others will be under consideration to ensure a *user-controlled* privacy. Our idea is to allow user-controlled views which will restrict the set of context-information visible to service providers.

The ethical issues, particularly that of *privacy*, will affect many aspects of the Youngster system. Storage of context (including location) information as well as storage of user profiles are two major examples. For this, the plan is to follow already established general principles (95/46/EC and 97/66 EC), which require that:

- Explicit consent of the subscriber is obtained.
- Complete information about the use and storage of the data is provided to the subscriber.
- The data is only used for the purpose for which it was collected.
- Personal data is erased after use or made anonymous.
- The user has the possibility to restrict transmission of location information.
- Network security is ensured.
- Data is not transferred to a third party without the consent of the subscriber.

Surveillance

While there is a great deal of useful and interesting information available on the Internet, there is also a great deal of objectionable content found there as well. While parents can filter information or sit down with their youngsters in front of a stationary computer, mobile youths will often be unsupervised while using their mobile terminals putting youngsters at a greater risk to encounter unacceptable content such as violence, hate/racism, pornography, etc. One can also add the possibility of manipulative content such as marketing, which is directly aimed at youths.

Functions such as content management will be important aspects to consider, as well as how much parental surveillance is necessary or acceptable to more independent and mobile youths.

Technical Challenges

What we have seen so far in the race towards UMTS and GPRS is the significance of technology platform choices, frequency licenses and an overall high bandwidth hype. Less has been done in order to create content and services suitable for the new mobile networks.

Today, services available via SMS (short message service) include traffic reports, flight/airline information, news, information, personal reminders and much more. With the introduction of GPRS and UMTS, there will be a demand for more specialised and complex services. Youngster will take these ideas and the technology a step forward.

Given the importance of location services, a presentation of the expected evolution of location services seen from the telecom operators' viewpoint will be presented in this article. For further details please consult [2] and [3].

Location Services

Location awareness will be an essential technical building block in the Youngster project. Youngster will not aim to invest work in developing these methods, but rather pilot location services and test interfaces for location services. The definitions throughout this article are:

- *Location services (LCS)* can be defined as the capability to provide the geographical location of a mobile terminal with a given QoS.

- *Location based services:* Use of geographical location provided by LCS to enhance other services.

During 2000 – 2002 mobile operators will test mobile-based systems and network-based systems such as *Cell-ID in combination with other methods like Timing Advance (TA)*. These are easy to implement in today's mobile communication systems, but offer limited accuracy, ranging from only 100 to 600 m. Multiple overlaying from different location estimates, or in the case of TA, the use of the measurement data before they are transformed to the TA standard parameter can help to improve accuracy. The multiple overlaying from different location estimates can still fail because too few base stations are accessible, especially in rural areas. These parameters are already implemented, offer a simple way to determine location, and can be used for applications with a low demand for accuracy. TA is also specified by ETSI [4] and proposed for assistance of other LCS (Location Services) methods, and as a fallback.

The first implemented Mobile based Terminal location systems will use SIM application toolkit algorithms for position calculations and SMS for transmitting its location to others (e.g. communication with location servers and applications). The development will move away from SIM toolkit based systems / SMS and instead towards systems based on new WAP versions (1.2 and beyond) and/or GPRS/UMTS.

Further, at the start of 2002 the strategy forecast for mobile operators will be to implement combined network-based systems like E-OTD (Enhanced Observed Time Difference) / GPS systems (Broadcast of assistance data from GPS). Combined network-based TOA (UL-ToA, Uplink Time of Arrival) / GPS assisted systems are also possible, but may eventually turn out to be too high an investment for operators.

These network-based systems will:

- Enable accuracy for GPS assisted systems from 10 – 100 metres;

- Offer an average accuracy of 100 – 200 metres for network-based E-OTD or TOA systems;

- Require a minimum of three base stations and certain time synchronization between these (E-OTD)

  - This time synchronization will introduce other benefits such as the possibility to reduce interference caused by adjacent sites;

  - The requirement of at least three base stations in sight will cause problems in rural areas where the ability to reach more than one base station is lower;

- Offer the possibility to locate mobiles without or only with software changes, and to implement a self-positioning system (E-OTD).

In general more investments are needed on the network side for the network-based positioning. Network-based is the only architecture that can locate existing mobiles, because normally no modification of the mobile phones is needed. This is a great advantage compared to the other architectures. In a mobile-based solution the positioning is part of the functionality of the mobile terminal. This translates into added complexity and cost due to added software and hardware for the terminals. The mobile calculates the position by use of signals from the network, e.g. base station(s).

All current network-based mobile location services exist on today's 2G mobile data and analogue cellular networks. No location technology is at this time commercially available for 3G digital air interfaces such as TDMA, W-CDMA and CDMA.

An interesting fact is that most of the systems described here will improve their accuracy as terminals are moving. In E-OTD and ToA this is due to the "de-correlation or whitening" introduced in the multipath channel by the higher mobile speed, and for other techniques (signal strength, direction finding) the average of measurements over a distance eliminates the problems caused by multipath fading and shadowing.

ETSI considers all GSM LCS (Location Services) a network feature, and not a supplementary service. This means that ETSI specifies LSC whether the Mobile Station (MS) supports LCS or not. So far ETSI has specified the following methods:

- TOA (Time of Arrival);
- E-OTD (Enhanced Observed Time Difference);
- TA (Timing Advance);
- GPS assisted.

GPS-based positioning location systems are currently under development and standardization [7] and should be available for the next mobile phone and mobile communication generation in 2002. However, power consumption, cost, and some technical challenges are yet to be considered. Some of these challenges are met with *Assisted GPS systems*. These utilize a GPS reference network (or a wide-area differential GPS network) that is connected with the GSM network. The assistance data from the reference network can be used by the MS to increase performance of the GPS sensor and

- Reduce the sensor start-up time;
- Increase the sensor sensitivity; and
- Consume less handset power than conventional GPS does.

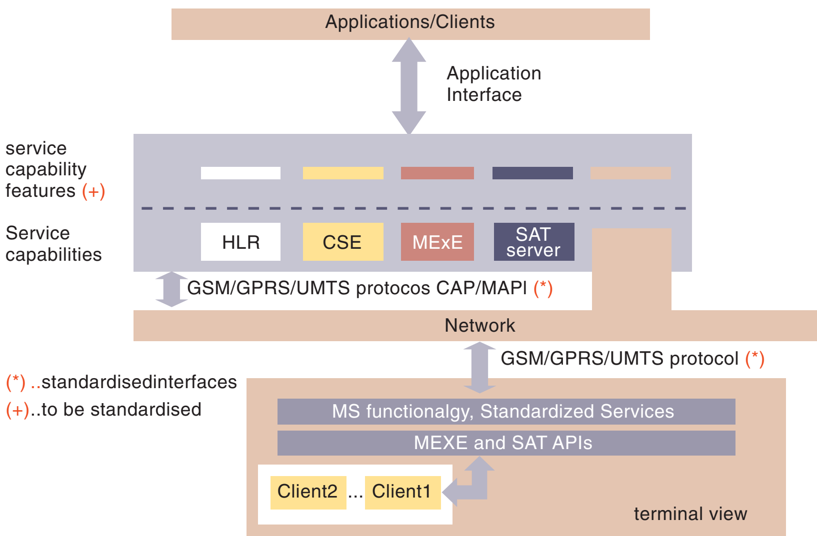
Service Platform and Interfaces

There is currently ongoing work on an OSA lab at Telenor R&D for testing the OSA architecture and interfaces.

The work from 3GPP on standardizing an open architecture for service platforms (OSA) in UMTS networks is highly relevant, and will therefore be presented as a reference architecture for services. The OSA (Open Service Architecture) describes an architecture that enables service providers to make use of network functionality through open standard interfaces [5], i.e. the OSA API. The network functionality is described as Service Capability Servers. One of the addressed Service Capability Features is User Location (SCF), specified in OMG (CORBA) IDL-format. The reference system chosen for the coding of locations is the World Geodetic System 1984 (WGS 84). The architecture is illustrated in more detail in Figure 4.

The purpose of the OSA API is to shield the complexity of the network, its protocols and specific implementation from the applications. This means that applications do not have to be aware of the network nodes a Service Capability Server interacts with, in order to provide the Service Capability Features to the application. The specific underlying network and its protocols are transparent to the application *enabling external service providers' access to the platform and its underlying networks by using an open API.*

Figure 4 OSA architecture



Within the OSA concepts the following Service Capability Servers are identified [6]:

- CAMEL Service Environment;
- WAP execution platform;
- Home Location Register (HLR).

The stage 3 of OSA methodology is based on UML and OMG Interface Language (IDL).

Conclusion

The demand for context-aware features in mobile communications is expected to rise dramatically with the introduction of GPRS and UMTS. The technical advancements, especially in location services, are new and exciting building blocks for telecom operators. There are however new challenges to be met, among them how to open the service platform to third parties, and with this enhance a dynamic service creation. In order to fully utilize these new techniques a new and more open service architecture will have to be investigated. The Youngster project aims to research how this service platform can be realised.

Young people are most likely the one group to adopt the new technical possibilities fastest. The Youngster approach is based upon service creation, interaction with the environment, personalization and context-awareness. All these aspects combined are expected to stimulate the creativity that exists in everyone and in turn contribute to the development of more innovative and useful services for the future.

However valuable these new possibilities may be, they do not come without concerns. Ethical and security issues are of great concern and will be significant aspects in the Youngster project.

As a result, Youngster hopes to inspire innovation among the younger user groups and by their feedback help develop services that all users feel enhance their everyday lives. Backed by a qualified team, Youngster has the means to successfully meet the expectations that are anticipated both by the Youngster partners and by IST.

References

1 http://www.cordis.lu/ist/

2 Schneider, C. *Mobile Location Systems.* Kjeller, Telenor R&D, 2000. (R&D Note N 23/2000.)

3 Sollund, A. *Location services for handheld devices.* Kjeller, Telenor R&D, 2000. (R&D Note N 56/2000.)

4 ETSI GSM 02.71, 03.71, 04.35

5 3GPP. 2000. *3rd Generation Partnership Project; Technical Specification Services and System*. (TS 23.127.)

6 3G. 2000. *Technical Specification Group Core Network; Open Service Architecture; Application Programming Interface.* (TS 29.198 version 1.0.0.)

7 http://www.etsi.org/news/mobile%20news/lcs.htm

Towards User-centric Communications with the Virtual Terminal

DO VAN THANH, ERIK VANEM AND DAO VAN TRAN

In an Information and Communication Technology age, the society is flooded with electronic devices. The number of devices that each individual needs to master in order to function properly, has increased dramatically from the last decade and will continue to increase with an accelerating speed. Without assistance the poor non-technical user will sooner or later succumb to the technology pressure. The Virtual Terminal has the goal of coming to the user's rescue. It will help the user unify all the heterogeneous electronic devices and make them serve him according to his own will and preferences.

Do van Thanh (43) obtained his MSc in Electronic and Computer Sciences from the Norwegian University of Science and Technology in 1984 and his PhD in Informatics from the University of Oslo in 1997. In 1991 he joined Ericsson R&D Department in Oslo after 7 years of R&D at Norsk Data, a minicomputer manufacturer in Oslo. Since 1999 he has been in charge of the User centric services research activities at Telenor R&D, with a focus on VHE, OSA and next generation mobile applications. He holds an associate professor position at the Center for Technology at Kjeller. He is author of numerous publications and inventor of a dozen patents.

thanh-van.do@telenor.com

Erik Vanem (28) holds an MSc in Physics from the University of Oslo in 1996. Before joining Telenor R&D he worked as geophysicist at PGS Reservoir, as research assistant at the Norwegian Defense Research Establishment (NDRE), and as a physics teacher at the Oslo University College. He is now working with User centric services at Telenor R&D.

erik.vanem@telenor.com

1 Introduction

The ultimate goal of communication services is to assist human beings in their communication with other human beings. They are aiming at either extending the possibility of communication or improving the quality of it beyond what the nature originally allows. Indeed, communication in nature can only be done over short distances. The telephony service allows people to lead a conversation while being far away from each other. Mobile telephony goes a step further by allowing people to talk to each other while on the move, anytime and anywhere. There is no doubt that these two services are so far the most successful ones in telecommunication. We believe that the reason for their successes is the fact that they succeed in helping the users with their communication. They are in fact user-centric and have their focus on the user. In order to be successful, future communication services should also be user-centric and not technology-centric. Although the current trend is to produce more advanced terminals with more fancy looks, we are convinced that it is also crucial to assist the user in the management of his terminals. Ideally in communication, terminals should be transparent. They are still there but the user does not need to pay attention to them and can concentrate on talking with his counterpart. In this article we present the virtual terminal concept, which is aiming at reducing the burden of terminal management for the users and contributing to make the terminals more transparent to them.

2 Why the Virtual Terminal?

2.1 Multiple Communications Devices

Nowadays the user is confronted with several different communications devices as for example a plain-old telephone, a mobile phone, a cordless phone, and a PC or a workstation that acts as a multimedia terminal. All these devices are autonomous and function independently of each other and without any co-ordination. In fact they are not even aware of the presence of other devices. As the owner the user is required to handle them all and does not always succeed since as a human being he cannot perform many tasks at the same time. For example, both the plain-old telephone and the mobile phone start ringing simultaneously and force the user to alternate between calls or terminate one of them. Another inconvenience is the repetition of the same tasks for each device. The user has to define his profile and preferences on every device. For example, he has to enable a voice-answering service both on his cellular phone and his plain-old telephone when he is busy and does not want to be disturbed.

From the user's point of view it is desirable to be able to handle all the mobile devices uniformly or preferably to consider them as "one big terminal". Collaboration and co-ordination between devices are hence required. With UPT (Universal Personal Telecommunications) [1], the user is allowed to register and make use of several terminals but only successively and not simultaneously.

2.2 Multiple Devices of Different Functions

The user may need and want to use several devices with different functions and uses such as a mobile phone, a laptop, a PDA (Personal Data Assistant), a digital camera, a printer, a scanner, an electronic paper, etc. These devices can be connected together through a cable, a physical plug or a wireless link such as Bluetooth [2,3] or IEEE 802.11 [4]. More than one device may have connections to different networks. The connections can be of different types such as wired or wireless, narrow or high bandwidth, single or multilink.

The devices must be integrated in such a way that they work together. It is also important that they collaborate to offer a coherent user interface to the user. It should be possible for the user to talk to a person on the mobile phone while exchanging pictures taken by a digital camera and hand-written notes, sketches from the electronic paper, etc.

Dao van Tran (50) obtained his MSc in Computer Science from the University of Oslo in 1996. He joined Telenor's Network Division in 1986 and Telenor R&D in 1996. Since then he has been involved in several projects, among them the EURESCOM project EU-P608: TINA Concepts for Third Generation Mobile Systems, Service Management system for SDH transport network, Management of Access Network, and 'Virtual Terminal'.

dao-van.tran@telenor.com

Existing solutions are based on master-slave architecture where the PC acts as the master which controls all other devices. Communications with the network are also done via the PC on a unique link. Actually only the PC is known by the network and if it stops working then communications with other devices will also stop. It is worth noting that the PC must also be configured to communicate with the other devices, and all the necessary communication software, i.e. drivers, must also be installed in the PC. The configuration is static and it is not possible to add or remove devices on the fly.

2.3 Combination of Stationary and Mobile Devices

The user may have both stationary devices such as PC, workstation, TV, loudspeaker, etc. and mobile devices such as mobile phones, PDA, display, microphone, etc. These devices are autonomous and are connected to different networks like for example ISDN, Ethernet, TV cable, GSM, etc. It is desirable to be able to move input and output streams from mobile devices to stationary ones and vice versa seamlessly without having to stop and resume the ongoing session. With such a feature the user can make use of all devices that are available at the places he visits. There must of course be security measures to prevent the use of unauthorized devices.

2.4 Unique User Profile

Nowadays with several devices the user has to set up and define his profile for each of them. When he wants to modify something he has to do it on all of them. Another inconvenience is that when replacing older devices with newer ones, the user must also move his profile. It is desirable seen from the user's point of view to have a unique user profile to deal with.

3 What is the Virtual Terminal?

If it were cumbersome to handle the devices separately, then the most natural solution would be to handle them together as one. With the Virtual Terminal concept all the different communications devices will be considered as one big terminal with multiple input and output capabilities. The services, e.g. telephony or basic call as it is technically called, will first be delivered to the Virtual Terminal which will select and deliver to the device that is most appropriate at any time.

The terminal is no longer an integrated and recognisable device but a set of distributed devices that allow access to services. The terminal is no longer present or the user does not have to pay attention to the terminal. The terminal is in the sense "transparent", i.e. it is still there but the user sees through it.

- With the Virtual Terminal concept the user can unify all his communication devices such as a plain-old telephone, a cordless phone, a mobile phone, and a multimedia terminal like an H323 terminal such that they together behave like one device.
- With the Virtual Terminal concept the user can unify all his computing and communication devices such that they can behave as one device with multiple capabilities.
- With the Virtual Terminal concept the user can make use of his mobile devices and also stationary devices at the visiting site.
- With the Virtual Terminal concept the user can move the input and output of a session from a set of devices to another one. For

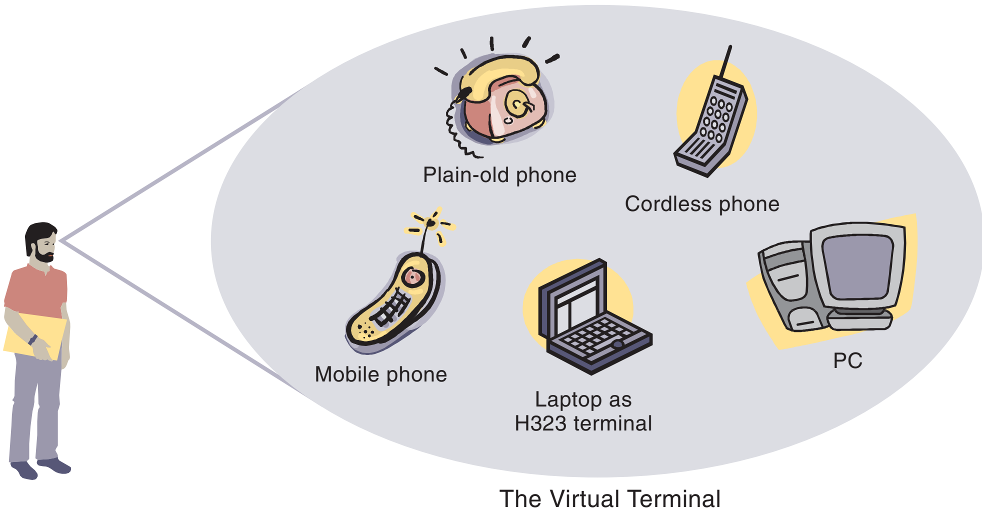


Figure 1 The Virtual Terminal Concept

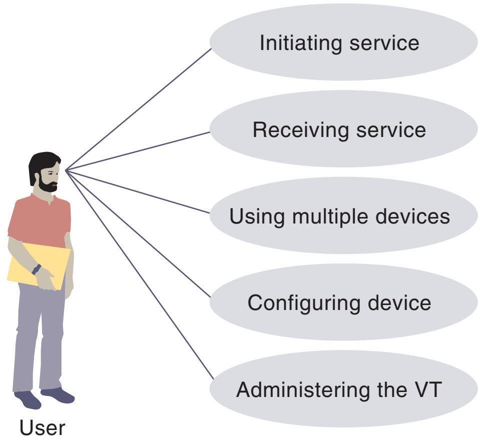


Figure 2 The first use case diagram for the Virtual Terminal

example, he can move visual output from a mobile device with small display to a larger and better stationary screen.

- With the Virtual Terminal concept the user can define, add or remove the devices that are included in the large Virtual Terminal.

- With the Virtual Terminal concept the user can set up and modify his preferences for all devices at one place.

4 What does the User want from the Virtual Terminal?

In a user-centric approach, we try to figure out what the user expects from the Virtual Terminal by elaborating use cases. Use cases are used to describe the outwardly visible requirements of a system [5]. Identifying the use cases of a system is in other words equivalent to identifying what the users require of the system.

For simplicity, we want to distinguish between two different types of services and two types of devices as follows:

- By communication services is meant human-to-human communication. They can be either synchronous (voice telephony, chat, multimedia services, etc.) or asynchronous (e-mail, SMS, etc.).

- Computing services or data services include communication with machines, and these services can also be synchronous (games, etc.) or asynchronous (word processing, etc.).

Furthermore, a communication device is a device offering communication services. A computing device is a device allowing data services and communication with other computing devices.

In addition there are other electronic devices and peripheral devices, which can handle different types of output and input streams, e.g. TV screens, loudspeakers, digital cameras, microphones, printers, scanners, etc.

The Virtual Terminal should be able to act like a regular terminal, i.e. the user should be able to initiate both communication and data services via the Virtual Terminal. Likewise, he should be able to receive incoming services on any device included in the Virtual Terminal. The Virtual Terminal should also be able to handle the cases where the user wants to use multiple devices for the input and output of his services and to control the input and output flows. In some cases it is also desirable to dynamically configure some of the devices belonging to the Virtual Terminal, which should allow the user to do this or even do it automatically for him. In addition it should be possible for the user to administer his Virtual Terminal in a straightforward way. We thus identify five main use cases, as represented by the use case diagram in Figure 2.

These use cases are quite general, but it is possible to go into more detail by identifying more specific use cases, which are said to extend the most general ones.

4.1 Initiating Services

As we have already distinguished between different types of services, we distinguish between *initiating communication services* and *initiating data services*. The following three use cases extend the *initiating communication services* use case: *initiating directly* (the user specifies directly the counterpart's address or identifier), *initiating using an address book* (this address book can be contained in the Virtual Terminal and thus be reachable from all the user's different devices), and *initiating a group* (the user specifies a group of persons he wants to communicate with). The extended use cases are illustrated in Figure 3.

4.2 Receiving Services

We define three use cases extending this use case, which cover the alternatives the user has when receiving an incoming service request: *accepting a service request* (the user wants to accept the offered service and get it delivered to an appropriate device – be it a local device, a remote device or even to several devices simultaneously), *receiving in mailbox* (the user does not want to receive the service at the time of request, but wants to record a message in his mailbox and look at it some other time), and *refusing a service request* (the user does not want to receive the service and the request is turned down. No connection to any of the user's devices is being made). These use cases are illustrated in a use case diagram as shown in Figure 4.

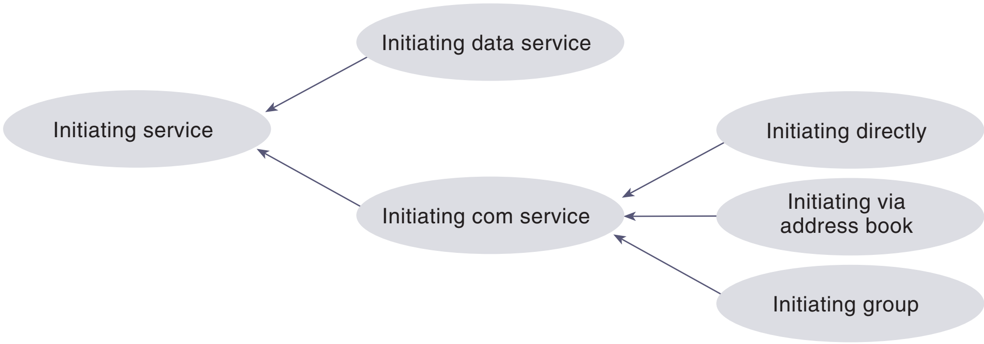


Figure 3 Use case diagram for initiating services with extensions

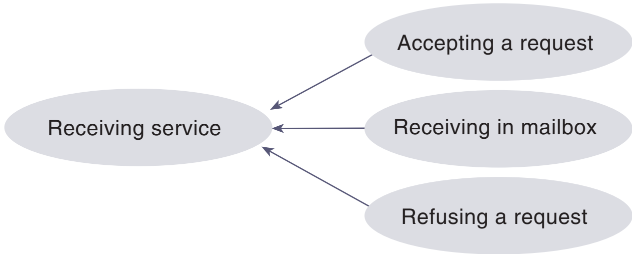


Figure 4 Use case diagram for receiving services

4.3 Using Multiple Devices

Since the Virtual Terminal may contain several devices, it opens the possibility for the user to use multiple devices, both simultaneously and successively switching from device to device in a dynamic way.

The user may want to change from one device to another or to use multiple devices at the same time. These options are covered by this use case. *Changing device* includes changing from one device to another during an ongoing session and changing offline, i.e. changing which device to receive subsequent incoming service requests on. *Adding/releasing device* covers the cases where the user wants to use multiple devices during the same session in a dynamic way, i.e. the user can add and release other electronic devices or peripheral devices to or from his ongoing session. In this way, the user can specify where the input and output streams are to be delivered. In order to allow the user to change devices and use multiple devices at the same time, we introduce the use cases *activating a device* and *deactivating a device*. Activating a device during an ongoing session will include it, and deactivating a device will exclude it from this session. This switching between different devices can be done either manually by letting the user specify explicitly which devices to use at any given time, or it can be done automatically, determined by the Virtual Terminal when nearby devices are discovered. The use case diagram corresponding to the use cases extending *using multiple devices* is shown in Figure 5.

4.4 Configuring the Devices

Different devices sometimes need some sort of configuration in order to be able to perform their tasks. For example, this can be installation of drivers or other necessary software. The *configuring the devices* use case lets the Virtual Terminal help the user do these configurations, either automatically or initiated explicitly by the user. The necessary software needs only be available somewhere in the network for the Virtual Terminal to be able to find, download and install it on the device in question.

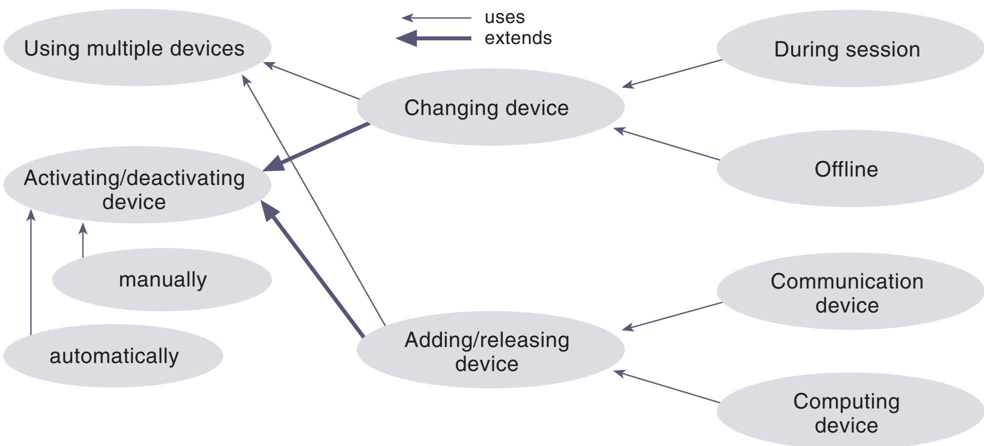


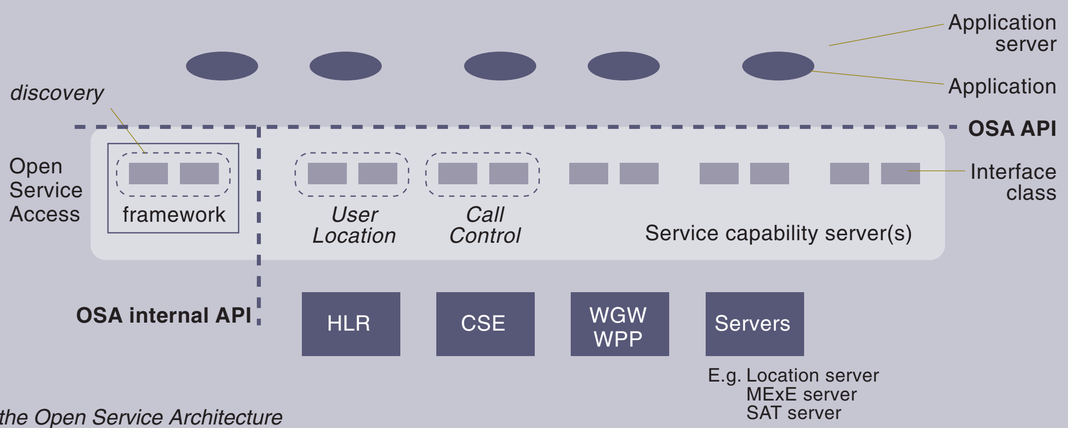
Figure 5 Use case diagram for using multiple devices

Box 1 – The Open Service Architecture (OSA)

3GPP has standardized an Open Service Architecture. This architecture defines an open API for design, implementation, control and execution and application provided by network operators and third party providers.

Applications can be network/server centric or terminal centric. Terminal centric applications reside in the Mobile Station. Network/server centric applications are running outside the core network and make use of service capability features offered through the OSA API. (Note that applications may belong to the network operator domain although running outside the core network. Being outside the core network means that the applications are executed in Application Servers that are physically separated from the core network entities.)

The Open Service Architecture (OSA) defines an architecture that enables operator and third party applications to make use of network functionality through an open standardised API (the OSA API). OSA thus provides the glue between applications and service capabilities provided by the network, and applications become independent of the underlying network technology. The applications constitute the top level of the Open Service Architecture (OSA). This level is connected to the Service Capability Servers (SCSs) via the OSA API. The SCSs map the OSA API into the underlying telecom specific protocols (e.g. MAP, CAP, etc.), and are therefore hiding the network complexity from the applications.



Overview of the Open Service Architecture

The SCSs provide their features to applications in Service Capability Features (SCFs). For Rel'99 the following Network SCFs are defined: Call Control, Data Session, User Location, User Status, Terminal Capabilities and Message Transfer. Note that call control only supports one-to-one calls and the applications cannot initiate a call.

All interfaces are specified in OMG IDL, which makes OSA API independent of programming language, platform and Operating System. It is up to the implementers to map the OSA API to other proprietary protocols like H.323 and SIP.

Current Release of OSA is Release 4 and the specification and IDL files are available for free download at www.3gpp.org.

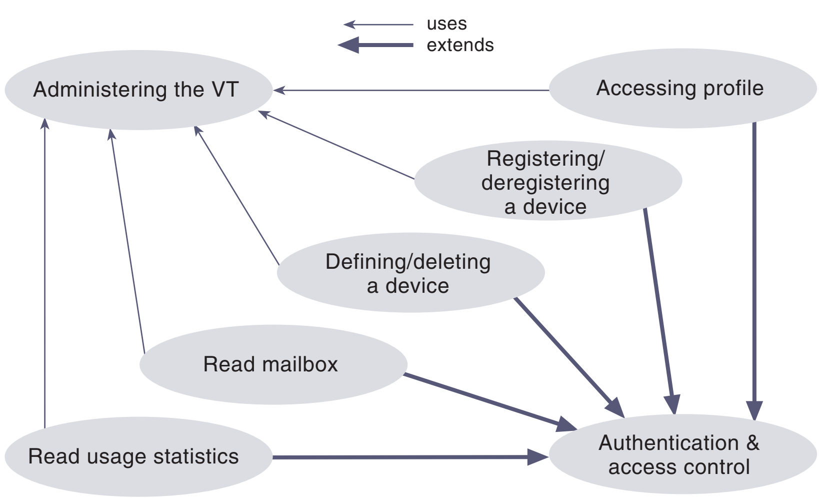


Figure 6 Use case diagram for administering the Virtual Terminal

4.5 Administering the Virtual Terminal

This article identifies five extensions of this use case: *accessing profile* (which allows the user to read and make changes to his profile), *defining/deleting a device* (where the user specifies which devices to be constituents of the Virtual Terminal, i.e. which devices are known by the Virtual Terminal and hence can be used for future sessions), *registering/deregistering a device* (where the user can specify for any given time which specific device to be registered at), *reading mailbox* and *reading usage statistics*. Admission to the Virtual Terminal should be restricted to the user only. This means that all of the five use cases extending *Administering the Virtual Terminal* mentioned above involve a use case *authentication and access control*. These use cases are shown in Figure 6.

5 How should the System Architecture be?

In order to realise the Virtual Terminal concept as described in the aforementioned visions and demands a terminal management and co-ordination function called Virtual Terminal is needed. In short, such a function should have the following capabilities: It should continuously maintain, monitor and update the configuration of the Virtual Terminal, i.e. it knows exactly what devices are present and active in the Virtual Terminal. It should also be able to multiply and deliver streams from applications to respective devices and to unify and deliver streams from devices to respective applications.

There are three alternatives concerning the location of this Virtual Terminal:

- On a user's mobile device;
- On a user-owned stationary device such as a PC;
- On a server somewhere in the network, owned by a service provider.

The first alternative has the advantage that it is closer to the devices and hence easier to implement. The major drawback lays in the fact that the mobile device might lose contact with the system and it is more exposed to damage or theft. The second alternative is not convenient since it requires the user to have an expensive stationary device, which is not always the case. Finally, although most challenging to realize, the last alternative proves to be the best since it is the most flexible. It does not require that the user owns an expensive terminal, and roaming is provided to allow the user to move away from his home domain while keeping in touch with his Virtual Terminal. This conclusion regarding location is in agreement with other research projects' conclusions [6].

Figure 7 shows the proposed system architecture of the Virtual Terminal. Other applications are only aware of the Virtual Terminal and not of the user's many devices. Applications will deliver services to the Virtual Terminal, which again has the responsibility to ensure that the delivery to the user is done via the most appropriate device. From a technological viewpoint, the Virtual Terminal object could reside on a standard server that is connected to an IP-based network as shown in Figure 8. All the networks are connected together by gateways.

6 How could the Virtual Terminal be Implemented?

In this section we will study successively the five main use cases and discuss the possible implementation alternatives.

6.1 Initiating Services

By service is meant both communication and computing services. The user has at his disposal a set of devices and can initiate services from any of them. For the use cases Initiating data service and Initiating directly communication service no additional functionality than the existing ones on the device is required. The user can just select the device he wants to and initiate service directly from there without going through the Virtual Terminal. However, if he later on wants to swap to another device the initiation must have been done through the Virtual Terminal in the first place or this will not be possible.

Further, for the use cases Initiating communication service via address book and for a group, additional functionality is required if a common address book for all devices is wanted.

Figure 7 The Virtual Terminal Architecture

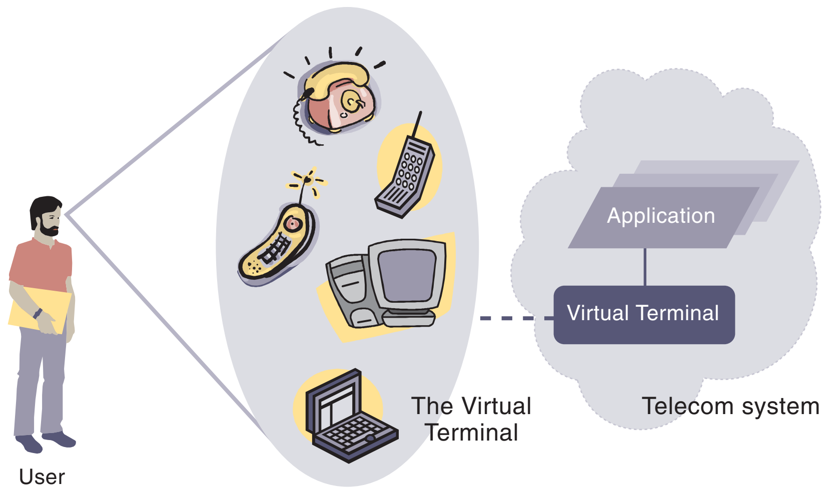
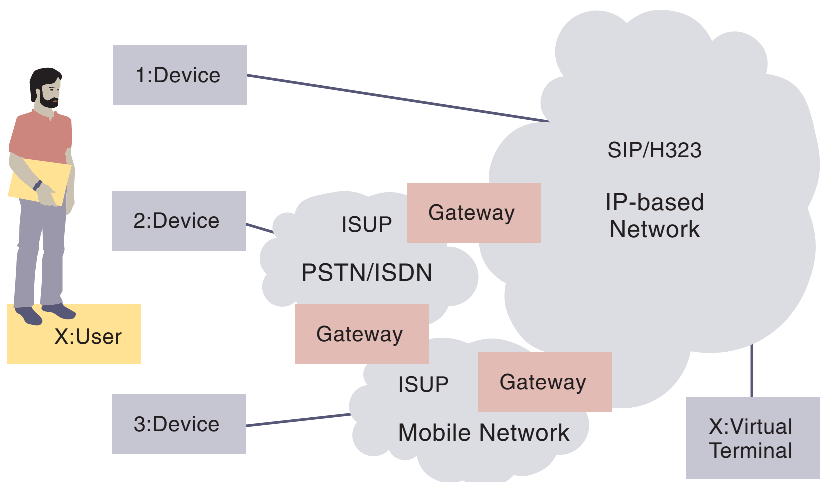


Figure 8 Physical network connection of the Virtual Terminal



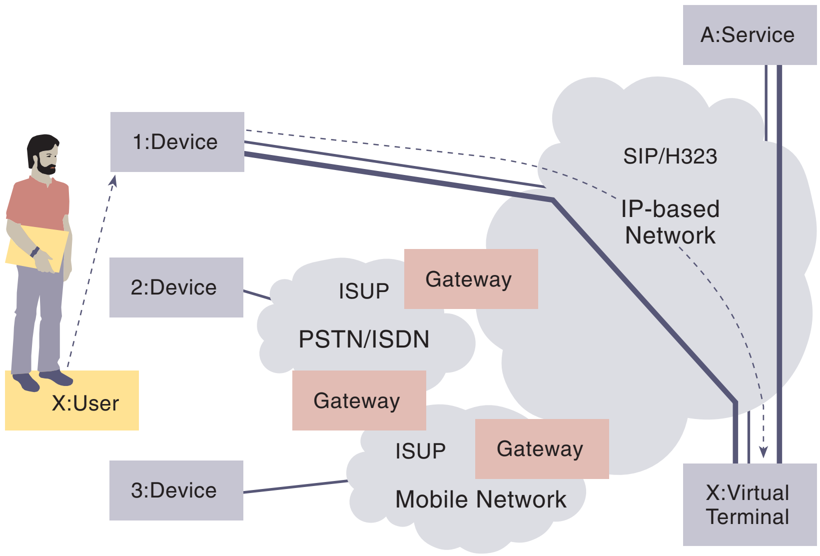


Figure 9 Initiating service from an IP-based device

In order to realise the Use Case Initiating Service the following conditions must be met:

- Every object Virtual Terminal must have an identity that is visible to all devices, i.e. it should be possible to address the Virtual Terminal from any device. Each Virtual Terminal may have several identifiers:

  - An MSISDN number (telephone number);

  - An email address;

  - Some kind of unique identifiers conformed to OSA/Parlay [7,8].

  If the Virtual Terminal has several identifiers, a mapping between these identifiers is required.

- It should be possible to establish a connection from any device to the Virtual Terminal and a connection from the Virtual Terminal to any service or any other device. The Virtual Terminal must be capable of acting as a connector that connects the streams together.

The user may have different devices connected to different networks. One device can be an IP phone or a PC directly connected to an IP-based network, another device can be a fixed phone or cordless phone connected to the PSTN/ISDN network, and a third device can be a cellular phone connected to a mobile network such as GSM.

The protocols used to establish and terminate a connection vary for different networks. For IP-based networks either SIP [9] or H323 [10] protocols should be used but for PSTN/ISDN and mobile networks, the protocol is SS7 ISUP. When the user initiates a service from a chosen device a connection must be established from this device to the Virtual Terminal. Different protocols should therefore be used depending on the type of device, i.e. IP-based device contra PSTN/ISDN or mobile device.

Let us consider these two device types successively.

6.1.1 Initiating Services from an IP-based Device

If the user uses a device that is connected to an IP-based network, either SIP or H323 protocol is used for signalling. In addition, information on the requested service or device must be conveyed to the Virtual Terminal so that it knows which service or device to connect to. This case is illustrated in Figure 9.

Two implementation alternatives are identified:

Alternative 1:
- Use **SIP/H323** for establishment of connections.

- Use CORBA/Java RMI for control and information exchange with the Virtual Terminal [11,12].

Alternative 2:
- Use OSA/Parlay API for establishment of connections.

- Use CORBA/Java RMI for control and information exchange with the Virtual Terminal.

The most straightforward implementation will be to incorporate either SIP or H323 in the Virtual Terminal. However, such a solution does not imply portability since the code of the Virtual Terminal needs to be changed for each protocol.

A better solution might be to use OSA/Parlay API since a generic and network independent Call control API is offered. Portability is hence ensured and the Virtual Terminal will function on either SIP or H323.

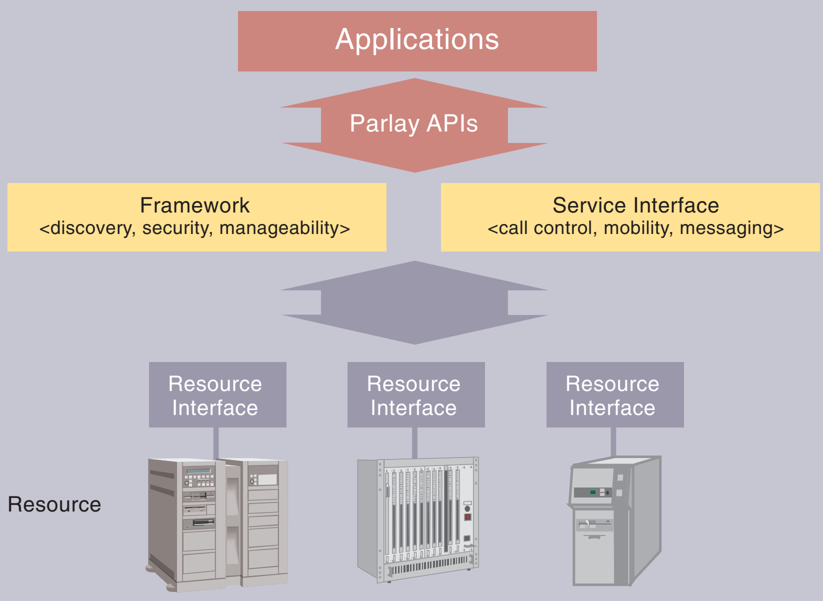
6.1.2 Initiating Services from PSTN/ISDN or Mobile Devices

If the user uses either a device connected to the PSTN/ISDN network or a device connected to a mobile network such as GSM, the SS7 ISUP is used to establish connection.

Assume that a user wants to make a telephone call to another phone. He dials the number of the addressed phone. The signalling must somehow be routed to the Virtual Terminal, and there is a need for an IN service, which sends the signalling to the Virtual Terminal. This IN service

Box 2 – Parlay API

The Parlay Group is an open, non-profit, multi-vendor forum organized to create open, technology independent Application Programming Interfaces (APIs) that enable IT companies, ASPs, ISVs, Internet Companies, E-Business Companies, software creators, service bureaux, and large and small enterprises as well as network providers, network equipment vendors and application suppliers to develop applications across multiple networks. Furthermore, the Group promotes the use of Parlay APIs and ultimate standardization. The Parlay API specifications are open and technology-independent, so that the widest possible range of market players may develop and offer advanced telecommunication services.



The Parlay architecture

Functions provided by the service interfaces allow access to traditional network capabilities such as call management, messaging, and user interaction. The service interfaces also include generic application interfaces to ease the deployment of communications applications.

In the current specification, Parlay API 2.1, the following Service interfaces are specified: Mobility, Call Control, User Interaction and Connectivity. All interfaces are specified in both OMG IDL (CORBA) and Microsoft MIDL (DCOM), which make the Parlay API independent of programming language, platform and Operating System. The call control service interface in Parlay 2.1 specifies one-to-one call, multiparty call, multimedia call and conference call. It is also possible for applications to initiate calls ("Call from the blue").

The Parlay API does not specify any mapping from service interfaces to proprietary interfaces. The implementers of the Parlay API have to perform the mapping according to the wanted underlying network technology such as SIP, H.323, CAMEL and ISUP, etc.

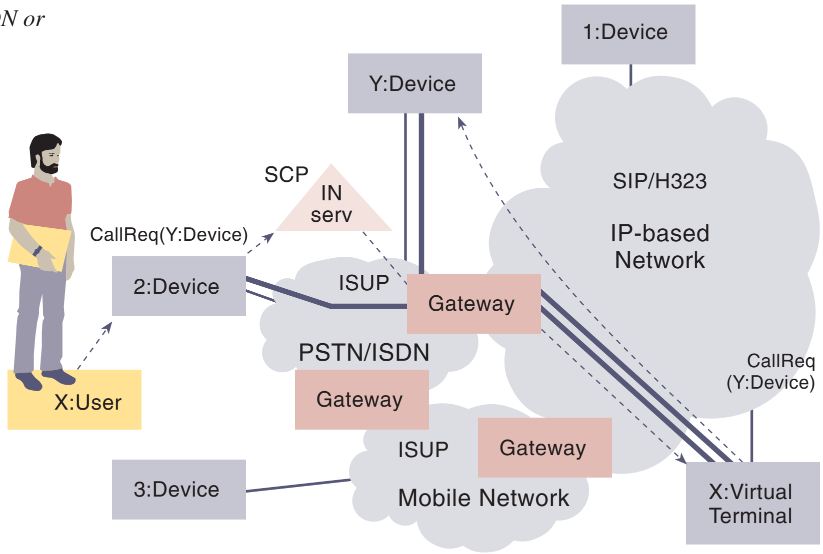
The 3GPP OSA API is based on the Parlay API. The specification is available for free download at http://www.parlay.org/.

is instantiated when the user registers a device as a component of the Virtual Terminal. The Virtual Terminal will then forward the signalling towards the addressed phone. If the party at the addressed phone accepts the call, two connections, one between the device and the Virtual Terminal and one between the Virtual Terminal and the addressed phone, will be established. This is shown in Figure 10.

The connection between the PSTN/ISDN or mobile device and the Virtual Terminal is established using the existing signalling mechanism, i.e. SS7 ISUP. However, the gateway between PSTN/ISDN and the mobile network must be capable of converting the signalling to either SIP or H323.

For establishing the connection between the Virtual Terminal and the addressed phone there are

Figure 10 Initiating call from PSTN/ISDN or mobile phone



two alternatives similar to the ones mentioned in the previous section: To use SIP/H323 or an open API like OSA or Parlay.

6.2 Receiving Service

When a service is addressed to the user the service request is sent to the Virtual Terminal. The Virtual Terminal will deduce the most appropriate device by going through the list of registered devices and the user profile. The Virtual Terminal is initiating signalling at this device. At this point there are three use cases: Accepting a request, receiving in mailbox and refusing a request.

6.2.1 Accepting a Request

If the user accepts the service, two connections will be established: one between the requesting service or device and the Virtual Terminal and another between the Virtual Terminal and the selected device.

It is worth noting that there are again three solutions that are similar to the ones considered earlier in the use case Initiating service.

6.2.2 Receiving in Mailbox

If the user is busy or does not want to receive the service at the time of the request he can specify that the service should be delivered to his mailbox. Again there are several alternatives for the mailbox, e.g. voice mail, e-mail and SMS.

6.2.3 Refusing a Request

If the user receives a service request he does not want to accept, he can simply refuse the request and no connection is made. The user should be able to keep a list of services, phone numbers, IP addresses etc. that he does not want to receive in his Virtual Terminal, or he can refuse the request manually after a notification of the service request.

6.3 Using Multiple Devices

After a service is initiated or a service request is accepted, a connection is established. During a session, the user might want to use multiple devices for input and output simultaneously and to dynamically change which devices to use according to what kind of services he is using and according to his needs and available devices at any time. He may also wish to add new devices to his session on the fly. The Virtual Terminal will manage this for him without requiring the user to have any technical skills.

6.3.1 Activating and Deactivating Devices

This use case is used by the use cases *Changing device* and *Adding/Releasing device*, which will be described later. When the user is changing devices, what he is really doing is to activate one device and deactivate the previous one. Similarly, when adding or releasing devices to or from an ongoing session, one is simply activating or deactivating the devices in question. During an ongoing session, activating a device will include it in the session and deactivating it will exclude it. No device is considered active in idle mode.

There are two ways of activating/deactivating a device:

- **Manually**: The user tells the Virtual Terminal which device or devices to be used explicitly. Such activation should be easy to perform or else the user will not use it. For PCs, PDAs, it is no problem letting the user enter his selection. For mobile phones, which have a more restricted user interface, it is more difficult. One solution is to assign a number to each device. The user can use any phone with a # key to register a device by simply enter * then the number assigned to this device and terminate with #. In this scheme USSD (Unstructured Supplementary Services Data) is used to send the command to the IN service, which forwards it to the Virtual Terminal.

- **Automatically**: Bluetooth may be used to detect and activate devices. The user still needs to activate an initial device as for example his mobile phone. When entering his office where a multimedia PC is available or when entering his living room where a stereo system or a TV is available, the mobile phone will detect the presence of these devices. The mobile phone asks for the device IDs, e.g. IP address etc. and also what types the devices are. It is necessary to have sufficient authentication and access control schemes to prevent illegal use of the equipment. It is worth noting that with automatic registration it is not necessary to have a definition of a device in the user profile since the device type can be provided on the fly by the device itself. It is, however, more important to have sufficient security. The mobile phone can then send the device ID and the device type to the IN service, which forwards it to the Virtual Terminal. The Virtual Terminal can then activate the device. It is necessary to elaborate a strategy for selection and activation of devices because approaching a device does not necessarily mean that the user wants to change over to this device.

A combination of automatic and manual activation of devices might be the preferred solution: Devices are detected automatically, and the user is simply asked whether or not he wants this new device to be activated or not by his active device. In this way the user will gain control over which devices are active, but minimal user interaction is required.

6.3.2 Using Multiple Devices Simultaneously

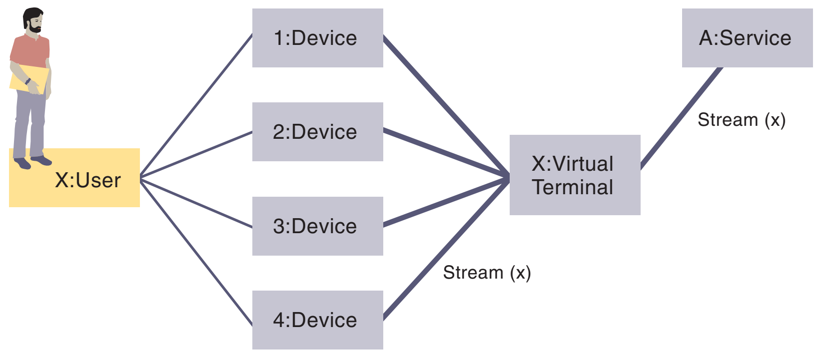
If a user wants to use a device during a communication session, it is required that the appropriate input and output streams are delivered to this device. In the case where the user wants to use multiple devices simultaneously, the need arises to deliver multiple streams to the different devices simultaneously. In order to achieve this, there must be some sort of splitting or multiplying of the incoming stream that is to be delivered to several output devices and some sort of merging of the outgoing streams that are collected from different input devices.

We will look at the two different situations where the connections are established and routed through the Virtual Terminal and where they are routed directly between the device and the service.

6.3.2.1 Connection through the Virtual Terminal

We now assume that a connection is being made between one or more of the user's devices and a service, and that this connection is routed through the Virtual Terminal.

Figure 11 Multiplying the incoming stream in the Virtual Terminal and delivering identical streams to all active devices



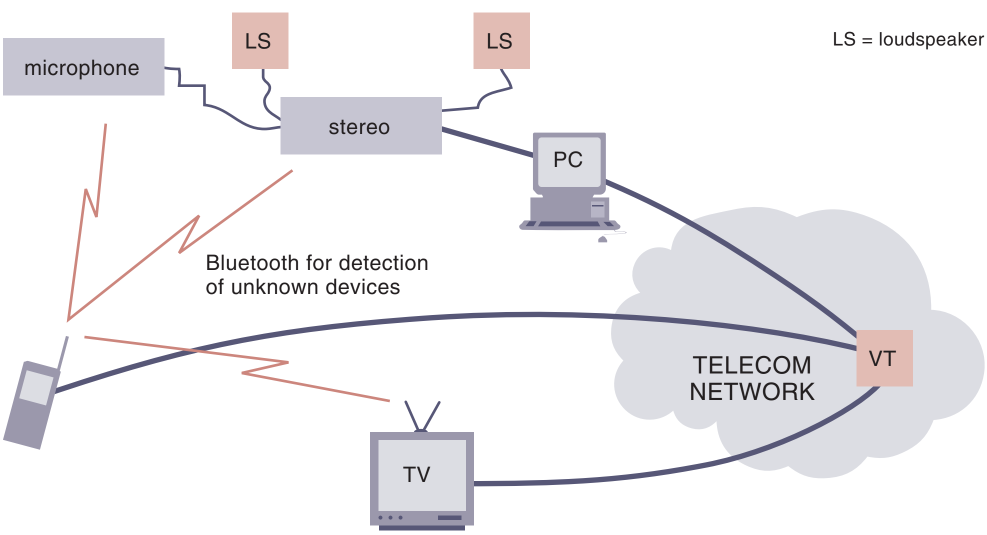
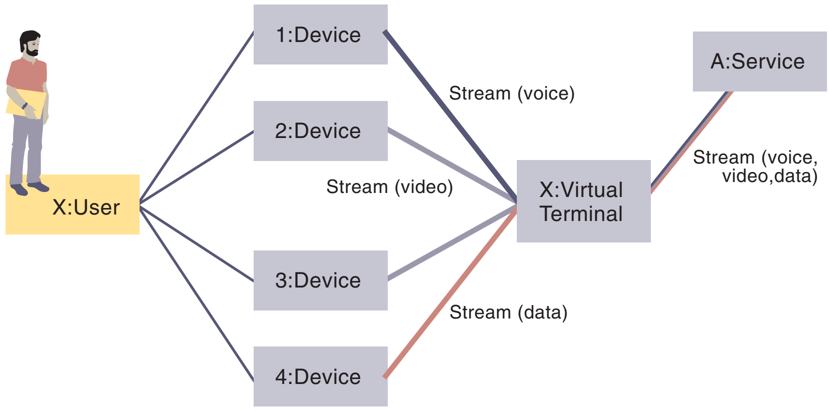


Figure 12 Virtual Terminal as a set of communication devices + other electronic devices connected to the network

Figure 13 Splitting the incoming stream into different types of streams: voice, video or data and delivering the streams to appropriate devices



There are two possibilities for delivering the service to multiple devices. One is to multiply the incoming stream and hence get several identical streams that are delivered to the different active devices. When the devices are connected to the service through the Virtual Terminal, the Virtual Terminal is responsible for multiplying the streams. This situation is shown in Figure 11.

As an example of this case, consider the case when the Virtual Terminal contains a set of communication devices + other electronic devices such as microphones, loudspeakers, TV screens, etc. A Bluetooth link can be used to detect unknown nearby devices that are available, but communication with them can be handled by the Virtual Terminal and go through the network.

When the devices are all connected to the network, the mobile device can move far away from them but still communicate with them. In Figure 12, the situation where the electronic devices are connected to the network is shown. The thick lines indicate where the network connections are.

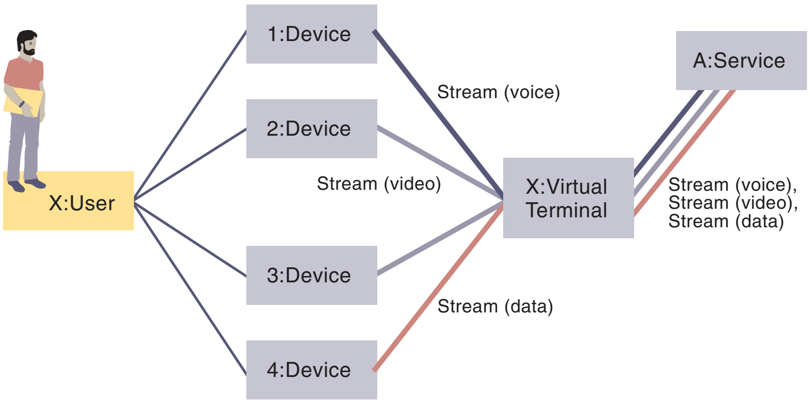
The other solution is to split the stream into different parts containing voice only, video only or data only respectively, and deliver different streams to different devices. Again, the Virtual Terminal handles the splitting of the streams into different types of streams when the connection goes through the Virtual Terminal. This situation is shown in Figure 13.

The first solution of multiplying the streams is probably the most straightforward one to implement but it obviously does not make very efficient use of the network resources by sending the same bits more than once and to devices that do not need them. This solution would mean that for example video packets are delivered to a set of speakers that only need the voice stream. A situation like this would not be symmetric with respect to incoming and outgoing streams either.

Splitting the streams into pure voice, video or data streams would make more efficient use of the network resources, but this solution might cause problems with timing and synchronisation during a multimedia session, if for example the voice and the video streams arrive out of sync. Today, there is no functionality in the lower layers to handle the splitting, and there would have to be an application taking care of this. This again would involve more processing.

A third solution could be possible that would be a more "true multimedia" solution. This would require that the service is actually delivering three separate streams containing only voice, video and data respectively. In this case there would be no reason for splitting up the streams in the Virtual Terminal since the streams are already divided into voice, video and data streams. Simply redirecting the different streams to the appropriate devices would be sufficient. One could however suspect that there are situations where one would like to multiply one or more of the streams, for example if there is a need to get the video on two separate screens at the same time. Nevertheless, this true multimedia situation is not the situation in today's networks even though it might become a reality in the near future. The true multimedia situation is shown in Figure 14.

Figure 14 A true multimedia situation



With either of the solutions described above, the user should be able to dynamically change one or more of the output and input devices during a session by communicating with his Virtual Terminal.

6.3.2.2 Direct Connection between the Devices and the Service

Another solution is to establish a connection directly between the utilized service and the user's devices. During the initiating phase a "signal link" between the service and the Virtual Terminal is established. The Virtual Terminal will then tell the service how the user wants the service to be served, e.g. which devices the user wants to use for different types of streams, which devices to be used for voice, which for video and which for text. These streams will then go directly to the devices, instead of via the Virtual Terminal.

This solution can make more efficient use of network resources, especially when the user and the utilized service are both located on a visiting network. But on the other hand the dynamic use of multiple devices and changing of devices during a session is difficult because the Virtual Terminal cannot do any rerouting of the streams when the user wants to use multiple devices. For example, it will be difficult for the user to switch from a mobile device connected to the GSM network to an IP telephone connected to an IP network if the mobile phone has a direct connection with the service. On the other hand, this can be achieved if the mobile device and the service are connected via the Virtual Terminal as described

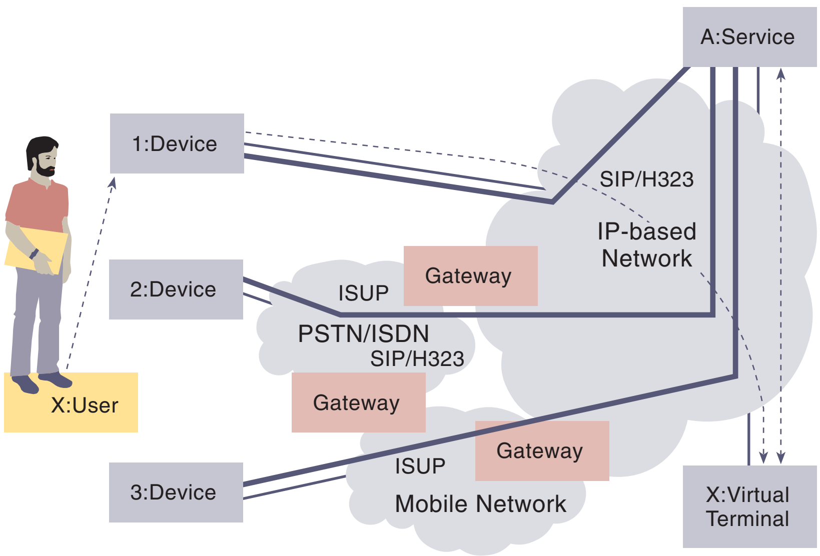


Figure 15 Connection directly between devices and services

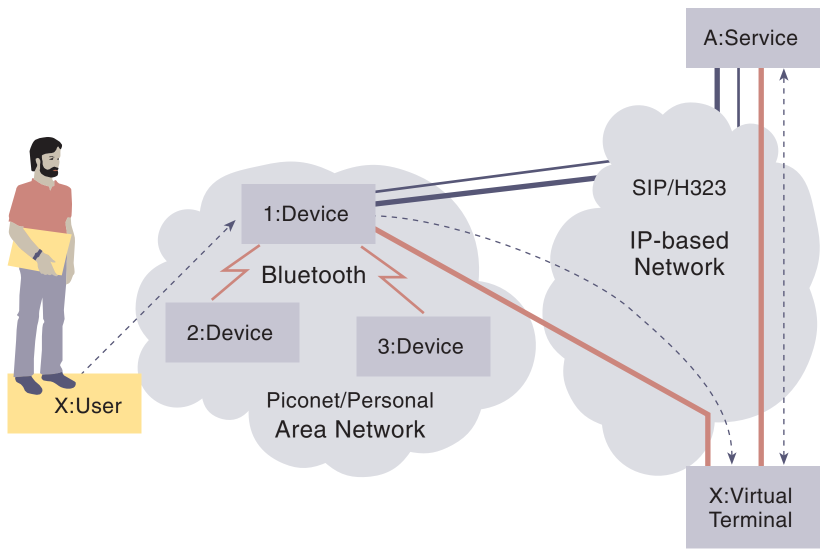


Figure 16 Some devices do not have network connection

in the previous section. Also when the user offers services using multiple devices (outgoing service) to other users, it is necessary with a server somewhere in the network that converges the streams before sending them to the user on the other side. Figure 15 shows this scenario where the Virtual Terminal plays the role similar to an HLR in GSM networks, but where the Virtual Terminal contains more information than an HLR. The "Signal link" between the Virtual Terminal and the service can be disconnected after the signalling phase.

6.3.2.3 Connection to Devices without Network Connection

In cases where some of the devices do not have any network connection, as illustrated in Figure 16, Bluetooth may be used to make a piconet or personal area network. In this case the main active device could contain the necessary functionality, e.g. converging/splitting of streams depending on the service. The main stream may go directly between the main active device and the other side's service or via the Virtual Terminal as illustrated by the dashed lines.

Figure 17 shows an example of such a situation.

6.3.3 Switching between Devices

With either of the situations described above for using multiple devices, the user should be able to change active devices whenever desired. This will be an important part of the functionality provided by the Virtual Terminal. The different situations will give rise to different solutions for changing the devices. In this section we will give a brief overview of how the user can switch between devices. We distinguish between changing device and changing output/input.

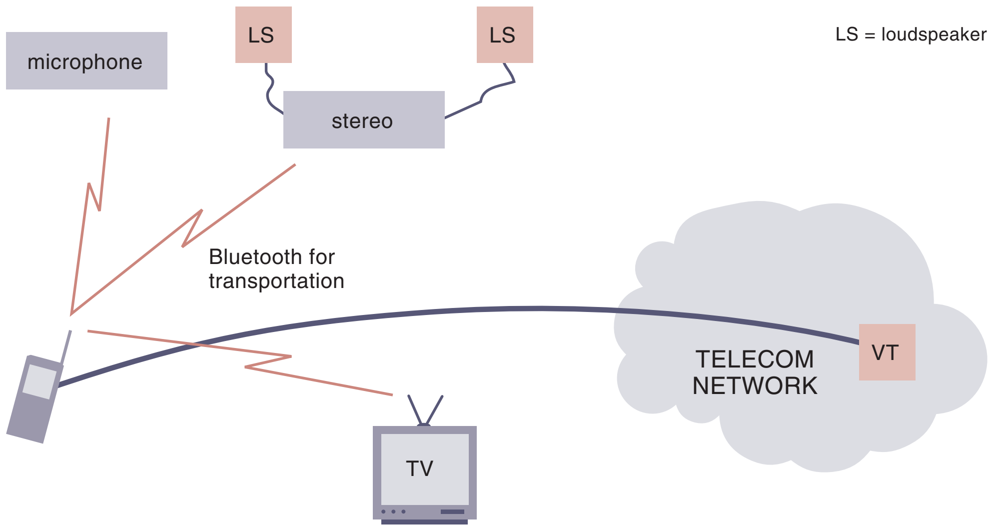


Figure 17 A set of communication devices + other electronic devices not connected to the network

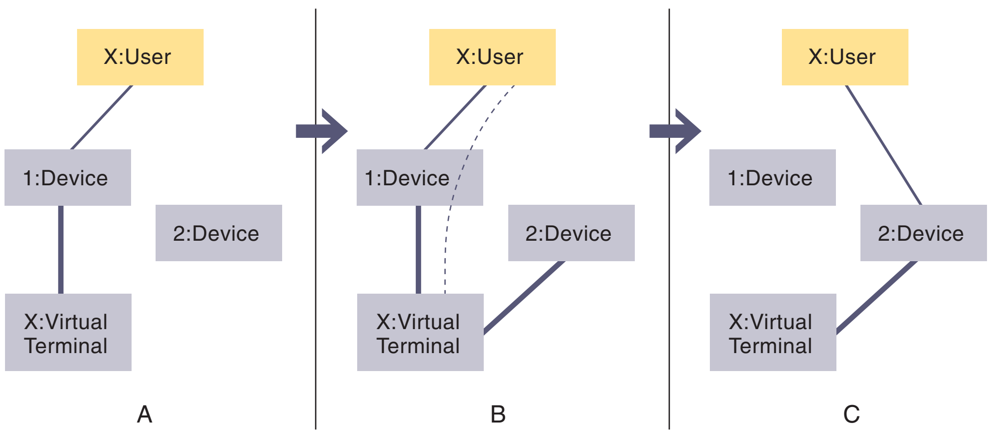


Figure 18 Changing the active device during a session

Box 3 – IP Telephony protocols

H.323

H.323 is the globally accepted standard for audio/video/data communication. It specifically describes how multimedia communications occur between user terminals, network equipment, and assorted services on Local and Wide Area Internet Protocol (IP) networks.

Activity around H.323 is especially high due to the unified support of a global coalition of companies, including personal computer and communications systems manufacturers, and operating systems makers. H.323-compliance has also been promoted and accepted by Internet Phone and Voice-Over-IP manufacturers as the standard for interoperability.

The H.323 standard
H.323 is sometimes referred to as an "umbrella" specification, meaning that in the document itself there are references to other recommendations. Other recommendations in the H.323 series include H.225.0 packet and synchronization, H.245 control, H.261 and H.263 video codecs, G.711, G.722, G.728, G.729, and G.723 audio codecs, and the T.120 series of multimedia communications protocols.

Together, these specifications define a number of new network components – H.323 terminal, H.323 MC, H.323 MP, H.323 Gatekeeper and H.323 Gateway) – all of which interoperate with other standards-compliant end points and networks by virtue of an H.323/H.32X gateway.

SIP

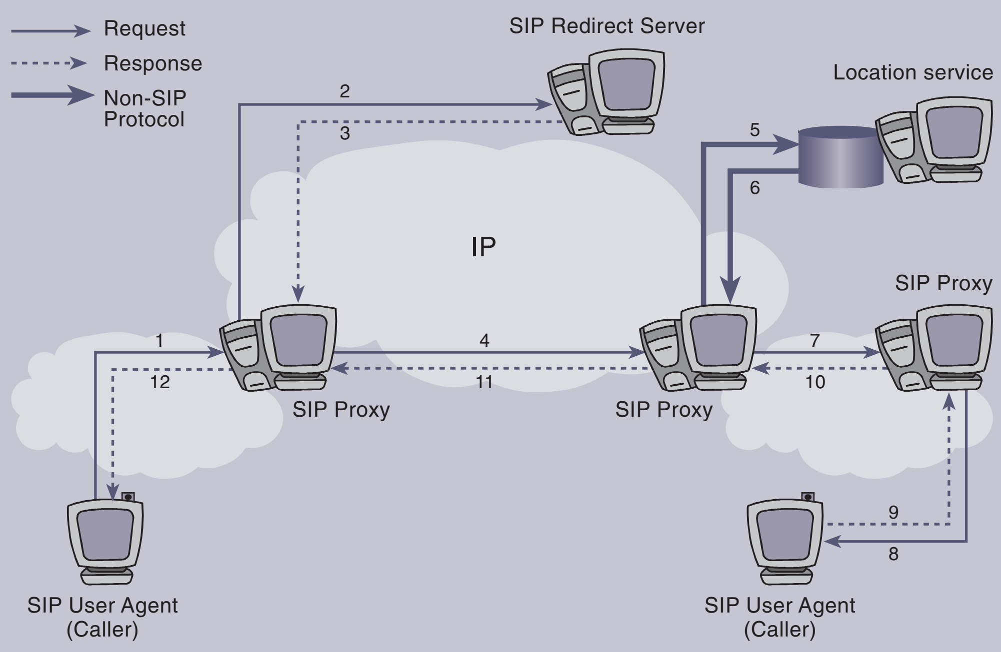
The Session Initiation Protocol, or SIP, is a new IETF signalling protocol for establishing real-time calls and conferences over Internet Protocol networks. Each session may include different types of data such as audio and video although currently most of the SIP extensions address audio communication. As a traditional text-based Internet protocol, it resembles the hypertext transfer protocol (HTTP) and simple mail transfer protocol (SMTP). SIP uses Session Description Protocol (SDP) for media description.

SIP is independent of the packet layer. The protocol is an open standard and scalable. It has been designed to be a general-purpose protocol. However, extensions to SIP are needed to make the protocol truly functional in terms of interoperability. Among SIP basic features, the protocol also enables personal mobility by providing the capability to reach a called party at a single, location-independent address.

SIP Architecture
SIP's basic architecture is client/server in nature. The main entities in SIP are the User Agent, the SIP Proxy Server, the SIP Redirect Server and the Registrar.

The User Agents, or SIP endpoints, function as clients (UACs) when initiating requests and as servers (UASs) when responding to requests. User Agents communicate with other User Agents directly or via an intermediate server. The User Agent also stores and manages call states.

SIP intermediate servers have the capability to behave as proxy or redirect servers. SIP Proxy Servers forward requests from the User Agent to the next SIP server, User Agent within the network and also retain information for billing/accounting purposes. SIP Redirect Servers respond to client requests and inform them of the requested server's address. Numerous hops can take place before reaching the final destination. SIP's tremendous flexibility allows the servers to contact external location servers to determine user or routing policies, and therefore does not bind the user into only one scheme to locate users. In addition, to maintain scalability, the SIP servers can either maintain state information or forward requests in a stateless fashion.



Box 3 continued

Session Description Protocol (SDP)
The Session Description Protocol, or SDP, is a protocol for describing audio, video and multimedia sessions. SIP, MGCP (Media Gateway Control Protocol), SAP (Session Announcement Protocol) and RTSP (Real-Time Streaming Protocol) all use SDP.

Relationship Between SIP and H.323

Both SIP and H.323 define mechanisms for call routing, call signalling, capabilities exchange, media control, and supplementary services. SIP is a new protocol that promises scalability, flexibility and ease of implementation when building complex systems. H.323 is an established protocol that has been widely used because of its manageability, reliability and interoperability with PSTN. There is a general consensus among standards organizations, companies and technology experts that standardized procedures need to be specified to allow seamless interworking between the two protocols. Bodies such as TIPHON (ETSI), aHIT (IMTC) and IETF are working to address this topic.

6.3.3.1 Changing Device
We assume that sometimes the user will only use one device for a service, for example his cellular phone. If the user recognises a more appropriate device, for example his desktop telephone when he enters his office, he should be able to switch to this device even during an ongoing session. He must then in some way tell his Virtual Terminal that he wants to change, and the Virtual Terminal will redirect the input and output streams to this new device.

The changing of the active device during a session is shown in Figure 18, where the user has an ongoing session delivered to device 1, then tells the Virtual Terminal to change active device to device 2, and then finally the Virtual Terminal delivers the service to device 2 where the user can continue his session. The communication between the user and the Virtual Terminal will always be via the active device. This changing of devices should be very easy to use for the user and preferably seamless with respect to the running applications.

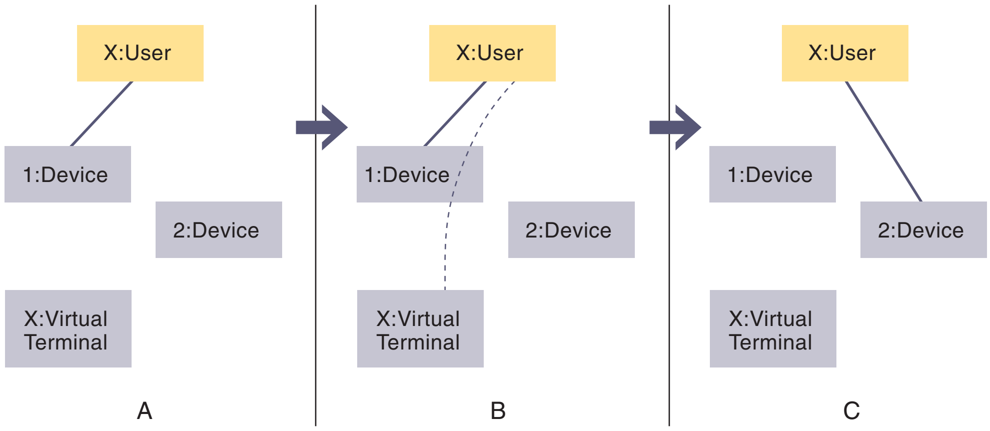


Figure 19 Changing device offline

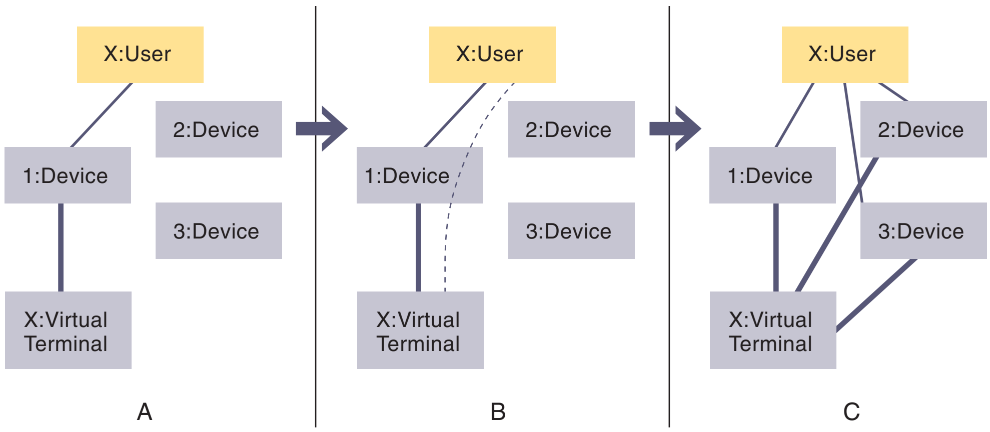


Figure 20 Redirecting the input and output streams

Even when the user does not have an active session going on, he might still be registered at a device. This is specified so that any incoming service requests will be delivered to this device. If the user wants to change which device to receive a possible service request on, he may communicate this to the Virtual Terminal, which will consider the new device as where to reach the user. This situation is shown in Figure 19.

It should also be possible for the Virtual Terminal to automatically change which devices to be registered at according to the user profile and depending on for example time of day and the location of the user. In this case the user will not have to tell specifically which device to use as long as he has already configured his profile correctly.

6.3.3.2 Changing the Input and Output

If the user wants to change only some parts of the input or output streams, the Virtual Terminal should allow him to do so. These changes will be relevant to the user in cases where he arrives at places where more appropriate devices are located or if his services are changing and hence require different output or input quality.

The user should be able to tell the Virtual Terminal through his active device which devices he wants services to be delivered to at any time as shown in Figure 20. He should be able to switch back and forth as many times as he likes and to any available devices he has access to.

The Virtual Terminal should always know what type of devices the user wants to use. The user can either specify this whenever he switches to a new device, or if it is a previously known device, this information can be contained in the profile within the Virtual Terminal so that it already knows what type it is, i.e. voice, video or data, and whether it is offering output or input capabilities, etc.

6.3.4 Adding and Releasing new Devices

Since each device often has a limited user interface, the combination of several devices will offer the user better user interfaces and hence better services. During a service session a user may want to add one or more elementary devices with just input or output functionality, e.g. a big screen display, speakers, microphone etc. or more smart devices like PC, smart phone, etc. to the session. We may consider the cases where the new device is trusted or known by the Virtual Terminal and where it is not.

6.3.4.1 Adding known Devices to a Session

In the case where the new devices are trusted or known to the Virtual Terminal, e.g. devices at the user's home or office, the Virtual Terminal already knows which services those devices can offer and how they can work in combination with the main active device. One can include those devices in the user service profile on the Virtual Terminal in advance, and with a few simple key presses on the active device one can request the Virtual Terminal to add or release one or several of these devices to or from the service.

This requires that the new devices have network connection and the Virtual Terminal is responsible for the rerouting and if required, for the splitting of the stream. It means that the stream should go via the Virtual Terminal. In the cases when new devices do not have network connection, Bluetooth may be used to add new devices to the service, but then the Virtual Terminal does not need to be involved in the adding process.

6.3.4.2 Adding unknown Devices to a Session

If the new devices are unknown, the establishment of a multi-device session is more complicated. If a new device is unknown to the Virtual Terminal, the IP address and the identity of the device as well as information about which kind of services the device can offer must be found out. This can either be done manually by the user or automatically by the active device. To do this manually is often not preferred since it requires user involvement. Another problem is that unknown devices are also un-trusted by each other, so some kind of mutual authentication and access control has to be performed. With an automatic process one is faced with general problems like:

- Automatic discovery of devices. Both networked and not networked devices;
- Ability of device to announce its presence to the network;
- Ability to describe its capability as well as query/understand the capabilities of other devices;
- Seamless interoperability with each other and self-configuration without administrative intervention;
- Security issues.

Some new technologies like Universal Plug and Play [13], Salutation [14], Bluetooth [2], Jini, etc. [15,16,17] may be used to solve these problems. When the identity and other characteristics of the new devices are known, it is possible to add them to the ongoing service session.

Figure 21 A laptop and peripheral devices connected to the network

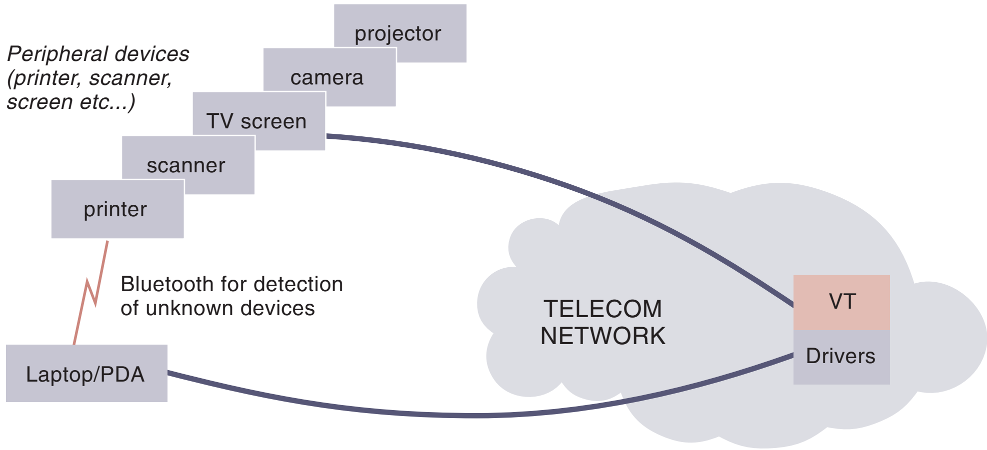
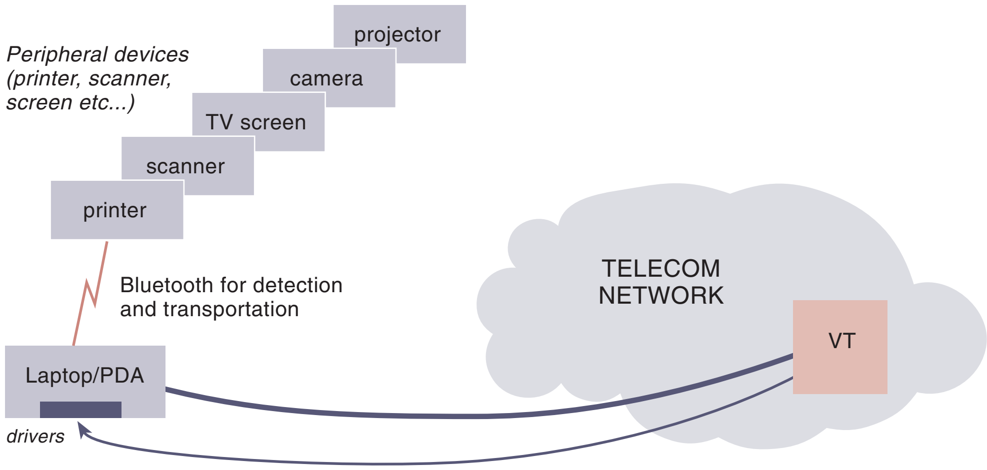


Figure 22 A laptop and peripheral devices not connected to the network. In this case the necessary drivers can be downloaded and installed on the laptop using the Virtual Terminal



6.3.4.3 Releasing Devices from a Session

Releasing a device from an active session should be a very straightforward task. The user can either communicate this to the Virtual Terminal via one of his other active devices or simply turn off the device to be released.

6.4 Configuring the Devices

Another functionality the Virtual Terminal can offer is to help the user configure his different devices when this is required. This will be illustrated by the following examples.

Assume the Virtual Terminal contains a set of communication and computing devices + peripheral devices. Again, there are two different cases similar to the ones mentioned above, with or without network connection. We consider a laptop and peripheral devices such as a printer, a scanner, a digital camera, etc.

When the peripheral devices are connected to the network, the communication with these devices can go through the network. A Bluetooth link can be used to detect available unknown devices and to find out which drivers they require, etc. The different drivers for the printers, scanners, etc. can be contained in the Virtual Terminal and do not need to be installed on the laptop. Another advantage is that the laptop and the peripheral devices can be far away, for example the user can send data to his home printer even if he is away.

In the case where the laptop wants to communicate with foreign peripheral devices that are not connected to the network, a cable or a cable substitute such as a Bluetooth link is needed for direct communication. In this case the laptop must have the appropriate drivers installed. The Virtual Terminal can find the necessary drivers for the devices and then send and install them on the laptop. It can also clean up when the mobile device moves on – i.e. delete the drivers when they are no longer needed.

6.5 Administrating the Virtual Terminal

As specified earlier this use case consists of five sub use cases: Accessing profile, Defining/deleting a device, Registering/deregistering a device, Reading mailbox, Reading usage statistics. In this section, we will examine each of these use cases successively.

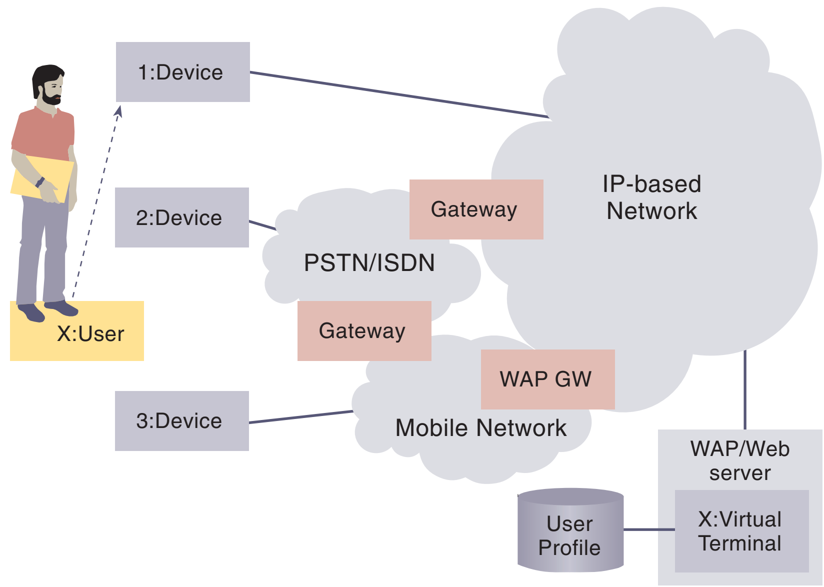


Figure 23 Profile access from WAP/Web

6.5.1 Accessing Profile

Each user has a profile where he may store all his set up and preferences. The profile may incorporate things like bookmarks, address list and also items that the user uses frequently. At this stage it has not been analysed what should be part of the user profile. However, it is important that the profile contains a list of all components of the Virtual Terminal, i.e. all the devices that the user has at his disposal. Such a list should contain both the device unique ID, e.g. MSISDN number, IP address, etc., and also the device type, e.g. GSM phone, SAT (SIM Application Toolkit) phone, WAP phone, PDA, PC, TV, etc.

The profile might also contain a list of possible mailboxes to where incoming messages can be sent. The user should be able to choose what kind of mailboxes he wants in this list and which mailbox to use at any given time. The user's personal address book should also be placed in the Virtual Terminal so that he can reach it from any device he might be using.

Another important issue is that the profile should be easily accessible to the user. The user should have the possibility to access his profile from any Web browser and any WAP phone and to change elements in it if desired. It is therefore necessary to have sufficiently strong authentication and access control to prevent fraudulent access to the user profile. This use case and also the other three use cases will hence use the use case Authentication & Access Control. The Virtual Terminal should also have access to a certain database in order to read and write the user profile.

As shown in Figure 23, the Virtual Terminal should reside on a Web/WAP server or alternatively reside on a separate application server but accessible from the Web/WAP server. There are several commercial application server products available such as Bea Weblogic, Cold Fusion, NetDynamics, Netscape Application Server, etc. The evaluation and selection of appropriate application server, however, are out of the scope of this article.

The Web/WAP can be a standard public domain Web server such as Apache, a W3C XML (Extended Markup Language) bean server such as Bea Logic (Weblogic). Depending on whether the terminal has a Web browser or a WAP browser HTML (Hypertext Markup Language) or WML (Wireless Markup Language) will be generated from XML.

6.5.2 Defining and Deleting Devices

All the devices that the user has at his disposal and may use in a specific situation must be defined in his profile so that the Virtual Terminal knows how to handle them. Such a definition can be done manually and the user must then update the list of defined devices contained in the Virtual Terminal with all necessary information about the new device to be defined. Deleting an already defined device manually is naturally a much easier task. New devices can also be defined in the Virtual Terminal automatically when they are discovered by some service discovery protocols that detect all necessary information, but to define too many devices automatically might result in a long list of devices that are hardly ever used. A better approach could be to let all necessary information about new devices be detected automatically, and to simply

Box 4 – Bluetooth

Bluetooth is named after the Scandinavian Viking king Harald Blåtann, and it was the Scandinavian communication companies Ericsson and Nokia who joined up with IBM, Intel and Toshiba to form the Bluetooth Special Interest Group (SIG) and specified this new technology. Several other companies joined afterwards, and today, the SIG consists of more than 2000 members. The current main members are: 3com, Ericsson, IBM, Intel, Lucent, Microsoft, Motorola, Nokia and Toshiba.

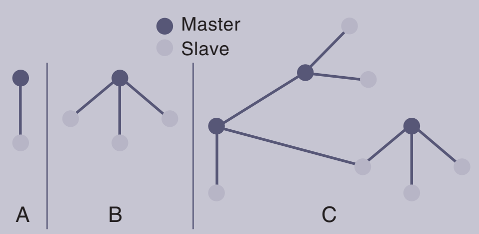
The Bluetooth technology is still at its early stage, but there are already some Bluetooth devices being produced. The devices range from access points for LANs, PCMCIA cards for laptops and USB devices, to mobile phones and mobile phone accessories like headsets. Some of these devices can be used in existing devices and make them participate in a Bluetooth piconet. On the software side, several implementations of the Bluetooth stack already exist.

Short-range radio communication

Bluetooth is a short-range radio communication between various devices, like computers, mobile phones and other personal devices. Since radio links are used there is no need for a clear line of sight between the devices. This makes communication through walls, clothes and briefcases possible, and tasks like automated synchronization of PDAs and modem usage can be completed without lining up the devices. Bluetooth radio communication operates on the public Industrial Scientific Medical (ISM) band at 2.4 GHz and utilizes frequency hopping which makes the links robust and secure. The standard chip is constructed to transmit over ranges up to 10 metres, but by increasing the power the range can be extended to more than 100 metres.

Piconets

Bluetooth devices form ad hoc networks called piconets, where one device acts as the master, and the rest as slaves. A device can be a slave in multiple nets, but only master in one net at a time. The master controls the communication in the piconet; it decides the clock, frequency-hopping pattern, and keeps track of active and sleeping devices. Multiple piconets can be interconnected and form what is called scatternets as shown in the figure below.



Piconets and scatternets

ask the user to include them in the list of defined devices or not.

6.5.3 Registering and Deregistering Devices

Through a registration the user tells the Virtual Terminal which device or devices, among the defined devices, should be used at a certain moment. This will primarily be important for idle devices since it specifies where the user will receive incoming service requests, i.e. on what device he can be reached by the Virtual Terminal. Registration can be classified as pre-registration and on-the-fly registration.

- **Pre-registration**: Prior to usage, the user defines the device or the set of devices that should be used. Pre-registration can be further divided into:

  - **Default registration**: A default device or a group of devices is defined as default.

  - **Timetable registration**: A timetable is defined to indicate which device or devices to be used at what time and what day. Such a timetable should be included in the User profile.

Box 5 – List of abbreviations

API	Application Programming Interface
ASP	Application Service Provider
CAMEL	Customized Applications for Mobile Network Enhanced Logic
CORBA	Common Object Request Broker Architecture
DPE	Distributed Processing Environment
IDL	Interface Definition Language
ISP	Internet Service Provider
ISV	Independent Software Vendor
ISUP	ISDN User Part
MExE	Mobile Executed Environment
OSA	Open Service Architecture
OMG	Object Management Group
SCS	Service Capability Server
SDP	Session Description Protocol
SIM	Subscriber Identity Module
SIP	Session Initiation Protocol
SS7	Signalling System number 7
USSD	Unstructured Supplementary Services Data
XML	Extended Markup Language
3GPP	3rd Generation Partnership Project

- **Location table registration**: If somehow the Virtual Terminal is able to detect the location of the user, a location table can be used to find which device or devices to be used at what time and what day. Such a location table registration should be included in the User Profile.

• **On-the-fly registration**: The user decides and informs the Virtual Terminal about which devices he wants to be registered at "On-the-fly". These registrations will typically override any existing pre-registrations previously defined.

6.5.4 Reading the Mailbox

The user should have the possibility to access his mailbox. For voice mail, no additional functionality is required other than the one for existing voice mail. For text mail the user must be able to access from a PC with a standard Web browser or a WAP phone. The same requirements apply here as for accessing the user profile described in earlier section.

6.5.5 Reading Usage Statistics

The user should be able to read usage statistics from a PC with a standard Web browser or a WAP phone. The same requirements apply here as for accessing the user profile described in an earlier section.

7 Conclusion

In this article we present the Virtual Terminal, which is not only a vision but also a realisable concept. However, the use cases and also the implementation alternatives, although considered thoroughly, are only sufficient for carrying out a proof of concept. In order to be fully deployed as a real service for end-users, many obstacles must be surmounted. The technologies that the Virtual Terminal relies on must be mature. The OSA API should be implemented and optimised for heterogeneous networks. The SIP protocol used in the establishment and termination of connections needs to be expanded to support point-to-multipoint connections and also multi-stream connections. IP networking on Bluetooth, which was originally meant to be only a cable replacement, needs to be specified. For automatic configuration the Virtual Terminal also requires discovery protocols where service types or device types are defined universally. Last but not least, there are the challenges of integrating multiple systems together from fixed to mobile systems, from circuit-switched to packet-switched systems and from cellular to IP-based systems. Perhaps it is precisely these challenges that make the Virtual Terminal quite exciting as a research project.

References

1 ITU-T. *Principles for Universal Personal Telecommunications (UPT)*. 1993. (ITU-T Rec. F.850.)

2 *Bluetooth SIG*. (2000, November 24) [online] – URL: http://www.bluetooth.com

3 Muller, N J. *Bluetooth Demystified*. New York, McGraw-Hill Telecom, 2000. ISBN 0-07-136323-8.

4 *IEEE Standards Wireless Zone – Overview*. (2001, March 09) [online] – URL: http://standards.ieee.org/wireless/overview.html-802.11

5 Schneider, G, Winters, J P. *Applying use cases*. Reading, Mass., Addison-Wesley, 2000. (Object Technology Series, ISBN 0-201-30981-5.)

6 *The ICEBERG project*. (2001, March 09) [online] – URL: http://iceberg.cs.berkeley.edu/

7 3GPP. *Virtual Home Environment; Open Service Architecture (Release 1999)*. Valbonne, 2000. (3G TS 23.127, V3.1.0, 2000-06.)

8 *Parlay Group*. (2000, October 24) [online] – URL: http://www.parlay.org

9 IETF. *SIP: Session Initiation Protocol, March 1999*. New York. (RFC2543)

10 ITU-T. *H.323 Recommendation*. (2000, October 26) [online] – URL: http://www.itu.int/itudoc/itu-t/rec/h/h323.html#dl

11 OMG. *The Common Object Request Broker: Architecture and specification*. Revision 2.3, 1999. (2001, March 16) [online] – URL: http://www.omg.org/

12 *Remote Method Invocation Specification*. (2001, March 09) [online] – URL: http://java.sun.com/products/jdk/1.1/docs/guide/rmi/spec/rmiTOC.doc.html

13 *Universal Plug and Play Connects Smart Devices*. (2000, November 15) [online] – URL: http://www.axis.com/products/documentation/UPnP.doc

14 *Salutation Architectures and the newly defined service discovery protocols from Microsoft and Sun*. (2000, November 14) [online] – URL: http://www.salutation.org/whitepaper/JINI-UPnP

15 *Discovery and Its Discontents – Discovery Protocols for Ubiquitous Computing*. (2000, November 14) [online] – URL: http://www.ncsa.uiuc.edu/People/mcgrath/Discovery/dp.html

16 *Service Location Protocol, Version 2*. (2000, November 15) [online] – URL: http://www.rfc-editor.org/rfc/rfc2608.txt

17 *An Architecture for a Secure Service Discovery Service*. (2000, November 15) [online] – URL: http://ninja.cs.berkeley.edu/dist/papers/sds-mobicom.pdf

Giving Mobile Users Access to Location-aware Services – Opportunities and Challenges

RICCARDO PASCOTTO AND GEORG NEUREITER

Location-aware services will play an important role, particularly in providing services for mobile users. Furthermore, market studies have shown that these services have a lot of opportunities and challenges.

The IST Youngster project is developing an enhanced active mobile user environment that supports the creation of new and innovative types of services including context-aware solutions based on sensing, collecting, transmitting and processing context information such as terminal characteristics, network status, user location, situation, etc. More details concerning the European research project Youngster are available in the paper "Youngster: Focusing on future users in a mobile world" in this issue of *Telektronikk*.

This paper will give an introduction on location awareness as well as its enabling technologies, discuss key R&D research issues and outline the approaches of the Youngster project.

Riccardo Pascotto (30) is Project Manager, International Co-operations at T-Nova Deutsche Telekom Innovationsgesellschaft mbH Berkom. He received his diploma in technical computer science from the University of Applied Sciences, Berlin in 1996. He then worked as an engineer for consultation in multimedia communication and software development. Since 1998 he has been Project Manager at T-Nova Berkom working in a number of European research projects contributing to the 4th R&D Framework Programme of the ACTS and TAP. His area of specialisation is mobile multimedia. Currently he is project co-ordinator of the IST 25034 Youngster project.

riccardo.pascotto@telekom.de

Georg Neureiter (29) received his B.Engr. in 1991 from the Hoehere Technische Bundeslehranstalt für Nachrichtentechnik in Salzburg and started work at Professional Data Systems AG (Salzburg) where he was responsible for planning, installation and maintenance of structured networks. He resumed his studies in 1995 and graduated from the University of Salzburg with an M.Sc. in telecommunications. He joined T-Nova Deutsche Telekom Innovationsgesellschaft mbH in 1998, and is now involved in international projects as a specialist engineer in the field of GSM network evolution towards next generation networks and location aware services, applications and its associated technologies.

georg.neureiter@telekom.de

Introduction

Imagine you are a youngster aged 16 standing on top of a mountain in the Swiss Alps with some friends. In your backpacks you and your friends have your snowboards along with your outdoor-proof handsets of the next generation and would like to share the thrill of surfing down the exciting slopes. The handset itself fulfils two functions: One is that your girlfriend, at home with a broken leg from your latest experience and unable to come along, can watch the live video you provide her through the wireless network. The second function is that there are some hazards in front of you like steep and rocky parts you would prefer to avoid in order not to release an avalanche. All the handsets used are capable of tracking you and your pals, and if something happens your friends will be able to track you and find you in case you get lost or end up under some serious amounts of snow.[1)]

Key Challenges

Current and upcoming technologies will soon allow the establishment of so-called location based services, LBS. Although some of these technologies were developed a long time ago in the military field, it took quite a long time before they found their way to a larger community. Unfortunately, all of these enabling technologies have their advantages and disadvantages, and therefore it seems difficult to rely on only one technology. In addition, they strongly depend on whether the customers are moving around; for example it makes a great deal of difference if the customer is in a dense urban area, in an open landscape, or in a vast unpopulated area like mountains, deserts or the sea. Global mobility also raises another important issue of roaming of location based services. Few network operators have put LBS into operation yet and there might be no way of telling how these few applications will react if the handset is taken abroad.

As already mentioned, there is a handful of heterogeneous technologies to determine the location, so if future location based services are to be developed, there is a need for an open interface to "plug-in" the appropriate technology. Thus, a flexible middleware approach seems to be the key issue to future location based services to meet the customer demands in terms of security, rapid and flexible software adaptation towards third parties, and finally support of various positioning technologies.

Although the architecture to be developed in the EU-funded Youngster project aims to support context aware services in general, this paper focuses more on location based services, which are on the one hand part of the Youngster project but on the other hand are among the applications to be deployed on such a platform. The various imaginable service scenarios derived from an enabled location based platform are opening a

[1)] *At the moment, there are professional devices available to mountaineers to take along when they go into such an area. Extending the capabilities of a handset in order to find somebody in the snow would add a great deal to saving many lives because despite the danger few take the professional devices with them.*

new door towards the upcoming future generation of wireless services; it even promises to add a great deal of value to existing services (i.e. emergency services, location dependent billing, local information about weather reports, restaurants, shopping, and many more).

The name Youngster also reveals the additional focus to a market which looks to be extremely promising in terms of using such services in the near future. Not only do young people take to using new services faster than the older generations, but also their lifestyle is naturally highly mobile, very communication oriented, and they are interested in personalised information. By creating tools and services that are attractive for young people to use, we aim to inspire them to develop the skills necessary to put these new technologies and services into use, and to help in the development of new services through rapid feedback and adapting ideas to their own preferences and requirements. This should have various resulting benefits – both for them and for society in general.[2] In particular the penetration of mobile phones among young people is remarkably high. By extending this to other aspects of Information Society technology and making these attractive to young people to access and use, a growing momentum could be created amongst young people which would have a significant effect on the Information Society revolution. This would also create new opportunities for services geared specifically at youth.

Various Positioning Technologies

There is a large number of developments related to mobile location services of the future generation. A broadening range of different position determining systems such as OTD[3], TOA[4] and wireless assisted satellite based positioning (GPS[5]/GLONASS[6]) among others are adding a confusing amount of heterogeneity which seems to be difficult to address. In addition, the first GPS chips have successfully been integrated into GSM handsets[7], but its services are still too limited.

From a technological point of view, the aforementioned technologies are grouped into terminal centric and network centric position determining systems. The difference is where the position is actually determined in terms of measuring some physical signal values like time differences, signal strength and angle of the received signal. Another important issue – also facing future possible location based applications – is the power to decide if the location information is forwarded to another party or if it is kept private. This feature is naturally allowed by its design on handset centric technologies. Related also to marketing issues is the increase in costs related to each technology. A brief definition of each mentioned localisation technology is provided in the following paragraphs.

Cell ID

Depending on the base station with the best signal strength budget, the serving base station cell area coverage is used as a rough estimation of the caller. Accuracy highly depends on the actual cell size like macro/micro/pico cells (average size of macro cell = 30 kilometres, micro cell = 5 kilometres, pico cell = 0.5 kilometre). Due to optimisation purposes, cells are more often sectored into three or less often to other topographic related values which allows to narrow down further the coverage area and thus serve more customers with only one physical site. Cell ID is sometimes also referred to as Cell of Origin (COO).

Timing Advance

Within one serving cell, a time is allocated to the determined time difference of actual arrival of a signal from the mobile handset. The accuracy also depends – like Cell ID – on the cell size and the sectoring factor.

OTD (Observed Time Difference)

The signals from at least three geographically distributed base stations are received by the handset and optionally[8] also by a reference station called LMU (location measurement unit). The handset and the LMU forward the time differences of arrivals to a central computer centre, and the location is then calculated there using triangulation formulars and applying the co-ordinates of the known location of the involved base stations.

---

[2] *eEurope: An Information Society For All*

[3] *OTD-Observed Time Difference such as E-OTD (Enhanced Observed Time Difference) and OTDOA (Observed Time Difference of Arrival)*

[4] *TOA-Time Of Arrival*

[5] *GPS-Global Positioning System*

[6] *GLONASS-GLObal NAvigation Satellite System*

[7] *http://www.benefon.com/*

[8] *In the case of unsynchronised base stations.*

	Accuracy	User-controlled privacy	Speed of Response	Costs to Mobile Operator	Handset production Cost Increase	System Availability
Cell ID (optionally sectored)	Depends on cell size/ sectoring	No	3 sec	Minimal	Nil	1999
Timing Advance	1 km	No	5 sec	Minimal	Nil	1999
OTD (E-OTD/ OTDOA)	150m	Yes, partly[9)]	not known	Low/ Medium	Medium	2001
Wireless Assisted GPS/ GLONASS	5m (ideal conditions)	Yes	up to 1 min	Low	High (style, battery life)	2000
TOA	125m	No	10 sec	High	Nil	Late 2000
AOA	125m	No	10 sec	High	Nil	Late 2000

Table 1 Overview of enabling key technologies in the mobile communication market[10)]

Wireless Assisted Satellite based Systems (GPS/GLONASS)

Signals from at least four visible satellites as well as one earth based reference station are received by a GPS/GLONASS receiver unit incorporated into the handset. The fixed reference station helps to assist the handset to improve accuracy. There are nowadays different forms of implementation, i.e. the actual place where the location is calculated. In some cases, a snapshot of the satellite's positions is forwarded to a central server which in turn determines the location. Some other assisting systems help the GPS/GLONASS receiver to lock on the appropriate satellites which should be in the receiver's line of sight.

TOA (Time of Arrival)

This technology measures the differences between the times of arrival of the signal transmitted by a handset and received by several base stations (at least three) and forward the results to a central computer centre for further processing.

AOA (Angle of Arrival)

Similar to TOA, AOA determines the direction of the received signal transmitted by the handset at several base station sites. Sophisticated and bulky antenna arrays are needed to achieve this.

If the introductory vision of the snowboarder is taken, it will be clear that it might not be possible to just equip the handset with a GPS/GLONASS chip, because when buried under a certain amount of snow, a system will certainly not be able to determine the exact position of the poor person. The reason why this example has been used is that each technology has its advantages and disadvantages. If these technologies could be combined in a proper way, the advantages could be combined and would outnumber the disadvantages. Some promising and obvious combinations have already been proposed (i.e. sectored Cell ID in combination with TA, supported also by GPS/GLONASS in remote areas), but there are by far too many factors which make it difficult to choose the right combinations. However, the focus of this paper is not to discuss the appropriate combination of positioning technologies to meet the demands of cellular carriers, rather it is to point out the need to support hybrid technologies.

Table 1 provides a short overview of some key positioning technologies in the field of mobile communication systems, emphasising some of their more important preferences. It should also be mentioned that these technologies are currently still under discussion regarding how they might be implemented at a larger customer base.

Mobile Service Platform

The MSP (Mobile Service Platform) to be developed in the Youngster project will be a unique platform for mobile location services incorporating a wide range of context and personalisation features. These include location information, time/date, social context (for example, whether the youngster is currently in the classroom or at home) and various other parameters relating to the current context combined with a high degree of personalisation. This platform will enable the

9) *Depending on implementation*

10) *This table was partially taken from www.cursor-system.com/newsite/html/prod/tech.htm.*

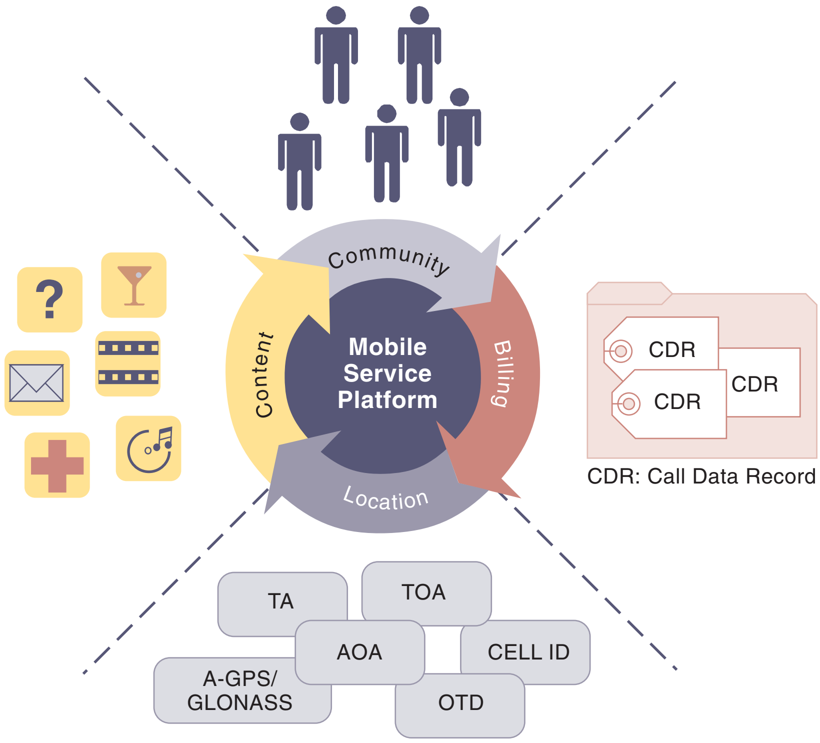


Figure 1 The Mobile Service Platform

easy development of context aware services. The MSP will consist of a set of traditional servers as well as open technologies such as Web servers, WAP gateways, Java engines and other technologies like LDAP[11]. There are plans to enhance this system through several parts: Among them are a User Information Retrieval System, modules for personalisation of services, servers for active services, support modules for location-based services, tools for content generation, and more. The platform also supports active services which are services brought to the user, known as push services, compared to the pull services known today. Working together with young users, Youngster will identify their needs and generate and demonstrate services specialised for young users. Figure 1 introduces the key concepts of the Youngster approach.

Content providers offer generic content, of different media types, which can be accessed as such, or combined with third-party content.

A key development on which this Mobile Service Platform depends is an enhanced user model, which includes not only conventional attributes relating to the user's location (such as plain co-ordinates, velocity, moving direction, etc.) but also a wide range of context attributes such as how this information is to be interpreted in the context of the user. These attributes will vary from relatively static ones to highly dynamic ones. Some of this information will be held by the user, some by the application service provider and some by the underlying support services. These attributes may be used in the selection of appropriate services for the user, by the services themselves to adapt to the user's needs, and by the advanced content formats. Consequently, appropriate levels of security and privacy in relation to the user profile are essential for any services that use it since the user may only be willing to grant access to parts of the user profile to different services. This raises questions as to where the user profile should be held and how access to it should be controlled. These issues will be addressed.

The enhanced user model and the MSP that contains it will be capable of creating services with greater potential to take account of and respond to the user's context than services that are currently available.

Key R&D Issues

Given the increasing availability and acceptance of wireless access networks, consumers (service providers and users) are eager to explore their applications to take advantage of the freedom of mobility. These applications can be made effective and user-friendly most appropriately by means of a middleware architecture; thus demanding an open mobile service platform for wireless mobile applications.

With particular respect to providing solutions giving mobile users easy and personalised access to web-based services, the following key impacts are relevant:

- *Localised and up-to-date information*
  The information should be given at the right moment and for the right location. This will be achieved by providing a mobile service platform (MSP), including access into various types of mobile devices, such as notebooks and PDAs which now allow for accessing net-based services even via third generation mobile networks such as UMTS.

- *Increased efficiency*
  The determination of the current position of an entity (user or goods) is an essential requirement in order to provide specialised services. Furthermore, positioning allows more efficient management of fleet and dispatch services and improved traffic management.

- *Personalised value-added access to net-based services*
  The convenience and capabilities of the access to net-based services, or even the possibility

[11] *Lightweight Directory Access Protocol.*

to create individual services will be improved. Users may download up-to-date application parts and specify where and when they may be reached or how they want to access a service. The information is filtered according to the specific needs of the customer: user preferences for certain stores or restaurants would automatically be registered in the customer database. Whenever the caller initiates a request for the nearest restaurant or hotel, they would automatically be notified of the closest one meeting their requirements.

Besides the above mentioned R&D issues security and personal privacy aspects are very important in the area of context-aware services, but also for community-based services.

Youngster will therefore make a special effort in this area by incorporating security features. For example, user controlled views will restrict the set of context-information visible to service providers. Context-information gathered by one participant (e.g. a service provider) will not be generally visible, only if the service provider grants access. Mechanisms like logging, automatic negotiation of privacy agreements and others are under consideration to ensure a user controlled privacy. Functions such as content management will be important aspects to consider, as well as how much parental surveillance is necessary or acceptable to more independent and mobile youths. This affects issues such as storage of location (or other context) information as well as storage of user profiles. For this we plan to follow the established general principles of the European Commission (95/46/EC and 97/66 EC).

Another important challenge – and benefit to the Youngster Project – are regulations such as set by the Federal Communications Commission (FCC) for 911 Automatic Location Identification (ALI) in the US or the European initiative 112, which furthermore put a lot of pressure on content-, service- and network providers, ensuring that location technology is implemented. The E911 directive stated: "By December 31, 2005, achieve 95 percent penetration of ALI-capable handsets among its subscribers. Furthermore, FCC adopted the following revised standards for Phase II location accuracy and reliability:

- For handset-based solutions: 50 metres for 67 percent of calls, 150 metres for 95 percent of calls;
- For network-based solutions: 100 metres for 67 percent of calls, 300 metres for 95 percent of calls."[12)]

According to the Mobile Location Services conference held in November 1999, the European Commission is also considering introducing a location enabled emergency system for mobile network carriers around 2003.[13)]

Hence, a crucial aspect in the evaluation process and acceptance testing is ensuring that the demands for protection of privacy are fulfilled and that the systems developed comply with current ethical rules.

Market and Services

Market studies forecast that by 2006 mobile location services will generate revenues more than $20 billion. But due to the investments that are required to collect location information, it is difficult for a single player to market a service on its own. Thus, the market structure in the mobile location service market is very segmented. Partnerships between mobile operators, content and service providers are therefore essential for the development of systems and services, for the specification of standards, and for the effective delivery of solutions to the market.[14)]

One example for such a "win-win-co-operation" is DoCoMo's I-mode service, which currently attracts more than 17.1 million mobile users as per the end of year 2000.

On the other hand appropriate business models are missing to address the specific user needs and service provider requirements. Even though the Youngster project is strongly focusing on a young user group there is a need for examining new business models.

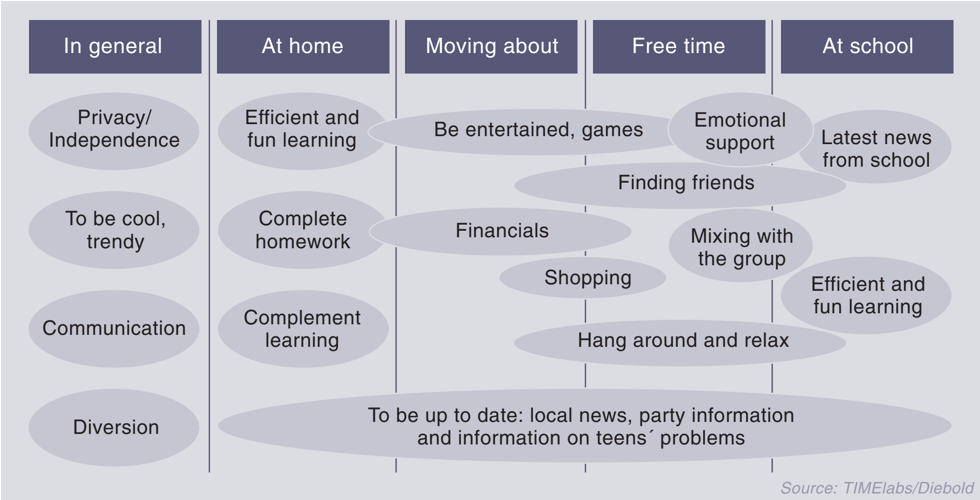
Some of the new models will include service usage in exchange for display space (e.g. for displaying advertisements), sponsoring of service usage and equipment, lower prices for active community members as well as analysing new cost neutral approaches, relying on a close relationship and gain benefits in another form.

Analysing the life style and behaviour of young people, one will identify the needs of users in general to implement location based services.[15)]

---

12) *FCC WIRELESS 911 REQUIREMENTS, January 2001.*
13) *IBC – Conference, Mobile Location Services, 16-17/11/1999, Mr Leo Koolen, CEC & ETSI.*
14) *Partially taken from Ovum Ltd., Mobile Location Services, 1998 & Mobile Location Services: Market Strategies, Dec. 2000.*
15) *Timelabs/Diebold: "Winning in Mobile eMarkets", Oct. 2000.*

Figure 2 Needs and Related Services of young people



The Location information itself enhances many applications, such as the so-called "Lifestyle Mobility Services":

- "Find the next" Applications – Restaurant, Pub;
- Offers a huge range, as Travel Information or Route Planning;
- Developing of innovative Applications – Friend Finder;
- Supports specialised "community interest" applications.

A common complaint from users is that while it is wonderful to be able to access all kinds of information on-line, it is not so wonderful when an enquiry or search turns up over a thousand hits, many of which are irrelevant. Over the mobile network, where bandwidth is a scarce and expensive resource, this is an even more pressing issue. Location dependence allows information to be targeted more precisely.

The above mentioned services are important for users because they offer more relevant personal information. Businesses and individuals are becoming increasingly reliant on electronic access and delivery of information to improve their efficiency and quality of life. Furthermore, there is a huge potential to generate additional revenues for service operators, e.g. through advertising and subscription. One can think of many other examples combining personalisation and location awareness that could exploit this approach for a wide range of users. One obvious area is tourism where it may be used to create services for tourists providing information on places of interest near them based on the preferences in their profile and their current location. This information can be presented with appropriate multimedia clips (based on their interests) and directions on how to get there.

In general, location based services will offer a lot of opportunities, but to implement security/privacy aspects is a crucial point for the expectancy by the customer.

To fill these gaps the project will develop technologies for an open active mobile multimedia service using the Mobile Service Platform. These services will include the following features:

- Sensing of and smart interaction with other devices in the user's environment;
- Personalisation and adaptation of services, user awareness;
- Integration of location aware services.

Afterwards the prototypical services will be validated in a Norwegian Field trial with a suitable test community of young people. The results of the trials will be evaluated, and success will be assessed in terms of the response of the young people who participate and the use that they make of the services.

Conclusion

In this paper we have argued that location-aware services will play a particularly important role in providing services for mobile users.

All enabling positioning technologies have drawbacks, either of cost, complexity or accuracy. We assume that a combination of a satellite based system like GPS and a flavour of a network-based positioning determining systems will establish a "hybrid solution" which is able to provide location information anywhere.

Although the market will develop through early trials and implementations using early positioning systems, market development will be limited until there is a clear technological choice.

User requirements as well as upcoming R&D issues show that there is a clear need for personalised and location-based services accessible anytime and anywhere. Furthermore, the provision of an open service platform infrastructure will provide an important advance for the success of the market. This will extend and enhance the potentials of mobile scenarios for citizens and will enable existing services to be improved and new ones to be developed. These new services will be attractive and easy to use as they will combine personalisation with context awareness to provide a form of support not previously available. However, the introduction of location-aware services require new business models. Security, personal privacy and ethical aspects are very important in the area of context-aware services and should be taken into account.

The Youngster approach is to combine innovative technology and service developments together with an enthusiastic user group: young people. Consequently, a test community of young people will be involved from the definition of requirements up to the evaluation of the prototypes in a European field trial.

In general, the developments that will emerge from the Youngster project will provide a step towards the creation of a computing and advanced networking technology which is aware of "the user's presence, his personality, his needs, and is capable of responding intelligently ..." (IST Advisory Group[16)]). These developments will enhance the quality of citizens' lives by providing universal and personalised access to services.

References

For further information please visit http://www.ist-youngster.org/.

[16)] *Information Society Technologies Advisory Group, Orientations for Workprogramme 2000 and Beyond.*

Status

Introduction

PER HJALMAR LEHNE

Per Hjalmar Lehne (42) obtained his MSc from the Norwegian Institute of Science and Technology in 1988. He has since been with Telenor R&D working with different aspects of terrestrial mobile communications. 1988 – 1991 he was involved in standardisation of the ERMES paging system in ETSI as well as in studies and measurements on EMC. His work since 1993 has been in the area of radio propagation and access technology. He has participated in the RACE 2 Mobile Broadband Project (MBS), COST 231, and COST 259. From 1998 he was leader of Telenor R&D's smart antenna project. He is currently involved in work on 4th generation mobile systems and the use of MIMO technology in terrestrial mobile networks.

per-hjalmar.lehne@telenor.com

The International Telecommunications Union – ITU – is facing increasing competition in having the global leadership in telecom standardisation. Several regional and more industry-based fora have taken the lead in some aspects. Among other things, this is caused by the way the ITU works and the fact that standards traditionally have taken a very long time to develop. Other fora have found ways to work that make the lead time from idea to standard much shorter. This is probably the greatest threats to ITU's position.

In this issue of *Telektronikk*'s Status section, we focus on the work in ITU and the significance it still has. The first paper is written by Anne Lise Lillebø and gives a comprehensive report of the major results from the *ITU World Telecommunications Standardisation Assembly (WTSA),* which was held in Montreal, Canada in September – October 2000. An important outcome is the introduction of the so-called Alternative Approval Process (AAP) – also called the "fast-track approval process". ITU reforms were in general an important issue here.

In the second paper, Arve Meisingset gives a personal view on *Standardisation Policy*, especially focused on ITU-T, based on his broad experience from several years of active participation. He draws a complex and confusing picture of international technology development and standardisation, and points out the need for better co-ordination between the different actors on this scene.

The third paper is of a technical character. Terje Henriksen gives an introduction to *Network Level Modelling* in ITU. He describes the Reference Model for Open Distributed Processing (RM-ODP) from different viewpoints: the enterprise, the information, the computational and the engineering viewpoint, as well as the functionality viewpoint.

WTSA 2000

ANNE LISE LILLEBØ

Anne Lise Lillebø (52) is Adviser to the Corporate Management of Telenor. Her main responsibilities include general and policy matters related to international organisations. She holds a Master of Arts degree from the University of Oslo.

anne-lise.lillebo@telenor.com

This article presents an overview of the major results of the ITU World Telecommunication Standardization Assembly (WTSA) in Montreal, 27 September – 6 October 2000.

1 Introduction

1.1 Tasks

The World Telecommunication Standardization Assembly (WTSA) is held every four years by the International Telecommunication Union (ITU), a specialised agency of the United Nations, to define general policy for ITU's Telecommunication Standardization Sector (ITU-T Sector). WTSA adopts the work programme, sets up the necessary Study Groups (SGs), designates chairmen and vice-chairmen of the Study Groups, reviews the working methods and approves Recommendations.

The main issues of the WTSA in Montreal were:

- Adoption of the ITU-T work programme and new structure of Study Groups;
- Establishment of a new Special Study Group to deal with IMT-2000 and beyond;
- Appointment of chairmen and vice-chairmen of the 14 Study Groups;
- Working methods of the ITU-T, including the so-called Alternative Approval Process (AAP) or fast-track approval process;
- ITU Reform;
- Rights of Associates;
- Cost sharing of international Internet connections.

1.2 Participation

The WTSA is open for participation both from Member States and Sector Members (operators and industry) of the T-Sector. Telenor is a Sector Member of the ITU-T and participated in its capacity as Sector Member, whereas Norway was represented by the Norwegian Post and Telecommunications Authority. A number of participants from Sector Members from other countries were included as members of their national delegations and did not represent their companies under their own names.

1.3 WTSA Structure and Management

WTSA was skilfully chaired by Mr. Mike Israel, Canada. The work was divided into seven committees:

Committee 1	– Steering Committee
Committee 2	– Budget control
Committee 3	– Working methods of ITU-T
Committee 4	– ITU-T Work Programme and Organization
Committee 5	– Telecom Network Infrastructure
Committee 6	– Telecom Services and Tariff Issues.

The discussions in the committees were often dominated by the wish to obtain consensus instead of seeking the majority view. Consensus building is the normal way of work in a Study Group, but a WTSA could also seek a majority view in order to obtain results in a timely manner. Although Sector Members are very active in the work of the ITU-T Study Groups, the discussions at the WTSA were dominated by the Member States.

1.4 Opening Speeches

Mr. Yoshio Utsumi, ITU's Secretary-General, gave a strong opening speech encouraging the WTSA delegates to take bold decisions on reform in order to maintain ITU-T's position as a recognised world-wide standards body. ITU-T tries to streamline its working methods and image in order to speed up the development of global standards, to keep its present Sector Members and especially to attract new players in the IT field.

1.5 European Preparations

For the first time CEPT ("Conférence Européenne des Administrations des Postes et des Télécommunications" – a body of European policy-makers and regulators) had made co-ordinated preparations for the WTSA in co-operation with ETNO (European Public Telecommunications Network Operators' Association – a trade association for European operators) and EICTA (European Information and Communications

Technology Industry Association). Telenor is a member of ETNO and took part in the European preparations together with the Norwegian telecommunications regulator. The work resulted in the submission of five European contributions:

- Establishment of an ITU-T Project Group to address IMT-2000;
- A draft resolution on ITU Reform;
- Proposed amendment to the Alternative Approval Process (AAP) to secure Sector Member participation;
- Proposal on improving ITU-T's consideration of strategic issues;
- A revised draft Recommendation on international Internet connection.

Telenor co-signed the proposals on IMT-2000, Reform, AAP and Internet.

2 Work Programme and New Study Group Structure

Standardisation work is carried out in the various Study Groups (SGs) where Member States and Sector Members of ITU-T develop Recommendations on the basis of the study of Questions (areas for study). The WTSA focused on a large range of topics including Internet Protocol (IP) based networks, international mobile telecommunications, optical network infrastructure and new technologies related to new multimedia systems.

WTSA allocated a total of 193 Questions to the Study Groups. ITU-T will have fourteen Study Groups in the next four year period 2001 – 2004 with the general areas of responsibility as shown in Box 1. The new SG structure started on 1 January 2001.

WTSA agreed to disband the former SG8 on "Characteristics of telematic systems".

The continuation of SG7 was challenged and Telenor supported the Canadian proposal to amalgamate most of SG7 with SG 11. In addition, Telenor proposed the transfer of certain questions from SG7 to SG10. However, the continuation of SG7 was agreed but the need for coordination between SG7, SG10 and SG11 was emphasised.

There was general agreement that SG13 should be responsible for conducting IP-related studies and its mandate now encompasses multi-protocol and IP-based networks and their interworking.

Box 1 – New ITU-T Study Groups, Study Period 2001 – 2004

- SG 2 – Operational aspects of service provision, networks and performance
  Chairman: Mr. R. Blane (Inmarsat, UK)
- SG3 – Tariff and accounting principles including related telecommunication economic and policy issues
  Chairman: Mr. R. Thwaites (Australia)
- SG4 – Telecommunication management, including TMN
  Chairman: Mr. D.J. Sidor (Nortel Networks, USA)
- SG 5 – Protection against electromagnetic environment effects
  Chairman: Mr. R. Pomponi (Cselt, Italy)
- SG 6 – Outside plant
  Chairman: Mr. F. Montalti (Telecom Italia, Italy)
- SG 7 – Data networks and open system communications
  Chairman: Mr. H. Bertine (Lucent Techn., USA)
- SG 9 – Integrated broadband cable networks and television and sound transmission
  Chairman: Mr. R. Green (Cable Labs, USA)
- SG 10 – Languages and general software aspects for telecommunication systems
  Chairman: Mr. A. Sarma (Deutsche Telekom, Germany)
- SG 11 – Signalling requirements and protocols
  Chairman: Mr. Y. Hiramatsu (NTT, Japan)
- SG 12 – End-to-end transmission performance of networks and terminals
  Chairman: Mr. J.-Y. Montfort (France Télécom, France)
- SG 13 – Multi-protocol and IP-based networks and their internetworking
  Chairman: Mr. B.W. Moore (Lucent Techn., UK)
- SG 15 – Optical and other transport networks
  Chairman: Mr. P. Wery (Nortel Networks, Canada)
- SG 16 – Multimedia services, systems and terminals
  Chairman: Mr. P.-A. Probst (Swisscom, Switzerland)
- SSG – Special Study Group "IMT-2000 and beyond"
  Chairman: Mr. J. Visser (Nortel Networks, Canada)

2.1 New Special SG IMT-2000 and Beyond

The main objectives of the European proposal on IMT 2000 was to create an entity with efficient working methods meeting the requirements of the market and avoiding duplication of work. The European contribution proposed the establishment of a new type of project group within the ITU-T that could work in a more flexible manner independently from the present working procedures applying to a regular Study Group. It was also proposed that such a group should be authorised to develop a new kind of output, e.g. technical specifications. Duplication of work with other SDOs (Standards Development Organisations) should be avoided and provision should be made for referencing outputs of SDOs in its output documents.

Although many participants agreed that there was a need to adopt a flexible approach tailored to rapidly respond to market requirements, no agreement could be reached whether this necessitated a new set of working methods different from those of a regular SG.

A number of important non-European players supported the establishment of a separate Study Group for IMT-2000 and could not endorse the European contribution. The compromise reached was to set up a new Special Study Group on IMT-2000 and beyond (SSG). The primary responsibility of this special SG is the overall network aspects of IMT-2000 and beyond, including wireless Internet, convergence of mobile and fixed networks, mobility management, mobile multimedia functions, interworking and interoperability. The SSG will be tasked with enhancing network interoperability among existing IMT-2000 systems specified by ITU-T and external standards development organisations (SDOs), Partnership Projects (PPs), IETF and relevant external forums and to provide a migration path regarding network aspects and mobility from existing IMT-2000 systems towards systems beyond IMT-2000. Care must be taken to avoid duplication of work being performed in the ITU-R, in other SDOs, in PPs (such as the Third Generation Partnership Project – 3GPP) and the IETF.

The special SG on IMT-2000 and beyond (SSG) is special in the sense that the SSG will be allowed to apply the provisional working procedures set out in Recommendation A.9 approved by WTSA. The Special Study Group is encouraged to work electronically to the maximum extent possible. The Special Study Group is not bound by the rules governing the regular Study Groups concerning the frequency of physical meetings. Physical meetings may be announced electronically with a minimum of one month's notice and contributions may be input up to five working days before the meetings.

The Special SG on IMT-2000 and beyond is authorised to develop and approve Recommendations in the same manner as other Study Groups. The Special SG may also consider other deliverables of a status below that of an ITU Recommendation such as normative technical specifications or interim Recommendations. Proposals for such output should be submitted to the Telecommunication Standardization Advisory Group (TSAG) for consideration and approval. The provisions for such deliverables must, however, be incorporated in the A-series Recommendations (containing ITU-T working procedures) and approved by TSAG following the traditional approval procedure including consultation of Member States.

3 Appointment of SG Chairmen and Vice-Chairmen

The candidates for chairmen and the vice-chairmen of the Study Groups are discussed at the meetings of heads of delegations of Member States (where Sector Members are not authorised to take part) and are finally appointed by the WTSA. Following a request from Europe to the TSB Director (Telecommunication Standardization Bureau, the secretariat of ITU-T), a list of all the candidates with their CVs and affiliation was available to the ITU members before the WTSA. It is the first time that the membership as a whole had access to information about the candidates and the number of candidates to the various posts prior to the WTSA and this transparency was much appreciated by Telenor and other participants.

Although there is an evident need for ITU-T to have competent chairmen and vice-chairmen, the geographical balance and national interests are often given more weight when a candidate is chosen. For the new special SG on IMT 2000, there was a list of 11 candidates for the vice-chairmanship. Unfortunately, WTSA was unable to select a limited number – instead all 11 candidates were appointed as vice-chairmen!

In Study Group 2, there were two competing candidates for the chairmanship both coming from European members – UK and France. As no agreement on one candidate could be reached, it was decided to share the chairmanship and vice-chairmanship by splitting the functions in periods of two years each. Such a solution has never before been practised in the ITU-T. It is disappointing that it was impossible to decide on one chairman, for example by using a vote. This is certainly a weakness of the present system.

The WTSA also designated the chairman (Mr. G. Fishman, Lucent Techn., USA) and vice-chairmen of the Telecommunication Standardization Advisory Group (TSAG) and agreed to limit the terms of office of all SG chairmen and vice-chairmen including those of TSAG to eight years (two SG periods).

4 Role of Telecommunication Standardization Advisory Group (TSAG)

A number of members of the ITU-T sector had raised their concern about the four-year cycle of the WTSA and questioned the sector's ability to reach decisions responding to market needs. It was concluded that TSAG could be used as a tool in between WTSAs to solve part of these problems.

WTSA passed a resolution delegating authority to TSAG to act on behalf of WTSA in between two WTSAs. The authorisation now includes the ability to restructure and establish ITU-T Study Groups and to assign chairmen and vice-chairmen to act until the next WTSA, to issue advice on Study Group schedules and to advise the Director on financial matters. TSAG is also requested to consider establishing two new permanent groups: a strategy group to develop a policy and strategy for the Sector and an operational group to develop working methods.

Telenor supports the delegation of more power to TSAG, which meets once or twice a year. We find that enhancing the role of TSAG is a good way of progressing necessary decisions, which cannot wait for another WTSA. TSAG is also a forum where both Member States and Sector Members are actively involved in the discussions and where there is a good basis for obtaining compromise on reform.

5 Changing Role of the WTSA

The participants at the WTSA in Montreal found that as TSAG was delegated increased authority, it was logical to reconsider the role of the WTSA. A resolution on the more general high level role of WTSA was agreed, stating that there is still a need for WTSA to meet in order to address overall policy and strategic issues such as for example future work direction, approval of Recommendations on an exceptional basis and to discuss matters of common interest to the Sector. It was also recognised that future meetings of WTSA might be of a shorter duration. The resolution encourages Member States to make contributions to the Plenipotentiary Conference to continue the modernisation process.

6 New and Revised Working Procedures – Alternative Approval Process (AAP)

The Plenipotentiary Conference in Minneapolis 1998 (PP98) had tasked the T-Sector to develop an alternative approval process (AAP) for Recommendations that have no regulatory or policy implications – so-called "technical Recommendations" – based on the PP98 changes to the Constitution (CS) / Convention (CV) and PP Resolution 82, which lays down that such Recommendations should be approved by Member States and Sector Members "acting together". The intention was to give the Sector Members a direct influence on the approval of technical Recommendations and to speed up the approval process in order to further reduce time-to-market delivery of standards.

TSAG had elaborated a carefully balanced proposal for the AAP relating to technical Recommendations. When draft Recommendations have been developed to a sufficiently mature state (the "consent" stage), the last call will be initiated. The draft Recommendation will be sent out for comment both to Member States and Sector Members. After a period of four weeks, the draft Recommendation is considered approved if no comments of opposition are received. If comments of opposition are made in this last call, a series of processes are foreseen to enable members to reach an agreement.

However, in the very last call, if an unopposed agreement on the approval of the draft Recommendation has not been achieved, only the Member States present at the SG meeting are consulted and one Member State may veto the approval of the Recommendation. This veto solution can hardly be interpreted as Member States and Sector Members "acting together".

The European proposal on AAP was to change the proposed final approval procedure for AAP in the very last call from "unopposed agreement of Member States present in the meeting" to "unless objected to by (3) of the present Member States and Sector Members". A number of influential Member States considered that any involvement of Sector Members in the final approval of Recommendations implied giving them the right to vote. According to the present basic instruments (Constitution/Convention), the right to vote in ITU is the prerogative of Member States only, and the European proposal was turned down.

WTSA adopted the text developed by TSAG as Recommendation A.8. It is a positive sign that the procedure is given in an A-series Recommendation and not in a WTSA Resolution. TSAG is responsible for the updating of the A-series Recommendations which contain various working procedures for the T-Sector and it is possible to change such A-series Recommendations on a continuous basis without waiting for another WTSA.

The introduction of AAP might be considered as a first positive step to allow the Sector Members a more active participation in the decision-making. If no opposition to the draft Recommenda-

Box 2 – ETNO Declaration

We, the 45 member companies of the European Public Telecommunications Network Operators' Association, welcome efforts by the ITU-T sector to develop global standards which could have a strong impact on the future of our industry. We have been in Montreal this week to underline our support for the work of the ITU, which provides for a unique global partnership between governments and industry.

However, we believe that the ITU-T should be a more commercially oriented organization designed to promote telecommunication services of value for both developed and developing countries. Necessary changes should be implemented rapidly as our member companies must deal with the realities of today's fast-changing marketplace.

At week's end, we remain seriously concerned over the exceedingly slow pace of change that does not reflect the "drastic changes" called for by the Secretary General of the ITU at the outset of this assembly. For the ITU to remain relevant, the industry sector must be given a stronger voice. Far too much debate during this assembly has been taken up with procedural matters rather than focusing on technical standardisation issues.

For the first time ever, some of our member companies are not participating in this assembly out of frustration with the slow pace of the ITU reform process. The next plenipotentiary gathering in Morocco will be critical to the future of the ITU. Unless we see more evidence of meaningful reform, our member companies will increasingly consider other standardisation bodies to develop specifications and standards.

Still we rely on the understanding and sense of compromise of a majority of visionary member states that will allow the ITU-T to develop into a more efficient standardisation body without detracting from the ITU's strength and values also fully recognised by ETNO Members. We would also like to express our sincere appreciation for those member states who support us in finding a good way for acting together in the future.

Montreal,
October 2000

tion is expressed in the various phases of the procedure, a Recommendation might be approved within two months from the time when the text is mature.

The AAP will be applicable upon the termination of the WTSA. It remains to be seen whether the AAP will be successfully applied in the T-Sector. The TSB will make use of electronic means in the consultation of Member States and Sector Members.

7 Associates

The Plenipotentiary Conference, Minneapolis, 1998, introduced a new category of participants known as Associates to enable small entities or organisations to take part in the work of ITU's three sectors. WTSA confirmed that Associates are entitled to take part in the work of one selected single Study Group of ITU-T. The WTSA participants hope that this new category of membership will encourage smaller companies, institutions and organisations to take part in ITU-T activities. The Council has decided that the annual amount of financial contribution for Associates will be 10,500 Swiss Francs (1/6 of the 2001 financial contribution of a Sector Member).

8 Resolution on ITU-T Reform

It was agreed that WTSA should offer an opinion and advice on the future organisation of global standardisation activities as an input to the ongoing work of ITU's Working Group on ITU Reform (WGR) to consider a new global standardisation entity under the umbrella of ITU.

The European contribution endorsed the idea that all technical standardisation in the ITU should take place in one single body and secondly, that a possible new entity dealing with global standardisation should be within the ITU, but not necessarily within the ITU-T as proposed by Canada.

The WTSA adopted a resolution focusing on issues relating to technical standardisation and lists a number of key attributes considered important for the successful continuation of a global standardisation body such as openness, transparency, visibility, consensus based approval and responsiveness to the needs both of the market and the developing countries. TSAG is encouraged to support the work of the Working Group on ITU Reform and Member States are requested to make input on these issues to the next Plenipotentiary Conference.

9 Draft Recommendation Diii on International Internet Connection

In exceptional cases draft Recommendations are submitted to WTSA for approval. There is a tendency that this procedure is used for contentious draft Recommendations. At a WTSA approval by formal vote can be done by simple majority.

The draft Recommendation Diii on cost sharing for international Internet traffic was one of the most controversial issues of the whole Assembly. The Recommendation put forward by SG3 aimed at setting down a principle for the equitable cost sharing of international Internet connections. The opponents to the Recommendation feared that the Recommendation might impact on the development of the Internet if it imposed a given result on commercially negotiated agreements. There was a sharp divide between the US and a number of industrialised countries and the rest of the world. Europe's proposal of an alternative text was rejected. It would have been preferable to have the text sent back to the SG for further work.

The resulting solution was a redrafted text slightly more watered down than the European text. The approved Recommendation states that "administrations involved in the provision of international Internet connections negotiate and agree to bilateral commercial arrangements enabling direct international Internet connections that take into account the possible need for compensation between them for the value of elements such as traffic flow, number of routes, geographical coverage and cost of international transmission amongst others". The overall interpretation of this text was that international Internet connections remain commercial agreements. The US and Greece made reservations that they did not intend to apply this Recommendation in their international charging arrangements.

It should be emphasised that all ITU-T Recommendations are voluntary and do not form part of any intergovernmental treaty. The Member States are not formally bound to observe the Recommendations, but in practice quite a few players will normally refer to ITU-T Recommendations as the recognised way forward.

10 European Co-ordination during WTSA

Both CEPT and ETNO had extended co-operation meetings during the Assembly. For the first time CEPT had made a concerted effort to prepare European views ahead of the WTSA. However, during the Assembly it became evident that a number of European Member States had quite differing views and it was extremely difficult to obtain the necessary commitment from all the CEPT members to pursue a specific direction of action. Although five European contributions were formally submitted to the WTSA, some of the European players were not in agreement and Italy and Greece spoke against the contributions in the formal meetings. The disagreement within CEPT gave a signal to the outside world that CEPT was not united in its approach to the WTSA issues and this might have given other regions the opportunity to divide Europe and take advantage of the situation.

ETNO issued a declaration at the end of the WTSA expressing their concern on the slow pace of change adopted by the Assembly that does not meet the call for rapid change as underlined by the Secretary General in his opening speech. Much of the debate at the Assembly revolved around procedural matters and did not concentrate on strategic standardisation issues. ETNO aims at giving private industry (i.e. Sector Members) a more active role in the overall management of the ITU-T.

Box 3 – Abbreviations

AAP	Alternative Approval Procedure
CEPT	Conférence Européenne des Administrations des Postes et des Télécommunications
EICTA	European Information and Communications Technology Industry Association
ETNO	European Telecommunications Public Network Operators' Association
IETF	Internet Engineering Task Force
IMT-2000	International Mobile Communications 2000
IP	Internet Protocol
IT	Information technology
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector
PP98	Plenipotentiary Conference 1998
PPs	Partnership Projects
SDO	Standards Development Organisation
SG	Study Group
SSG	Special Study Group
TSAG	Telecommunication Standardization Advisory Group
TSB	Telecommunication Standardization Bureau
WGR	Working Group on ITU Reform
WTSA	World Telecommunication Standardization Assembly
3GPP	Third Generation Partnership Project

Standardisation Policy

ARVE MEISINGSET

Arve Meisinset (52) is Senior Research Scientist at Telenor R&D. He is currently working on information systems planning, and has previously been engaged in Case-tool development and formal aspects of human-computer interfaces. He has been involved in several network management projects, and has a particular interest in languages for data definitions and mathematical philosophy. He is ITU-T SG10 Vice Chairman, Working Party Chairman for WP3/10 Distributed Object Technologies, and the Telenor ITU-T technical co-ordinator.

arve.meisingset@telenor.com

ITU is a specialised agency of the UN. Having participated for some years at all levels of ITU-T, in addition to some participation in other bodies, I will summarize some of the main questions and conflicts in global standardisation – as I see them.

The members of a global standardisation organisation, like ITU-T, will necessarily have different roles, different backgrounds and different interests, which make it difficult to establish effective consensus views. For example, West European telecommunication operators have experienced a liberalisation where they are no more restricted to operate within national borders, and they can act independently of the state, while they experience a strong competition in their original national markets. The situation is similar in the USA, but the state seems to exhibit a stronger role at the regional level, as seen in the divestiture of AT&T and the lawsuit against Microsoft. Similar regulation of business is attempted in Europe; however, here each EU member state is an ITU Member State, while in USA the entire union of states only is a Member State.

While the developed countries in the past have used high prices on international connections to subsidise the development of their national infrastructure, they are now ready to compete all over the world. Less developed countries have not had the opportunity to build their infrastructure and adjust to a deregulated market before the competition starts. Hence, we see different attitudes to policy questions between developed countries, to developing countries, between operators, between vendors and between the various types of ITU members, i.e. Member States or Sector Members (telecommunication operators, industry and others). Example issues where these conflicts come to the surface are cost sharing for international telephone traffic, cost sharing of Internet traffic, call back, satellite terminals, and the role of State Members versus Sector Members.

The standardisation sector of ITU has in the past served as an interest organisation of traditional telecommunication operators. Agreed ITU-T Recommendations have proved to be a powerful means to force vendors to deliver products according to the operators' wish. With the developer, i.e. vendor, participation, this has often proved to be an effective way of bringing forward interoperable products from different vendors. However, new telecommunication operators have not always given participation in ITU-T a high priority.

As the computing technology has been introduced to implement communication services, there has been a need to co-ordinate standardisation work in the various global standardisation organisations, like ITU-T, ISO and IEC. Use of so-called common texts, has resulted in standards being carried between these bodies and rubber-stamped by them. This is not a very efficient way of co-ordination.

Regional standardisation organisations have tried to work faster and more flexible than the global standardisation organisations, and in many cases they have succeeded. National standardisation organisations have most often not been that ambitious, but have seen their role as co-ordinating and expressing their national interests in the regional and global organisations, and to promote the international standards at the national level, including translation.

However, in some cases the regional standardisation organisations have global aspirations, as expressed in the ETSI Future Role. ETSI had a success story with GSM, and is trying to copy the success for UMTS. This is done by establishing the 3GPP outside ETSI, with participation from other regional standardisation bodies, while the results are adopted as ETSI Deliverables and by the other bodies, as well. This way, 3GPP acts as a global standardisation body in the mobility domain and rivals the current international standardisation organisations, such as ITU-T. It is this conflict that comes to the surface with the establishment of the ITU-T Special Study Group for IMT-2000 and beyond. The Europeans do not want it, as they are happy with UMTS as the global standard, while others want interoperation between several regional standards. This is also a question of where work should be done and where we should send the experts. 3GPP is an asset to the ETSI aspirations, while being a challenge to ETSI's current role by moving the most important work item outside of ETSI.

Attempts to co-ordinate work in global and regional standardisation bodies have been made through Joint Standards Collaboration meetings. Also, there is a range of lobbying organisations trying to move the work in certain directions.

ITU has identified some 600 fora and consortia within the computing and telecommunication domains. ITU-T has established a formal relationship to some tens of these bodies. The fora and consortia are often established by product developers. Some have become very successful, and in some cases they are competing with and outperforming the traditional standardisation organisations. The Internet Engineering Task Force (IETF) is an example of this, and has challenged ITU-T on protocol standardisation. The Object Management Group (OMG) has had great success on standardisation of higher level protocols and software. Often, however, fora and consortia need the standardisation organisations to give credibility to their results and ensure usage of them. TINA-C and EURESCOM are organisations that have used ITU-T for this purpose. It should be noted that the vendors attending the fora and consortia are not a homogenous body, but may have conflicting interests. Traditional telecommunication operators, not developing their own products, have a less central role in many of the fora and consortia. There are exceptions, as well. TeleManagement Forum (TMF) is heavily influenced by traditional telecommunication operators, while the main beneficiaries may be new telecommunication operators. Often the same experts or companies play key roles in several of these bodies simultaneously, making it difficult to overview the relationships between the fora and the standardisation organisations, as is the case between TMF and ITU-T Study Group 4.

Fora and consortia have frequently been cleverer to promote their results than the traditional standardisation organisations. However, the standardisation organisations are catching up. ITU-T encourages its Study Groups to develop promotion plans. Study Group 10 has started to give tutorial presentations to newcomers and outsiders in conjunction with every meeting. Also, Study Group 10 members have established an independent forum, the SDL Forum Society, which organises conferences and seminars. Together with the ETSI MTS group, these organisations provide a triangle organisation which uses each other's strengths. This way, Study Group 10 has already achieved some of the flexibility that is sought through the ITU Reform.

It is frequently claimed that standardisation organisations are too slow and bureaucratic. However, the statistics show that in ITU-T the time from initiation of work to approval of the Recommendation is shorter than the comparable time in many recognised fora. This is impressive, since ITU-T often puts more emphasis on ensuring the technical and formal quality of documents than the fora. Experience shows that results from fora have to be considerably reworked to become an ITU-T Recommendation.

Also, the strife to provide open standards for implementation by anybody, implies that ITU-T puts more emphasis on formal specification of protocols, test suites and languages than in many other organisations. I fear that continuing pressure to increase the production speed will reduce the quality of ITU-T Recommendations. Contrary to the strife for openness and high quality of ITU-T Recommendations, some fora may with their low quality output documents be suspected to have a hidden agenda to provide some competitive advantage to their fora members. On the other hand, while some fora provide their documents free of charge, ITU-T Recommendations have a relatively high price, which may reduce their widespread use. On an experimental basis, ITU-T will in 2001 allow any users to download 3 Recommendations free of charge.

Given the different roles of ITU members and high membership fees, it is difficult to reorganise ITU to take on new challenges and attract the right newcomers to the work. It seems to be much easier to establish a new forum with a clear objective and appropriate membership to carry out its mission. The membership fee for Associates in ITU-T is still too high to attract university members. Capability to make rapid reorientation is the area where ITU-T necessarily comes short compared to fora and consortia. Maybe this situation could be improved by trying to develop rolling (bottom-up) technology strategies for ITU-T.

Fora can be more flexible than ITU-T by producing technical specifications not satisfying the same formal requirements as ITU-T recommendations. This could be a convenient division of roles between technology development and standardisation. However, there is a danger that the experts will attend the technology development and not standardisation if these two phases are split on different organisations. Even establishment of fora work practises within ITU cannot outrule the establishment of independent fora. Therefore, ITU-T should continue co-operation with these fora and focus on its role as the pre-eminent global standardisation organisation for telecommunications. The use of Focus Groups within ITU-T and the establishment of the Special Study Group on IMT-2000 and beyond are attempts to provide fora work practices within ITU-T.

In ITU-T we have seen needs to distinguish between standardisation and

- regulatory issues;
- radio spectrum and space management;
- technology development;

and corresponding procedures and financing of these different pieces of work. The new Alternative Approval Process within ITU-T applies to technical Recommendations only, while the Traditional Approval Process applies to Recommendations of a regulatory nature. The current Radio sector (ITU-R) does management of radio spectrum and space as well as standardisation of radio and broadcasting technology; attempts to move the standardisation aspects to ITU-T have failed. The ongoing work on ITU Reform tries to separate technology development from the intergovernmental procedures and financing of ITU. Currently all financial contributions go to ITU as a whole, and are then split on the three sectors, with the Development sector (ITU-D) using the largest portion, then the Radio sector (ITU-R); and the Standardisation sector (ITU-T) receives the smallest portion, even if this is the sector having the largest production – and sector member participation. Sector Members are seeking greater independence for the individual sectors of ITU, especially for the ITU-T, direct financial contributions to this sector only, and they seek greater Sector Member influence on decision-making, finances and management.

The above provides a complex and confusing picture of international technology development and standardisation. There is a need for better co-ordination between the global standardisation organisations and to the regional organisations. It is my belief that some fora have attracted too much of the attention of the telecommunication operators, and that these operators would benefit from maintaining ITU-T as their main interest organisation. This strategy will require that they continue to carry new work items to ITU-T and to provide their best expertise to develop the solutions. I also believe that ITU-T should extend its scope to cover more software and business software standardisation that support the operation of the telecommunication business.

Finally, many fora participants have experienced getting lost in the process and having no real influence on the result. The working procedures of ITU allow every voice to be heard during the development of a Recommendation, independently of the member being small or large, or being very involved in the subject matter or not.

Network Level Modeling in ITU

TERJE HENRIKSEN

Terje Henriksen (58) is Research Scientist at Telenor R&D. He has been involved in standardization and R&D projects within the area of network level modelling since 1991; from 1996 to 2001 as the rapporteur for Question 18 of SG4 within the ITU-T.

terje-fredrik.henriksen@telenor.com

1 Introduction

In the early nineties, influenced by the emerging object oriented methodology, SG XVIII (later SG 13) in ITU developed the overall network architecture defined in G.805 [1]. It provides a high level view of the network function based on a small set of architectural entities (functional blocks) interconnected via reference points. Two main network representations may be provided on the basis of this architecture:

- Topology[1)] in terms of links, subnetworks and access groups[2)];

- Connectivity in terms of trails, link connections, subnetwork connections, ports and reference points.

The topological view describes the geographical distribution of the resources of a layer network. The access group is a container for a number of co-located access points. The subnetwork represents the routing capabilities within a site or grouping of sites and the link represents the transport capacity between subnetworks. Layering is a method for splitting the overall transport function into a hierarchy of layer networks on top of each other, each of which utilizing the service from the server layer to provide its own service. A topological view of a layer network consisting of three sites is shown in Figure 1.

Following the definition of the overall network architecture, the development of the generic (i.e. technology neutral) network model started in 1994 in Question 30 of SG 15. As a consequence of the decision to make SG 4 the "Study Group for Management" within the ITU organisation, network level modeling became Q.18 of SG 4 in 1996.

This status paper describes the foundations for the generic model, the functionality currently supported and the extensions planned. To demonstrate the capabilities of the model, the modeling methodology is described in some detail. The creation of a trail termination point is used as a modeling example to illustrate the usage of the different modeling constructs.

The existing model is applicable to connection-oriented core network technologies like SDH, ATM and WDM. Work has started to include the access network and also connectionless communication such as IP.

2 Modeling Methodology

The modelling methodology [2] was developed by Q.18/4 on the basis of the Reference Model for Open Distributed Processing (RM-ODP) which is specified in [3, 4, 5, 6]. The overview of the resulting methodology given in [7] is the basis for this description.

RM-ODP consists of five viewpoints:

- The Enterprise Viewpoint;
- The Information Viewpoint;
- The Computational Viewpoint;
- The Engineering Viewpoint;
- The Technology Viewpoint (not addressed by the Q.18/4 group).

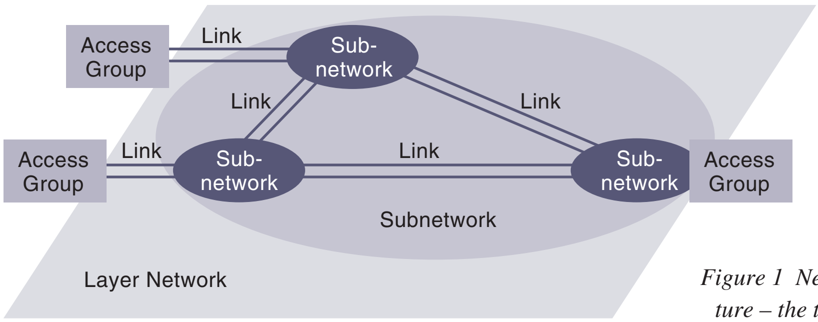


Figure 1 Network Architecture – the topological view

---

1) *of a layer network*

2) *When modeling the topology of a subnetwork rather than a layer network, one may replace the access group with the connection point group containing a number of co-located connection points.*

Each viewpoint constitutes a self-contained specification of the system from a particular perspective. In addition, certain mappings between the viewpoints need to be established for the integrity of the overall system to be maintained.

Because the target application is a model for management, only the system aspects subject to management, i.e. the management requirements, need to be represented in the model. The management requirements are expressed in terms of actions with associated policies (enforcements or restrictions). This is done in the Enterprise Viewpoint and implies that the requirements become an integrated part of the model itself. Another, equally important implication is that the Enterprise Viewpoint becomes a repository for management requirements, i.e. the management specification *per se*.

2.1 The Enterprise Viewpoint

When compared with the original version of the RM-ODP framework, most of the changes pertain to the Enterprise Viewpoint. A finer granularity in the modelling approach is allowed, namely that of the network resource types defined in G.805. All the modelling constructs in the other viewpoints are provided with unique labels for backwards traceability to the functional requirements in the Enterprise Viewpoint. This is a fundamental mechanism for the support of conformance testing of implementations and also for requirement assessment during the specification phase.

As previously mentioned, functional requirements are modelled as actions invoked by a caller role[3)] and carried out by a provider role. In this context, an object is an entity that can be distinguished from other entities, has a separate existence and is described in terms of actions, policies and relationships with these other entities. Actions are enforced or restricted by action policies, that is – Permissions or Obligations respectively – to provide additional level of detail to the management specification. The caller has no knowledge of how the action is implemented. He only knows the action signature, i.e. the action name and the action policies.

Actions naturally belonging together are grouped into functional units called Communities. Actions may also be combined in ordered series, Activities, to realize more comprehensive functionality. The provider may decide whether the full action sequence or an abstracted view in terms of a single high-level action is to be exposed to the caller. It is possible to create new communities by combining existing ones, by adding new functionality or deleting parts of the existing functionality. By taking advantage of existing actions when defining new communities, substantial reuse of specification and possibly implementation may be achieved. Policies may also be defined on the community- and the activity levels.

For usage outside the context of its initial definition, the functionality provided by a community may be expressed through the Service Contract, essentially a listing of the functional contents. When making models on the Service Management level, the service contract may be used to define the technical part of the Service Level Agreement (SLA), the Service Level Specification (SLS).

In addition to capturing the functional requirements the Enterprise Viewpoint also serves as the road map towards the other viewpoints. Actions in the Enterprise Viewpoint map to interface operations in the Computational Viewpoint. The client and provider roles map to computational objects. Enterprise actions are normally concerned with the manipulation (create, delete, associate, etc.) of G.805 network resources such as trails, access groups, link connections, etc. Network resources map to objects, attributes or relationships in the Information Viewpoint.

When passing from the less formal architectural description of network resources in G.805 to a formal network model, additional element behaviour needs to be settled. Rec. G.852.2 [8] does that for the elements in G.805 and also provides definitions for some important elements currently missing in G.805.

An extract from the Enterprise Viewpoint specification for trail management showing the community policy, role, action and service contract definitions for trail termination point creation is shown below.

COMMUNITY trail management (tm)
COMMUNITY_POLICY
OBLIGATION signalId
"Each resource in the community shall have the same signal identification".[4)]
ROLE
tm_caller
"This role reflects the client of the actions defined within this community. One and only one caller role occurrence must exist in the community."

3) In RM-ODP, a role is a fraction of the behaviour of an object, in this case an Enterprise object.
4) This is the same as requiring each resource to be part of the same layer network domain (LND).

tm_provider
"This role reflects the server of the actions defined within this community. One and only one provider role occurrence must exist in the community."
layer network domain
"This role represents the layer network domain resource defined in Recommendation G.852.2. One and only one layer network domain role occurrence may exist in the community."
trail termination point
"This role reflects the trail termination point resource defined in Recommendation G.852.2. Zero or more trail termination point role occurrences may exist in the community."

ACTION
Create trail termination point
"This action is used for the creation of a trail termination point. The caller has the ability to provide a unique user identifier to identify the trail termination point that has been created."
ACTION_POLICY
OBLIGATION inputDirectionality
"The caller shall specify the directionality of the trail termination point to be created."
PERMISSION inputUserId
"The caller may provide a user identifier for the requested trail termination point."
OBLIGATION rejectUserIdNotUnique
"If PERMISSION inputUserId is part of the contracted service and if the user identifier is not unique in the provider context, then the provider shall reject the action."
OBLIGATION provideUserId
"If PERMISSION inputUserId is part of the contracted service, then the provider shall use the user identifier as the unique identifier when communicating with the caller."
OBLIGATION successReturnId
"If PERMISSION inputUserId is not part of the contracted service, the provider shall, upon success of this action, return the unique identifier for the created trail termination point."
PERMISSION inputUserLabel
"The caller may provide a user label for the requested trail termination point."

SERVICE CONTRACT tm_src
COMMUNITY_POLICY
OBLIGATION signalId;
ROLE
tm_caller, tm_provider, layer network domain, trail termination point ;
ACTION
Create trail termination point {OBLIGATION inputDirectionality, rejectUserIdNotUnique, provideUserId, successReturnId; PERMISSION inputUserId, inputUserLabel};

2.2 The Information Viewpoint

The Information Viewpoint defines the static behaviour of the system in terms of information elements, i.e. objects, attributes and relationships. The behaviour of the information objects is made up of invariants, attributes and mandatory relationships. It may either be provided as structured English text or be described using the formal language Z [9].

Rec. G.853.1 [10] is a library constituting the information elements that may be directly defined on the basis of G.805 and G.852.2, that is, without taking functional requirements pertaining to any specific management area into account. When producing an Information Viewpoint specification for a specific functional area, the elements from the library are used as superclasses and specialisations are provided taking the additional requirements of the functional area in question into account. New elements may be added as well.

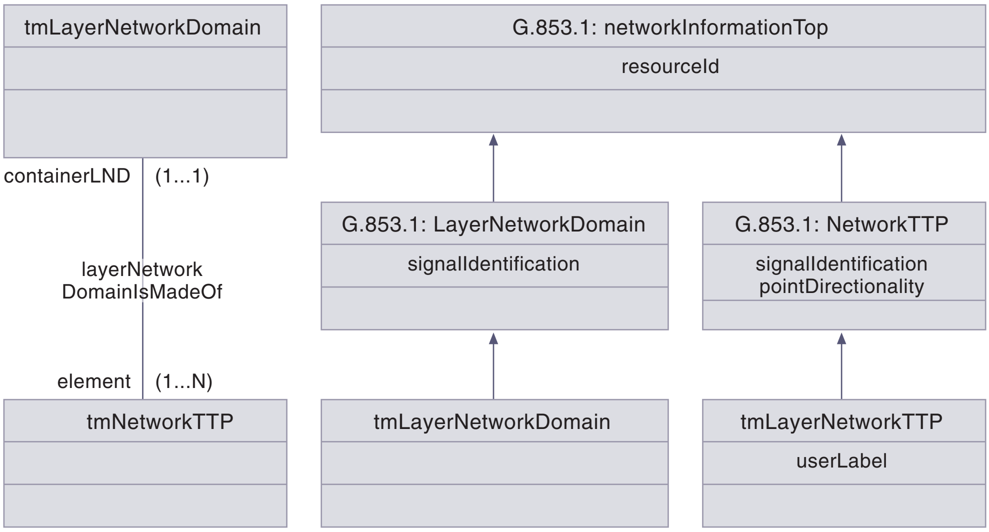
Information elements are tagged back to the corresponding Enterprise Viewpoint elements, alternatively to elements in G.852.2.

The Information Viewpoint is the ultimate source for the definition of the information elements within the system. This is reflected by the Parameter Matching clause in the Computational Viewpoint which maps the input- and output parameters to the corresponding information objects, attributes and relationships.

The UML class diagrams, the information object-, attribute- and relationship definitions relevant to networkTTP creation are shown below.

tmNetworkTTP
<COMMUNITY: trail management,
ROLE: trail termination point>
DEFINITION
"This object class is derived from
<G.853.1: networkTTP>."
ATTRIBUTE
<userLabel>
"<COMMUNITY: trail management,
ACTION: create trail termination point,
ACTION_POLICY: inputUserLabel>
This attribute is imported from G.853.1 and is used as a user friendly label for the networkTTP."
RELATIONSHIP
<layerNetworkDomainIsMadeOf>

Figure 2 UML class diagrams for trail termination point



tmLayerNetworkDomain
<COMMUNITY: trail management,
COMMUNITY_POLICY: signalId>
DEFINITION
“This object class is derived from <layerNetworkDomain>.”
ATTRIBUTE
-- *none additional*
RELATIONSHIP
<layerNetworkDomainIsMadeOf>

layerNetworkDomainIsMadeOf [5)]
DEFINITION
“The layerNetworkDomainIsMadeOf relationship class describes the relationship that exists between a layerNetworkDomain and the objects that compose it.”
ROLE
containerLND
“Played by an instance of the <layerNetworkDomain> information object type or subtype.”
element
“Played by an instance of the subtype of the <networkInformationTop> information object type.”
INVARIANT
inv_containerLNDRoleCardinality
“One and only one instance of the role *containerLND* must participate in the relationship.”
inv_elementLNDRoleCardinality
“One or more instances of the role *element* must participate in the relationship.”
inv_signalIdentification
“The *containerLND* and the *element* must contain the same signalIdentification information.”

2.3 The Computational Viewpoint

In the Computational Viewpoint, the dynamic system behaviour is described as interactions between computational objects constituting operational – as well as notification interfaces. The computational objects represent the finest granularity possible for computational objects in the mapping to objects in the Engineering Viewpoint.

For each computational object, mappings are provided to the appropriate caller and provider roles in the Enterprise Viewpoint.

The major part of the Computational Viewpoint deals with the specification of the operations belonging to each interface. The methodology used is commonly known as “Design by contract” [11]. Each operation request carries with it a number of input parameters and upon the successful execution, a number of output parameters are returned. Each parameter has a name, a type specifier and a value assigned. Every parameter is mapped to the corresponding Information Viewpoint element in the Parameter Matching clause.

The invariant state of the system before and after the execution of each operation is specified by defining the state of the system components, i.e. the relationships and attributes supported. This is done in the form of a set of explicit pre- and post conditions. When violating any of them, specific exceptions are raised. For each exception, an explanatory text as well as a type specifier is provided.

5) *This relationship defined in G.851.1 is included for the convenience of the reader.*

For every successful operation, a notification is generated to inform external recipients of the outcome of the event. The notification may contain additional information as required for the recipients to take advantage of the result.

The operations are described in a communication protocol neutral fashion. Protocol specific constructs for the actual communication protocol implemented are added in the Engineering Viewpoint. The parameters are defined using the ASN.1 description language.

An extract from the Computational Viewpoint specification for trail management describing the create networkTTP operation and the associated ASN.1 types, is shown below. The corresponding "Report networkTTP creation" notification is not shown.

Create networkTTP

<COMMUNITY: trail management,
ACTION: create trail termination point>
OPERATION createNetworkTTP {
  INPUT_PARAMETERS
    layerND: LayerNetwork-DomainChoice;
    pointDir: PointDirectionality;
    suppliedUserIdentifier: UserIdentifier;
    -- zero length string or 0 implies none supplied.
    suppliedUserLabel: GraphicString;
    -- zero length implies none supplied.

  OUTPUT_PARAMETERS
    networkTTP: NetworkTTPChoice;

  RAISED_EXCEPTIONS
    userIdentifierNotUnique: UserIdentifier;
    failureToCreateNetworkTTP: NULL;
    failureToSetUserIdentifier: NULL;

BEHAVIOUR
SEMI_FORMAL
  PARAMETER_MATCHING
    layerND: <INFORMATION OBJECT:
        tmLayerNetworkDomain>;
    suppliedUserIdentifier:
    <INFORMATION ATTRIBUTE:
        resourceId>;
    pointDir:
    <INFORMATION ATTRIBUTE:
        pointDirectionality>;
    networkTTP:
    <INFORMATION OBJECT: tmNetworkTTP>;
    suppliedUserLabel:
    <INFORMATION ATTRIBUTE: userLabel>;

  PRE_CONDITIONS
    inv_uniqueUserIdentifier
      "**suppliedUserIdentifier** shall not be equal to *resourceId* of any *element* in a *<layerNetworkDomainIsMadeOf>* relationship where **layerND** refers to *containerLND*."

  POST_CONDITIONS
    inv_existingNetworkTTP
      "**networkTTP** and **layerND** must respectively refer to *element* and *containerLND* in a *<layerNetworkDomainIsMadeOf>* relationship."
    inv_agreedUserIdentifier
      "*resourceId* of *tmNetworkTTP* referenced by **networkTTP** is equal to **suppliedUserIdentifier**, if it is supplied."

  EXCEPTIONS
    IF PRE_CONDITION inv_uniqueUserIdentifier NOT_VERIFIED RAISE_EXCEPTION
      userIdentifierNotUnique;
    IF POST_CONDITION inv_existingNetworkTTP NOT_VERIFIED RAISE_EXCEPTION
      failureToCreateNetworkTTP;
    IF POST_CONDITION inv_agreedUserIdentifier NOT_VERIFIED RAISE_EXCEPTION
      failureToSetUserIdentifier;
}

Supporting ASN.1 productions

  LayerNetworkDomainChoice ::= CHOICE {
    tmLayerNetworkDomainQueryIfce
        TmLayerNetworkDomainQueryIfce,
    userIdentifier
        UserIdentifier };

  NetworkTTPChoice::= CHOICE {
    tmNetworkTTPQueryIfce
        TmNetworkTTPQueryIfce,
    userIdentifier
        UserIdentifier};
-- *A query interface is an interface on the implemented object where properties such as the identifier of the object may be retrieved.*

2.4 The Engineering Viewpoint

The Engineering Viewpoint describes operations for specific interfaces based on a given communication protocol. Specifications exist already for CMIP [12, 13] and Corba IDL [14, 15, 16], and others will be provided. With CMIP, Managed Object classes representing network resources map to the network resources in the Enterprise Viewpoint. Actions map to operations in the Computational Viewpoint and name bindings and attributes map to the corresponding elements in the Information Viewpoint. The mapping scheme for Corba IDL is not finally decided yet.

Another important feature is that of distribution. Within RM-ODP, there are mechanisms included to support a number of distribution types by implementing the corresponding transparencies.

The functionality describing the network TTP creation for a CMIP interface is spread across a number of constructs, e.g. the MO class definition, the conditional packages, the name bindings and the error parameters. It is not readily separable from other functional elements and is not presented here for that reason.

2.5 The Technology Viewpoint

The Technology Viewpoint is concerned with implementation issues only and will not be discussed here.

3 Existing Functionality

Following the approval of the enhanced RM-ODP framework in 1996, a range of recommendations applicable to technologies such as SDH, WDM and, to a certain extent ATM, were approved in 1999. The functional areas covered were:

- Topology management (creation and deletion of topological resources such as subnetworks, links and access groups) [17, 18, 19].
- Pre-provisioned adaptation management (interlayer adaptation including the multiple client layer case) [20, 21, 22];
- Pre-provisioned link connection management (resource reservation) [23, 24, 25];
- Pre-provisioned link management (management of links with and without server layer support) [26, 27, 28];
- Trail management (set-up, modification and release of trails, creation and deletion of trail termination points, a number of associations between elements) [29, 30, 31].

Work is in progress to include partitioning, protection, routing, and failure propagation and also completing the model for ATM.

When assessing the applicability of a technology for the model, the specific requirements of the technology are analyzed. In some cases the non-matching requirements may be modeled on the NE level, while others may lead to technology specific extensions on the network level. For example, the assessment of ATM resulted in two major changes, the redefinition of link capacity to be either bandwidth or the number of link connections, and secondly, the notion of dynamic creation of termination points during connection setup.

The applicability of the model to WDM technology was demonstrated in the ACTS project Mephisto and shown at the public presentation in October 1999 [32]. The setup of a protected trail in the OCH layer on a prototype system consisting of three OADMs in a ring structure was shown. The network model used in Mephisto is based on the generic network model. The Configuration Manager developed provides the operator with one operation for each action in the communities supported, that is, topology management, pre-provisioned link connection management, subnetwork connection management and trail management.

It is argued in the Mephisto project [33] that the generic model may be utilized for the management of SDH over WDM, provided that an enhanced version of the route discovery function is used in the OCH layer. An alternative solution to limit noise accumulation is the introduction of a digital frame like the digital wrapper on top of the OCH layer combined with FEC inside each network operator domain involved.

4 Functionality under Development; Connectionless Communication

So far, only technologies for the core network have been considered. There are, however, increasing concerns to provide an end-to-end network view including access network technologies as well. A combined network and network element model for ATM PONs is under approval in ITU-T [34].

The current versions of G.805 and the technology specific extensions all presume connection – oriented communication, that is, prior to traffic flowing, a connection has to exist. With connectionless communication there is no connection setup in advance. To cope with that, a novel network architecture with the working title “g.cls” [35] is currently being defined in SG.13. The scope is not limited to connectionless technologies, so there is a potential for g.cls to become the replacement of G.805 rather than a complement. Developing a network model for connectionless technologies is part of the work plan for Q.18/4.

If IP is to become the major network technology of the future, enhancements in the routers as well as in the hosts are necessary. A comprehensive network model is a powerful tool for analyzing as well as managing such an enhanced IP based network.

One of the major enhancements necessary relates to the communication paradigm. It is widely agreed that the current “Best Effort” approach should not be the only level supported. Traffic Engineering is a toolbox that is being proposed to provide differentiation to the communication paradigm and thus to the QoS[6)], utilizing mechanisms such as RSVP [36], IntServ [37], DiffServ [38], MPLS [39], and COPS [40]. In addition, a control framework is necessary to verify compliance between planned and actual traffic and also for specifying how to deal with excess traffic. At the same time, one is trying to preserve the structural simplicity of the existing Internet.

Differentiated service quality should lead to differentiated charging. Consequently, new charging models must be developed to complement the existing flat rate charging scheme.

The second argument for an extension of the model is the interworking between IP and other technologies. Primarily, client-server schemes with IP in the client role are foreseen, but peer-to-peer configurations are expected too. A variety of interworking schemes need to be considered, ranging from gateways supporting one specific management function to the full interworking between IP and other technologies.

Another concern is that of handling the vast forecast increase in traffic as discussed in papers by Anderson [41] and others . This traffic will be aggregated and handled by large capacity routers. Unfortunately, data traffic on the one hand and voice and video traffic on the other pose quite different requirements to the network in terms of packet loss, delay and delay variation. This makes the traffic grooming function more complex. There are also issues relating to the work split between routing on the client level (IP) and the server level (OTN/WDM).

5 Conclusion

The development of the generic network level model in Q.18/4 is based on two major foundations:

- The Generic Network Architecture defined in G.805;
- The Enhanced Reference Model for Open Distributed Processing (RM-ODP) defined in G.851.1.

The modeling methodology is described in Chapter 2. The model for the creation of a trail termination point is provided as an example to illustrate the usage of the various modeling constructs available throughout the three viewpoints constituting the protocol neutral part of the model.

The functionality provided by the existing model is described in Chapter 3 and the future extensions, the access network and connection-less communication, i.e. IP, are discussed in Chapter 4.

To the knowledge of the author, the main application of the model so far has been to serve as the basis for the development of generic as well as technology specific models in fora such as the Tele Management Forum (TMF), the ETSI TMN, the SONET Interoperability Forum (SIF) and the ATM Forum. It is expected, though, that the inclusion of technologies for the access network and connectionless communication will open up for wider application, in particular when end-to-end issues, i.e. QoS, in combination with requirements for an abstract, high level network view are important.

6 References

1 ITU-T. *Generic functional architecture of transport networks*. Geneva, 11/95. (ITU-T rec. G.805.)

2 ITU-T. *Management of the transport network- application of the RM-ODP framework*. Geneva, 03/99. (ITU-T rec. G.851.1.)

3 ITU-T. *Basic reference model for Open Distributed Processing – Part 1: Overview*. (ITU-T rec. X.901.)

4 ITU-T. *Basic reference model for Open Distributed Processing – Part 2: Foundations*. (ITU-T rec. X.902.)

5 ITU-T. *Basic reference model for Open Distributed Processing – Part 3: Architecture*. (ITU-T rec. X.903.)

6 ITU-T. *Basic reference model for Open Distributed Processing – Part 4: Architectural Semantics*. (ITU-T rec. X.904.)

7 Henriksen, T. The generic network model – an ITU approach for interoperability. In: *Proceedings of the 5th IFIP TC6 International Symposium, Interworking 2000,* Bergen, Norway, October 2000. Rao, S et al (eds). Berlin, Springer, 2000.

8 IUT-T. *Management of transport network – Enterprise viewpoint description of transport network resource model*. Geneva, 03/99. (ITU-T rec. G.852.2.)

---

6) *QoS should be interpreted broadly including every kind of parameter affecting the perception of the end user of the service provided.*

9 Potter, B et al. *An introduction to formal specification and Z.* New York, Prentice Hall, 1992.

10 ITU-T. *Management of transport network-Common elements of the information viewpoint for the management of a transport network*. Geneva, 03/99. (ITU-T rec. G.853.1.)

11 Meyer, B. *Object oriented software construction*. Upper Saddle River, NJ, Prentice Hall, 1997.

12 ITU-T. *Generic network information model Amendment 1*. Geneva, 03/99. (ITU-T rec. M.3100 amd.1.)

13 ITU-T. *Management of transport network – GDMO engineering viewpoint for the generic network level model*. Geneva, 03/99. (ITU-T rec. G.855.1.)

14 ITU-T. *TMN guidelines for defining CORBA Managed Objects*. London, 05/00. (ITU-T draft. rec. X.780.)

15 ITU-T. *CORBA based TMN services*. London, 05/00. (ITU-T draft. rec. Q.816.)

16 ITU-T. *CORBA generic Network and NE level information model*. London, 05/00. (ITU-T draft. rec. M.3120.)

17 ITU-T. *Management of transport network-Enterprise viewpoint for topology management*. Geneva, 03/99. (ITU-T rec. G.852.3.)

18 ITU-T. *Management of transport network-Information viewpoint for topology management*. Geneva, 03/99. (ITU-T rec. G.853.3.)

19 ITU-T. *Management of transport network – Computational viewpoint for topology management*. Geneva, 03/99. (ITU-T rec. G.854.3.)

20 ITU-T. *Management of transport network – Enterprise viewpoint for pre-provisioned adaptation management*. Geneva, 03/99. (ITU-T rec. G.852.8.)

21 ITU-T. *Management of transport network – Information viewpoint for pre-provisioned adaptation management*. Geneva, 03/99. (ITU-T rec. G.853.8.)

22 ITU-T. *Management of transport network – Computational viewpoint for pre-provisioned adaptation management*. Geneva, 03/99. (ITU-T rec. G.854.8.)

23 ITU-T. *Management of transport network – Enterprise viewpoint for pre-provisioned link connection management*. Geneva, 03/99. (ITU-T rec. G.852.10.)

24 ITU-T. *Management of transport network – Information viewpoint for pre-provisioned link connection management*. Geneva, 03/99. (ITU-T rec. G.853.10.)

25 ITU-T. *Management of transport network – Computational viewpoint for pre-provisioned link connection management*. Geneva, 03/99. (ITU-T rec. G.854.10.)

26 ITU-T. *Management of transport network – Enterprise viewpoint for pre-provisioned link management*. Geneva, 03/99. (ITU-T rec. G.852.12.)

27 ITU-T. *Management of transport network-Information viewpoint for pre-provisioned link management*. Geneva, 03/99. (ITU-T rec. G.853.12.)

28 ITU-T. *Management of transport network – Computational viewpoint for pre-provisioned link management*. Geneva, 03/99. (ITU-T rec. G.854.12.)

29 ITU-T. *Management of transport network – Enterprise viewpoint for trail management*. Geneva, 03/99. (ITU-T rec. G.852.6.)

30 ITU-T. *Management of transport network – Information viewpoint for trail management*. Geneva, 03/99. (ITU-T rec. G.853.6.)

31 ITU-T. *Management of transport network – Computational viewpoint for trail management*. Geneva, 03/99. (ITU-T rec. G.854.6.)

32 Invitation to the MEPHISTO public demonstration Marcoussis, France, 20–24 September 1999. [online]. – URL: http://www.infowin.org/ACTS/NEWS/CONTEXT_UK/ 990899fr.htm.

33 Bertelon, L et al. *OTN management interworking with SDH- Specifications*. ACTS project no. AC209 Mephisto, deliverable D20 (restricted), the Mephisto Consortium, October 1999.

34 ITU-T. *Management services, object models and implementation ensembles for ATM-PON system*. Geneva, January 2000. (ITU-T draft rec. m.xxxc.)

35 ITU-T. *G.cls functional model*. Kyoto, 03/00. (ITU-T draft. rec. g.cls.)

36 IETF. *Resource ReSerVation Protocol (RSVP) – version 1 Functional Specification*. September 1997. (RFC 2205.)

37 http://www.ietf.org/html.charters/intserv-charter.html.

38 http://www.ietf.org/html.charters/diffserv-charter.html.

39 http://www.ietf.org/html.charters/mpls-charter.html.

40 http://www.ietf.org/html.charters/policy-charter.html.

41 Anderson, J et al. Protocols and architectures for IP optical networking. *Bell Labs Technical Journal*, 4 (1), 105–124, 1999.

7 Abbreviations

ACTS	Advanced Communication Technologies and Services
ASN.1	Abstract Syntax Notation no.1
ATM	Asynchronous Transfer Mode
COPS	Common Open Policy Service
DiffServ	Differentiated Services
CTP	Connection Termination Point
CORBA	Common Object Request Broker Architecture
CMIP	Common Management Information Protocol
cls	connectionless service
FEC	Forward Error Correction
IDL	Interface Description Language
IETF	Internet Engineering Task Force
IntServ	Integrated Services
IP	Internet Protocol
ITU-T	International Telecommunication Union – Telecommunications sector
LND	Layer Network Domain
MPLS	MultiProtocol Label Switching
MO	Managed Object
NE	Network Element
OADM	Optical Add and Drop Multiplexer
OCH	Optical CHannel
OTN	Optical Transport Network
PON	Passive Optical Network
QoS	Quality of Service
Rec.	Recommendation
RM-ODP	Reference Model for Open Distributed Processing
RSVP	resource ReSerVation Protocol
SDH	Synchronous Digital Hierarchy
SG	Study Group
SLA	Service Level Agreement
SLS	Service Level Specification
TTP	Trail Termination Point
UML	Unified Modeling Language
WDM	Wavelength Division Multiplexing