telektronikk 1.04

100 YEARS ANNIVERSARY

eAccessibility

Telektronikk
Volume 100 No. 1 – 2004
ISSN 0085-7130

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

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

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

Editorial board:
Berit Svendsen, CTO Telenor
Ole P. Håkonsen, Professor
Oddvar Hesjedal, Director
Bjørn Løken, Director

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

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

Prepress and printing:
Gan Grafisk, Oslo

Circulation:
3,000

Networks on networks

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

Per H. Lehne, Editor in Chief

Contents

Guest editorial

KNUT NORDBY

Knut Nordby

Information and Communication Technologies (ICT) have primarily been developed and implemented to simplify tasks and make life easier – or that is at least what most people seem to believe. When citizens in technologically advanced societies are required to use complex information and communication technologies (PCs, text editors, e-mail, Internet, telephones, mobiles, etc.) without making concessions to the needs of various user groups, the new ICT services can often create more difficulties than solve problems for many users – and many people will become 'virtually disabled'. In technologically less advanced societies, with no requirements to handle new technologies, most people will manage and can lead meaningful lives. Even the requirement of such basic skills as reading and writing will 'disable' many people, especially in a highly literate society such as Norway. It is not that most dyslectics cannot read the words, but they cannot comprehend the meaning of what they read, and when as many as one in five, to a greater or lesser extent, are dyslectic, this will obviously cause problems in performing tasks and making decisions based on text information.

As new ICT services are introduced and it is required that all citizens must use them, many people will be affected in various ways by not being able to cope with them. It is a bit like building a skyscraper office building without putting in any lifts, stairs or escalators, and then expect all workers to be able to access their offices, irrespective of on what floor it is. Only a few select mountain climbers will possess the technical skills and physical strength necessary to scale the façade and enter their upper floor offices – or be wealthy enough to afford a helicopter. For the vast majority of workers, however, anything beyond the ground floor will be more or less inaccessible, thus rendering them 'virtually disabled', even though they may not have any physical impairment. Now, if we put in stairs, many people will be able to walk up to their offices, but there is a limit for most people on how many flights of stairs they will be able to walk up on a regular basis; anything beyond 10–15 storeys will severely limit the number of people who will be able to get to their offices. If we also put in lifts most people will be able to enter their offices, and if we make the lifts large enough to accommodate wheel chairs or even beds, nearly everyone will be able to access their offices. However, even one single step will effectively bar people in wheel chairs, using crutches or walking frames; so gently graded ramps will also be necessary.

In this rather obvious, but highly hypothetical, metaphor it is easy to spot the problems and suggest solutions, but most ICT products and services are much more complex, and it is not always easy to see what the obstacle is and how to put it right. *'Design for All'* or e*design* thus comprises the task of 'putting in the stairs and lifts', metaphorically speaking, in ICT products and services to make them fully accessible to all users. Very often, a product or service that was created especially for a particular group of disabled users also shows itself to be of huge benefit to non-disabled users. To take one example: the 'talking book', which was initially introduced as a special service for blind people, has now become a mainstream product because of its obvious useful qualities. There are hundreds of other examples.

To ensure that good accessibility designs are actually brought to use, four major approaches are available: *Standardisation, procurement, regulation* and *legislation*. We shall not deal with procurement, regulation and legislation here. However, many of the articles presented in this issue of *Telektronikk* are based on standardisation work carried out by ETSI (European Telecommunications Standards Institute) under the auspices of its Technical Committee Human Factors (TC HF) and funded by the Commission of the European Union and EFTA (European Free Trade Association) over the *e*Europe Action Plan; see e*Accessibility for all: The role of standardisation in shaping the end-users' tel*-e*Europe*.

Sometimes a small, simple and inexpensive addition will make a product or service accessible to many more users under various conditions: e.g. the addition of a tactile marker (raised dot) on the '5'-key on a dialling key-pad to assist blind people will also make it more useful to fully sighted people under adverse visual conditions, such as dialling in darkness. Many more examples of inclusive design are given by Walter J. Mellors in the article *Design for All*.

Often it is possible for older or disabled people to use ordinary ICT products, terminals or services by connecting some form of *supplementary* or *assistive technology*; e.g. a special Braille keyboard or Braille reader for blind users, an inductive coupling to hearing aids for hard of hearing users or a text-to-speech

 ISSN 0085-7130

device or a supplementary large display for visually impaired users. This is further elaborated by Walter J. Mellors in the article *Requirements for Assistive Technology devices in ICT*.

We also take a look at accessibility from the other side, i.e. from the perspective of a disabled user. In his informative article *Information under your fingertips* the blind computer engineer Morten Tollefsen gives us a glimpse into the world of visually impaired people having to cope with ICT-services and equipment designed for sighted people.

In two papers; *Usability challenges in user interface standards development: Expanding the character standard for the 12-key telephone keypad* and *Generic user interface elements for mobile terminals and services*, Bruno von Niman tells about the standardisation work that has been carried out to make mobile telephones simpler to use and more accessible to all users, and especially to older and disabled people.

3G (3rd generation mobile) or UMTS (Universal Mobile Telecommunications System) has been much hyped the last years. In a very frank article, *The 3G saga, so far: Evolution, promises, challenges and its end user reality*, Bruno von Niman gives us both an informative historic overview of the development of mobile telephony and an outline of the near future prospects of 3G from a usability point of view.

Frequently, a new ICT service, which initially was developed for general use, turns out to be especially valuable to older or disabled users. One example is the use by deaf people of SMS over GSM (Global System for Mobile) phones. SMS (Short Message Service) has been a true blessing to all post-lingual deaf people who can communicate by text (most pre-lingual deaf people now use sign language). However, when they have to re-charge their pre-paid SIM (Subscriber Identification Module) cards, they must interact with an Interactive Voice Response (IVR) telephone system, which of course is impossible for deaf people. Bruno von Niman describes a standardised spoken command vocabulary for all ICT devices and services in his article *Development of an ETSI standard spoken command vocabulary for ICT devices and services*. This may provide a solution for those post-lingual deaf people who have retained their speech.

The Universal Communications Identifier (UCI) is intended to make telecommunications simpler to use for all users by replacing the ubiquitous telephone numbers and multiple address formats with the *names* that people, companies and organisations already possess. When introducing such a far-reaching new system, it is imperative that it is designed to be accessible both by disabled and non-disabled users. This is described by Mike Pluke in his article *ETSI's Universal Communications Identifier (UCI) – from its origins to its diverse benefits*. UCI also has significant security implications; e.g. to avoid spam and viruses and to protect vulnerable users such as children, which brings us to the other main theme of this edition of *Telektronikk* – e*security*, which is just as important as e*accessibility*.

As ICT-systems become more and more widespread and complex an increasing number of users are put in a vulnerable position. PIN-codes (Personal Identification Number) and passwords have become the latest curse and a source of much agony and misery for many people. This "conspiracy against human memory" (Donald Norman, 1988, in *The Psychology of Everyday Things*) requires us to memorise numerous irrelevant numbers and unintelligible combinations of letters, digits and symbols. It is easy to see why many people violate basic PIN-code and password security practice by writing down their PIN-codes or passwords and by choosing passwords and PIN-codes that are easy to guess or to crack. People who will normally lock their doors and not leave their wallet unattended in a public place, may thus inadvertently leave their computers and bank accounts wide open to trespassers and thieves.

In an evocative article, *Designing for all* e*World inhabitants using risk analysis as a design tool*, Erik D. Wisløff gives us an intriguing view of design for all and security issues with an economic twist – are we in our short-sighted thoughtlessness designing only for a minute minority market while throwing over board the majority mass market?

In an in-depth treatment of ICT security, Jan Audestad in his article *Information security and human frailty*, even evokes reminiscences of NASA's lax safety thinking, as unveiled by Nobel laureate physicist Richard Feynman in the aftermath of the Challenger disaster in 1986.

In his article; *On the mobile, its security issues and applicability potentials* Tor Hjalmar Johannesen gives us some interesting and novel ideas on the security, the technology and the potentials of the mobile phone.

The *commercial success* of a new product or service will depend more and more on *ease of use* and *accessibility*. Take as an example the exceptional success of the GSM mobile phone and the ease of sending SMS-messages: GSM-phones are now in effect available to everyone in Norway and are so simple to use that virtually any child can handle them. In contrast to this we have the unhappy failure of the UPT (Universal Personal Telephone). UPT was a brilliant concept, but was implemented in such a terribly user-unfriendly way that no credits can be awarded to those who instantly predicted its demise: you do not need a university degree in Human Factors to realize that entering long strings of digits each time you want to interact with the UPT-system was not a particularly well chosen user interface. There are documented cases of companies spending tens of millions of dollars advertising a new product or service only to find that it failed in the marketplace because the extra last half million needed to ensure its usability was never conceded.

Commercial success ought to depend even more on *security* and *safety of information*, but here the customers have been much more negligent. The Internet, originally conceived of as an open communication architecture, owned by no one, has turned out to be extremely vulnerable, with evildoers showering us with an ever-increasing flood of spam and viruses to disrupt the use of this otherwise fine system. And the ubiquitous ID-card with the even more ubiquitous four-digit PIN-code has been regarded as safe by most users, even in the face of an increasing deluge of fraud and robbery. In addition to ICT-services that are accessible by all, we also need ICT-services that are safe and can be trusted if ICT-systems are to survive as useful technologies for all.

Knut Nordby (61) holds the degree Magister Artium in psychology (equiv. to PhD). 1966 to 1985 he was Assistant Professor and Research Fellow at the Institute of Psychology, University of Oslo, doing research in vision. In 1985 he was invited to join the Research Institute of the Norwegian Telecom Administration (now Telenor R&D), where he is now Senior Research Scientist working in international standardization and developing telecom services for older and disabled people. Since 1987 he has lectured on Man-Machine Interaction at Centre of Technology at Kjeller, University of Oslo (UNIK). He is a founding member of ETSI Technical Committee Human Factors (TC HF) and was its Chairman from 1997 to 2003. He has been Rapporteur of ITU-T Study Group 4/2 Human Factors and he is a founding member of COST 219. Knut Nordby has over 90 publications to his credit.

knut.nordby@telenor.com

ISSN 0085-7130

*e*Accessibility for All: The role of standardisation in shaping the end-users' tel-*e*Europe

KNUT NORDBY

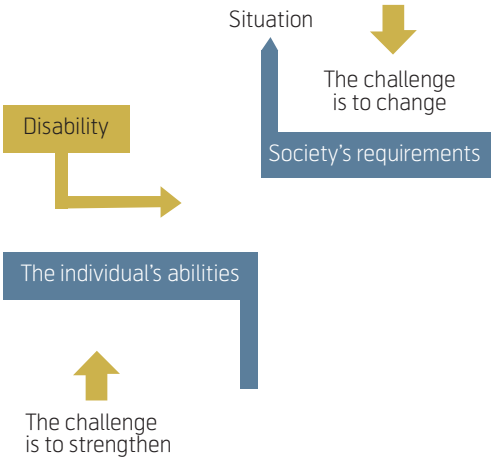
Knut Nordby

The growing need for access to communication and information services is not limited to able-bodied adults. The requirements of the elderly, the disabled and children must also be met. The potential for ICT to improve life is enormous, but there is grave concern about whether the new ICT products and services, which offer so many opportunities, are fully accessible to all people. An effective *e*Society relies on the ability of all citizens to access the new technologies. Standardisation is one way to achieve the goal to make ICT services accessible to as many users as possible. Through its Technical Committee Human Factors, ETSI has taken up this challenge and with funding from the EC and EFTA has produced a number of standards and guides to improve *e*Accessibility for all.

ICT (Information and Communication Technology) is gradually permeating nearly every aspect of life in modern society as more and more of our daily activities must now be performed electronically; *e*Mail, *e*Banking, *e*Commerce, *e*Health, *e*Education, *e*Voting, *e*Zines (e-magazines), entertainment, etc. – we are moving fast towards an esociety.

However, an esociety presupposes that all citizens actually can access the new electronic services and devices; e.g. mobile phones, PCs, PDAs, Internet, etc., but this is not always the case. Many new ICT technologies may actually exclude large user groups from participating fully in society; especially children, disabled and elderly people.

In many areas there are no *alternatives* when local bank branch offices are closed, when local shops disappear, and when essential societal services are computerised, and many older and disabled persons may therefore be 'virtually disabled', both at work and in their leisure time, if the new *e*Services are not accessible to them.



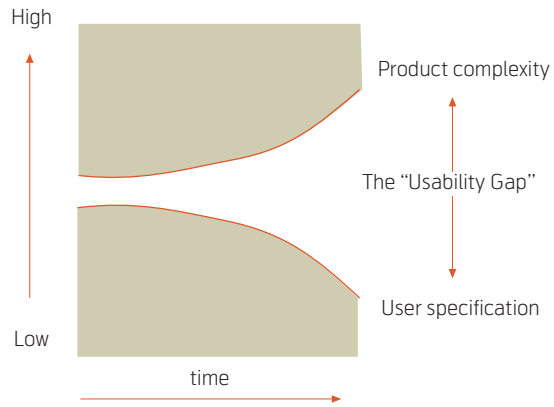
Closing the gap. The figure depicts the mismatch between the individual's abilities and society's requirements. This gap defines what we call a disability *and the challenge of 'Inclusive Design' is to close this gap*

A disability or handicap arises when the requirements imposed by society do not match the abilities of an individual. Young adults with no sensory, physical or cognitive impairments will usually be able to meet new requirements; e.g. understanding the operating procedures of new ICT services, manipulating the minute controls of small hand held ICT terminals, remembering numerous PIN codes and passwords, etc. But persons with sensory, physical or cognitive impairments, and these impairments increase with advancing age, will have difficulties or will not be

e*Accessibility for whom? Many new ICT services are not accessible to all users due to inconsiderate design*

"There is no reason anyone would want a computer in their home."
– Ken Olson, Digital Equipment Corp., 1977.

ISSN 0085-7130 © Telenor ASA 2004



The 'Usability Gap'*: As product complexity increases, user specialisation diminishes, thus widening the* 'Usability Gap'. *The objective of 'Design for All' is to narrow and eliminate this* 'Usability Gap'

'People with special needs' are the well educated, affluent, physically fit young males who require all the latest functions in their lap-tops, PDAs and mobiles, such as MP3, GPRS, WAP, GPS, camera, radio, video, DVD, games, etc. etc.

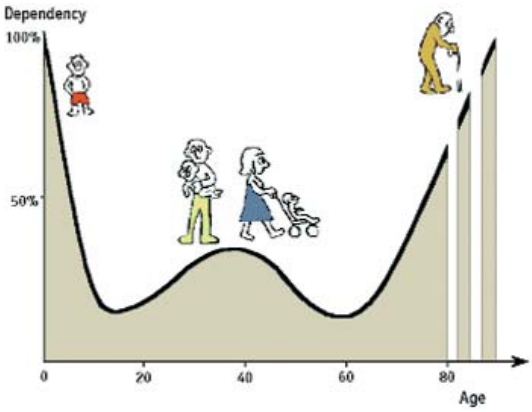
People with ordinary needs are all the others, also children, older and disabled people, who do not want special functions. They need terminals that are *simple-to-use*, *efficient* and *safe*, but industry keeps adding new functions – the 'Swiss-Army-Knife syndrome'.

able to meet society's requirements. Thus, a person who can manage quite well in a technologically less advanced society with no demands on being able to access ICT services, may actually be severely handicapped in a society where such skills are required; i.e. using ICT tools at work and for managing daily activities such as getting money from ATMs or paying bills and buying necessities over the Internet, etc.

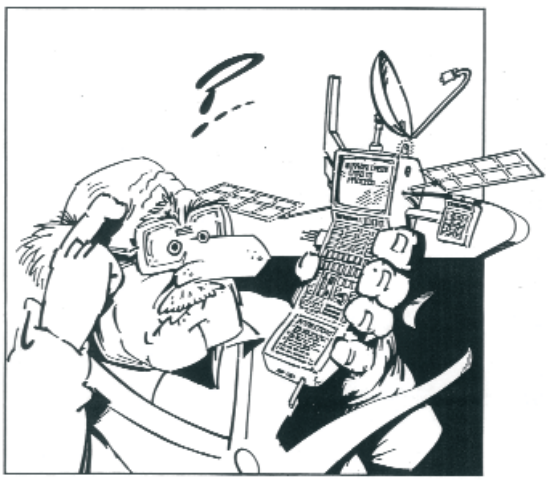
While ICT products and services are getting increasingly more complex, user specialisation, relatively speaking, is not following suit. Although most users' general ICT skills are better today than they were, say ten years ago, the complexity of new ICT products is increasing at a much faster pace, and we are witnessing a widening 'Usability Gap'.

The *e*Society may thus severely handicap many people and deprive them of basic *human rights*, such as the right to work, the right to education, the right to information and the right to communicate. It ought to be self-evident that those who create the problems should also rectify them, but it would be rather naïve to expect that industry would do so voluntarily.

Equal opportunities legislation that is now being passed in a number of countries will make it simpler to take service providers and manufacturers to court if they do not offer products that can be readily accessed. In some countries, e.g. Australia, Japan, UK and the USA, legislation has already forced the industry to attend to the needs of the special user groups. *Americans with Disabilities Act (ADA)* of



Dependency on others as a function of age (from ETSI Guide 202 048 'Guidelines on the multimodality of icons, symbols and pictograms')

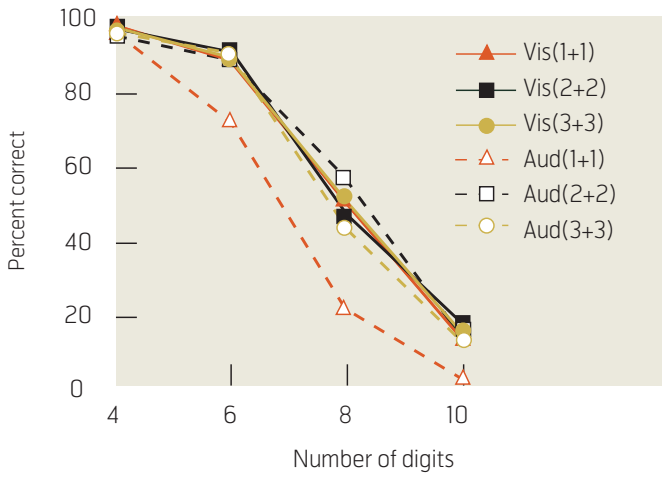


The 'Swiss-Army-Knife Syndrome'*: We need terminals and services that are simple-to-use, but industry keeps on adding unwanted functions*

ISSN 0085-7130

The 'Conspiracy Against Human Memory'*: PIN codes and passwords have become the curse of modern life, requiring people to remember non-contextual information such as random numbers or letter combinations*

1990, *Section 255* of the *Telecommunications Act* of 1996 and *Section 508* of the *Rehabilitation Act* of 1973 are three fine examples of legislation that has significantly improved accessibility to telecommunications, terminals, transport, and other services in USA. There are plans for similar Pan-European legislation. We have only seen the first few cases so far. While little binding case law exists, there have been several prominent U.S. Department of Education (DOE) rulings and out of court settlements that indicate the future course that case law will follow. The highest profile cases to date have been the cases of the American Federation for the Blind versus America Online and the Sydney Olympic Website, after the 2000 Sydney Olympic Summer Games in Australia.



Short-term memory for numbers. Percent correctly reproduced digits in their correct positions is shown as a function of number of digits, presented one-by-one (triangles), two-by-two (squares) or three-by-three (circles); either visually (solid lines) or auditorily (broken lines) (Nordby et al. 2002)

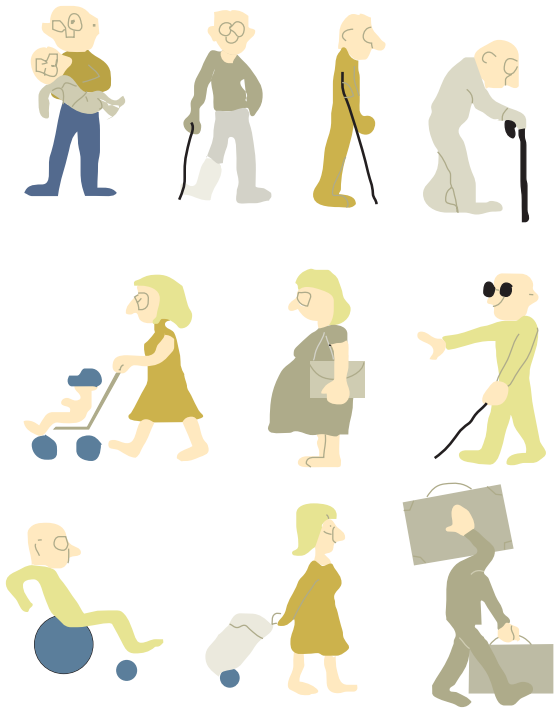
In both cases the defendants have moved to make their sites accessible. A number of organizations for senior citizens, for people with disabilities and consumer groups are preparing legal actions against companies marketing products that are not accessible to these groups.

Inclusive Design – Design for All

Human abilities vary for a great number of reasons and may be temporary or permanent. Abilities vary with age, education, health and intellectual powers. Some disabilities are permanent: blindness, deafness, paralyses, underdevelopment and many others. When we are ill or have had an accident, or under adverse conditions, e.g. in darkness, in intense noise, in very high or low temperatures, when tending to other tasks like driving in heavy traffic we can temporarily be disabled or disadvantaged.

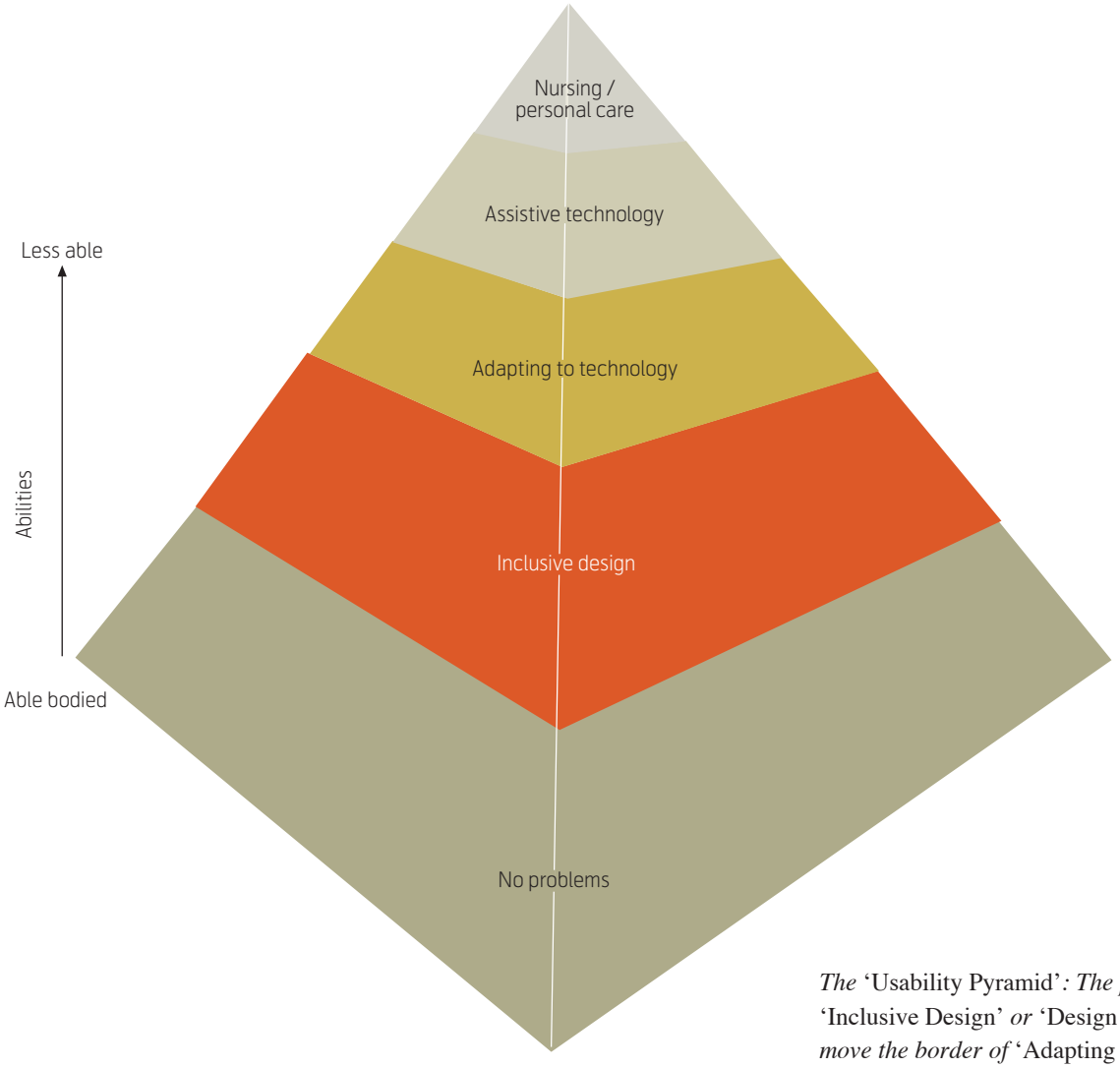
We can classify ICT-users by depicting all the users as a pyramid, where abilities vary along the vertical axis, from good abilities at the bottom to poor abilities at the top, At the bottom of the pyramid we have all the able-bodied and able-minded users who have *no problems* in using all ICT devices and services as they come.

One level up, we have those users who can manage to utilize mainstream ICT products with some form of *adaptation to technology*; such as writing down PIN codes and passwords to relieve their memories,

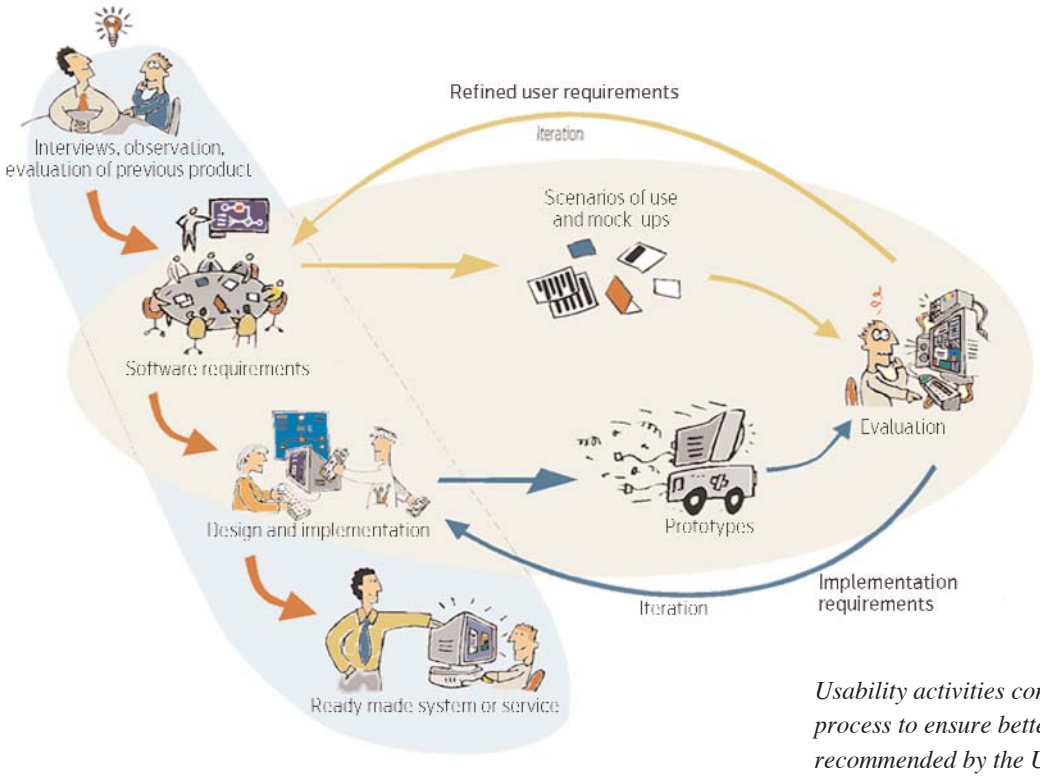


People with reduced mobility (after CRID; Consorci de Recursos i Documentació per a l'Autonomia Personal)

ISSN 0085-7130 © Telenor ASA 2004



The 'Usability Pyramid'*: The purpose of* 'Inclusive Design' *or* 'Design for All' *is to move the border of* 'Adapting to technology'



Usability activities comprising the design process to ensure better accessibility (as recommended by the USINACTS Project)

ISSN 0085-7130 © Telenor ASA 2004

ETSI headquarters, Sophia Antipolis, France

getting up close and using a magnifier to read small print, cupping their hands behind their ears to hear sound or spoken instructions, memorising the sequences of operations for getting cash from an ATM, etc.

At the next level up in the pyramid we have the users who cannot use ordinary ICT products and services without some form of *assistive technology*, such as extra large visual displays for those with low vision, tactile Braille reading devices, special supplementary keyboards for the blind, text-to-speech converters, etc.

At the top of the pyramid we find those most severely affected users who cannot use any form of ICT devices without the help of a *personal assistant*.

The purpose of *Inclusive Design* or *Design for All* is to move the border between those who have *No problems* and those who can use ICT when *Adapting to technology* as far up as possible. This can often be done with very simple means such as larger fonts and keys, better contrast, simpler operations and better-designed cognitive metaphors.

The figure on page 7 (bottom) depicts the main stages in the design process to achieve the best design possible, recommended by the USINACTS Project. One important element is the *iterative* (repeated) loops of *evaluation* and *redesign* to fine-tune the final design.

Interaction

"The wheel is an extension of the foot, the book is an extension of the eye, clothing, an extension of the skin, electric circuitry, an extension of the central nervous system."

– Marshall McLuhen and Quentin Fiore, The Medium is the Message, 1967

Standardisation

What is a standard?

A standard is an *agreement* by the industry to make or do something in a specified way. Standards are created by *consensus*.

There are two kinds of standards:

- *formal* (or *de jure*) standards are produced by mandated standards organisations
- *industrial* (or *de facto*) standards arise in the industry and are most common.

We should bear in mind that most formal standards are *voluntary*, i.e. no one is forced to use them. Until appropriate legislation has been implemented, procurement and regulations may be the most efficient instruments to enforce standards to improve accessibility for all.

There are three levels of standards:

- *National standards*. In Norway: NS (Norsk Standard) and NEK (Norsk Elektroteknisk Komité).
- *Regional standards*. In Europe: CEN (Comité Européen de Normalisation), CENELEC (Comité Européen de Normalisation Electrotechnique) and ETSI (European Telecommunications Standards Institute).
- *Global standards*. ISO (International Standards Organisation), IEC (International Electrotechnical Committee) and ITU-T (International Telecommunication Union).

Why do we need standards?

We need standards for:

- *Compatibility* of terminals and services from different suppliers,
- *Interoperability* between terminals and services from different suppliers,

- *Transfer of learning* between terminals and services from different suppliers,
- *Better accessibility* to terminals and services,
- *Increased safety* of terminals and services.

Further, we need standards for:

- *Legislation*; standards are needed by legislators to write laws to improve accessibility for all citizens,
- *Regulation*; standards are needed by regulators to specify good accessibility when granting operating licences,
- *Procurement*; standards are needed by government and civic authorities to specify good accessibility in public calls for tenders.

In this case the stupid design is easy to spot, but there is much stupid design in ICT that is not so easily spotted

ETSI's role

The *European Telecommunications Standards Institute* (ETSI) is the recognized ESO for standardisation of telecommunications and related fields of broadcasting and information technology. Established in 1988 and now a leading player globally, ETSI produces a wide range of Standards, Guides, Technical Reports and other technical documentation as Europe's contribution to world-wide standardisation. ETSI operates through the voluntary consensus of its members, taking full account of the views of all interested parties.

ETSI now has over 800 members from nearly 60 countries inside and outside Europe, and represents network operators, manufacturers, administrations, service providers, regulators, research bodies and users. The involvement of all these players ensures that the standards ETSI produces meet the needs of a rapidly developing market.

Within ETSI, the main focus of activity on accessibility is in *Technical Committee Human Factors* (TC HF). TC HF is responsible for considering the ease of use and accessibility of telecommunication equipment and services for all users, including the requirements of special user groups such as children, elderly and disabled people. It contributes to work on user interfaces, specifically for mobile communication, multimedia, text communication and user identification.

*e*Europe

In year 2000 the Commission of the European Union (EC) and EFTA (European Free Trade Association) launched a new initiative – the *e*Europe Action Plan 2002 *An Information Society For All*. To promote computer literacy and to create a partnership environment between the users and providers of Information and Communication Technologies (ICT) solutions, *e*Europe aims to secure equal access to digital systems and services for all of Europe's citizens.

ETSI is the European Standardization Organization (ESO) that is officially responsible for standardization of telecommunications, broadcasting and certain aspects of information technology (IT) within Europe. With its sister ESOs CEN and CENELEC, ETSI produces Standards and Guides for the European market. In recognition of this, the EC asked these ESOs to undertake essential standardisation work to achieve the goals of *e*Europe. As a result, Europe has seen a concerted effort over the last four years in support of 'Action Lines' crucial to the *Information Society*. One of these, which are central to the *e*Europe strategy, is *e*Accessibility.

ETSI's expertise in this area is based within its *Technical Committee* for *Human Factors* (TC HF), responsible for ease of use and accessibility of telecommunication equipment and services for all users, including children, elderly people and disabled people. It contributes to work on user interfaces,

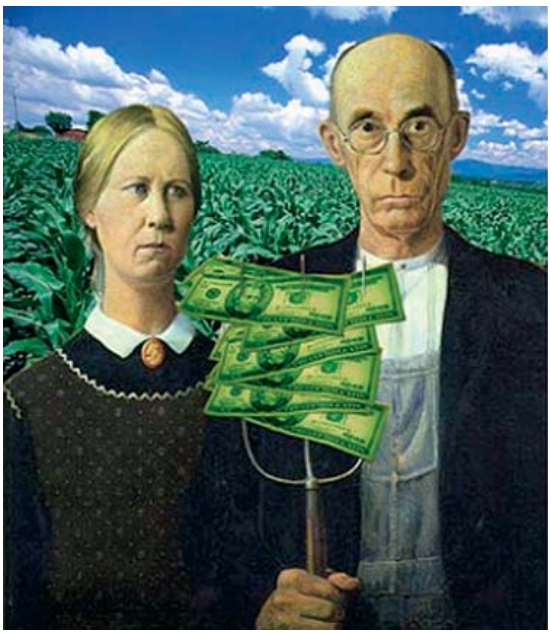
Human Factors is the application in product design of scientifically based knowledge about the *capabilities* and *limitations* of humans, with the aim of making terminals, services and environments *simpler-to-use*, *more efficient* and *safer*. Human Factors is thus a key factor in the commercial success of new ICT products and services.

"Everything that can be invented has been invented."
– Charles H. Duell, US Office of Patents, 1899

specifically for mobile communications, multimedia, text telephones and user identification. By adopting the Design for All approach and ensuring that *Assistive Technologies* are considered as part of the design process, it will be possible to improve access to the Information Society for people who might otherwise be excluded.

When the *e*Europe initiative was launched, TC HF emerged as a key player in the development of standards to ensure *e*Accessibility. EC and EFTA have provided funding so far to set up no less than seventeen teams of experts, known as Specialist Task Forces (STFs), to work under TC HF auspices. Each STF brings together leading experts from various ETSI members to work for limited periods to accelerate urgently needed work. The following will be mentioned:

- *STF 180* has produced an ETSI Guide *EG 202 067* on the basic concepts of the UCI (Universal Communications Identification) system. This proposes that every individual is given a personal identification identifier that can be used irrespective of changes in communications systems (see *ETSI's Universal Communications Identifier (UCI) – from its origins to its diverse benefits* by Mike Pluke).



Many elderly people are much better off than before and have become an important consumer group (paraphrase on 'American Gothic' by Grant Wood)

- *STF 181* has produced an ETSI Technical Report *TR 102 068* giving guidance as to how Assistive Technology devices such as special displays, special keypads and text-entry devices can be interfaced with ICT systems via wired or wireless transmission technologies to aid older and disabled people to be able to utilise mainstream ICT technology such as mobile phones, PDAs, lap-tops, etc. (see *Requirements for Assistive Technology devices in ICT* by Walter J. Mellors).

- *STF 182* has created an ETSI Standard *ES 202 076* on the generic spoken command vocabulary for basic telephone services and ICT devices. This opens up the possibility for disabled people to access electronic devices and services using speech. Products based on this standard are expected to be on the market very soon (see *Development of an ETSI standard spoken command vocabulary for ICT devices and services* by Bruno von Niman).

- *STF 183* has produced an ETSI Guide *EG 202 048* giving advice on the use of alternatives to visual icons, symbols and pictograms in multimodal interfaces to serve the needs of disabled and elderly people.

- *STF 184* has produced an ETSI Guide *EG 202 116* containing *'Design for All'* guidelines for ICT products and services aimed at the working design engineer. It sets out the characteristics of users and their disabilities and describes the human-centred design process (see *Design for All* by Walter J. Mellors).

- *STF 199* has produced an ETSI Guide *EG 203 072* on the use of the UCI (Universal Communication Identification) in Next Generation Networks (NGN). The Guide gives the various technical solutions for applying UCI to future networks.

- *STF 200* has produced an ETSI Guide *EG 202 249* on the usability aspects of UCI based systems. Further development of the UCI system will be continued by the recently created ETSI TC TISPAN (an amalgamation of TIPHON and SPAN) (see *ETSI's Universal Communications Identifier (UCI)* – from its origins to its diverse benefits by Mike Pluke).

- *STF 201* has examined access to ICT by children under 12, and has produced an ETSI Technical Report *TR 102 133* – the first time an STF has been set up to look into children's requirements. The accessibility requirements to participate in ICT by young people have not been clearly identified or catered for, since until now there has been no

account of the developmental maturation of their physical, cognitive or social abilities that can readily be applied to product design. Not taking the specific needs of children into proper account may result in inability to access services, service abuse, online vulnerability to exploitation and possibly physical harm.

- *STF 202* has produced an ETSI Standard *ES 202* 130 on the allocation and ordering of the characters of the different European alphabets on the 12-key telephone keypad. This addresses the cultural diversity in Europe where special letters are used in the different languages: To be able to write one's mother tongue should be a basic human right (see *Usability challenges in user interface standards development: Expanding the character standard for the 12-key telephone keypad* by Bruno von Niman).

- *STF 203* has produced an ETSI Technical Report *TR 102 202* on the Human Factors aspects of work in Call Centres. This Report addresses some of the problems in Call Centres where business is conducted via telephone while using a display screen. Disabled or elderly people are often employed in Call Centres and the Report gives guidance on the design of the tasks to be undertaken and for the working environment.

- *STF 204* has produced an ETSI Guide *EG 202 191* on the design of multimodal interaction, communication and navigation at the user interface of ICT systems and terminals. This should aid disabled users, as well as all other users under adverse conditions, to interact with ICT systems and devices.

- *STF 230* is currently looking into the use of UCI systems to assist young, elderly and disabled people and has already produced its first Technical Report *TR 103 073*, an ETSI Guide giving more detailed guidelines will be published in 2004 (see *ETSI's Universal Communications Identifier (UCI) – from its origins to its diverse benefits* by Mike Pluke).

- *STF 231* is currently working on the harmonization of basic Man-Machine Interaction (MMI) of user interfaces in mobile phones and services (Draft ETSI Guide *DEG 202 132*). The creation of basic interaction elements will make it easier for users to switch from one make of terminal devices or services to another, to improve overall usability of the entire interactive mobile environment and to encourage the uptake of new technologies and services (see *Generic user interface elements for mobile terminals and services* by Bruno von Niman).

Children down to the age of four are already using ICT devices and services; phoning Mom and Dad (photograph: Bruno von Niman)

In September 2003, ETSI TC HF received funding for another four STFs over the new EC/EFTA *e*Europe Action Plan 2005 programme. These four STFs started work in March 2004 and are scheduled to finish in June 2005. These new STFs deal with the following topics:

- *STF 264* to produce an ETSI Technical Report *Tele-care in Intelligent Homes; Issues and recommendations*. The STF will specifically work on identifying and building consensus around the technical standardisation activities necessary for the Human Factors of services, applications, contents, infrastructure and security for Telecare delivery into intelligent homes. The STF will provide a Technical Report identifying the key issues and standardisation actions for ETSI and its members to facilitate the growth of care service delivery into intelligent homes.

- *STF 265* to produce an ETSI Guide on *User Profile Management*. Automatic activation of user profiles is seen as a key method of relieving users of the task of manually activating different user profiles as their situations change (an approach seen as very important in the implementation of the Universal Communications Identifier, UCI). This form of support for users could be a significant factor in exploiting user profiles in a way that makes man-

"640K ought to be enough for anybody."

– Bill Gates, 1981

ISSN 0085-7130

agement of the communications easy to understand and implement. The STF will identify users' behaviour when starting or changing activities and investigate how corresponding user profiles may be automatically activated.

- *STF 266* to produce an ETSI Guide on *Access to ICT by young people; Guidelines for standards developers*. Children (12 years and younger) are rapidly becoming a significant consumer group for advanced ICT services. In some cases, children as young as four or five are using and are increasingly dependent on ICT products and services. Products often take the form of or are 'disguised' as toys, but far too often they are designed for the generic user; i.e. able-bodied adults. Children are expected to use equipment and services designed for adults and that usually have physical and cognitive ergonomics that are inappropriate for their needs. The accessibility requirements for children's participation in ICT are currently not clearly identified or catered for, partly because there is no developmental account of children's physical, cognitive or social maturation that can be readily applied to product design. Standards bodies are therefore unable to provide needed guidance for designers and developers. If the characteristics and capabilities of children are not taken into account, this may result in problems such as; physical damage from prolonged use of inappropriately designed terminals, inability to access services, service abuse or on-line vulnerability to exploitation.

- *STF 267* to produce an ETSI Guide on *Duplex Universal Speech and Text* (DUST) communication. The scope of the work will include an identification of existing text-telephone and chat systems, including call set up methods used, and the identification of user requirements for efficient text and audio conversation in the communication phase of a call. Methods of providing these requirements will be identified as will methods of migration from current text telephone systems. The proposals will identify necessary extensions of multimedia communication systems and identify any remaining unsolved problems and areas where further protocol work is required.

The Council of the European Union has designated 2003 as the *European Year of People with Disabilities*. To mark this occasion, the three European ESOs CEN, CENELEC and ETSI together organized a major conference, *Accessibility for All*, in Nice, France, 27–28 March 2003. The event focused on the role of European standardisation in accessibility to products, services and environments, covering the full range of standardization issues, from transport services, construction design and ergonomics through to intelligent homes and telecommunications. From ETSI TC HF alone there were four presentations. The variety and quantity of projects in CEN, CENELEC and ETSI, all intended to improve *e*Accessibility, were quite impressive.

The conference clearly demonstrated how standardised technical solutions can help to achieve a better social inclusion of disabled people, children and elderly people, and showed what has been achieved so far. The event recognized the needs of these forgotten millions of our society and drew together a wide-ranging list of recommendations to widen the access to the Information Society, to transport and other aspects of modern life. These recommendations include collaboration between users, industry, standardisation bodies, regulators and others. ETSI has taken up the challenge and has already committed itself to taking serious account of these recommendations in its future work.

Erkki Liikanen, European Commissioner responsible for Enterprise and Information Society, delivered the final address of the conference. He said:

> *"We should aim for all citizens to be able to use electronic communications, whether they have less digital skills, are living in remote regions, have less income, or have special physical or mental needs. Everyone should share the benefits of the Information Society in terms of access to services and of greater choice, lower prices and higher quality."*

References – ETSI publications

All publiucations are available free of charge at www.etsi.org.

ES = ETSI Standard, EG = ETSI Guide,
TR = Technical Report, SR = Special Report

Human Factors (HF); Procedure for registering a supplementary service code. ETSI ES 201 384

Human Factors (HF); User Interfaces; Generic spoken command vocabulary for ICT devices and services. ETSI ES 202 076

"I think there is a world market for maybe five computers."
– Thomas Watson, Chairman of IBM, 1943

Human Factors (HF); User interfaces; Character repertoires, ordering rules and assignment to the 12-key telephone keypad. ETSI ES 202 130

Human Factors (HF); Guidelines on the multimodality of icons, symbols and pictograms. ETSI EG 202 048

Universal Communications Identifier (UCI); System framework. ETSI EG 202 067

Universal Communications Identifier (UCI); Placing UCI in context; Review and analysis of existing identification schemes. ETSI EG 202 072

Human Factors (HF); Guidelines for ICT products and services; 'Design for All'. ETSI EG 202 116

Human Factors (HF); Multimodal interaction, communication and navigation. ETSI EG 202 191

Universal Communications Identifier (UCI); Guidelines on the usability of UCI based systems. ETSI EG 202 249

Universal Communications Identifier (UCI); Using UCI to enhance communications for disabled, young and elderly people. ETSI EG 202 301

Universal Communications Identifier (UCI); Results of a detailed study into the technical areas for identification harmonization; Recommendations on the UCI in NGN. ETSI EG 203 072

Human Factors (HF); Access to telecommunications for people with special needs; Proposals for improving and adapting telecommunications terminals and services for people with impairments. ETSI ETR 029

Human Factors (HF); Requirements for Assistive Technology devices in ICT. ETSI TR 102 068

Universal Communications Identifier (UCI); Maintaining the usability of UCI based systems. ETSI TR 102 077

Human Factors (HF); Potential harmonized UI elements for mobile terminals and services. ETSI TR 102 125

Human Factors (HF); Access to ICT by young people; Issues and recommendations. ETSI TR 102 133

Human Factors (HF); Human Factors of work in Call Centres. ETSI TR 102 202

Human Factors (HF); Guidelines on real-time person-to-person communication services. ETSI TR 102 274

Human Factors (HF); Two surveys on Assistive Technology. ETSI TR 102 279

User Group (UG); User interoperability criteria. ETSI TR 102 308

Universal Communications Identifier (UCI); Improving communication for disabled, young and elderly people. ETSI TR 103 073

Human Factors (HF); An annotated bibliography of documents dealing with Human Factors and disability. ETSI SR 001 996

Human Factors (HF); Generic user interface elements for mobile terminals and services. Draft ETSI DEG 202 132

References – other publications

Nordby, K, Raanaas, R K, Magnussen, S. The expanding telephone number I: Dialling briefly presented multi-digit numbers. *Behaviour & Information Technology*, 21 (1), 27–38, 2002.

Norman, D A. *The Psychology of Everyday Things*. New York, Basic Books, 1988. ISBN 0-465-06709-3

Raanaas, R K, Nordby, K, Magnussen, S. The expanding telephone number II: Age variations in short-term memory for multi-digit numbers. *Behaviour & Information Technology*, 21 (1), 39–45, 2002.

For a presentation of the author, turn to page 3.

Design for all

WALTER J MELLORS

Walter J. Mellors

As the proportion of elderly and disabled people in the population inexorably rises, the 'grey pound' becomes a significant force which must be taken into account by every designer. One approach to satisfying this market is through Design for All, a general approach that designers can use to ensure that their products and services address the needs of the widest possible audience, irrespective of age or ability.

This paper describes a project carried out by the European Telecommunications Standards Institute (ETSI) under the umbrella of the *e*Europe initiative which aimed to assist designers by producing guidelines for integrating the concept of Design for All into ICT products and services.

Introduction

Europe is at the start of an information revolution that is changing the way companies do business and the way in which its citizens obtain the goods and services that they need. These changes are making telecommunications and Information and Communications Technology (ICT) an essential part of the economic, educational and social life of all. As the broadband world approaches, Telecommunications and Information Technology are converging and new technology offers unprecedented opportunities for modernisation throughout society. Unfortunately, this trend is a two edged sword in that it can exclude elderly and disabled people by failing to take their needs into account. Furthermore, as the Telecommunications and the Information Technology industries converge, the products on offer become more complex and feature rich, and the need to ensure they are easy to use becomes increasingly important and challenging to the designer.

As the population of Europe is ageing it has been noted that the requirements of ICT devices have tended to exclude this growing population suffering from their age related impairments and disabilities. This challenge has been recognised in Europe by the *e*Europe initiative which aims to bring the benefits of the Information Society within the reach of all.

As government services and important public information become increasingly available on-line, ensuring access to this information for all citizens becomes as important as ensuring access to public buildings. Achieving this accessibility requires the integration of all users into the information society, i.e. the inclusion of older people, people with disabilities and also people placed in impairing environments. This will only come about as a result of designing mainstream products and services to be accessible by as broad a range of users as possible. This approach is termed "Design for All".

Within the European Telecommunications Standards Institute (ETSI) the Human Factors committee (TC HF) recognised this growing need. Human Factors is the scientific application of the knowledge about the capacities and limitations of users with the aim of making products, systems, services and environments safe, efficient and easy to use. As the technical body within ETSI responsible for human factors issues TC HF decided to update previous work and to produce comprehensive guidelines for the design for all of ICT products and services.

Funded by the *e*Europe initiative a Special Task Force (STF184) was set up to undertake the necessary research and produce a new and completely revised version of ETR 116 [1] which was a handbook covering most aspects of terminal design, providing guidance on human factors issues, good human factors design practice and standards that relate to telephones, fax machines, videotelephones and multimedia terminals.

The work of this STF was carefully co-ordinated with the work of two other STFs set up under the same *e*Europe initiative. STF 181, which was working on the Requirements for Assistive Technology in ICT [2] and STF 183, working on Guidelines on the multimodality of icons, symbols and pictograms [3].

The work

The STF 184 project team comprised seven experts drawn from both the ICT industry and the academic community. It included members with wide experience in industry and with extensive knowledge of Human Factors and usability.

ISSN 0085-7130 © Telenor ASA 2004

Work commenced by making a detailed analysis of three base documents, ETR 116 [1], ETR 029 [4], and ETR 166 [5], examining the advice given in each to see if it was equally applicable to all users and to current and predicted technologies.

The advice was amended and brought up to date where necessary and new guidelines were developed in keeping with the widened scope and the changes in technology. Completely new sections were produced on design for all, on the user population, their characteristics and disabilities and on the human centred design process.

To assist usability of the guide, the design issues were grouped into a number of sections, covering general issues such as assistive technology, multimedia and user support and into design guidelines covering input and output components of user interfaces and product and service specific items. The whole document was extensively indexed and cross-referenced.

The result of the work was a document published as EG 202 116: Human Factors; Guidelines for ICT products and services; "Design for All" [6] which is aimed at the working design engineer rather than at Human Factors experts. It contains the details of requirements needed to satisfy Design for All and gives means of assessing whether the requirements have been met. It is therefore hoped that the document will be able to be used by industry as a means of demonstrating compliance with any possible future legislation requiring Design for All.

Design for All

The philosophy of Design for All is best summarised as "The design of products, services and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialised design". This does not mean that designers are expected to design every product to be usable by every consumer as this would be an impracticable, if not an impossible target. It has to be acknowledged that there will always be some people who, because of their severe impairments, need specialist equipment or assistive technology to modify the method of making input to, or receiving output from, some piece of mainstream technology.

Adopting "Design for All" when designing ICT products and services results in a three level model:

1. Mainstream products designed according to good Human Factors practice, incorporating considerations for people with impairments, that can be used by a broad range of users;

2. Products that are adaptable to permit the connection of assistive technology devices;

3. Specially designed or tailored products for very disabled users.

The guide written by STF 184 focused on the first level of this model, and provides designers with the information they need in order to increase the range of people who can successfully use mainstream products and services. It notes that as long as "Design for All" is adopted from the start of the design process, then it is possible to design products that are accessible to a significant number of disabled and elderly people, with minimum effort and cost.

Designing more usable mainstream products based on the "Design for All" philosophy is not only of benefit to the end user – it can also offer benefits to business. Considering the needs of older and mentally disabled people can help create simple and error-friendly products, which can reach a much broader market of people who at the moment could be described as "technologically abstinent".

Designing mainstream products for users with special requirements can also often bring benefits to other users and thus increase a product's marketability. Most users may benefit from the increased usability that "Design for All" brings. For example, control of volume amplification in telephones was originally developed for people with hearing problems but has also been found useful for anyone using a telephone in a noisy environment such as a dealing room, train station or factory. Thus a telecommunications design that kept the needs of the greatest number of people in mind (including people who are hard of hearing) provided an attractive feature for all users. Furthermore, when volume amplification is built into the original design of a telephone, the cost is inconsequential.

Taking account of the needs of people with visual impairment can also help those users trying to read

ISSN 0085-7130

a display in poor lighting conditions or without their reading glasses to hand. A product designed to be easily used by those with restricted movement or strength can also help those struggling with children or luggage.

It is important to remember that there is no clear boundary between people who are categorised as "disabled" and those who are not. Performance, or ability distribution, for a given skill or ability is generally a continuous function. For example, for every person who has severe visual problems there are numerous others who wear glasses or who could benefit from a larger label on a product that is easier to see in the poor light. In addition there are many people who pass through periods of temporary disability due to some injury.

The guide also notes that in order to maintain and develop sales in the US market, all European companies need to be aware of the growing amount of legislation in this field. For example, under regulations which took effect in the US in June 2001 issued under Section 508 of the Rehabilitation Act of 1973, all technology purchased by federal agencies in the US must be accessible to disabled users, with few exceptions.

This is having the effect that large American corporations are already incorporating accessibility, or "Design for All", into products that they sell into Europe. This is because they do not wish to develop separate products for the European market and they also see the demand for accessible products expanding to state governments and schools. European companies wishing to maintain/develop their US market also need to adopt a "Design for All" approach for their products.

Any prudent company should assume that similar legislative trends are likely to follow within Europe.

Users

In Human Factors terms, "user" refers to any person who uses, maintains or is affected by the use of the system under consideration. With a computer or a cashpoint terminal there is only one user at a time, whereas in a normal telephone call, there are usually two users; the initiator of the call and the receiver of the call.

An understanding of the intended user must be at the core of the overall design process. A proper analysis of the user requirements is essential and should always be included in the initial requirements specification.

It is instructive to note that the group of users that is all too often the apparent model for equipment and service designers is likely to be male, late 20s to late 40s, an engineer or at least a university graduate, and familiar with technology and its potential benefits. In order to have achieved this status, it is likely that they will have average hearing, sight and manual dexterity, as all education systems unfortunately tend to select in these areas by default. They will, however, have a 7 % (i.e. 1 in 14) chance of being red/green colour blind.

Reliance on this model, however, can be seen on second thoughts to be quite obviously defective. It fails to recognise the differences within the business community in the face of those social changes which have increased the number of female members of this group over recent years. It fails also to recognise the movement of labour within the European Union, and the changes that occur with age. Furthermore, since age changes begin to take effect even from the mid 40s, assumptions of ability at age 30 may not apply so readily at 48.

Within the Design for All process, the designer must consider the likely user populations and their characteristics. A PC might be aimed at the whole population, whereas a game system might be aimed primarily at younger users. For the intended users it is necessary to consider what is the spread of their characteristics and what are their individual differences. Within this process, the designer should aim to ensure that all user requirements are addressed and should give positive support towards integrating the requirements of children, the elderly and other people with special needs.

There are a large number of attributes that can be used to distinguish between people in a population and this section of the guide was organised so as to align with the classification of human abilities and the consequences of impairment given in CEN/CENELEC Guide 6 [7]. The ones that should be considered to have direct impact on the successful use of ICT products and services include:

- *Sensory abilities* such as seeing, hearing, touch, taste, smell and balance;
- *Physical abilities* such as speech, dexterity, manipulation, mobility, strength and endurance;
- *Cognitive abilities* such as intellect, memory, language and literacy.

Allergies can also be a significant factor in some products.

Each of the identified abilities and disabilities was identified and described in sufficient detail to give an introduction to the understanding of the resultant handicaps.

The effect of aging will cause a change or degradation of some characteristics. In general, most functional abilities will change. For example, older people tend to lose their ability to detect higher frequency sounds (see Figure 1) and many use a hearing aid.

The incidence and severity of visual impairment increases with age and the changes in the physical structure of the eye will lead, among other effects, to loss of visual acuity (the ability to see fine detail), the inability to accommodate changes of focus from short to long distances and a loss of speed of adaptation to changing light levels. Manual dexterity, mobility, strength and endurance decline. These effects are often accompanied by a slowing of the brain's ability to process information, causing difficulty in taking in, attending to and discriminating sensory information. This has the effect of causing an overall slowing of "behaviour" and the phenomenon which is generally referred to as a "loss of memory" which is sometimes referred to as "a senior moment".

Recognising the variance in ability across a sample of population, there is clearly a point at which ability becomes so far from the expected range for the population that it has to be considered outside (above or below) the expected range. Disability, by its definition, occurs where some ability falls below the expected range. Population figures are, however, very difficult to collect because of difference between the various national views of the onset of disability and the differing methods of collecting national statistics. In some countries, for example, soldiers wounded in action are not counted as 'disabled', for reasons such as pension schemes, even though their physical or cognitive impairments make them just as disabled as those who are registered as disabled.

Even the Statistical Office of the European Communities (Eurostat) states in its disabled persons statistical data [9] that "in spite of the large number of disabled persons, there are still no reliable European-level statistics in this field". This publication gives access to the information that exists, but it is limited in the countries covered and the information provided. Clearly much of the existing and often quoted published data must have been derived by extrapolation.

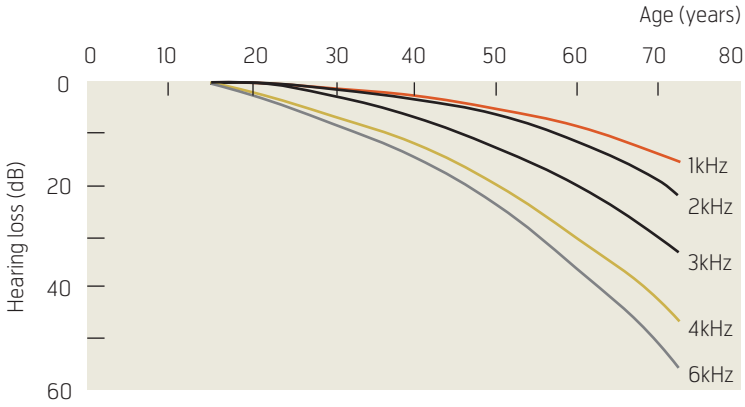


Figure 1 Hearing level as a function of age for non-noise exposed subjects (after Glorig [8])

In 2001 Eurostat published the results of a survey on disability [10] in 14 European countries (the EU 15 minus Sweden for which no data was available). This survey showed that, of the population aged from 16 to 64, 4.5 % reported that they suffered from severe disability and another 10 % suffered moderate disability. Unfortunately, this survey did not include those under 16, and more importantly, those over 64.

Figure 2, taken from the same survey, shows how the percentage of people reporting disability in any age

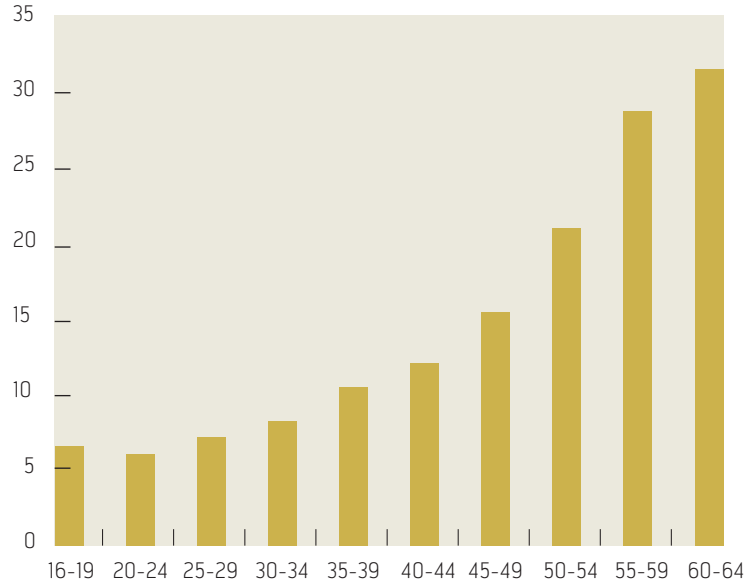


Figure 2 Age-specific percentages of persons reporting disability

ISSN 0085-7130

group increases with age. It can therefore be expected that the population aged 65 and over would report significantly increased percentage of disability. Reference [9] suggests that in the over 80 age group some 50 % to 60 % are disabled.

Thus, assuming some 20 % of the population to be over 65 [11] and taking as a minimum that an average of 33 % are disabled, this would suggest that approximately at least an additional 6 % of the population are disabled and elderly, giving a total of about 20 % of the population with moderate or severe disability.

This figure indicates the potential size of this market as a proportion of the entire market for ICT products and services.

The Human Centred Design process

There are many good reasons to adopt a Human-Centred Design approach to ICT equipment. One important reason is the legal regulation on the minimum safety and health requirements for work with display screen equipment. (EU Directive 90/270 [12]), which among its articles requires the employer "to perform an analysis of workstations in order to evaluate the safety and health conditions to which they give rise for their workers, particularly as regards the possible risks to eyesight, physical problems and problems of mental stress". It also requires that "consultation and participation of workers ... shall take place ... on the matters covered by this Directive". These are a set of very human centred requirements.

According to ISO 13407 [13], the incorporation of a human-centred approach is characterised by:

a. the active involvement of users and clear understanding of user and task requirements;

b. an appropriate allocation of functions between users and technology;

c. the iteration of design solutions;

d. multidisciplinary design.

The users and developers should interact throughout the design process. The nature of user involvement varies depending on the design activities that are being undertaken noting that the type of the product has an effect. When designing custom-made products, the actual users can be directly linked to the design process. When designing consumer products, appropriate representatives of the planned user groups should be involved in the design process.

The aim of the technology is to assist the user to carry out selected tasks. The design should identify all the tasks to be carried out and define which parts of the tasks are taken care of by technology and which parts are the user's responsibilities. Decisions cannot just be based on letting the technology do what it is capable of doing and allocating the remaining functions to the user. The human functions should form a meaningful set of tasks.

In iterative design, feedback from the users is a critical source of information. The exact user requirements cannot be defined at the beginning of a design process. On the one hand, the designers may not have a clear idea of what the users might want. On the other hand, the users may not have a clear idea what the technology could make possible. The current context of use is only the starting point of the design. The planned new system may change the context of use and then again the new context of use may change the user requirements for the technology. The design process should support this iteration by visualising the design decisions and evaluating them with the users in the planned context of use. As the result of the evaluation, both the context of use and the design may be refined.

Human-Centred Design requires a variety of skills. Depending on the nature of the system to be developed, the multi-disciplinary team may include end-users, management, application field experts, system designers, marketing experts, visual designers, human factors experts and trainers. An individual team member may represent different skill areas and viewpoints. The minimum team consists of the designer and the user.

The Guide describes means of evaluating the designs using either an analytical checklist approach or by usability testing, and details are given of both of these methods. An Annex to the Guide provides details of checklists and suggests methods of presenting the results.

Guidelines

The document then moves on to provide a set of detailed guidelines which for ease of use are logically grouped and all treated in a consistent form with definition, cross references, recommendations and sometimes comments. It commences with general design issues such as adaptability, colour, consistency, error management, feedback, flexibility and response times and then covers dialogue styles, assistive technology, multimedia presentation, labels, national variations, security and user support.

The next section gives specific guidelines dealing with input components starting with tactile inputs such as keyboards, pointing devices, switches, variable controls and software controls. Guidance on acoustic inputs covers microphones and speech recognition and visual input deals with cameras, head and eye movement and scanners. Iris recognition and fingerprints are the only biometric inputs dealt with as they are considered the most advanced. The section covering electronic input deals with card readers, machine readable cards, contactless cards and Bar code readers.

The section on output components treats visual outputs such as visual displays of various types, their characteristics, and quality requirements, and also visual indicators such as simple optical signals and icons. Acoustic outputs deal with non-speech audio such as tones, earcons, ring signals and music. Speech output and auditory menus also come in this section. Final clauses cover tactile outputs such as markers and Braille, vibrotactile indication and force feedback together with a short treatment of printed output.

The final parts of the guide give additional product specific guidelines for items such as cords, casework, connectors, handsets, portable eqipment and videophones. Service specific guidelines cover such matters as addresses, call handling, transmission, dialling, phone based interfaces, supplementary services and voice transmission.

Wherever possible, the design recommendations give objective data specifying the requirements of given features. The whole guide is fully indexed to assist the user to find his way around what is a fairly large document of some 200 pages.

Acknowledgements

The author would like to acknowledge the contribution to the work made by his colleagues in the STF, Martin Böcker, Laureano Cavero, Anne Clarke, Anita Cremers, Susan Jones and Helen Petrie.

References

1 ETSI. *Human Factors (HF) : Human factors guidelines for ISDN Terminal equipment design.* Sophia Antipolis, 1994. ETSI ETR 116 (1994-06).

2 ETSI. *Human Factors (HF) : Requirements for assistive technology devices in ICT*. Sophia Antipolis, 2002. ETSI TR 102 068.

3 ETSI. *Human Factors (HF) : Guidelines on the multimodality of icons, symbols and pictograms*. Sophia Antipolis, 2002. ETSI EG 202 048.

4 ETSI. *Human Factors (HF) : Access to telecommunications for people with special needs : Recommendations for improving and adapting telecommunication terminals and services for people with impairments*. Sophia Antipolis, 1991. ETSI ETR 029.

5 ETSI. *Human Factors (HF) : Evaluation of telephones for people with special needs : An evaluation method*. Sophia Antipolis, 1995. ETSI ETR 166.

6 ETSI. *Human Factors : Guidelines for ICT products and services : Design for All*. Sophia Antipolis, 2002. ETSI EG 202 116.

7 CEN/CENELEC. *Guidelines for standards developers to address the needs of older persons and persons with disability*. Brussels, 2002. CEN/CENELEC Guide 6.

8 Glorig, Ward, Nixon. Damage risk criteria and noise induced hearing loss. *NPL conference on Control of Noise,* 1961.

9 European Commission. *Disabled persons statistical data*, 2nd Ed. Luxembourg, Eurostat, 1995.

10 European Commission. *Disability and social participation in Europe*. Luxembourg, Eurostat, 2001.

11 Roe, P R W (ed.). *Bridging the GAP? Access to telecommunications for all people*. Brussels, The Commission of the European Communities, 2001.

12 Directive 90/270/EEC. *On the minimum safety and health requirements for work with display screen equipment*. Luxembourg, OJ L 156, 14–18, 1990.

13 ISO. *Human-centred design processes for interactive systems*. Geneva, 1999. ISO/IEC 13407.

Walter (Wally) Mellors (72) is a Chartered Electical Engineer, a Fellow of the Institute of Acoustics and a Member of the Institution of Electrical Engineers. For many years he was Head of the Telephone Laboratory in the GEC where he worked with BT on the Human Factors of Telephone systems. Since his retirement he has been involved in ETSI on Committees connected with telephony, hearing impairment, speech transmission and human factors and has been rapporteur for a large number of published ETSI documents. mellors_wmserv@compuserve.com

Requirements for assistive technology devices in ICT

WALTER J MELLORS

Walter J. Mellors

When disability makes it difficult to use the growing range of Information and Communications Technology (ICT) products and services, how does the user cope? This paper describes some work undertaken in the European Telecommunications Standards Institute (ETSI) under the umbrella of the *e*Europe initiative that attempted to describe a harmonised interface between Assistive Technology devices and ICT devices.

Recommendations were made for some preferred interfaces and protocols and a proposal made for an Interface Accessibility Initiative similar to the Web Access Initiative which could encourage the production of tools and aids for disabled users.

Introduction

In Europe today, people are living longer than in the past and so the population is ageing and the number of people with impairments and disabilities is increasing. There is a growing recognition of the need to keep these older people and people with disabilities in active touch with society to enable them to remain independent for as long as possible. ICT can play an important part in this process. Access by elderly and disabled people to mainstream technology and technology-based services has become a major issue in enabling and facilitating their integration into the new Information Society.

Under pressure of cost reduction, there is a growing tendency towards automation of many activities and, for example, many ticket machines and entry barriers are today unmanned. This can create increasing problems for disabled users, whose rights to participate in the normal activities of society should not be excluded by lack of required manual assistance. In general, disabled users should not be barred from equal access by the growing dependence on ICT in many areas of life. For example, any use of electronic voting should not prejudice a disabled person's right to vote.

Motivated by the changing market and also to some extent by the trend towards regulatory requirements, parts of the European ICT industry are trying to develop solutions for making their products usable for all users, including elderly users and those with disabilities.

This aim is exemplified by the Design for All approach, which attempts to make products usable by all people, to the greatest extent possible, without the need for specialized adaptation. In reality, Design for All is unable to satisfy the needs of some severely disabled users and must in practice remain as "Design for Most" [1]. There will always remain some people who, because of their severe impairments are unable to operate well-designed mainstream ICT products and services because there still remains a gap between their capabilities and the requirements of the user interface.

Where a competent "Design for All" solution cannot satisfy the requirements of all users, manufacturers should offer technical interfaces for the connection of so called assistive devices that fill the gap between the requirements of the user interface of the device and the abilities of the user.

Thus it is recognised that the provision of technology-based solutions to enable disabled and elderly people to lead full and independent lives requires two complementary approaches, the Design for All approach and the Assistive Technology (AT) approach. Even then there will remain a few users who cannot physically manage and who will probably require assistance from a carer to be able to cope.

The ETSI Human Factors committee decided that that elderly and disabled users would benefit from stan-

ISSN 0085-7130 © Telenor ASA 2004

dards for the assistive technology interfaces whether formal or de-facto so that one assistive device, e.g. a display for the presentation of information in an enlarged form, could be used for the widest possible range of products from different manufacturers. The manufacturers themselves would benefit by complying with possible European and other national regulations if they could offer a compatible interface even if they left the production of the assistive devices to third party manufacturers.

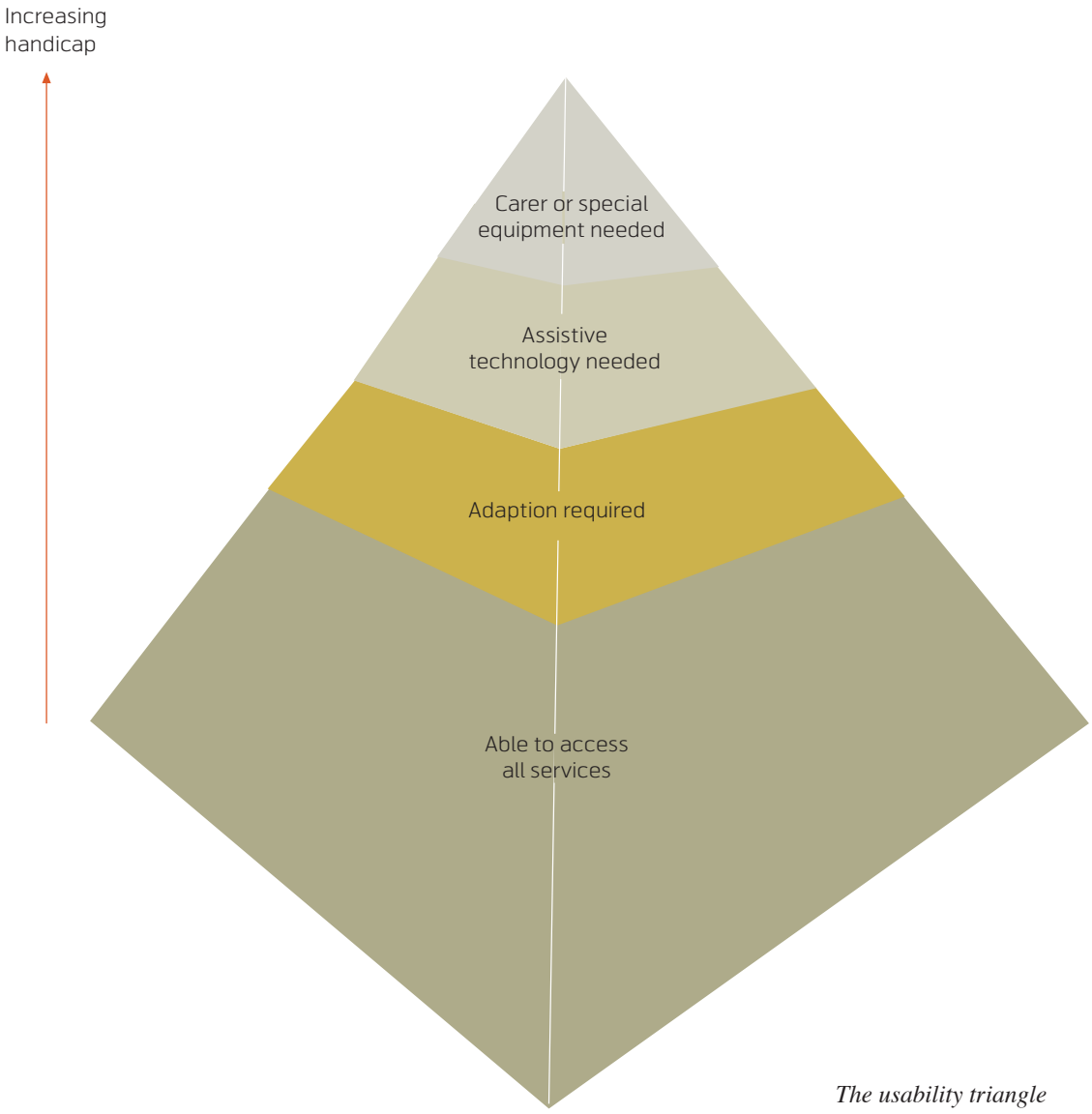
Early in the year 2001, under the *e*Europe initiative, the ETSI Human Factors committee therefore decided to set up a Special Task Force (STF 181) to try and bring about a consensus among key manufacturers in order to obtain a basis for the technical specification of these interfaces. The STF was asked to:

- Identify and document the user requirements for assistive technology solutions in ICT;
- Identify technical solutions considering different technologies (e.g. Bluetooth);
- Try to initiate and build a consensus among the key manufacturers; and
- Feed the identified solutions into other Technical bodies for technical specification.

The work of the STF was to be recorded in a technical report [3] and be carefully co-ordinated with that of STF 184 on Design for All guidelines for ICT products and services so as to avoid unnecessary duplication.

The team

The STF comprised seven experts drawn from the ICT industry and the disability community.
It included members with experience of fixed and mobile telephony, work with blind and deaf users and



The usability triangle

ISSN 0085-7130 © Telenor ASA 2004

with wheelchair users as well as general Human Factors experience. The members were drawn from industry, academia and from disability charities. There were three working members from industry and three with particular interest in disability.

There was some uncertainty at the beginning of the project as each of the groups sought to determine the other's point of view, but after some tentative discussion they rapidly got down to the initial tasks.

Following a period of preparatory research, which included a review of the existing literature, one group set out to identify and document the user requirements whilst the second tackled the problem of identifying possible technical solutions.

Background

Assistive technology is required by a user of ICT technology whenever the person's disability is such that they cannot operate the technology safely and efficiently. The increasing use of the design for all philosophy should mean that in the future, the majority of disabled people will not require assistive technology as they will be able to use mainstream technology successfully.

This situation can be illustrated by a usability pyramid where the broad base represents the bulk of users who can access most services without help, rising to those who need some sort of adaption such as enlarged font, then up to a smaller number who need assistive technology to use a service, leaving a small number at the peak of the pyramid who cannot manage without the assistance of a carer.

Design for All can move upwards the boundary between those able to access all services and those who require adaption. It can even move up the lower boundary of those who require assistive technology to access a service. Unfortunately neither Design for All nor assistive technology can assist those few at the peak of the pyramid.

Thus in those cases where "Design for All" is unable to provide an acceptable solution it will be necessary to incorporate a standard method of connecting any user's own assistive technology device which has an appropriate user interface.

If neither of these approaches is able to provide a satisfactory solution, then a carer or specially built equipment will be needed.

The ICT equipment should be designed in such a way that:

a. A person who is operating the device via assistive technology can use all of its relevant functions.

b. It can be easily and simply connected to the assistive technology device.

c. It has a standard method of interfacing with the assistive technology device and uses standard control commands.

d. It is designed to maximise the number of people who can operate it with standard assistive technology devices.

A further assumption was made that no assistive device should demand special additional power from the device with which it was working. If it required power additional to that normally at the interface, it should provide it itself.

User requirements

Definitions of user requirements can be approached from two directions. One is approached from an expert's technical view of satisfying the particular demands arising from a given disability. The other might be termed a "wish list" that is expressed by users.

From a technical point of view, a person with any disability will require assistive technology to use ICT equipment whenever they

- Cannot operate the controls (for instance because of their physical disability);
- Cannot obtain information from the device (for instance because of their sensory disability);
- Cannot understand how to operate the device (for instance if they have a cognitive disability).

There may also be occasions where the operation of the device would be possible but would cause the user pain or take too much effort. These are the physical needs that require to be satisfied.

A section of the report dealt with user handicaps, describing those resulting from sensory, physical and cognitive disabilities. This section was largely derived from the work of STF 184 as given in EG 202 116 [4]. The listing was similarly organised and in each case examples were given of those aspects that should be dealt with by Design for All and some examples were given of useful assistive technologies.

To determine what the disabled users themselves actually felt they needed, two questionnaires were

sent out. The first to a number of professionals working in the disability field and the second to a representative sample of the manufacturers and designers of assistive technology devices. The aim was to find out if there was any consensus on which of the disabled users would require assistive technology and which technology they would use it with.

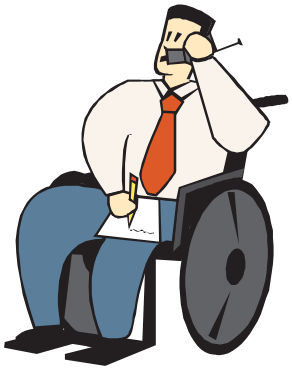
The first questionnaire was sent out to a list of 76 people and there were 35 respondents from 22 different countries. The replies showed that there were some different interpretations of the intent of the question structure, some users answering separately for each aspect of disability and some giving one answer that appeared to be intended to cover all aspects. This made it somewhat difficult to analyse the results.

The questions in the second questionnaire were open ended and so the replies again did not lead themselves to simple analysis. On the other hand they were simple and straightforward and produced valuable answers in a reasonably logical form making the information contained accessible to the user. There were 49 respondents to this questionnaire from 10 different countries. Of these, 35 provided answers to the questions in the survey.

In spite of the difficulties in close analysis of the two questionnaires it was felt that the replies contained a significant amount of useful information in the field of disability which it was worth preserving and making public for other workers in the field. They were therefore recorded in an ETSI technical report TR 102 279 “Two surveys on assistive technology” [2].

In the first questionnaire there was a range of answers to a question of who should be using assistive technology but the majority of respondents felt that Design for All was important and should cover many disabilities. Many respondents pointed up the value of speech recognition and text to speech translation facilities. Cost was one of the problems repeatedly identified by many respondents although weight and portability were felt important for assistive devices. Most responses identified particular features that were desired on various types of ICT equipment and a general need was expressed for real access to the coming world of electronic information and commerce.

The majority of manufacturers who responded wanted standard interfaces for the connection of assistive devices. There was a majority opinion that the protocols should be independent of the transmission system although some expressed the caveat that where an interconnection technology had its own



standards, they should be followed. Many said that there should be two-way communication between the ICT device and the AT device. A wish was expressed that any such standard should not stifle future innovation. Some suggested that creating awareness of a standard was an important factor.

Assistive technology

The team listed and classified the input and output requirements of the user interface of ICT devices and identified the reasons for specific needs of disabled users. This work led to characterisation of those input and output requirements of ICT devices that would

Assistive device / service	Data			Audio	Video
	Control and status	Text	Graphics		
Braille display	↔	→			
Tactual graphics display	↔	→	→		
Synthetic speech display	↔	→			
Enhanced visual display	↔	→	→		
Keyboard / pointer	↔	←	←		
Speech recognition	↔	←			
Hearing aid	↔			→	
Tactile hearing aid	↔			→	
Alarm/monitor system	↔			→	→
Smart house	↔	↔	↔	↔	↔
Navigation system	↔	↔	→	→	→

NOTE 1: → indicates information to assistive device;
← indicates information from assistive device;
↔ indicates information in both directions.
NOTE 2: Some systems may use fewer modes than the possible ones indicated.

Table 1 Examples of information exchanged with assistive device

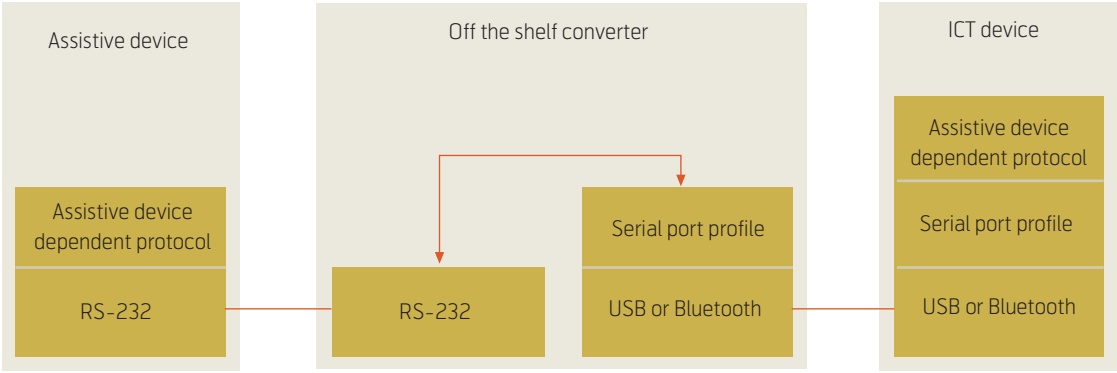


Figure 2 RS-232 interface transmission over USB or Bluetooth

need to be replaced by assistive devices at the man-machine interface.

Assistive devices were classified briefly according to ISO 9999. Although this standard covers a vast number of devices ranging from abacuses and abdominal hernia aids to zip pullers and zippers, only a few of the devices listed in that standard have the potential to be interconnected to information and communication technology (ICT) systems and they were identified as examples.

The listing notes that in some cases, the assistive device may be a mainstream device such as a cordless or mobile telephone, which can provide a really liberating tool for a wheelchair-bound user.

Technical solutions

As the work on technical solutions progressed, close liaison was kept with industry so as to ensure ready acceptance of the resultant recommendations and to bring about a consensus among key manufacturers in order to obtain a sound basis for the specification of these interfaces. Liaison between Ericsson, Siemens and Nokia representatives revealed just what was feasible within the requirements of the telecommunications industry.

At a very early stage, this liaison unfortunately revealed that it would not be possible to standardise on one single interface, whether wired or wireless. On the other hand, there was a remarkable co-operation between industrial representatives which demonstrated a willingness to provide status and device information at whichever interface was provided.

The signals crossing the assistive device interface were broadly classified into those control and status signals necessary to the ICT device and into communications signals. Control signals would include the mouse/pointer output, a translated speech command from a voice operated assistive device and number information in a communications device. Status signals from any device could indicate its readiness to operate or accept signal input. Communication signals were assumed to be the fundamental information intended to be input to and output from the device such as speech, text or video information.

Table 1 shows some examples of the type of information that might be exchanged between various ICT equipment and assistive devices.

The simplest possible interfaces should always be used. These can be used with a number of, wired or wireless, transmission technologies, supporting these basic connections.

Both the wired and wireless transmission technologies that may be used at the interface between an ICT device and an assistive device were reviewed by the STF and simple descriptions were given of their differing main features. A brief introduction to protocol stacks was provided for those new to the subject and a number of useful existing standard protocols were noted.

Command	Possible response(s)
+CASS=<sub-command>[,<parameters>]	+CASS: <response> +CASS ERROR: <error code>
+CASS=?	+CASS: (list of supported <sub-command>s)

In order to encourage harmonisation, recommendations were made of the preferred interface and protocols to be used for various types of connection.

USB is the currently preferred solution for the wired interface, HiperLAN2 or the IEEE 802.11 family for wireless local area networking and Bluetooth (IEE 802.15) for wireless personal area networking and access. It was considered that DECT is not likely to achieve wide use due to limitations in its global availability.

The RS-232 interface (Figure 2) is recommended for data. The report gives illustrations of connections using various interfaces and protocols showing how standard systems can be used. The example shows how RS-232 interface transmission can be arranged over a USB or Bluetooth connection using an off the shelf converter.

In the telecommunications industry, the AT command set has become the standard protocol for the transfer of control and status information. The AT command set is defined in ITU-T V.250 [5] and in ETSI TS 100 916 [6]. The STF has also proposed a special AT command to be used to identify any command string originating from an assistive device. It is necessary that the definition of the +CASS (arbitrarily chosen where +C stands for the generic cellular prefix, and ASS stands for assistive device) command string is coordinated within ETSI to prevent this command string being assigned for other purposes. If the +CASS command string is reserved for use by assistive devices, additional sub-commands can be added without the need for special coordination.

It is recommended that the audio interface should be similar to that used in Personal Computers but no recommendations could be made for video due to the current rapid development in this area.

The demonstrator

Arising from the work of the STF, a demonstrator model was produced of an assistive device (a Braille personal assistant) interworking with a mobile telephone. The personal assistant provided the normal personal assistant functions of a diary, telephone directory etc. but with a Braille input keyboard and a text to speech translator providing the output. When coupled to a mobile telephone, the dialling of numbers was supported as well as full text messaging facilities. This model was supported by a video in a form suitable for use over the Web that can be used to illustrate the value of harmonized interfaces.

Figure 3 The demonstrator

Conclusions

An *e*Europe community that allows everyone fair access to advanced information and communication media must include those citizens whose disabilities are such that they cannot use devices designed for all. For this group of users, often with multiple disabilities, it is crucial that affordable, effective and usable assistive devices be available.

These devices must be able to interact with a multitude of fixed and mobile ICT devices. For the development of these assistive devices, the standardization of the interfaces to ICT devices is a requirement. Standardization of these interfaces must begin at the earliest possible point.

For assistive devices to become affordable and effective, the significant players in each field need to agree on a set of protocols to be used in the communication between assistive devices and the relevant ICT devices. As the report identifies, in general it is not necessary to develop either new protocols or new hardware interfaces. The interface and protocol standards should be chosen from those already available so as to form a coherent set which covers all major aspects of information exchange between the two sets of devices.

The work to achieve this will include the upgrading of existing standards where the necessary commands do not exist, or the writing of new standards where no existing standard is relevant. Consensus on this set of interface standards must be reached in the appropriate standards fora in a process which involves manufacturers of mainstream devices, manufacturers of assistive devices and the groups representing the user with different special needs.

Recommendations

The STF noted that in the past, many worthwhile proposals in the field of disability had foundered due to the fact that they had not been actively brought to the attention of those most able to use them. They recognised that further progress in harmonisation and the wider access to assistive technology would be inhibited unless there was some continued action aimed at fostering the exchange of information and promoting the adoption of a common set of operating protocols.

It therefore proposes the European Commission should celebrate the European Year of People with Disabilities by creating a small team charged with the responsibility to set up and maintain a database of existing assistive technology interfaces and protocols, to identify areas where new work is needed and to stimulate and co-ordinate such work in the relevant European and international organisations. Organisations for people with disabilities could usefully be involved in the guidance of the work of such a team.

Such a database, working in an open way similar to the Web Access Initiative, would promote the exchange of experience of good practice both within Europe and globally. It would encourage the production of tools and aids accessible to people throughout Europe and would improve communication regarding disability on a worldwide scale.

In its report the STF made firm proposals that a team be set up charged with the responsibility to perform the following actions.

- *Create and maintain a public database of relevant interfaces and protocols;*
- *Identify the application areas where new work is needed;*
- *Stimulate and co-ordinate the development of the necessary standards and amendments;*
- *Make provision for the testing of the interoperability of new protocols and products;*
- *Identify areas where the new standards can improve the quality of life of elderly and disabled users;*
- *Disseminate the information to the interested parties;*
- *Stimulate and promote the use of the appropriate standards in ICT systems.*

Acknowledgements

The author would like to acknowledge the contribution to the work made by his colleagues in the STF, Jose Collado Vega, John Gill, Matthias Schnieder Hufschmidt, Bruno von Niman, Gill Whitney and Frits Wiarda.

References

1 Mellors, W J. Design for all guidelines for ICT products and services. In: *Proc. 18th International Symposium on Human Factors in Telecommunications*, 285–288, Bergen, 2001.

2 ETSI. *Human Factors : Two surveys on assistive technology*. Sophia Antipolis, 2004. ETSI TR 102 279.

3 ETSI. *Human Factors : Requirements for assistive technology devices in ICT*. Sophia Antipolis, 2002. ETSI TR 102 068.

4 ETSI. *Human Factors : Guidelines for ICT products and services : Design for All*. Sophia Antipolis, 2002. ETSI EG 202 116.

5 ITU. *Serial asynchronous automatic dialling and control*. Geneva, 1999. ITU-T V250 (05/99).

6 ETSI. *Digital cellular telecommunications system (Phase 2+) : AT commands set for GSM Mobile Equipment (ME)*. Sophia Antipolis, 2003. ETSI TS 100 916.

For a presentation of the author, turn to page 19.

Information under your finger tips

MORTEN TOLLEFSEN

Morten Tollefsen

I long for an information society for all, and with a bit of luck, it will not take another millennium to achieve this, or else I will have to rely on colossal medical advances to be able to obtain access to all the information I would like to have. Today I cannot have a tutorial on Microsoft® Office® in Braille, nor a usable electronic edition of the manual for my mobile phone. Thus, the potential for improvement in accessibility is vast.

Blind people can in fact use PCs and standard software. The information, however, must be presented in a totally different mode than on a visual screen. Both Braille and artificial speech are used, with a so-called 'screen reader' managing the output to these assistive devices. In principle, electronic information is excellent for those who cannot see. The problem, however, is that electronic information can also be totally useless. I will deal further with this issue below, but first I will give a brief description of the technical aids that can be used to make 'the information highway' driveable also for blind people. After all, it is safer for blind people to 'drive' on the information highway than to sit down behind the steering wheel in a real car and step on the accelerator.

The Braille display and artificial speech

A tactile display makes it possible to read screen contents in Braille. A Braille display consists of a flat unit with a row of 40 or 80 Braille character cells at the front. It is placed under the PC keyboard (Figure 1). Each Braille character cell consists of six movable small rounded pins in a two-by-three matrix. The pins can be raised in any pattern to display all letters and characters of the Braille alphabet. Although 80 Braille characters can be shown simultaneously, the fingers can only feel a couple of characters at a time. With such a small 'window' to the world, the overview of the screen contents is rather limited. Efficient software that both simplifies navigation and 'guesses' correctly what can be 'shown' on the Braille display is vital. Pure graphics cannot be shown on a Braille display, but icons, screen buttons and symbols can be given text descriptions or names and made 'selectable'.

You may have heard the joke about blind people finding interesting stories on the crust of bread with poppy seeds on top. To most people it may just look like a jumble of raised dots, but to a trained blind Braille reader it may contain interesting combina-

Figure 1 The Braille tactual display is usually placed under the normal PC keyboard.

ISSN 0085-7130

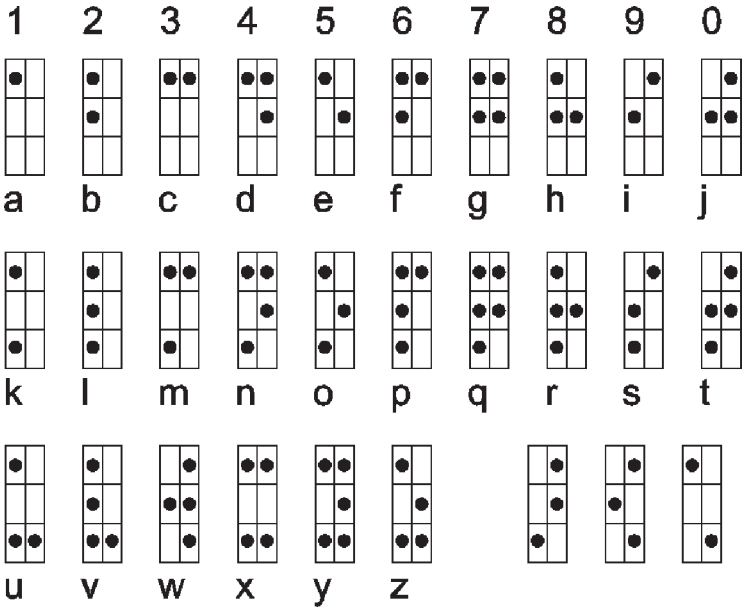


Figure 2 The Braille alphabet

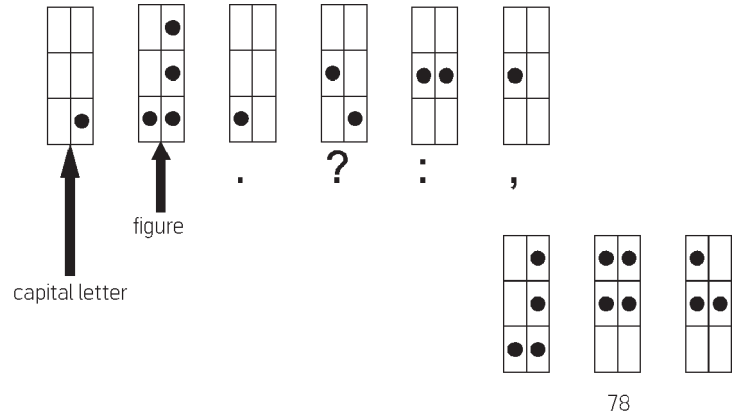


Figure 3 Special Braille characters

tions. The system to read with the fingertips is actually quite simple and straightforward as such, but it requires training – much training – to become a proficient Braille reader. Most people, however, can quite easily learn the Braille letters and characters in a couple of hours.

A simple egg-container for six eggs is often used to explain the basics of Braille. The Braille characters are made up from combinations of raised dots in a two-by-three matrix (see Figure 2). The combination of six dots gives only 64 possible patterns, but with the introduction of IT a special eight-dot Braille was introduced, giving 256 combinations. In 6-point Braille there are special prefix characters to denote numbers and capital letters (see Figure 3). Thus, the number 1, lower case a and upper case A are all represented by the same combination (i.e. one raised dot in the upper left corner of the matrix), but the special prefix character gives the actual meaning to the basic Braille character.

Braille is often combined with artificial speech. Work is in progress in developing so called 'diphon'-based speech, which combines sounds into signs, words and sentences. The sounds are recorded human speech (normally from one speaker). The Nordic languages still have some way to go before they can match the quality of the best American products, but we are hard on their heels.

Poor overview and too much graphics

The main challenge in today's technical IT-aids is that the overview is very much reduced, and that all the fun bits – graphics, pictures, and animation – have been peeled away. The one thing that information vendors do understand is that the real fun is in graphics, pictures, and animation … How boring and dull to surf the net without a screen. I, for one, would not mind being 'screened' against all the graphics.

But in a more serious vein, I must confess that I wouldn't mind being exposed to some of the creativity on the web. Good adverts are of course mercantile, but the 'very best ones' can be quite nice. Graphics and animation contribute to make adverts more exciting for most net surfers. Many blind people say that graphics are a waste, and turn them off since they slow down their surfing. Of course they slow down surfing and some users turn off the graphics, but I believe that this is a minority. Most users soon find themselves turning the graphics back on again, either to check something, or permanently.

Various web sites address different socio-demographic groups; a sombre news service or a bank service would probably require less fancy graphics than an out-and-out entertainment site. Critical studies have shown that practised newsreaders do not look at the pictures first, but go straight for the text. The various implements on a web page have different functions, a fact that information vendors should take notice of. Many of the challenges in presenting information on a PDA (Personal Digital Assistant) or on a WAP-phone (Wireless Application Protocol) are similar to those in presenting information as speech or on a one line Braille display.

The important, rational and appropriate thing is that alternatives to graphics and other inaccessible web elements are offered. This does not cost much, and it gives the user a choice and makes it possible for

blind users to find their way and get hold of the information that is out there on the web.

However, even web sites with no graphics can be quite unusable. The positioning of links, link names and the general structure of the pages are only some of the issues that separate the good web sites from the bad.

WAI

Any information provider who wants to can easily learn what is needed to make web sites more accessible to disabled users. *WAI* (Web Accessibility Initiative), which is part of the *W3C* (World Wide Web Consortium), has worked out a number of excellent accessibility guidelines (see http://www.w3c.org/WAI/).

Often only small modifications are needed: An example is when pictures are used as links or when pictures are important for comprehension if they are not described. An alternative text ('Alt'-tag) is then required. 'Regular' web surfers can display the 'Alt'-tag by positioning the mouse pointer over the picture. Some WAI guidelines depend on redundancy as a means to make the information available, and another example is to have a text version of spoken information to allow hearing impaired users access to the contents.

Why is WAI not enough?

One of my children has just learned to tell time. By observing the clock hands she is usually able to tell the correct time. Nevertheless, she has much left to learn: i.e. how long is it before the 'children's hour' on TV, is one-and-a-half hour longer than one hour, etc. Being able just to get the time right does not necessarily mean that she grasps the full concept of hours, minutes and seconds.

As worst-case, WAI can act in a similar way. The guidelines themselves are quickly learned, but if they are to be used in a sensible way, one needs to know more: How do normal technical aids work? What technical solutions are problematic? How many links is it reasonable to have on one page? Where on the page should links be placed? etc.

Designers of technical aids, both hardware and software, usually have as their explicit goal that blind users shall be able to utilize ordinary solutions. They must therefore acquaint themselves with what is actually going on to know how web sites are organized. One striking example relates to frames. Conventional wisdom maintains that frames are best avoided. However, most ordinary screen readers now seem to manage frames quite sensibly. Frame names (which normally are not visible to seeing users) are essential in allowing blind users to navigate properly. This is

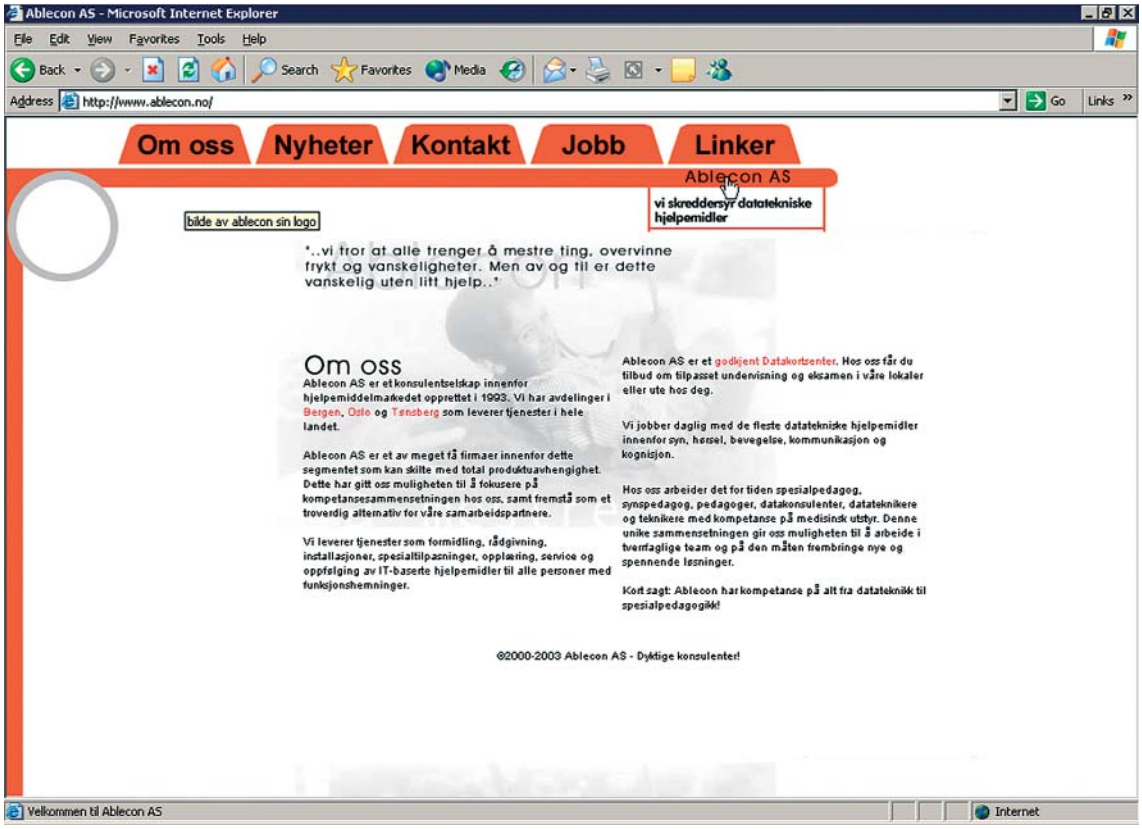


Figure 4 www.ablecon.no

ISSN 0085-7130 © Telenor ASA 2004

actually so important that appropriate standardised names should be proposed for various frames. In Jaws, users may jump directly to a frame (selected from a list). Frames with names such as '234454354544354-GGG', 'Upper left corner' or 'Frame3', provide very little assistance to blind users to jump to a desired spot, and for users who are well acquainted with a specific web site, even such badly chosen frame names may give some guidance. However, it is much better to use meaningful names such as 'Menu', 'Contents' or 'Contacts'.

One important principle in the WAI-guidelines is to describe pictures by text (this can usually be obtained via the 'Alt'-tag). In some cases this may make the site worse! The most typical example is the description of 'finery'. For blind users it is much better to be spared detailed text descriptions of these 'decorative elements' than to be interrupted by these text descriptions.

A striking example of incorrect use of WAI can be found at the web site (www.ablecon.no) of a company that teaches IT to disabled people in Norway. Here we get text descriptions such as: 'Empty picture', 'Picture of a circle', 'Part two circle', 'Orange line vertically', etc. I do not know if the page looks good, but these picture descriptions are meaningless. The page would be much better without these descriptions, and would then be displayed to me who am blind as if the graphics were not there.

Even 'good' web sites can be improved. On the RNIB (Royal National Institute for the Blind) home page we find picture descriptions such as 'RNIB Helping you live with sight loss' and 'Helping you live with sight loss'. I do not know the difference between the pictures or what they represent, but it does say something about what RNIB is. For me it would be even better if I did not have to read these text descriptions – it is obvious what RNIB is doing from reading the texts. Some of the link names are also somewhat inappropriately chosen: 'Jump to page content', 'Jump to search', 'Jump to site tools' and 'Jump to news and features'. It would be much simpler if the words 'Jump to' were deleted since that would allow using the first letter to jump straight to the links (see 'Ten suggestions for designing better web pages' below).

This being said, I must hasten to add that the examples of pictures that preferably should be explained are far more numerous than the pictures with descriptions that should be dropped. It is of course much more difficult when the pictures are used as links if the descriptions are missing.

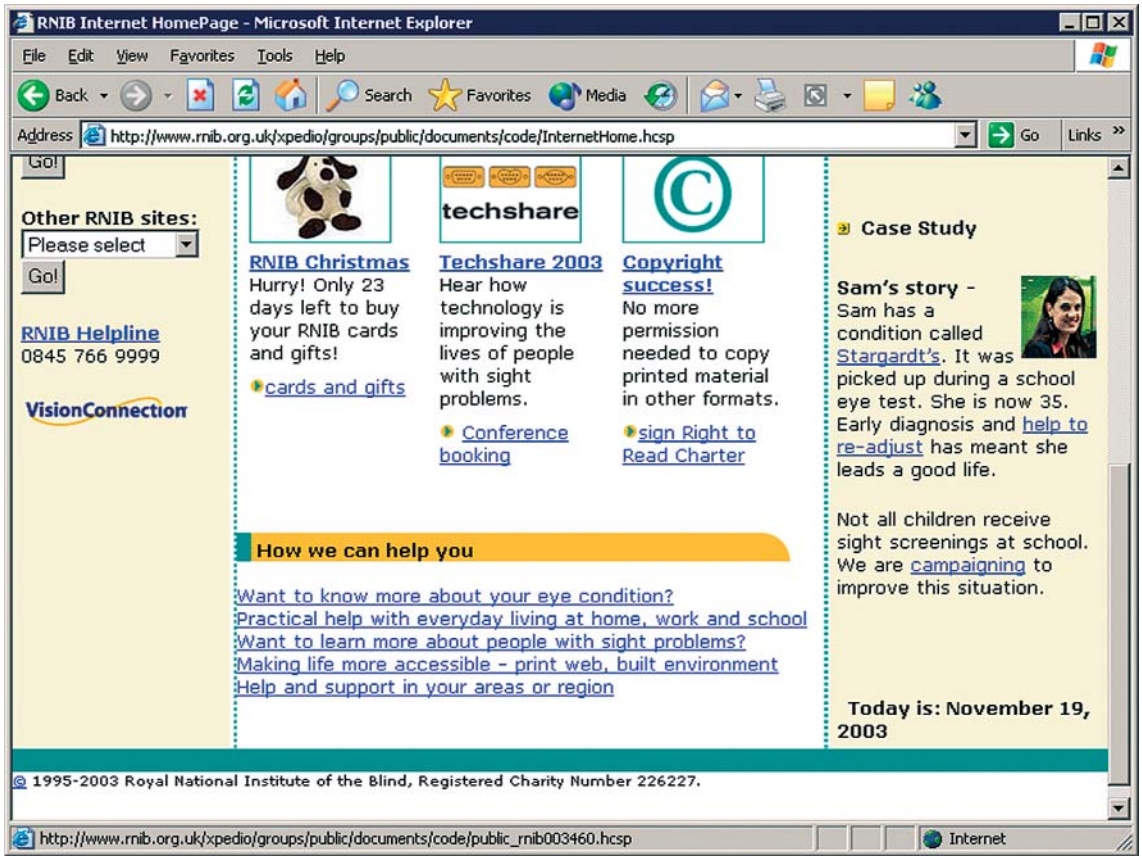


Figure 4 www.rnib.org

Many people fancy wine, and blind people are no exception. Vintners should therefore make their home pages fully accessible. I once went to one of the big winemakers, Lindemans. As anticipated, I found a number of pictures, which are also used as hyperlinks. Some picture descriptions were poor, but comprehensible; e.g. 'winemakers winemakers'. When 'home_images/shim' appears three times, things get a bit more unpredictable. I can of course select each link in turn, read the attached pages, and hope for the best, but if I return to the same producer's home page next year, I may not remember which 'home_images/shim' is which. When using the screen reader Jaws®, the screen appears like this to me:

Lindemans Wines
aboutUs/aboutUs
home_images/shim
winemakers/winemakers
home_images/shim
home_images/shim
home_images/shim
home_images/contactus
Can You Survive A Month on Lindemans Island?
Click here to see what it takes to WIN!

Ten suggestions for designing better web pages

I have no ambitions to describe all the guidelines in WAI and all those principles I believe are important. In the list below I have given some examples of bad design that often recur, but that can easily be set right. I have limited my list to the following ten items:

1. *Names on links:* Names are important. Screen readers can often show a list of links. The user may then use the initial letter to access a link quickly. However, if many links have the same name, e.g. 'Click here', this is not possible.

2. *Names on frames:* The principle is the same as for links; a descriptive name makes it easier for blind people to find their way.

3. *Descriptions of pictures* are essential, especially when pictures are used as clickable links. However, descriptions should be avoided for pictures that are used as purely decorative elements. If graphics are used to show codes or similar information, e.g. in net-bank services, alternatives must be offered.

4. *Be aware of the number and the placement of links on a web page*. Normally a web page becomes more cluttered if there are too many links. Most screen readers will work best if the links are placed at the bottom or to the right on the page. This is because the pages are reformatted and presented as a single long column. It is actually very easy to implement this when knowing style sheets and screen reader functionality.

5. *Be very careful about introducing completely new web-functionality*. It takes time before assistive technology aids work properly with new concepts. Recent examples have been Java® and Flash®.

6. *Use standard forms with informative text*. Many forms are difficult to use for blind people. The only way to ensure that a form will work well is to test it with a screen reader, and preferably with other technical aids.

7. *Consider offering alternatives to complex tables*. Also, take care to use clear row/column labels for all significant tables.

8. *Use standard HTML styles*. This will make navigation much quicker. Blind people should be offered e.g. a list of all headings and be able to jump directly to anyone of them. This will also make it considerably easier to present information in a more structured way; e.g. in Braille printouts.

9. *Do not rely on parallel versions of web sites*, e.g. graphical and text versions. It is only in some special cases that this is wise. The result of an additional 'text only' version is often more work and that the different versions are not synchronised after a while. Exceptions are when all information is retrieved from databases (see next item).

10. *Information vendors who are really serious about accessibility should have their web sites tested by the intended users*. Preferably, the web pages should be assessed with the usual technical aids. In MediaLT (where I work) we have discovered that this kind of enhancing usually makes a web site more useful to all users, not just to people with disabilities.

Enhancing web sites will not only benefit blind people!

MediaLT was engaged to work with the Olympic games information system in Salt Lake City. Novadata AS did the programming while MediaLT designed the interface for blind users. Because of the

demands of the IOC (the International Olympic Committee) we had to make a separate version of the web site for blind people. This was not too difficult since all information resided in a database. Much of the effort lay in devising alternatives to the complex tables (lists of competitors, results, etc.). We also had to make changes in the structure of the forms. The final solution worked very well for visually disabled people, but what surprised us, however, was that the logs showed that the version for blind users was visited by a very large number of other people.

The fully internet-based bank Skandiabanken has designed a separate interface for blind customers. The bank is recognized for a very tidy and clear interface for all, but the version for blind customers is even tidier. The system works really well. When I show other customers of the bank how to access the version for blind customers, they usually respond: “this is really super!”, and they often prefer to use the enhanced version.

These are just two examples; I could list many more! My proposition is that all information suppliers should take accessibility more seriously. This can make much information and many services much more user friendly for blind people, but enhancing a web site will most often make it more user friendly to all users.

Morten Tollefsen (37) is Cand.Scient. from the University of Oslo, Department of Informatics. He has written several papers on software and disabled users. Tollefsen worked as a researcher in the field of informatics and disabled persons (University of Oslo) and as a chief engineer (municipality of Oslo), before establishing MediaLT together with Magne Lunde in 1999. Both Lunde and Tollefsen are blind. MediaLT has managed several research projects, and the company is partner in two European projects.

morten@medialt.no

Usability challenges in user interface standards development:

Expanding the character standard for the 12-key telephone keypad

BRUNO VON NIMAN

Bruno von Niman

The European Telecommunications Standards Institute (ETSI) has recently published a new standard, ETSI Standard ES 202 130 – *Human Factors (HF): User Interfaces; Repertoires, ordering rules and assignment of characters to the 12-key telephone keypad (European languages)*, specifying the assignment of characters on the 12-key telephone keypad for a range of European languages. For the first time, a standard now exists for letters, digits and special characters, including European language-specific letters (Latin, Greek and Cyrillic scripts) and other common characters (such as the Euro symbol and punctuation marks).

This paper describes the development of the new standard during 2002–2003 and reflects upon related usability challenges addressed during the work by Specialist Task Force (STF) 202 of the ETSI Technical Committee Human Factors (TC HF).

1 General

Devices with telecommunication functionality represent the largest consumer product segment in the world. Telecommunication, converging with information processing and intersecting with mobility and Internet technology, is leading to the development of new interactive applications and services, offering global access.

At present, finding the characters necessary to enter a name in the terminal's phone book, searching for a name, writing an SMS (text) message or logging on to a mobile Internet portal cannot always be performed easily, because different manufacturers apply different entry mapping and ordering for the characters on the keypad. Usage varies sometimes even between devices and applications from the same manufacturer. Standardizing the use of the characters on keypads will give users easier access to different communication devices and services, with simple, correct and efficient text input, search and retrieval. It will also broaden market opportunities for manufacturers and suppliers and reduce their development costs.

The original reason for assigning letters to the rotary dial pad and later to the numeric telephone keys was to provide alphabetic 'aliases' for digits, as mnemonics in dialling. The emergence of and need to use a telephone keypad for entering data was not envisaged. Neither was imagined the concept of 'phone books' stored inside a telephone, nor the successful facility to transmit short text messages, SMS, as an alternative to voice communication.

The only standard available previously, addressing assignment of characters to the 12-key telephone keypad, is limited to the assignment of the basic 26 Latin letters (a to z). Language-specific letters (e.g. ü, é, å, ä, ö) as well as other characters (e.g. the Euro sign) were not addressed.

The lack of addressing such typically European issues has led to diverse and inconsistent solutions for European languages, obviously creating accessibility barriers to basic communication access in *e*Europe.

Europe has 230 indigenous languages; worldwide there are close to 7000. The largest number of languages presently supported by a specific ICT device or service is approaching 50. Cultural and linguistic diversity is one of the key strengths of Europe. However, in ICT, it raises issues that need to be considered and solved in order not to limit access to services, their availability and usability, on the basic as well as more advanced levels.

This development was aligned with the European Commission's initiative *e*Europe, a program for accelerated uptake and inclusive deployment of new, important, consumer-oriented technologies (http://europa.eu.int/information_society/eeurope).

1	ABC 2	DEF 3
GHI 4	JKL 5	MNO 6
PQRS 7	TUV 8	WXYZ 9
*	0	#

Figure 1 The only standard available previously, addressing assignment of characters to the 12-key telephone keypad, was limited to the assignment of the basic 26 Latin letters (a to z)

ISSN 0085-7130

2 Scope of the work

The standard specifies the minimum repertoire and assignment of graphic (letter, digit and special) characters to standard 12-key telephone keypads on ICT devices with telephony functionality. It applies to public or private, fixed or mobile network terminals, without an alphanumeric keyboard but providing a 12-key keypad in hardware form (e.g. as push button keys) or software form (e.g. as soft keys on a visual display). It also applies to network-based services accessed through such terminal devices.

The new standard complements and is thereby compatible with ETS 300 640 by additionally including European language-specific letters (Latin, Greek and Cyrillic scripts) and other common characters (e.g. the Euro sign and punctuation marks). It specifies solutions for both language-independent and language-specific keypad assignments, mapped to the 12-key telephone keypad, also providing common and language-specific information on character repertoires and ordering.

The standard is fully applicable to the languages of the European Union (EU) member states as of 2004 (also covering the official languages of the European Union) and those of near-term enlargement candidate countries and, additionally, to the official languages of the EFTA countries and Russian.

Letters specific to some minority languages, e.g. in particular those recognised in ratifications of the Council of Europe charter on regional/minority languages, e.g. Sami, have been included in the tables as suitable.

Figure 2 Sample of terminals to which the new standard applies

In anticipation of future expansions, the language-independent repertoires and keypad assignments specified also include letters needed in some of the remaining European official languages. Future revisions of ES 202 130 document may include the letters of other languages and other characters.

Figure 3 Some terminals to which the new character standard does not apply

This ETSI Standard does not cover any implementation related issues, e.g. specifics of predictive text input or user interface design.

3 User requirements

Intended users of the standard are those implementing it, for example interaction designers and other developers of ICT devices and services, designing user interfaces deploying text input and output, applied to 12-key keypad arrays provided in hardware form (e.g. as push button keys) or software form (e.g. as soft keys on a visual display) and telecommunication network-based services accessed through such terminal devices.

Intended end users addressed are the consumers (end users) of the ICT devices and services mentioned above, ranging from first time to experienced advanced users, who can produce tactile stimuli in the form of a key press and perceive written text. The end users' main goal is to efficiently use ICT devices and services under circumstances intended by these.

The deployment of ES 202 130 will enable users to reapply knowledge and previous experience between different ICT devices and services using a 12-key standard keypad array and a display. Control of common functions such as entering of characters and retrieval of text in a certain order will be simplified. Well-established services which rely on alpha mnemonics (e.g. '800 DOCTOR' rather than '800 362867' are not negatively influenced as the standard only complements ETS 300 640.

For certain end users with special needs, especially those with sensory or physical disabilities, ES 202 130 will proove very helpful due to consistent implementations (same character always found in the same position, regardless of the terminal manufacturer). For certain disabilities, e.g. in the case of temporary or permanent difficulties caused by cognitive problems or the lack of necessary level of proficiency in the respective language and other communication impairments such as: visual impairments, the inability to produce distinctive tactile stimuli or difficulties in handling, distinguishing and understanding textual information, the present document is not expected to have any impact.

For detailed guidance, including specifics of user impairments and resulting handicaps, possible solutions on access-for-all achievable through assistive technologies, design for all and multi-modal interfaces, see Human Factors (HF); Requirements for Assistive Technology Devices in ICT, Human Factors: Design for all: guidelines for ICT products and

Figure 4 All official languages of the EU are covered. This includes the new members as of May 1, 2004 as well as the candidate members. In addition, the EFTA languages and Russian are covered

services and Study of multi-modality of Icons, Symbols and Pictograms.

Uniformity in the basic interactive elements increases the transfer of learning between devices and services and improves the overall usability of the entire interactive environment. Such transference becomes even more important in a world of ubiquitous devices and services.

Guiding principles during the development of the ordering and assignments of the alphanumeric characters have been:

1. Consistent and harmonised across different devices and services;

2. Easy to learn and remember;

3. As natural as possible, matching previously acquired knowledge;

4. Redundancy (multiple solutions possible to reach desired input).

4 Methodology

4.1 Initial survey

As start of the work to developing the standard, an informal survey of the key assignments in a number of mobile phone models was carried out on several major manufacturers' handsets. The survey was based mainly on specifications and user manuals downloaded from Internet but also on 'hands-on' investigation.

4.2 Future-proofness

The standard is expected to have considerable impact, not only in the sheer number of users that it could affect, but also in the functionality it may enable or – conversely – limit.

The original reason for assigning letters to the numeric telephone keys was to provide alphabetic 'aliases' for digits, as an 'aide-memoir' (mnemonics) in dialling. That there would emerge a need to use a telephone keypad for inputting data was something that nobody envisaged. Neither was the concept of 'phone books' stored inside a telephone imagined, nor the – unexpectedly successful – facility to trans-

mit short text messages as a complement to voice calls.

4.3 Characters required

Approximately 240 Latin-repertoire letters are needed to cover the major European languages. With Greek and Cyrillic letters added, the number increases to well over 350.

This can be compared to the 75 Latin-repertoire letters (mix of capital and small) supported by the present GSM 03.38, 7-bit scheme generally implemented in today's mobile phones and networks (only 85 letters all-in-all when the Greek capital letters of that scheme are included).

Letter	GSM 03.38 7-bit coding	ISO/IEC 6937 coding	ISO/IEC 10646 identifier	ISO/IEC 10646 name
a	6/01	06/01	U+0061	LATIN SMALL LETTER A
A	4/01	04/01	U+0041	LATIN CAPITAL LETTER A
á	—	12/02 06/01	U+00E1	LATIN SMALL LETTER A WITH ACUTE
Á	—	12/02 04/01	U+00C1	LATIN CAPITAL LETTER A WITH ACUTE
à	7/15	12/01 06/01	U+00E0	LATIN SMALL LETTER A WITH GRAVE
À	—	12/01 04/01	U+00C0	LATIN CAPITAL LETTER A WITH GRAVE
ă	—	12/06 06/01	U+0103	LATIN SMALL LETTER A WITH BREVE
Ă	—	12/06 04/01	U+0102	LATIN CAPITAL LETTER A WITH BREVE
â	—	12/03 06/01	U+00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX
Â	—	12/03 04/01	U+00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
å	0/15	12/10 06/01	U+00E5	LATIN SMALL LETTER A WITH RING ABOVE
Å	0/14	12/10 04/01	U+00C5	LATIN CAPITAL LETTER A WITH RING ABOVE
ä	7/11	12/08 06/01	U+00E4	LATIN SMALL LETTER A WITH DIAERESIS
Ä	5/11	12/08 04/01	U+00C4	LATIN CAPITAL LETTER A WITH DIAERESIS
ã	—	12/04 06/01	U+00E3	LATIN SMALL LETTER A WITH TILDE
Ã	—	12/04 04/01	U+00C3	LATIN CAPITAL LETTER A WITH TILDE
ą	—	12/14 06/01	U+0105	LATIN SMALL LETTER A WITH OGONEK
Ą	—	12/14 04/01	U+0104	LATIN CAPITAL LETTER A WITH OGONEK
ā	—	12/05 06/01	U+0101	LATIN SMALL LETTER A WITH MACRON
Ā	—	12/05 04/01	U+0100	LATIN CAPITAL LETTER A WITH MACRON
æ	1/13	15/01	U+00E6	LATIN SMALL LETTER AE
Æ	1/12	14/01	U+00C6	LATIN CAPITAL LETTER AE
b	6/02	06/02	U+0062	LATIN SMALL LETTER B
B	4/02	04/02	U+0042	LATIN CAPITAL LETTER B
c	6/03	06/03	U+0063	LATIN SMALL LETTER C
C	4/03	04/03	U+0043	LATIN CAPITAL LETTER C
ć	—	12/02 06/03	U+0107	LATIN SMALL LETTER C WITH ACUTE
Ć	—	12/02 04/03	U+0106	LATIN CAPITAL LETTER C WITH ACUTE

Figure 5 Latin-script language-independent alphabet

Char	GSM 03.38 7-bit coding	ISO/IEC 6937 coding	ISO/IEC 10646 identifier	ISO/IEC 10646 name
	2/00	02/00	U+0020	SPACE
!	2/01	02/01	U+0021	EXLAMATION MARK
'	2/02	02/02	U+0022	QUOTATION MARK
#	2/03	02/03	U+0023	NUMBER SIGN
%	2/05	02/05	U+0025	PERCENT SIGN
&	2/06	02/06	U+0026	AMPERSAND
"	2/07	02/07	U+0027	APOSTROPHE
(	2/08	02/08	U+0028	LEFT PARENTHESIS
)	2/09	02/09	U+0029	RIGHT PARENTHESIS
*	2/10	02/10	U+002A	ASTERISK
+	2/11	02/11	U+002B	PLUS SIGN
,	2/12	02/12	U+002C	COMMA
-	2/13	02/13	U+002D	HYPHEN-MINUS
.	2/14	02/14	U+002E	FULL STOP
/	2/15	02/15	U+002F	SOLIDUS
:	3/10	03/10	U+003A	COLON
;	3/11	03/11	U+003B	SEMICOLON
<	3/12	03/12	U+003C	LESS-THAN SIGN
=	3/13	03/13	U+003D	EQUALS SIGN
>	3/14	03/14	U+003E	GREATER-THAN SIGN
?	3/15	03/15	U+003F	QUESTION MARK
@	0/00	04/00	U+0040	COMMERCIAL AT
[	1/11 3/12	05/11	U+005B	LEFT SQUARE BRACKET
\	1/11 2/15	05/12	U+005C	REVERSE SOLIDUS

Figure 6 Language-independent (European) repertoire of digits and special characters

It was found necessary to include in the language-specific repertoires more letters than are contained in the 'core' of those languages, called "Type A" letters. This is because in all languages there is a user need to input also foreign-origin words, some of them needing 'foreign' letters. Further, in all countries there exist user preferences in spelling of some names with 'foreign' letters, and possibly also a need to represent names – personal and/or geographical – correctly in recognised minority languages, such as Sami in Norway.

The repertoire tables therefore also include "Type B" letters. These parts of the tables shall be seen as 'best-effort' in the development of the standard, and may become modified in the future.

4.4 Character ordering

Ordering of characters is a highly complex problem, and has been the subject of very large amounts of work in several standardisation bodies, both national and international. Earlier ETSI and ISO/IEC standards specify principles based on a 'multi-level' approach for the ordering of strings of characters. However, it was found necessary to adopt a simplified 'single-level' method for this standard, considering the limited capabilities of telephone devices as compared to computer systems.

As regards letters, the two language-independent repertoire tables specify a deterministic ordering. For the language-specific repertoire tables, however, some additional criteria were applied because of established practices in telecommunications, e.g. for printed telephone directories.

In all European languages, the letters A–Z are considered part of the alphabet even if, in many of them, some of the letters are not used in any indigenous-origin words. Also some languages have special-shape letters, like the Icelandic Þ and the German ß (which remains in use, also after recent spelling reforms). Additionally, all languages use special variants of letters A–Z with diacritical marks, like the acute accent and the cedilla (e.g. É and Ç). For ordering, most languages consider such variants equivalent to the basic letter. In some languages, however, a few of them are considered letters of their own, and ordered differently. For instance, the letter Å is ordered in Norwegian as the last letter of the alphabet.

As far as possible, national conventions were followed for the language-specific repertoire tables. This may possibly cause 'non-deterministic' ordering in specific cases. Although unsatisfactory in principle, it was concluded that this could be accepted for the relevant applications.

4.5 Device dependencies

For mobile phones, there exist four interacting but independent cases of language dependencies:

1. User interface related settings (e.g. menu language)
2. Ordering (e.g. of lists such as the list of phone book entries)
3. Keypad layouts (i.e. character assignment to keys)
4. Dictionaries for 'predictive' text input (e.g. T9).

4.6 System and network constraints

The present generation of system network implementations for applications such as SMS messaging has different constraints, bounded by the ETSI standard GSM 03.38. Part of that standard was originally taken over from paging, with only a rudimentary set of characters defined: the "ASCII set" complemented by a few specifically European-language letters, amongst them the ten Greek capital letters not having a corresponding visual representation in the Latin alphabet. The Cyrillic alphabet was not covered at all.

The GSM 03.38 standard applies only to what is transmitted between a mobile phone and the "Mobile Switching Centre" (MSC), not to how SMS generation is handled inside the phone. Naturally it is however not very meaningful to generate SMS messages with characters that can then not be transmitted, so the character limitations of the standard also limits what needs to be generated. With the original – 'default' – SMS character set, multi-linguality is therefore completely unsatisfactory.

In GSM Phase 2, an alternative to the original character set was introduced, in principle permitting about twice the numbers of characters of the original SMS scheme. This alternative was designated "user-defined", i.e. no scheme was specified in the standard. It appears no user – i.e. operator – has utilized this possibility.

With GSM Phase 2+ and UMTS (3G), another alternative is introduced, namely the coding scheme of ISO/IEC 10646-1, also known as Unicode. With this scheme there is, in principle, no longer any limitation on the repertoire of characters that can be represented in SMS messages. European multi-linguality is therefore enabled, as far as representation of characters is concerned.

4.7 Keypad input sequences

In today's keypad-input implementations – foremost in mobile phones – the digits are generally placed as the last character in the key-press sequence, following not only the standardized letter assignments (ABC on key 2, DEF on key 3 etc.) but also all special letter variants assigned to the keys.

The same principle was considered for the present document. However, the special needs of visually impaired users make the principle questionable.

It was therefore decided to place, instead, the digits immediately following the present standardized letter assignments; i.e. as the fourth key-press on all keys except 7 and 9 (PQRS and WXYZ) where it is the fifth.

5 Outcome and adoption

Before approval and publication, ETSI Standards undergo a two-step approval procedure. The present standard was approved on the first level by ETSI TC HF at its 31st Plenary in June, 2003.

The ETSI Membership Voting Procedure followed, a two-month process during which all ETSI members

Key	Letter	ISO/IEC 10646 identifier	ISO/IEC 10646 name
2	a	U+0061	LATIN SMALL LETTER A
	b	U+0062	LATIN SMALL LETTER B
	c	U+0063	LATIN SMALL LETTER C
	2	U+0032	DIGIT TWO
	á	U+00E1	LATIN SMALL LETTER A WITH ACUTE
	à	U+00E0	LATIN SMALL LETTER A WITH GRAVE
	ă	U+0103	LATIN SMALL LETTER A WITH BREVE
	â	U+00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX
	å	U+00E5	LATIN SMALL LETTER A WITH RING ABOVE
	ä	U+00E4	LATIN SMALL LETTER A WITH DIAERESIS
	ã	U+00E3	LATIN SMALL LETTER A WITH TILDE
	ą	U+0105	LATIN SMALL LETTER A WITH OGONEK
	ā	U+0101	LATIN SMALL LETTER A WITH MACRON
	æ	U+00E6	LATIN SMALL LETTER AE
	ć	U+0107	LATIN SMALL LETTER C WITH ACUTE
	č	U+010D	LATIN SMALL LETTER C WITH CARON
	ċ	U+010B	LATIN SMALL LETTER C WITH DOT ABOVE
	ç	U+00E7	LATIN SMALL LETTER C WITH CEDILLA
3	d	U+0064	LATIN SMALL LETTER D
	e	U+0065	LATIN SMALL LETTER E
	f	U+0066	LATIN SMALL LETTER F

Figure 7 Language-independent assignment of digits and Latin-script characters

Key	Letter	ISO/IEC 10646 identifier	ISO/IEC 10646 name
1		U+0020	SPACE
	.	U+002E	FULL STOP
	,	U+002C	COMMA
	?	U+003F	QUESTION MARK
	!	U+0021	EXLAMATION MARK
	'	U+0022	QUOTATION MARK
	"	U+0027	APOSTROPHE
	-	U+002D	HYPHEN-MINUS
	:	U+003A	COLON
	;	U+003B	SEMICOLON
	(	U+0028	LEFT PARENTHESIS
	)	U+0029	RIGHT PARENTHESIS
	@	U+0040	COMMERCIAL AT
	_	U+005F	LOW LINE
	&	U+0026	AMPERSAND
	/	U+002F	SOLIDUS
	\	U+005C	REVERSE SOLIDUS
	[	U+005B	LEFT SQUARE BRACKET
	]	U+005D	RIGHT SQUARE BRACKET
	{	U+007B	LEFT CURLY BRACKET
	}	U+007D	RIGHT CURLY BRACKET
	¿	U+008F	INVERTED QUESTION MARK
	¡	U+00A1	INVERTED EXCLAMATION MARK
	~	U+007E	TILDE
	^	U+005E	CIRCUMFLEX ACCENT
	§	U+00A7	SECTION SIGN

Figure 8 Language-independent (European) assignment of digits and special characters

(e.g. Telenor) are provided the chance to cast a weighted vote. STF 202 worked hard to achieve consensus among the different manufacturers, to ensure the new standard is widely acceptable and will be implemented. As a result, the ETSI Members were unanimous in their support of ES 202 130. In addition, the standard was delivered on schedule, in time to meet urgent implementation-oriented market needs, which is unusual – if not unknown – in the case of character standards, where so many varied interests have to be taken into account.

References

ETSI references are available free of charge at www.etsi.org.

ETSI. *Human Factors; User Interfaces; Character repertoires, ordering and assignment to the 12-key telephone keypad (European languages)*. Sophia Antipolis, 2003. (ETSI ES 202 130)

ETSI. *Human Factors (HF); Assignment of alphabetic letters to digits on standard telephone keypad arrays*. Sophia Antipolis, 1996. (ETSI ETS 300 640)

ETSI. *Digital cellular telecommunications system (Phase 2+); Alphabets and language-specific information (same as GSM 03.38 version 7.2.0, Release 1998)*. Sophia Antipolis, 1998. (ETSI TS 100 900)

ETSI. *Requirements for Assistive Technology Devices in ICT*. Sophia Antipolis, 2002. (ETSI TR 102 068)

ETSI. *Design for all: guidelines for ICT products and services*. Sophia Antipolis, 2002. (ETSI EG 202 116)

ITU. *Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network*. Geneva, 2002. (ITU-T Recommendation E.161 (02/01))

EEC Council. Regulation No 1 of 1958 determining the languages to be used by the European Economic Community (EEC). *Treaty establishing the European Community*. Official Journal 017, 06/10/1958, 0385–0386, 31958, R0001. April 15, 1958. November 27, 2003 [online] – URL: http://europa.eu.int/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=en&numdoc=31958R0001&model=guichett)

EU Convention documents for EU Enlargement. Athens, Greece, April, 2003.

Note 1: The tables presented in this article are samples with a reduced size. All complete tables are available in ETSI ES 202 130 ETSI Standard ES 202 130 (may be downloaded free of charge from http://pda.etsi.org/pda)

Note 2: This article is based on the ETSI Standard (ES) 202 130, developed by ETSI STF202, consisting of the STF Leader Bruno von Niman and the contracted experts Martin Böcker, Karl Ivar Larsson, Ian Rattigan, Matthias Schneider-Hufschmidt and Aino Wihervaara, representing Ericsson, Siemens, LWP Consulting, Sony Ericsson, Siemens and Nokia.

Note 3: Figures and illustration sources (in order of appearance): ETSI EG 202 130, sonyericsson.com, rim.com, palm.org, nokia.com and eurunion.org.

Bruno von Niman (36) holds an MSc in Computer Science, Man-Machine Interaction profile from the Universities of Uppsala, Sweden and Stuttgart, Germany, 1993. After a brief stop in academia he worked for Ericsson in Stockholm 1994–2003 with concept and product development, user interaction design, user experience of Mobile Enterprise communication solutions, coordinating Group activities. Since 2003 he has been running an independent consultancy and represents Sweden (the NSO) in ETSI TC HF. Since 1997 he has been Vice Chairman of ETSI Technical Committee Human Factors (TC HF), he has been leader of several ETSI Specialist Task Forces, expert in several projects sponsored by the European Commission and member of several conference program committees related to the area of usability, communication and mobility. He has more than 30 peer-reviewed publications and is a frequent speaker and expert panelist at the most important global mobile communication, human factors and user experience events.

bruno@VONNIMAN.com

Generic user interface elements for mobile terminals and services

BRUNO VON NIMAN

Bruno von Niman

This article presents the development of and issues addressed by the draft ETSI Guide (DEG) 202 132, *Generic User Interface Elements for Mobile Terminals and Services*, under development by the European Telecommunications Standards Institute (ETSI) Specialist Task Force (STF) 231 of the Technical Committee Human Factors (TC HF), during 2002-2004.

1 General

Information and communication technologies (ICT) play a key role in the daily activities of many people. The mobile telephone is the most successful device ever invented which also corresponds to a deep human communication urge.

The number of mobile subscribers overtook the number of wire-line subscribers globally in 2003, mobile services growing six times faster than fixed-line services. In addition to the 1.28 billion presently subscribed, over half a million new mobile telecommunications users sign up each day and people talk more and increase their use of data services. The growth is particularly strong in China, India and Russia, partly driven by tariff reductions. Today, the world penetration is only 20 % (Asia-Pacific still only has 12 % penetration in mobile subscriptions while Western Europe and North America has 80 % and 51 % respectively).

Also the offered capabilities and services emerge, from only being able to make a call to downloadable personalization achieved through ring signals, software programs such as games and the introduction of multimedia information services such as mapping and directions, traffic information, e-mail access, quasi-cordless functionality or video telephony. The growth is driven by new, voice-centric customers but also by increasingly empowered users of advanced mobile services, such as mobile data services and applications. Mobile data services are also being brought to the personal computer desktop.

Connectivity and interoperability between telephony networks, personal computing, the Internet, and ever-smarter mobile devices and services offer enormous potential for improving life. However, complexity is on the increase, and there is concern about whether these new products, services and their content will be fully accessible to all people. An effective *e*Society relies on the fact that *all* citizens are granted access. Users who cannot get over the hurdle of the first installation of their devices and services will perpetually be excluded from the *e*Society. Ensuring access to mobile communication for all is a common goal of vendors, operators, service providers, user associations as well as politicians, often talking about the creation of the e-inclusive information society.

It is important to consider the use of market driven solutions that utilise technologies with forward-looking interoperability. Such an approach can provide users with increased satisfaction in the use of superior modes of communication devices and ICT equipment. A similar approach, the Digital Home Working Group, has recently been announced in the networked consumer electronics area in order to establish a platform of interoperability for digital media.

The draft ETSI Guide is based on the results and recommendations provided by ETR 102 125. The work is conducted in close and open collaboration with the industry, aiming at consensus building and implementation-oriented recommendations. It is being presented to the international community at various mobile communication conferences, workshops, symposia and other events such as the 3G World Congress 2003, the World Handset Forum 2003, the Human Factors in Telecommunications 2003 Symposium and the GSM World Congress 2004, thereby increasing the understanding for the benefits of such an approach, preparing the ground for implementations.

The work is aligned with and sponsored by the European Commission's initiative ***e*Europe**, a programme for inclusive deployment of new, important, consumer-oriented technologies, opening up global access to communications and other new technologies, for all – see http://europa.eu.int/information_society/eeurope.

The *e*Europe 2005 Action plan – following on from the *e*Europe 2002 initiative – aims to provide a favourable environment for the creation and uptake of new services and new jobs, to boost productivity, to modernize public services and to give everyone the opportunity to participate in the global information society. Thereby, a most competitive and dynamic

 ISSN 0085-7130

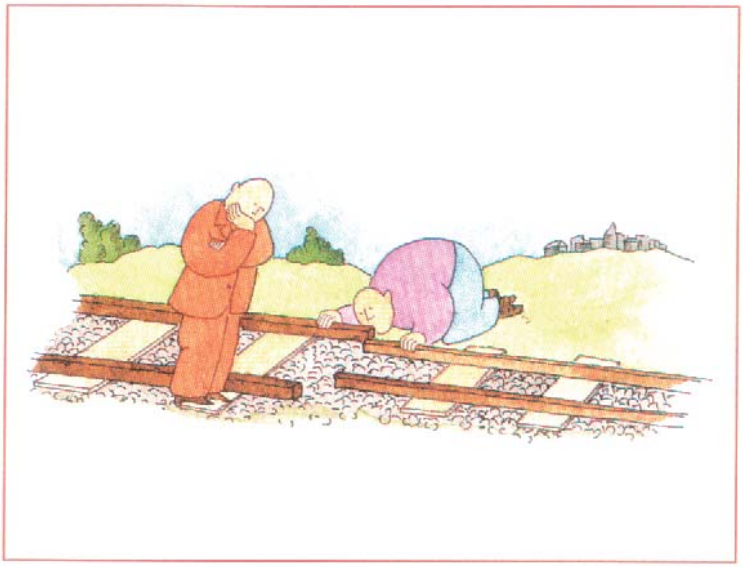


Figure 1 Standards are important to ensure interworking and easy access to all users

economy, exploiting the opportunities of new technologies, can be created. However, this will only happen if people have confidence in the commercial and public services offered to them electronically.

2 Scope of the work

The draft EG aims at simplifying end user access to information and communication services from mobile communication devices. It does not restrict the ability of market players to further develop their devices and services, nor does it limit their options to trademark user interface elements or profile the user experience of brand-specific user interface implementations as a competitive edge.

The draft EG addresses key issues from the end user's perspective, providing guidance on proposed generic user interface elements for mobile terminals, services and certain aspects of application handling. The aim is to provide simplified access to basic and selected advanced functions of mobile communication. User requirements and available results of standardisation work have been considered and integrated in the present document, providing implementation-oriented guidance.

Throughout the document, a Design-for-All approach has been adopted, taking special needs of children and elderly users with physical and sensory disabilities into account.

3 Rationale for generic user interface elements

Manufacturers, operators and service providers differentiate their products and services by trying to make them unique, or at least different from and better than those of their competitors. Areas in which such differentiation can be achieved include industrial and screen design, feature sets and also the user interface design. In this light, the user interface is not an obvious candidate for the definition of generic user interface elements across manufacturers and service providers.

Harmonised and generic user interface solutions have in the past found acceptance, in particular in those product types that raise specific safety issues. A good example of safety-motivated user interface harmonisation is the controls and indications used in cars. The user interface elements of cars (e.g. the gear, the steering wheel, the arrangement of pedals, the symbols and colour schemes used for many of the indications) are harmonized across manufacturers to such an extent that users expect to be able to immediately drive any car (e.g. a car picked up from a rental station) without reading any instructions.

A second type of products with generic user interface elements concerns those products expected to be used by many different people. These are typically products in public places (e.g. public telephones) or in work environments.

Figure 2 Beneficial generic user interface concepts (despite increasing complexity)

Figure 3 Public telephone boots benefitting from common user inteface approaches

Finally, there are established and accepted de-facto standards regarding the user interfaces of particular product types such as elements of graphical user interfaces in PCs and the design of musical instruments.

User interface harmonization or the emergence of generic user interface elements is the result of either de-facto standards (as in the case of the graphical user elements) or of standardisation (as in the case of the keypad arrangement on public phones). In either case, the harmonization potentially benefits end users, manufacturers and service providers. What user interface harmonization should not do is to restrict the manufacturer or service provider in expressing their brand identity or in coming up with particularly good solutions. Neither should user interface harmonization be an obstacle to novel and innovative solutions (see e.g. the emergence of new solutions for the gear change in cars, such as semi-automatic gearshifts.

In other product areas, no harmonized UI concepts have emerged, often with resulting difficulties for the users. One notorious example of a product with only limited user interface harmonisation is VCRs.

In the recent past, standards bodies have issued recommendations that are expected to facilitate the uptake of new and emerging types of user interfaces. For example, ETSI ES 202 130 presents the manufacturer with a clear instruction on how to assign European letters to telephony keypads and how to order lists in various languages thus saving the time and effort for finding individual solutions; and benefits the end user by generating consistent expectations on how characters will be handled in comparable devices. Another example is the generic spoken command vocabulary for ICT devices and services, ETSI ES 202 076, allowing the implementation of one standardised set of voice commands across a large range of heterogeneous devices and services.

Basic considerations of what makes a user-interface area a candidate for user-interface harmonization, are:

- The proposed harmonisation should not present any barrier to innovation;
- Neither should it present an obstacle to good product-specific user interfaces;
- Only the semantic of a harmonized user-interface element should be specified in most cases, not the actual design and implementation;
- End user aspects, such as learnability, familiarity, trust, configuration and access, should be considered;
- Commercial aspects (quicker uptake of new technologies, larger user base) as well as legal requirements and possible regulations should be taken into account.

4 User requirements in a mobile communication environment

The draft guide does not attempt to comprehensively document users' high-level goals as they can be very diverse and may appear unrelated to mobile terminals and services. It does not try to limit the range of users whose requirements should be considered and documented. The fundamental high-level goals relating to the human need to communicate are the same irrespective of the communication abilities of the user. Some users may be able to easily use all forms of voice and text communication whereas others may have impairments that make it difficult or impossible for them to use some means of communication. Such limitations, however, do not limit the users' requirement to communicate.

Some ways of satisfying a documented user requirement may be effective for some users but ineffective or less effective for users who have certain impairments. However, it may be possible to

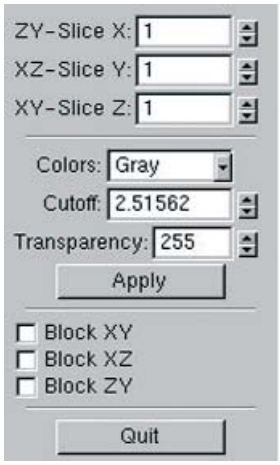


Figure 4 Example of a common PC GUI approach

Figure 5 Common terminology and symbols are often beneficial

find a means to satisfy a user requirement that is equally effective for all users. The achievement of such a solution is only likely to be achieved if the abilities of a wide range of users are taken into account when considering ways to satisfy a user requirement. This approach to satisfying user requirements with a single solution that will work for every user is termed 'Design for All'.

The adoption of a 'Design for All' approach does not guarantee that all solutions will work equally well for all users. However, failure to consider the needs of all users in attempting to satisfy user requirements will mean that solutions that are suitable for all users will be missed and solutions that restrict the range of potential users will be proposed instead.

Throughout the entire development work a 'Design for All' approach is taken in the evaluation of user requirements and the consequential proposal of technical solutions. This approach should maximise the number of solutions that are suitable for all users.

Concept	D	S/A	N	Recommended Name	Description	Comment
#-Key	X			#-Key ("Square-key")	Key of the 12-key keypad	Used for entering the character '#' and for additional functionality, in North America sometimes referred to as "Pound-Key" or "Hash".
*-Key	X			*-Key ("Star-key")	Key of the 12-key keypad	Used for entering the character '*' and for additional functionality
12-key keypad	X			12-key keypad	Telephone keypad with keys for '0' to '9', '*' and '#'	As defined by ITU-T Rec. E.161
Access code	X			Access code	In PBX-systems or MVPN implementation, the number to dial in order to get an external line	
Answering machine	X					
AOC		X				
Automatic answer	X					

Figure 6 Draft recommended terminology for mobile devices and services

However, not every solution in the document will be suitable for all users as some solution that attempts to meet the needs of every user will be very undesirable for many users.

The approach to user requirements taken in the ETSI Guide has been to work from the system features that are described throughout the rest of the EG and to try to identify the user needs that these features have been designed to satisfy. The system capabilities are what the system provides that can be seen to deliver what the user requirement asks for. It may be that a user requirement can only be satisfied by the provision of a number of system capabilities.

5 Terminology, symbols and acoustic signals

5.1 Terminology

This section in the ETSI Guide will present a table of recommended terms to be used in mobile ICT devices and services, in user interfaces and user documentation. These terms are related to terminal characteristics, services and applications, and network aspects.

The terminology presented contain the following columns:

- Concept: A term frequently used by users or specialists.
- Recommended name: The recommended term for the concept in question.
- Description: Short description of the concept (e.g. of the phone feature or i/o element).
- Comment: Additional information including references to synonyms and to terms that are frequently confused with the term in question.

If there is more than one popular term for one and the same concept, a preferred term will be recommended, based on domain knowledge and expert judgement.

5.2 Symbols

Symbols (in some cases also referred to as icons or pictograms) are used to denote the meaning of the control or indication of a device or service and have the advantage, compared to text, of being language independent (and therefore potentially universally understood) and space efficient. For this reason, they are particularly well suited for use in the context of telecommunications terminals and services. However, for a set of symbols to be successfully associated with

the underlying functionality, the symbols have to be carefully designed, evaluated and selected. Some de-facto standards exist in telecommunications (e.g. the symbols for Bluetooth and GPRS) while many of the core functions of telecommunications devices and services are represented by brand-specific symbols.

This section will deal with basic symbols to be used for representing the functions of telecommunications terminals as well as those of telecommunications services. References will be made to existing recommendations for symbols from various standards bodies followed by proposals for symbols to be developed and recommended by standards bodies. Finally, suitable methods for evaluating candidate symbols will be presented.

5.3 Acoustic signals

Acoustic signals (also sometimes referred to as earcons) used to denote the meaning of the control or indication of a device or service have the advantage, compared to printed or voice messages, of being language independent (and therefore potentially universally understood).

The availability of acoustic signals benefits all users but in particular those with visual impairments and those who cannot direct their visual attention to the device, e.g. when driving a car or when the display of a modular device is not available (e.g. a person wearing a Bluetooth headset and carrying his mobile phone in a pocket).

For this reason, acoustic signals are well suited for use in the context of telecommunications terminals and services. However, for a set of acoustic signals to be successfully associated with the underlying functionality, the signals have to be carefully designed, evaluated and selected.

This section will deal with with basic acoustic signals generated locally by terminals or transmitted by networks in order to be used in the context of telecommunications functions and services. References are made to existing recommendations for acoustic signals from various standards bodies followed by proposals for acoustic signals to be developed and recommended by standards bodies. Finally, suitable methods for evaluating candidate signals are presented.

6 Additional basic, generic elements and functions

Other areas regarded as basic, common and important, addressed in detail in the ETSI Guide, are presented below.

6.1 International access code

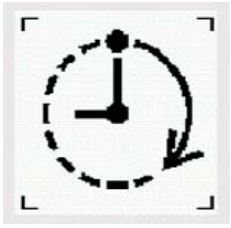


Figure 7 Symbols denote meaning

In today's telecommunication networks there are still different dialling requirements or options to start an international phone-call (e.g. '00<country-code>' in most European countries or '011<country-code>' in the US) (92/264/EEC and ITU-T Rec. E.164). Replacing the international access code by the symbol '+' has become the procedure of choice for entering dialling information or telephone numbers in address books or phone directories (ref. ETS 300 907 and GSM Association, Feb 2003). This trend has been further strengthened by Microsoft's and Palm's adoption of this solution on PCs and PDAs.

Phone numbers saved or dialled with the international access code may be used in the mobile home network since the service centre is able to interpret the number correctly.

6.2 National emergency services

Emergency calling is the most appropriate candidate feature for standardized usage patterns. Since often used in stress situations, users must be able to initiate an emergency call from every perceivable phone without further analysis or thinking, with their own or foreign phones, in the home network, and even without a SIM-card or without knowing the access code to a phone. Furthermore, legal requirements have to be taken into account by all manufacturers of devices.

SOS functionality has been the subject for standardization, e.g. in GSM, where access to emergency calling without a valid user subscription is required by regulations. Also, access to the European emergency call number (ECN), 112, has been (partly) harmonized in the GSM networks.

The user procedure to start an emergency call (keys to be pressed, dialogues for user confirmation) has

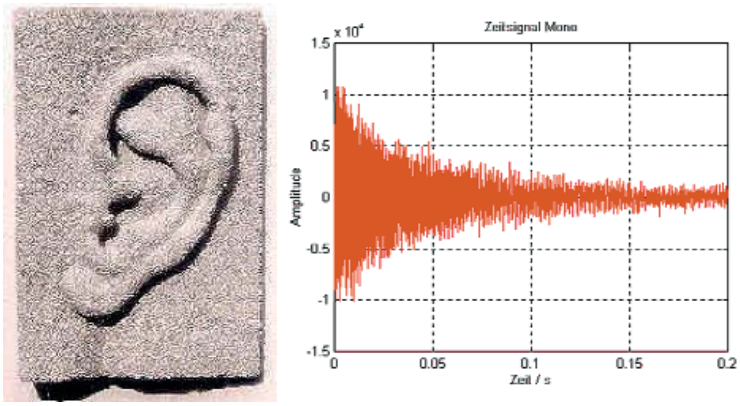


Figure 8 Acoustic signals (earcons) are language-independent

ISSN 0085-7130

already been subject to harmonization efforts. Especially in this area, a harmonized user interface seems to be important. Everyone has to be able to start an emergency call from every communication device while keeping the number of unnecessary emergency calls at the lowest possible limit. Also, in this area, there is no brand advantage to be gained from implementing non-harmonized user interfaces.

There are a number of requirements to be fulfilled by implementations of emergency call facilities:

1. Emergency calling must be possible without SIM card or valid SIM password;

2. Emergency calling must be possible with the keypad locked;

3. Emergency calling should be possible without knowing the emergency call phone number;

4. The number of inadvertent emergency calls should be kept to a minimum.

6.3 Safety and security indicators

Users must relearn indicators representing specific properties of the communication or the device itself whenever they switch from one device to the other. Also, since some of these indicators may be safety relevant it seems a valuable area of harmonization to address these indication symbols. Since several of these indicators are unused in the current generation of mobile devices, generic user interface elements can more easily be developed.

The list of possible indicators include:

- Network access / flight mode
- Secure/encrypted communication
- Network unavailable / SOS only
- Keypad protection
- Ringer indication (ringer off/vibrator/ringer on)
- Call diversion
- Battery low indication, and
- Unlock keypad.

6.4 Access to voice-based telephony services

Basic functionality in a handset is an important topic for UI harmonization, as it is used by most users. Possibilities for defining industry-wide basic, common conceptual procedures will be analysed with careful consideration.

For users with visual impairments who have problems locating the respective function in a menu, common access to basic voice services is extremely helpful. To achieve a maximum benefit for these users, experts in blind UI will be consulted.

Tele-services are often inaccessible to users with disabilities due to the lack of flexibility/modality in user interfaces, often requiring simultaneous audio, visual and dexterity competence. Alternative solutions for inclusion are possible through assistive devices or multimodal interface approaches and will be examined in detail.

6.5 Access to video telephony services

This section will focus on UMTS video calls and examine the possibility of recommending generic UI elements for:

- Basic configuration
- Availability and selection of call and connection types
- Symbols, terminology and acoustic signals, and
- Assistive device requirements.

6.6 Text entry, retrieval and control

Efficient and intuitive text entry and retrieval are one of the basic, key requirements – and stumbling blocks – in the contemporary mobile devices. It would be beneficial to the end users and operators to see efficient, intuitive, and also common solutions to text entry.

Additional complexity arises through the necessary control functionality for predictive text entry systems. Turning these systems on or off, input of new, unknown word and the selection between prediction alternatives are major obstacles to using these systems for many users. Easy-to-use command shortcuts, harmonized over many different devices, might broaden the possible user group of these predictive text entry systems.

ES 202 130 is taken as the basis for continuous harmonization efforts in this area.

6.7 Keypad layout and hardware issues

The availability and positioning of command and control keys (e.g. Send/End, On/Off, Volume) as well as physical layout accessibility enablers (e.g. the dot on the 5 key) will be addressed.

6.8 Accessibility and assistive device interfaces

User interface harmonization for the young, elderly and disabled users are beneficial. The end users will get devices that support them better in their tasks; achieving their goals while the manufacturers can more readily satisfy regulatory obligations to provide access to all users. ETSI STF 181 has dealt with these

issues and stated a set of requirements for these interfaces, found in TR 102 068. Finally, assistive technology manufacturers will get standardized interfaces and conventions that make it easier for them to attach their technologies, services and devices with the mobile telephones.

Europe-wide standardization of assistive technology device interfaces across all ICT devices is becoming an increasingly urgent topic and should be handled with high priority from the standardization bodies. As accessibility requirements vary widely between and within different categories of users, access via assistive devices often provide better solutions to the end user than access directly via the terminal device.

7 Configuration for service access, interworking and portability

Ideally, users should not have to be exposed to configuration procedures before access to services is possible. However, for several services users will still have to deal with some amount of configuration, which may often be complex and presents a barrier to getting started with a service immediately as well as utilizing its potential in the long term. In order to widen and simplify user access it is recommended to harmonize software configuration procedures (setup, installation, etc.) that enable users to access and use terminal- and network-based services.

Configuration procedures can be arranged along a continuum of decreased/increased user interaction and hence complexity for the user:

- Pre-configuration
- Guided configuration
- Manual configuration.

UI elements for configuration, common to services in general and according to the above procedures, will be detailed in the ETSI Guide. Other related areas addressed include:

- User-centric error messages, and
- Interworking and portability.

8 Advanced functionality-related interaction elements

The feature-richness of mobile networks, systems, services and applications is on the increase, enabled by the on-going convergence between traditional voice and data services and the migration from second to third generation mobile networks. A most important difficulty experienced by present mobile communication users is also most often the showstopper for the takeoff of new technologies: technical complexity. Hiding complexity away from the end user is one of our most difficult challenges at present, striving for an excellent user experience.

Therefore, in order to assist these users, the ETSI Guide addresses the areas in the below clauses.

8.1 Universal addressing in converging networks

At present, users of communication services have a range of communication addresses that vary from service to service. A user normally has a different telephone number for each separate telephony subscription. Similarly, if a user chooses to have multiple email services they will have a number of different email addresses as a consequence. As a user acquires more services, they will begin to acquire a wider range of communication addresses. This clause will look at the issues associated with this proliferation of communications addresses and recommend potential solutions that move from this situation of address anarchy towards a situation where the communications addresses a user has more fully meets the user's needs and is less dominated by technical and commercial limitations of communications networks and services.

8.2 Number format and portability

With appropriate terminal, service and network logic, it should always be possible to take an internationally

Figure 9 A future communicator concept

ISSN 0085-7130 © Telenor ASA 2004

formatted number and use it to make a successful communication. As such, solutions that use numbers stored in international format should be encouraged. Similarly, users should also be encouraged to capture and store numbers in the full international format. Such widespread adoption of the storage and processing of internationally formatted numbers should greatly reduce the incidence of miss-dialled numbers by people who are roaming. This reduction in errors will be a major benefit for network and service providers in reducing the amount of non-chargeable abortive call attempts.

8.3 Personal data format and portability

As the mobile terminal and service market expands and diversifies, people create, capture and display personal data in a number of different terminals and services. This personal data includes calendar, address book, v-cards and other organizer data. Currently, the information that is stored and the format in which it is stored may be different from terminal to terminal and between terminals and services. As a result it is often impossible to easily and reliably move this data between terminals and services.

To avoid the complications that arise because of the different ways in which data is stored, it is necessary to ensure that all internally held data is either stored in or can be automatically converted to a common set of standards. To minimise the complications associated with the synchronisation of data, it is important that a common data synchronisation scheme is widely adopted by the manufacturers of products and services.

The user interface of the address book, editing functionality etc. are explicitly excluded from this harmonization proposal, as we assume that these usability related issues are important for manufacturers' brand positioning.

8.4 User data privacy and security

With the increasing interconnectedness of society, issues of privacy and security will become ever more important. If the security of communications products, networks and services are such that they are vulnerable to attacks that allow the user's information to be observed, stolen or altered then users will lack confidence in these products, networks and services. Similarly, if private personal data is made available to the wrong people, then users will lose confidence in the products, networks and services that allowed that data to be misused.

Increasing threats to both privacy and security are emerging due to factors such as:

- the large amount of personal information that is now being collected in order to personalize products and services;
- communication convergence (the moving of data across different domains and the usage of transport mechanisms of lower security such as Wi-Fi);
- the storage of personal information on remote customer databases and the weak controls on how that information may subsequently be used;
- the low and misconceived level of understanding of privacy and security threats in the general public;
- the almost inevitable trade-off between security and usability (e.g. multiple security levels such as PINs, shared-secrets, etc.).

This ETSI Guide will provide guidance on how to address these most important aspects.

8.5 Payments, cost of services and content

When outside the home network environment (e.g. traveling abroad), users should be made aware about the roaming cost, when they are likely to be charged more than they expect or more than they are normally prepared to spend. The EG will also address:

- Cost aspects of access to ad-hoc services
- Non-telephony related payments (e.g. parking)
- User confusion, privacy and integrity aspects
- Call type / connection type selection with cost and quality advice.

Mobile terminals are becoming wallets also containing credit card capabilities. Areas we will examine for guidance include information on cost of voice and data calls, cost aspects of access to ad-hoc services, minor non-telephony related payment confirmations such as paying for parking, user confusion, privacy and integrity aspects also taken into consideration (e.g. paying for leisure parking with corporate SIM card).

8.6 Speech and voice user interfaces

Voice is a fundamental human paradigm for communications, forming an important foundation for universal access to the services and benefits of communications technology. Voice user interfaces are also a terminal, display and location independent user interface technology, enabled by speech recognition technologies.

In order to simplify the user's learning procedure and enable reuse of knowledge between different applica-

tions and devices, dialogue design and command vocabulary recommendations will be made in the EG.

8.7 Positioning services

This section will address the area of positioning, from the end user's point of view. Positioning is a very personal and private issue for the users but can also provide help depending on the positioning service, but the privacy issues are very important to take into consideration.

This clause will concentrate on things such as usable accuracy, what might be useful services for the users and how could this information be presented to the user so it would be beneficial for the user (e.g. coordinates X,Y,Z do not tell anything to the 'ordinary' user).

Positioning services include different types of services; e.g. there are location information services such as maps, tracking services and location based push services. General usability recommendations will cover these.

8.8 Messaging

Messaging applications are becoming increasingly complex and difficult to understand for the user because of the functionalities and formats they are providing for the user. Also, different evolutions of mobile devices are differently capable of supporting these different messaging services and users are not totally aware of these things. Operator dependent services and co-operation agreements also differ and the user is totally lost when and where all of these devices supported messaging services exist.

The messaging system should be as easy to use as possible. When taking the system into use for the first time, the configuration of settings for sending and receiving messages must be as simple as possible. The daily usage of the system must be straightforward. The composing and viewing of messages should not take too much effort. Also sending and receiving of messages should be highly automated. Most of the usability issues are not directly dependent on the technology, but depend more on the actual implementation of the system.

The user interface of messaging applications, editing functionality etc. will be explicitly excluded from this ETSI Guide, as these issues are important for manufacturers' brand positioning. Covered UI elements will include:

- composing messages (text, images, etc.)
- opening the messaging application/functionality
- accessing messages
- entering receiver or multiple receivers information
- sending messages
- delivery notification and storage.

8.9 Mobile instant voice

Mobile Instant Voice is a low-cost, near-future GPRS application, relying upon IP Multimedia Subsystem (IMS) with functionality similar to push-to-talk (walkie-talkies). The EG will provide recommendations on set-up and delay times, in order to make these expected mass-market applications easy, efficient and satisfactory to use for all.

References

ETSI references are available free of charge at www.etsi.org.

ETSI. *Potential harmonized UI elements for mobile ICT terminal devices and services*. Sophia Antipolis, 2002. ETSI TR 102 125.

ETSI. *Requirements for assistive technology devices in ICT*. Sophia Antipolis, 2002. ETSI TR 102 068.

ETSI. *Generic spoken command vocabulary for ICT devices and services*. Sophia Antipolis, 2002. ETSI ES 202 076 version 1.1.2.

ETSI. *Character repertoires, ordering rules and assignment to the 12-key telephone keypad (European languages)*. Sophia Antipolis, 2003. ETSI ES 202 130.

ETSI. *Guidelines for ICT products and services; Design for All*. Sophia Antipolis, 2002. ETSI EG 202 116.

ETSI. *Access to ICT by children; Issues and Guidelines*. Sophia Antipolis, 2003. ETSI ETR 102 133.

GSM Association *Requirements for Q4 '03 to Q1 '04 products*. 2003. GSM Association M-Services Phase II Evolution 3.2.0.

ETSI. *Requirements for communication of citizens with authorities/organisations in case of distress (emergency call handling)*. Draft, Sophia Antipolis, September 2003. ETSI SR 002 180.

European Commission, Directorate General Information Society. *Inclusive Communications (INCOM), Subgroup of Communications Committee, Working Document*. Draft, October 2003.

Council of the Euopean Communities. *Introduction of a standard international telephone access code in*

the Community. Decision 92/264/EEC of May 11, 1992. (URL: http://europa.eu.int/ISPO/infosoc/legreg/docs/92264eec.html)

ITU. *The international public telecommunication numbering plan*. Geneva, 1997. ITU-T Recommendation E.164.

ETSI. *Digital cellular telecommunications system (Phase 2+) (GSM); Man-Machine Interface (MMI) of the Mobile Station (MS)*. Sophia Antipolis, 1996. GSM 02.30 version 5.7.1 Release 1996.

Note 1: The latest draft of the ETSI Guide 202 132 may be downloaded free of charge from http://portal.etsi.org/STFs/HF/STF231.asp

Note 2: This article is based on the draft ETSI Guide (EG) 202 132, under development by ETSI STF231, consisting of the STF Leader Bruno von Niman and the contracted experts Martin Böcker, Riitta Jokela, Mike Pluke, Matthias Schneider-Hufschmidt and Kristoffer Åberg, representing Ericsson, Siemens, Nokia, Castle Consulting (supported by Telenor) and Sony Ericsson.

Note 3: Figures and illustration sources (in order of appearance): Unknown sources, ETSI DEG 202 132 and Siemens (used with permission).

For a presentation of the author, turn to page 38.

The 3G saga, so far: Evolution, promises, challenges and its end user reality

BRUNO VON NIMAN

Bruno von Niman

This article presents a brief overview of the history of mobile communication and looks into the promises and challenges of 3G, from the end user perspective.

1 Milestones in the evolution of mobile communication

Fixed and mobile telephony are satisfying a deep, natural, human communication need. With the advent of the third generation of mobile telephony, 3G, the digital divide can be reduced and solutions for many necessities – from telecare to *e*-government – made possible.

Samuel F.B. Morse developed the fully functional telegraph in 1837. A good decade later, supposedly between 1849–1857, Antonio Meucci invented a 'sound telegraph', a device for transforming electricity into sound, calling the invention a *teletrofono*, or electric telephone, and filed his first patent caveat (a notice of intention to take out a patent) in 1871. The patent caveat lapsed in 1874.

In October 1861, Phillip Reis demonstrated an 'electric ear' before the Physical Society of Frankfurt, Germany. Reis coined the word *telephony* during that demonstration. In 1876, Alexander Graham Bell made his first successful telephone experiment and filed for a USA patent on February 14, 1876, just two hours before Elisha Gray did the same!

In the beginning, telephony was limited to a one-set, fixed-device approach.

In 1899, Marconi made trials with radio communication from ships, reporting from America's Cup. Two years later, the first radio message was sent from England to Canada.

Mobile telephony dates back to 1910, when the Swede Lars Magnus Ericsson, founder of *Telefonaktiebolaget L.M. Ericsson* and his wife Hilda introduced the first 'car telephone', using a normal telephone set and two long poles to hook onto a pair of roadside telephone wires, thus connecting to an operator in the telephone exchange. As this implementation was never commercially launched, we cannot even call it 0G!

Public radio was introduced in 1921. Eight years later, in 1929, the Chicago Police Department used radio communication between the police station and patrol cars. In 1935, radio communication between cars was first used in Europe, by the Gothenburg police department in Sweden. Soon thereafter, two-way radio and handwriting recognition were invented.

In 1946, the first single-cell, manually switched telephone radio service was introduced by AT&T in St. Louis, USA. One year later, in 1947, Claude Shannon and Robert Pierce developed specifications for CDMA, the first cellular type mobile.

On December 3, 1950, Sture Lauhrén made the world's first cell phone call using a prototype system developed in Sweden by L.M. Ericsson and Televerket. In 1956, Mobile Telephony System A, MTA (or shall we call it 0G?) was publicly launched in Stockholm and Gothenburg, with a maximum capacity of 150 subscribers. A telephone set used in MTA weighed 35 kilogrammes!

In 1965, Mobile Telephony System B, MTB, was launched to solve the capacity problems in MTA. MTB supported six simultaneous calls and offered a capacity of 660 subscribers. As transistors were already introduced, the weight of a mobile telephone was reduced to 9 kg. In 1966, the first fax was sent through a mobile telephone line.

In 1973, Motorola vice presidents Marty Cooper and John Mitchell made the first public demonstration of a call from a handheld wireless phone.

ISSN 0085-7130

Mobile phones have come a looong way! Several generations of evolving telecommunication terminals

In 1981 the first generation analogue Nordic system for Mobile Telephony 450 MHz (NMT 450), nowadays also called 1G, was introduced in Saudi Arabia and Scandinavia. In the mean time, incompatible systems such as AMPS, NAMPS, TDMA and CDMA were developed and launched in the United States.

In 1982, eleven countries founded Groupe Spéciale Mobile (GSM) with the goal to define a global standard for digital mobile telephony. Seven years later, in 1989, it was hosted by ETSI.

In 1985, the IMT-2000 study began in ITU-T with the establishment of an Interim Working Party (IWP 8/13) and work continued in Task Group 8/1. IMT stands for International Mobile Telecommunications and the number 2000 had three meanings. It was supposed to represent the year 2000, when the ITU-T hoped the system would become available, data rates of 2000 kbps and frequencies in the 2000 MHz region.

In 1986, NMT 450 became too popular, leading to capacity problems and NMT 900 was launched. A year later, the first PDA (Apple Newton) and the first pocket phone (Ericsson 750g, talk-time 12 minutes, standby time 4 hours) were launched.

In 1988, the European Telecommunication Standards Institute, ETSI, was founded. In 1989, voicemail for mobile subscribers was introduced and wide area paging a year later.

In 1991, Radiolinja in Finland launched the first GSM system, called 2G, and the first really portable, pocket-sized handsets appeared. With the addition of General Packet Radio Service (GPRS) capabilities in 2001, GSM evolved to 2.5 G.

In 1993, Short Messaging Services, SMS, was launched, as feedback indication for voicemail waiting. The same year, subsidized GSM handsets appeared on the market. The first Internet browser with a graphical user interface, Mosaic, was launched the same year.

In 1994, commercial operation of D-AMPS (IS-54) in the US and PDC in Japan started.

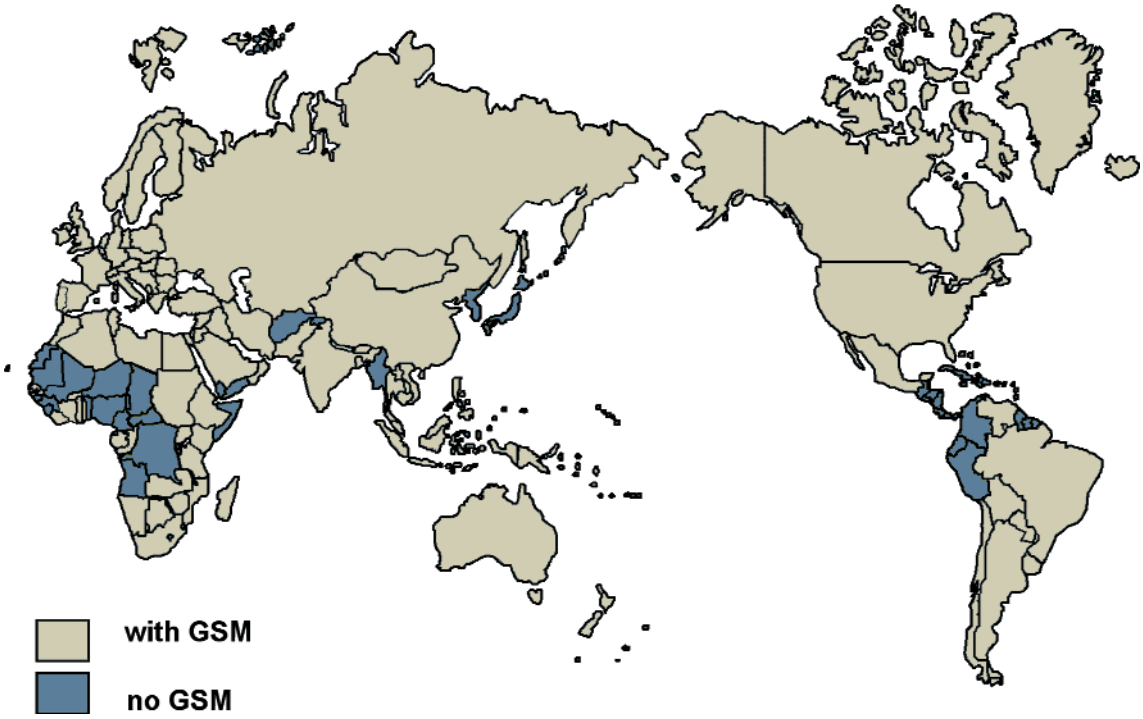
In 1996, the first Palm Pilot, Nokia Communicator and Motorola StarTac (90 g) were launched. The year after, in 1997, GSM 1800 MHz was launched to solve GSM capacity problems. The WAP Forum was founded and the first pre-paid subscriptions launched. In 1998, the first handset with a colour screen was introduced (Siemens S10).

In January the same year, WCDMA was agreed and selected as the main 3G radio standard by ETSI. In December 1998, ETSI SMG, T1P1, ARIB TTC and TTA created 3GPP in Copenhagen, Denmark.

In 1999, Wireless Internet Application Protocol (WAP) services and the first commercial GPRS networks were launched (first handsets: Nokia 7110 and Ericsson R520/320).

In 2000, Bluetooth and GPRS specifications were disclosed. In 2001, the first Java-enabled handset was launched (Motorola Accompli 008) as well as the first Bluetooth-enabled handset (Ericsson R520). The following year, in 2001, the first MMS-compatible handset (Ericsson T68i) was released.

In June 2001, NTT DoCoMo launched a trial 3G service; an area-specific information service for i-mode and in October, the first commercial Wideband Code-Division Multiple-Access (WCDMA)-based 3G mobile network.



The GSM global 'footprint' (ETSI, 2004)

Two years later, the company has more than two million subscribers and experiences an accelerating growth rate in the number of users. In September 2002, Mobilkom Austria launched Europe's first national, commercial UMTS network, followed by Vodafone KK in Japan.

Driven by voice-centric and text messaging usage patterns, by the end of 2003 the number of mobile subscribers globally was larger than the number of fixed-line subscribers. In addition to the 1.3 billion mobile users already subscribing, over half a million new mobile telecommunications users sign up each day – people talk more and continuously increase their use of data services.

In some countries, such as the Nordic and Western European, the penetration of mobile subscriptions is now well above 80 %.

North America still has some way to go from the present 50 %, while Asia-Pacific has a 12 % penetration. World penetration is currently estimated at 20 %, with growth particularly strong in China, India, South America and Russia.

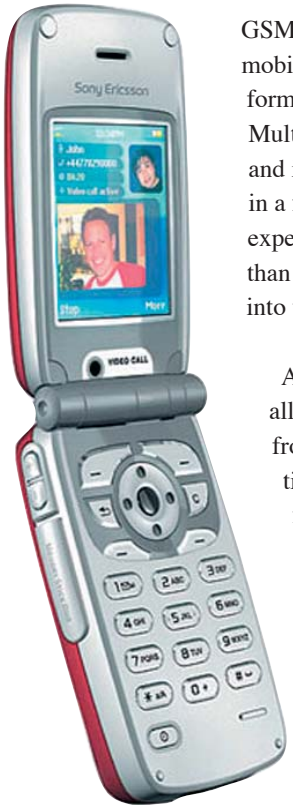
GSM celebrated 1 billion users in early 2004. CDMA and other technologies are being utilised by approximately 300 million users on a global level.

2 The magic of 3G

3G is the short form for the third generation of mobile telephony systems, universally also known as Universal Mobile Telecommunication Systems (UMTS). Technologically, it includes a number of releases (based on certain, time-stamped specification packages) and a family of radio standards, WCDMA and CDMA being the most supported (for further details, see http://www.umtsworld.com/ and http://www.3gpp.org/). The main improvements are related to the system capacity, data speed and multimodality of the user interaction, enabling and supporting mobile multimedia. Other capabilities evolve in parallel, such as the creation of personalized, location-dependent downloadable information and personalization items, games and the introduction of multimedia information services such as mapping and navigation, multimedia messaging and video telephony.

Seamless connectivity, interoperability and roaming between networks, terminals, services and applications are the basic user requirements that must be fully satisfied. These, in combination with ever-smarter mobile devices, ad-hoc networking, services and applications offer considerable potentials for an improved life quality – and some fun.

The main end user issues, in the longer term – beyond the era of early 3G launch problems, are complexity, accessibility and the cost-benefit ratio.



One of the most well regarded 3G terminals on the market

GSM is still the major global mobile platform. Other platforms such as Code Division Multiple Access (CDMA) and its variants have support in a few countries, but it is expected to remain less used than GSM globally, even into the 3G era.

Approximately 94 % of all operator revenues are from voice communication services, 5 % being from text messaging and less than 1 % coming from mobile data services and applications. The low penetration of mobile data service use in GSM and GPRS networks does not exactly provide 3G with a jump-start. However, in certain areas such as Northern Europe, Korea and Japan, mobile data services are beginning to take off. It is worrying, however, that even in Japan 3G, despite a strong growth, is only slowly catching on and, despite the i-mode success, the benefits and airtime of 3G data services remain somewhat uncertain.

While 3G does offer excellent infrastructure features, supports new user interface paradigms and is able to deliver a rich user experience, there is little chance that operators will be able to ask premium prices for services over this new infrastructure.

The feature-richness of mobile networks, systems, services and applications is on the increase, enabled by the on-going convergence between traditional voice and data services, new mobile media applications such as video streaming and games and the migration from second to third generation mobile networks. A key attribute of this evolution is the non-continuous, location-dependent access to and availability of certain services. For example, in 2004, international GPRS roaming between operators is still not always offered to roaming users, thereby limiting their access and influencing established communication patterns. This situation will persist and will be carried over to third generation mobile services (until full national or global coverage is offered), without the need to rely upon fallback access to GSM and GPRS networks.

Additional disturbing aspects expected to have an impact on the end user experience of 3G during the early years are coverage, service presence and continuity issues. Users of mobile communication services have traditionally had continuous access to a set of well-known, well-configured and always available services, at a well-defined cost. This is no longer the case, with the advent of 3G networks and ad-hoc networking. For voice communication, there are no problems foreseen, as 3G falls back to GSM (and GPRS) where there is no coverage. However, as these 2G and 2.5G technologies only support a subset of the 3G services, the users will experience discontinuity of service levels. For example, video calls may be available in densely populated urban areas but not in the countryside, where the fallback solution might be the provision of GSM voice connections. Depending on how well these service level differences are explained and indicated to the user, they will harm the trust placed on and reliability expected from the new technology.

Fixed wireless services are slowly being replaced by mobile services. Unreliable data indicates that between 5 and 7 % of mobile subscribers do not have access to a fixed line any more. The release of higher frequency spectrums in 3G allows for the development of new technologies and applications. Wireless local distribution technologies have shown potential, but there is a lack of solid business models. Global players will be very careful not to 'cannibalise' their existing mobile and fixed broadband services.

In Europe, there are approximately 30 3G systems in 'standby mode', ready for commercial launch and active promotion or already commercially launched. The 3G operator *3* has been most aggressive by means of launch and promotion activities in the UK, Sweden, Denmark, Austria and Italy during 2003. In

ISSN 0085-7130 © Telenor ASA 2004

2004, with the successful fine-tuning of subscription plans and tariffs, a considerable number of users are added monthly. There are also several CDMA-based 3G networks in operation or being launched, mainly in Asia and South America.

However, at the time of writing this, there is still a shortage of well designed and 'mature' 3G handsets and terminals, major manufacturers not shipping beyond-first-generation volume models yet. Motorola, NEC, Nokia and Sony Ericsson have all launched at least one model, but only the second-wave launch is expected to provide acceptable solutions and work around issues such as bulkiness, poor battery life, unreliable network-terminal interworking, handover and cost.

As a consequence, some leading operators have chosen to launch their first 3G offerings as a filed trial with friendly users or relying upon the benefits of a broadband, mobile PC connection, offered by a 3G PC card.

However, it is expected that the real large masses of customers will not be attracted at this early phase due to the inferiority of most handset models (at least compared to 2G/2.5G terminals). However, as major progress is expected with second-generation 3G handsets to be released in 2005, so is the number of customers going for 3G.

From a user's perspective, there are not many major differences between 2.5G and 3G. The most noticeable ones are video calls and video messaging offered in 3G, while the most important one, the parallel availability of voice and data channels, i.e. the multimodal user interaction enabler, is part of a specification package to be supported at a later phase. Additional services support quick downloads of larger software and applications such as games, provide real-time interactivity and multiplayer capabilities and improve and make the user experience more interactive.

The user's main concerns are usually not about technology, but issues such as the user experience, the type of service, coverage, costs, tariffs and terminals.

In theory, 3G offers full global mobility (supported by its multi-radio-mode terminals) with better technical capabilities and user interface technologies but at a higher cost than 2G. Early adopters of 3G services are expected to be less tariff sensitive and quality demanding corporate users and other heavy users of communication services. Many operators expect 3G to cover the user population with the presently highest potential revenues. Data speeds are sufficient to deliver good quality real-time video or downloadable media streaming solutions, but these are not for free – on the contrary: available price plans indicate a cost level of 3–4 NOK for video calls per minute and data traffic at around 15 NOK/Mb, if larger volumes are consumed. This means, to watch Solskjær score a goal in soccer would cost 5–10 NOK and to view *Aktuelt* would cost approximately 50–100 NOK, unless a fixed- or flat-rate, time-based price plan is provided.

As mentioned earlier, 3G provides excellent chances to improve the accessibility to Information and Communication Technologies (ICT), as multimodal user interaction is logically made possible by the capability of 3G systems to offer simultaneous voice and data channel connections (at least at a somewhat later implementation phase). Network operators are preparing to use this opportunity. For example, in the UK, the seven leading operators have, together with the regulator OFCOM, put together the *Mobile industry good practice for service delivery for disabled and elderly customers in the UK* (see http://www.ofcom.org.uk/consumer_guides/telephony_con_guides/gp_guide_eld_dis.pdf). This applies to the GSM world but will most probably contribute much to the accessibility of 3G solutions.

In this short article I have tried to raise, describe and discuss some relevant 3G issues and tried to anticipate the development and needs for improvement in a non-exhaustive way. A hidden ambition of mine was to provide a realistic picture and raise the interest of the reader beyond the often seen populist views found in the popular media.

3G launch in the USA

ISSN 0085-7130 © Telenor ASA 2004

The President of NTT DoCoMo making the first ever 3G call (in 2002, live on national televison)

This article was intended as an appetizer – those ready for a full meal and interested in the latest developments, trends, issues and products, I recommend attending the quality event of the 3G industry, the 3G World Congress, organized by IIR (the 9th 3G World Congress will be held in Hong Kong on November 15–19, 2004 – for details and information, see http://www.3gcongress.com/).

The 3G saga continues and will, most probably, have a real happy end, contributing to making our lives better – and somewhat more fun!

References

GSM Association M-Services Phase II Evolution 3.2.0 Requirements for Q4 '03 to Q1 '04 products. 2003.

UMTS World. May 13, 2004. [online] – URL: http://www.umtsworld.com/

3GPP. May 13, 2004. [online] – URL: http://www.3gpp.org

GSM World. May 13, 2004. [online] – URL: http://www.gsmworld.com/documents/index.shtml

Note: Figures and illustration sources (in order of appearance): Ericsson Photo Library, unknown sources, siemens.com, 3.com, ETSI, sonyericsson.com, vodafone.com, IIIR and additional unknown sources.

For a presentation of the author, turn to page 38.

Development of an ETSI standard spoken command vocabulary for ICT devices and services

BRUNO VON NIMAN

Bruno von Niman

This article describes the development of the new ETSI Standard (ES) 202 076: *'Generic spoken command vocabulary for ICT devices and services'*. Its basic approach focuses on simplifying the learning procedure for all end-users, thereby allowing for the reuse of basic knowledge between different terminal devices and services, leading to a faster and easier adoption of new technologies. The availability of common, basic interactive elements increases the transfer of learning between devices and services and improves the overall usability of the entire interactive mobile environment. Such a transfer becomes even more important in a world of ubiquitous devices and services in a multimodal environment, enabling access for children, the elderly and people with physical or sensory disabilities.

1 Introduction

Telecommunication, converging with information processing, and intersecting with mobility and the Internet, is leading to the development of new interactive applications and services, offering global, universal, inclusive access to all.

A technology enabling the most natural user interaction with these (often complex) systems and services is speech recognition. In recent years, speech recognition has become commercially viable on off-the-shelf devices and services; e.g. devices with telephone functionality (providing the dominant user interface in telecommunication). As the graphical user interface changed the way we interact with personal computers, voice user interfaces are shaping communication.

The results of this effort, an ETSI Standard (ES), will provide useful help to developers, leading to quicker and more consistent and cheaper UI development, addressing all users.

This work was aligned with and sponsored by the European Commission's initiative ***e*Europe**, a program for inclusive deployment of new, important, consumer-oriented technologies (http://europa.eu.int/information_society/eeurope).

2 Standardization of the spoken commands

2.1 Scope of the work

In early 2001, ETSI set up Specialist Task Force (STF) 182 to develop a new ETSI Standard: *Generic spoken command vocabulary for ICT devices and services*. The purpose of the standard is to simplify the learning procedure for end-users and to allow reuse of knowledge between different applications and devices. In all speech-controlled products and services adhering to the standard, one command will always mean the same thing to all users, even across different product or service ranges.

The scope of the work was to specify a minimum set of spoken commands, required to control the generic and most common functions of ICT devices and services that use speaker independent speech recognition. The ES specifies the necessary and most common vocabularies to be supported by ICT devices and services for voice input, including command, control and editing. It is applicable to the functions required for navigation, information retrieval, basic call handling and configuration of preferences. It also addresses the most common telecommunication services.

The ETSI Standard specifies user-tested commands for the languages with the largest number of native speakers in the European Union: English, French, German, Italian and Spanish, as spoken in their respective countries. Future revisions of the ES will hopefully include all European languages, language

 ISSN 0085-7130

versions and ICT commands.

The ES does not cover dialogue design issues, the full range of supplementary telecommunications services, performance related issues, natural spoken numbers covering more than one digit (other than 'double') or speech output.

2.2 Who and how to decide what becomes a standard?

Traditionally, ETSI has relied on experts, their experience and knowledge to standardise telecommunications user interfaces. Faced with three choices, STF 182 could:

- As the experts – decide the most suitable spoken commands, based on what would work best in a speech recogniser;
- Reuse spoken commands that are used in existing products and services (usually in English) and translate them to the other European languages; or
- Embark on user-centred data collection in the countries in our scope.

We chose a methodology combining these elements, with the focus on data collection from native speakers of each language, allowed by the available resource and time plan for the work. However, our domain knowledge and expertise on speech recognisers was still used to 'filter out' responses that cannot reasonably be used, and our knowledge of existing services and products was also applied where relevant.

3 Development methodology and its implementation

The two main perspectives of the methodology for collecting and validating spoken commands are:

- for users, the command words are both intuitive and easy to remember, while
- a speech recognition system requires commands to be easily discriminable.

In the scientific literature, authors such as Book, Goldstein, Guzman and MacDermid, which we have used in preparing the present document, have described several methods. The methodology consists of several distinct steps:

1. *Spontaneous generation of potential command words:* the purpose of this step is to make an inventory of words that humans would intuitively use, given the task that they want to complete.
2. *Confidence rating of the found potential command words:* the purpose of this step is to ensure that the words found in the first step are considered likely to complete the task when a test subject is given the choice to use the word.
3. *Phonetic discrimination:* the purpose of this step is to ensure that command words that can be active simultaneously in a dialogue context can be recognized correctly by the speech recognition system.

There are several ways of performing each step. In the following clauses we will explain the methodologies in further detail.

3.1 Spontaneous generation of command words

In order to ensure that command words for speech recognition enabled devices and services are intuitive, some evidence must be gained as to which word(s) a user would use without prior training or experience. It is not trivial to find this out, because in order to get this kind of information the (potential) user is likely to be primed for certain words or phrases. For instance, if a test is set up where test subjects are explained the task to be performed it is likely that the explanation contains some words that are candidate command words. If, on the other hand, dialogues of users of actual running systems are analysed it is likely that

ISSN 0085-7130 © Telenor ASA 2004

the command words found in these dialogues are the words that the system designer has chosen and that the user has learnt to use.

Two methods allowing for the collection of spontaneous command words from test subjects are presented below. In both methods, test subjects play a vital role. They must be recruited among people who understand the services for which the command words are sought and are familiar with the functionality, but are not actual users of speech-enabled implementations of such services.

First, for all services, the conceptual functionality to be supported must be determined and described (indicated in the second column of the various tables in clause 5 of the present document).

Secondly, for each of the functionalities the test subject must be explained which functionality is meant, without priming the test subject for particular words.

3.1.1 The storyboard method

In this method, described by MacDermid and Goldstein (1996), a professional artist makes a so-called storyboard, a set of illustrations or cartoons, for each function. The test subject is explained the background and shown the illustration (an example is shown in Figure 1), and is asked to say the command she would use in order to activate the shown functionality.

The advantage of this method is that the same storyboard can be used for several different languages, as long as there are limited cultural constraints involved.

The disadvantage is that some functionality might be very difficult to describe pictorially, and that there can be quite a lot of effort necessary from the artist.

3.1.2 Carefully worded descriptions method

In this method, described by Guzman, the functionalities are described textually in a paragraph of text, which is carefully constructed not to use any word, which might possibly be used as a command word.

The advantage of this method is that there is no need for a highly skilled professional for generating the textual descriptions.

The disadvantages are that the descriptions may sometimes turn out to be very clumsily constructed in order to prevent using an obvious command word, and that this effort must be carried out in all languages one wants to conduct the inventory in. Also, these must be carefully developed for each target language.

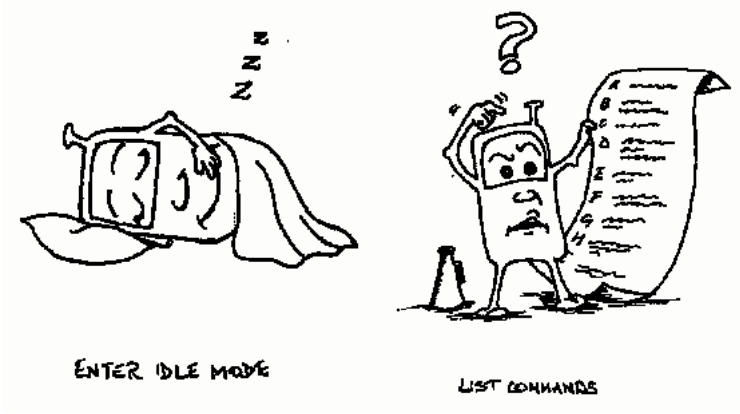


Figure 1 Examples of single picture storyboards, for the commands "List commands" and "Enter idle mode"

3.2 Confidence rating of command words

When the spontaneous words have been generated with a sufficient number of test subjects, a histogram of the word frequencies can be made. This histogram gives a lot of information about the variability in responses. Thus, it can be seen immediately what the most likely command words are. Integration of the sorted frequencies indicates what the coverage of spontaneous commands is, if the most frequent words are available for recognition.

For the confidence-rating test described, one can select the most frequent words that cover at least a given percentage (say, 80 %) of the spontaneously generated words. For instance, for 'confirmation' this might be 'yes', 'sure' and 'no problem', if these are the most frequently generated commands and cover 85 % or the responses for the functionality 'confirmation'.

The procedure described above does not guarantee that the command words imply the targeted functionality. For instance, for functionality 'confirmation' the command phrase 'why not' might have come up in the list of spontaneous commands, but reversibly it might not be obvious to a user that the command 'why not' implies confirmation; it might suggest the inquiry of a reason. A confidence rating of command words tests how likely it is that the given command word implies the correct functionality.

Because the 'spontaneous generation of words' test is an open response experiment, many different expressions for very similar command words can be obtained. Therefore, a manual check of the histogram can be necessary, where responses with the same essential term are grouped together. Thus, for each function, a set of candidates for the confidence test can be obtained.

A way of measuring the confidence of command words is the following, after the work of Guzman. A group of test subjects that are independent of the ones used to generate command words is presented the same data as in the first test, but are requested a different response. The test data can again be either from the 'storyboard' or 'descriptions' method. Instead of asking for a spontaneous command, the candidate commands from the first test are shown. For each of the commands, the test subject are asked how confident they are that the command word will imply the functionality, on a 5-point scale.

Instead of an explicit confidence measure, the subjects can be asked to choose the command word for which they have most confidence.

3.3 Phonetic discrimination

For a voice-enabled application, it is essential that command words are recognized correctly. Although a system can ask for confirmation for certain not undoable operations (e.g. 'delete subscription'), it is not acceptable if almost every command needs confirmation (e.g. "do you want to hear the next item? Please say yes or no.").

For a given application or service there will be several contexts defined in which certain command words will be available, e.g. in the context of 'confirmation question' words like 'yes' and 'no' will be active in the ASR vocabulary. The number of incorrectly recognized commands can be reduced if the available words in a given context are acoustically reasonably different.

There are several ways to test the discriminability. We will assume that for the service or application the ASR contexts are well defined. For a given context, there will be a number of words active.

3.3.1 Speech recognizer field test

One can find out the acoustic discriminability by a field test with a real speech recognition system. This test gives realistic discriminability measures, but the results are sensitive to many chosen parameter setting. Some of these are:

- The type of recognition system (brand, speaker dependency, noise robustness, etc.).

- The test database (recorded speech samples versus live speech from test subjects, speaking style, etc.).

The recording of test databases for voice commands requires quite a lot of effort, but there are several references (e.g. SPEECON, Dialog 2000 and SpeechDat-Car) in which databases for voice commands are gathered.

The test is conducted by preparing the ASR to recognize the required command set for each context, and then test each context with several instances of all the available commands within the context, uttered by many different test subjects.

The *confusability* of a command word *A* with respect to an alternative command word *B* can be defined as the fraction of times an utterance of word *A* is recognized as word *B* by the recognition system. A confusion matrix for each context containing the confusability of all active menu words with respect to each other, can indicate which command words pose particular problems to the recognition system. The *discriminability* of a set of command words is a measure that characterizes the whole confusion matrix.

If the test database consists of recorded speech, the test can be repeated for another ASR system. This will give insight in the recognition system dependency of the discriminability results.

3.3.2 Pronunciation dictionary test

An alternative to a field test is the analysis of the acoustic realizations of the command words. This can be performed without collecting speech databases or test subjects, but the predictions are not validated. The only thing necessary is a *pronunciation dictionary*, a tool that is used often by speech recognition ICT device or service developers. A pronunciation dictionary consists of a lookup table of words in terms of their phone (a separate unit of sound, similar to a phoneme in linguistics) sequences. For instance, the phone sequence for 'yes' may be specified as the sequences 'j eh s' or 'j ea' (where we have introduced a Latin character readable phone symbols 'j' 'eh' 'ea' and 's'). Typically for a Western language phone sets of 40 to 60 phones are defined for ASR systems.

The acoustic discriminability of two command words can be predicted on the basis of the phone sequences of the words. The number of different phones (order is important) might be called the first order prediction of the discriminability. For instance, in the context 'start' 'stop' the number of different phones is 2 for 'stop' and 3 for 'start'. This is a relatively low number compared to the number of phones in the words, respectively 4 and 5. As a contrast, the context 'begin' 'end' has no phones/positions in common, so the number of different phones is 5 and 3, respectively.

A more elaborate scheme takes into account the confusion probability of two phones: e.g. most ASR systems (as well as humans) have difficulties always distinguishing between 'm' and 'n'. For a particular ASR system these phone confusion probabilities may be measured, but this requires a quite elaborate test

set-up of the ASR system. If this information is not available, the phones in a language might be grouped, and the discriminability can be measured in terms of the different phone groups. For example, if 'p' and 't' are in the same phone group (plosives), the words 'top' and 'pot' have all phone group/positions in common, and the predicted discriminability is very low.

For some languages, pronunciation dictionaries are publicly available. However, the complete set of command words in a service or product is limited, and the individual pronunciation of the command words can be found by consulting an expert phonetician. This person can also help in specifying groups of phones that can be considered 'very similar'.

3.3.3 Applying the acoustic discrimination

The discriminability measure can be used to find the optimally performing command words for each recognition context, a complex procedure. An example can help to clarify the procedural difficulties. Suppose, for instance, that in a media browsing application the functions 'move to first message' and 'exit application' are available simultaneously. Suppose further that for the first function the commands 'top' and 'first' come out of the confidence test with preferences 70 % and 30 %, while for the second function the commands 'quit' and 'stop' appear to have preference levels 25 % and 75 %, respectively.

Without paying attention to acoustic discrimination, the command words 'top' and 'stop' would be the preferred ones. If the acoustic discriminability is taken into account, however, which word is going to be replaced by an alternative command with lower subjective preference? The percentages for the alternative words, 30 % and 25 % respectively, appear very similar, and moreover they may not be the only important factor. There might be other words for which discriminability plays a role (e.g. a command word 'quick' with high subjective preference). This means that discriminability optimization is a process that should be applied to the whole menu structure, possibly involving different contexts and even different applications.

It is difficult to formally define a procedure for optimizing the command vocabulary words, because many more factors should then be incorporated such as frequency of occurrence of the commands and likelihood that other applications will be available. A more pragmatic approach to the problem therefore is the following procedure:

a. For each context, start with the command words suggested by the confidence rating test.

b. Find possible pairs of commands that give rise to acoustic discriminability problems.

c. Choose an alternative for one of the command words, with minimum repercussion with respect to confidence rating.

d. Repeat step b) and verify that there are no other commands that clash acoustically with the alternative command word.

e. Repeat step a) to verify that all functions that have new alternative commands do not occur in other contexts or have no acoustic discriminability problems there.

This procedure assumes that there is a relatively low probability that two command words will have low acoustic discriminability.

3.4 Application of the methodology to the ETSI standard

In this clause, the work methodology selected and applied through the development of the ETSI Standard is presented. The specific choices were made on the basis of experience, available resources and expert opinion.

3.4.1 Identification, definition and selection of application areas and key functionality

In the very early phase of the work, key areas of typical ICT device and service functionality were defined, collected, listed and evaluated. Belonging commands were considered, grouped, categorized and reduced to a generic, minimum sub-set of functionality and belonging commands.

The considered input consisted of empirical data, off-the-shelf products, expert knowledge and previous work (mentioned in the ES in Annex B: Bibliography).

3.4.2 Spontaneous command generation test

This test was carried out through an interactive web survey for all considered languages. Subjects were acquired from various sources and either paid for their work or compensated for their efforts in another way (e.g. entering a lottery with mobile phones and wine offered as prizes to be won). The survey was placed on an unpublished web site so that there was control over who took part in the survey. The test leader provided subjects not having web access of their own, web access.

For the functions, carefully worded descriptions were generated. For all of the languages considered in the present document, an expert, knowledgeable in all of the languages, reviewed the descriptions in order to

ensure consistency across the languages. For each description, the subjects were asked to provide a command word or phrase. Optionally, they could give one alternative word or phrase.

The form was submitted to a central processing facility that collected all the subjective data in a database. Responses were first normalized in spelling and responses with the same essence were grouped under the essential word. For each of the commands a frequency histogram was made, weighing the first choice twice as much as the (optional) second choice. Then the two to six commands that contributed to the majority of the responses were selected for the confidence test.

3.4.3 Confidence rating ('multiple choice') test

The confidence was measured by a forced Multiple Choice test. Like the spontaneous generation experiment, the data was collected through an interactive web survey. Subjects were acquired along similar lines and it was verified that the subjects had not participated in the first survey.

The presented commands have been reviewed by the experts and in some cases complemented with other candidate commands in cases where strong evidence suggested these should be included.

For each of the functions in the Multiple Choice test, statistics of the responses were produced. This information was used to perform the acoustical discrimination verification procedure.

The total number of test subjects used in the spontaneous command generation and confidence rating 'Multiple Choice' tests was an impressive 329.

3.4.4 Acoustical discriminability

In order to determine acoustical similarity between two command words, the pronunciation of standard pronunciation lexicons were used. Similarity was based on grouping of phones in similar acoustic-phonetic classes.

3.4.5 Final decisions

The final choice of command vocabulary for each language was based on a joint expert opinion, considering a common weight of the following aspects:

1. The confidence test rating,
2. Evidence in literature,
3. Domain knowledge and personal experience, and
4. The acoustical discriminability, based both on pronunciation and ASR field experience.

First choice		Alternative choice (weight 1/2)	
Frequency	Responded command	Frequency	Responded command
6	Options	2	Menu
5	Menu	2	Help
3	List commands	1	Hi
2	Help	1	What options are available
2	Hello	1	What can I do
1	What options do I have	1	Ready
1	What can I do	1	Options
1	Wait query	1	Menu please
1	Starting up commands	1	Main menu
1	Options please	1	List
1	List options	1	Identify commands
0		1	Hold–what command
0		1	Commands help
0		1	Commands
0		1	Commands menu
0		1	Available commands

Table 1 Raw response histogram data for the command "List all commands and/or functions" (words that have been italicized are later counted as either 'options' or 'commands')

3.5 Example of data collection

As an example, the procedure for determining the command word for the ES function 1.1, 'List commands and/or functions', is given below.

In the web survey, subjects were asked what command word or phrase they would like to say in the following situation:

> *Command 1:*
> *The ICT device or service is waiting for you to say a command but you do not know which commands are available to you.*

Respondents could give two alternative commands for each description. The first suggestion was given a score of 1 and the second suggestion scored ½. Then the total score for each suggestion was calculated, combining suggestions that only have small variations on inspection. Thus, for the first description, the frequencies tabulated in Table 1 were obtained.

Analyses showed that both the command words 'options' and 'command' occurred as an essential word in various phrases in the tail of the histogram. These phrases have been grouped under the essential word, leading to the most frequently used words for the Multiple Choice test: Options (9), List commands (7), Menu (6½), Hello (2), What can I do (2).

The above commands were used as possible responses in the multiple choice survey, and one option was added as a special case because this is often used in commercial PC dictation systems: "What can I say". The accompanying question in the second experiment was:

> *Question 1*
> *'Speak-to-me' is waiting for you to say a command but you do not know which commands are available to you. Which one of the following commands is the most appropriate here?*

The Multiple Choice experiment gave rise to the following statistics, tabulated in Table 2.

Based on the multiple-choice statistics, discriminability evidence and predictions and joint expert opinions, the standardized command was chosen to be 'Options'.

Choice	Percentage answered
What can I say	17.6 %
What can I do	5.9 %
Choices	5,9 %
Options	29.4 %
Menu	0.0 %
List commands	35.3 %

Table 2 Response data for the Multiple Choice test

4 Conclusion, results and outcome

By using this methodology, we feel confident that designers of spoken command-based products and services will want to conform to the ETSI standard. They will know that the standard is based on users' own terminology combined with expert judgement. This will make their products easier to use than if they develop their own commands, based on a lower-level ambition effort, also requiring users to learn a new vocabulary for each new product.

The final results form the basis for the standardized spoken command set and are thereby included in the ETSI Standard itself.

Before approval and publication, ETSI Standards undergo a two-step approval procedure. The present standard was approved on the first level by ETSI TC HF at its 28th Plenary in June, 2002.

The ETSI Membership Voting Procedure followed, a two-month process during which all ETSI members (e.g. Telenor) are provided the chance to cast a weighted vote. STF 182 worked hard to achieve consensus among the different manufacturers, to ensure the new standard is widely acceptable and will be implemented. As a result, the ETSI members were unanimous in their support of ES 202 076. In addition, the standard was delivered on schedule, in time to meet urgent implementation-oriented market needs.

It is our belief that simplifying the learning procedure for end-users will allow for reuse of basic knowledge between different terminal devices and services and lead to a faster and easier adoption of new technologies.

References

ETSI references are available free of charge at www.etsi.org.

ETSI. *Human Factors (HF); Specification of user requirements for use in ETSI Deliverables*. Sophia Antipolis, 2001. (ETSI ES 201 930)

ETSI. *Human Factors (HF); Requirements for assistive technology devices in ICT*. Sophia Antipolis, 2002. (ETSI TR 102 068)

ETSI. *Human Factors (HF); Guidelines for ICT products and services; 'Design for All'*. Sophia Antipolis, 2002. (ETSI EG 202 116)

ITU. *Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network*. Geneva, 1996. (ITU-T Recommendation E.161)

ETSI. *Human Factors (HF); Human factors guidelines for ISDN Terminal equipment design*. Sophia Antipolis, 1994. (ETSI ETR 116)

ETSI. *Human Factors (HF); Guidelines on the multimodality of icons, symbols and pictograms*. Sophia Antipolis, 2002. (ETSI EG 202 048)

ETSI. *Human Factors (HF); Definitions, abbreviations and symbols*. Sophia Antipolis, 1997. (ETSI EG 201 013)

Cohen, M. Universal Commands for Telephony-Based Spoken Language Systems. *SIGCHI Bulletin*, 32 (2), 25–30, 2000.

Guzman, S et al. Determining a set of acoustically discriminable, intuitive command words. *Proceedings of the Applied Voice Input/Output Society (AVIOS)*, San José, USA, 242–250, 2001.

MacDermid, C, Goldstein, M. The Storyboard method : establishing an unbiased vocabulary for keyword and voice command applications. *Adjunct proceedings of Human-Computer Interaction (HCI'96)*, London, UK, 104–109, 1996.

von Niman, B et al. Generic vocabulary for spoken commands. *Proceedings of Human Factors in Telecommunications (HFT'01)*, Bergen, Norway, 305–306, 2001. (ETSI STF 182)

Note 1: The complete ETSI Standard ETSI ES 202 076 may be downloaded free of charge from http://pda.etsi.org/pda.

Note 2: The present paper is based on the ETSI Standard (ES) 202 076, developed by ETSI STF182, consisting of the STF Leader Bruno von Niman and the contracted experts Catriona Chaplin, Jose-Antonio Collado-Vega, Lutz Groh, Scott McGlashan, Walter Mellors, and David van Leeuwen representing Ericsson, Sony Ericsson Teléfonicá, Siemens, PipeBeach, WM Services and TNO Human Factors.

Note 3: Figures and illustration sources (in order of appearance): vonniman consulting, Siemens (used with permission), ETSI ES 202 076 and unknown sources/illustration libraries.

For a presentation of the author, turn to page 38.

ETSI's Universal Communications Identifier (UCI) – from its origins to its diverse benefits

MIKE PLUKE

Mike Pluke

With the expanding range of communications services and with an even greater range of organizations providing those services, the future communications environment is very exciting, but also potentially complex. As people start to use new services they acquire more and more telephone numbers, email addresses, instant messaging identities, etc., and it can be just as hard to contact someone because these different communications identifiers are unknown.

ETSI has looked at this issue and concluded there is a need to introduce a new communications identifier that can be used for all forms of current and future communications instead of this increasing mass of different identifiers. Not only will this new Universal Communications Identifier (UCI) solve the problem of coping with the increasing number of identifiers, it will allow the person you are communicating with to be clearly identified in a way that the user can trust, it will allow users more control of how and when they communicate, and it will help users protect themselves from spam and other online threats whilst at the same time improving the chances that wanted communications will be successful.

1 Introduction

Since 1999, the Technical Committee Human Factors (TC HF) of the European Telecommunications Standards Institute (ETSI) has been considering the issue of the identifiers that users of communications systems use when they communicate. This was seen as a present and developing problem, so ETSI began to describe those problems, identifying what was really required, and then proceeded to describe and elaborate a Universal Communications Identifier (UCI).

This article summarises the main elements of the ETSI analysis of identification issues and its proposed new UCI. As well as a description of what the UCI is, the paper looks at ways in which UCI might work and at its potential to enhance the communications experience of groups of users ranging from the communication intensive 'Road Warrior'" to elderly people, young children and people with disabilities who currently find participation in any form of electronic communication a difficult and sometimes hazardous experience.

2 Communicating with people

A primary assumption behind all of ETSI's work on a UCI is that someone trying to communicate focuses primarily on the person (or work role) they are trying to reach. Having decided whom they wish to reach, the choice of how to communicate (e.g. what service to use) is an independent secondary decision. Substituting general names for the specific name of a person or means of communication in the sentence "I want to phone Mike Miles" clearly illustrates this point. "I want to communicate with Mike Miles" makes perfect sense, whereas the sentence "I want to phone someone" sounds like a strange and very desperate statement and not a rational day-to-day communication choice.

The rationale in the previous paragraph points to the need for any identifier used in communications to relate to an individual or work role and not to any other concepts. This contrasts very strongly with the present situation where identifiers never relate clearly and solely to such person/role identities. Stanford University [1] has clearly described this situation:

> *"People are the outsiders in the current communications revolution. Computer hosts, pagers, and telephones are the addressable entities throughout the Internet and telephony systems."*

Figure 1 illustrates many aspects of this mismatch between current communications identifiers and the true person/role identifier concept.

In Figure 1 there are eight identifiers that have something to do with the imaginary character – 'Mike Miles'. Mike has a fixed phone and a fax at home, both of which are shared with other members of his family. Mike can be reached by phone or fax using the appropriate telephone numbers, but these numbers are not uniquely related to Mike as they also apply to other members of his household. Mike may have exclusive use of his work phone, but the work fax is usually shared. The work phone number currently has a unique relationship to Mike, but if he leaves the company that number will be assigned to somebody else, so the number only relates to Mike temporarily

 ISSN 0085-7130

Figure 1 Who is the 'real' Mike Miles?

in his present work role. The mobile telephone and pager numbers may currently have a unique relationship with Mike, but if Mike changes telephone provider then this number may change if he is unable or unwilling to make use of number portability. Any national renumbering exercise that affects any of Mike's numbers will also result in failure of calls made using the old versions of the numbers. With all these telephone numbers, the relationship between the number and Mike's personal or work role identity are not obvious. Anyone given one of these numbers would be unable to identify the person or role it relates to – there is nothing in the number that indicates the actual identity of its owner.

Mike's email addresses both contain text that relates to the true identity 'Mike Miles'. Whereas Mike's work email address has both of Mike's names, written in the order in which they are normally written, his home email address has a less obvious combination of Mike's family name followed by two initials. Mike could have been forced to use this variant of his name as another 'cheapnet.co.uk' subscriber had already taken 'Mike.Miles' or because 'cheapnet.co.uk' required the user part of the email address to contain only letters and numbers. If Mike Miles decides to change his Internet service provider (ISP), or if he gets a job with another company, the email addresses will no longer be valid and Mike will have to acquire new ones. Also, anyone seeing milesmr@cheapnet.co.uk may assume that this is the email address of the Mike Miles that they know, but there is no guarantee that the person using this email address is really a Mike Miles at all, it could be someone using an alias name to hide their true identity.

So, to sum up, the identifiers in the diagram on the left are all related to Mike Miles, but none of them is an adequate personal identifier. The identifiers either do not name an individual or work role, or they do so in a way that may be unclear and that can't be trusted. Also, each identifier only works with a single service. The telephone numbers are only used for telephony (or possibly SMS), the fax numbers for receiving faxes, the pager number for receiving paging messages, and the email addresses are only of any use for email. Knowing any one of these 'addresses' restricts the ways in which it is possible to contact Mike Miles. If Mike Miles' email address is known, this is of no help at all if there is an urgent need to telephone him. The only way to be fairly sure of contacting Mike Miles is to know all the 'addresses' shown in Figure 1. Even knowing all of them is only likely to be of use for a limited time as, over time, any one of them may change. This will result in the number of failed attempts to reach Mike growing over time unless the stored record of Mike's 'addresses' is updated every time Mike changes one of them.

3 Communicating in the real world and the virtual world

For the vast part of mankind's history communication has been in the 'real world' where interaction has either been face-to-face or by means of written messages carried by some third party between two people. Group communication has similarly been in the form of some public face-to-face communication with an audience or via written communication in some publication.

Over the centuries people have become familiar with the above means of communication. In face-to-face communication, people:

- can visually recognize the other person if they have previously met them;
- can form a judgement about people with whom they are communicating by assessing their appearance, asking for a means of identification, relying on third parties to identify the other person, etc.
- can easily identify if a stranger claims to be a person with whom they are familiar.

Many conventions for resolving potential threats in face-to-face communication have been evolved to ensure that both parties to a communication feel at ease. The identity card that many reputable organizations issue to their employees who call at people's doors are widely accepted and provide reassurance to the person being called upon and also remove potential suspicion from the person making the call. Such mechanisms rely on an identity card issued by an

organisation known to and trusted by the person viewing the card. For postal communication, sealed envelopes and transport by reputable message handlers provide reassurance that a communication has been received in an unaltered state.

It is only very recently in the history of mankind that people have started to use electronic forms of communication. These communications in the 'virtual world' cannot use the same mechanisms to ensure that a communication can be trusted. In the early days of telephone communication, most communication was between people who already knew each other or between an individual and a well-known company. In these circumstances, with the inherent security of telephone networks, most people could be certain that they were speaking to the people that they expected to be speaking to. Today, with the rapid increase of companies that 'cold-call' people, it is not possible to be certain that a call is coming from the person that claims to be calling. Many of these 'cold-calls' involve selling, so it is increasingly likely that telephony fraud is happening because the person being called cannot reliably confirm the identity of the person calling.

In most forms of Internet communication, such as email, instant messaging and chat forums, the communicating persons have much less chance of knowing whether the person they are communicating with is who they claim to be. Unlike telephony, many forms of Internet communication allow the person initiating the communication to state an identity (e.g. in the email address or in the "From" field). However, the mechanisms for ensuring that the claimed identity accurately describes the true identity of the initiator of the communication are all but absent.

For many years, when the Internet was a non-commercial communication mechanism, this uncertainty about personal identities was seen as unthreatening and had been praised as a positive contributor to freedom of speech. Certainly anonymity, or falsified identity, can be of benefit to freedom of speech for people using the Internet in an oppressive society. However, in recent years this unchecked ability to hide or falsify identity has been heavily exploited by people distributing viruses and *spam*[1)] mail. Spam mail also relies on many identity-related mechanisms (e.g. rapidly changing email addresses, mail redirection services, falsified "From" addresses) to both ensure that the messages are accepted by the recipient and to prevent the recipient from blocking further messages from the same originator.

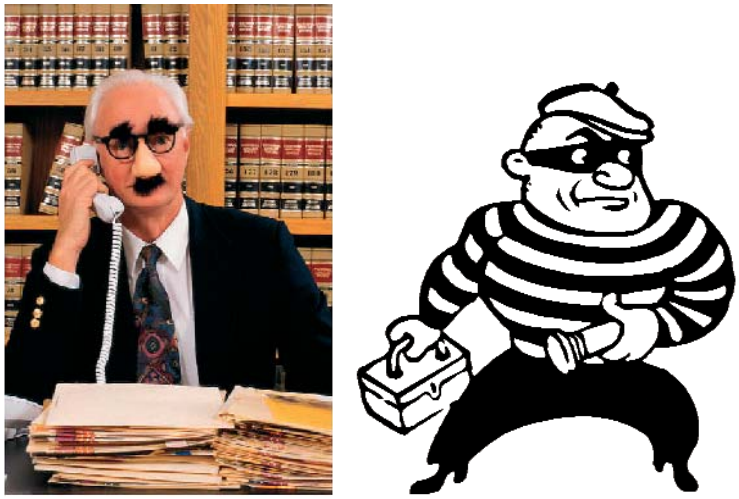


Figure 2 Aliases and anonymity in the 'real world'

If the mechanisms that we take for granted in the 'virtual world' are translated into the 'real world', the absurdity of them becomes very obvious. Accepting communications from an anonymous source is equivalent in the real world to being approached in the street by an individual wearing a mask and totally trusting them. In the 'real world', the fact that someone was wearing a mask would be seen as suspicious – this is not the normal behaviour of a trustworthy person. In the 'real world' it is easy to walk away from a person who you distrust and with whom you do not wish to communicate. The automatic delivery of email into mailboxes makes it difficult to escape this unwanted communication.

Someone masquerading as another person in the 'real world' would also be seen as very suspicious and it could also potentially be an illegal act. In the 'real world', convincingly disguising yourself as another person is difficult and the previously mentioned methods such as identity cards have been introduced to protect people who are unlikely to detect such deceptions. In the 'virtual world' it is impossible for the average Internet user to detect masquerading and thus all Internet users are commonly exposed to threats that would be very rare in the "real world".

There is currently a lot of international publicity about threats due to spam, viruses and the *grooming*[2)] of children by paedophiles masquerading as children.

[1)] spam *usually refers to unsolicited communications that have been very widely distributed in an attempt to derive commercial gain.*

[2)] grooming *usually refers to the attempt to build a (online) relationship with a child in order to gain the child's confidence prior to attempts to sexually abuse the child.*

At the same time, the nuisance from 'cold-calling' over the telephone is increasing. Many of these callers withhold their Calling Line Identity (CLI) and hence they cannot be identified before answering the call. Thus, barring calls from the calling number cannot block future calls from them. All of these threats and nuisances owe much of their success to the weak model of identity on the internet and in telephony. What is needed in the 'virtual world' is a means to allow people to be more sure who is trying to communicate with them and then more able to deal with communications from unwanted sources. If this can be done whilst still allowing anonymous communication and the use of alias identities used in legitimate ways, then a much safer, but at the same time more powerful communications environment, will be created.

4 What is an identifier?

For the purposes of understanding the thinking behind ETSI's Universal Communications Identifier (UCI) it is useful to distinguish two different functions that are part of the general topic 'identifiers'. In the rest of this article these two functions will be described as 'identity descriptors' and 'identity pointers'.

An 'identity descriptor' is a mechanism used to identify the person (or organizational role) that has originated an incoming communication, or to confirm the identity of the person (or organizational role) that has received an outgoing communication. Examples of attempts to fulfil the 'identity descriptor' function are calling line identification (CLIP) and connected line identification (COLP) services and also the 'From' and 'To' fields of email.

An 'identity pointer' is the numeric or alphanumeric string that is entered, or recalled from a store, in order to specify the intended recipient of an outgoing communication. People are using identifiers as 'identity pointers' when they key in a telephone number on a keypad or when they select an email address from a contact list when sending an email.

These two functions have different requirements and much of the early ETSI work was concentrated on identifying these requirements (although this ETSI work did not explicitly distinguish between these two distinct functions of identifiers).

4.1 Identity descriptor requirements

There are a number of essential properties that an identifier needs to possess before it can effectively fulfill the 'identity descriptor' role. The most important properties are listed in the following sections.

4.1.1 Descriptiveness

Requirement:

The form of the identifier should be such that it can clearly describe the entity (person or organisational role) that it represents. Only alphanumeric identifiers can possibly meet this requirement. Rules that place constraints on the form of the name used in the identifier will undermine the ability of the identifier to meet the descriptiveness requirement, e.g. requirements that the same name can only be used once in a naming domain are unsuitable as two or more people with the same name may exist within that domain and they have a legitimate right to use that name.

Current practice:

The 'From' field of an email has the potential to provide a free-text alphanumeric description and thus, it is close to meeting the descriptiveness requirement.

In contrast, the telephone number provided with CLIP (Calling Line Identity Presentation) and COLP (Connected Line Identity Presentation) fails badly in fulfilling this role. Firstly the telephone number is numeric and therefore cannot give a clear description of a person or an organizational role. Secondly a telephone number is usually associated with a telephony service subscription and, as the same person may not always be using the telephone, the telephone number is sometimes a very poor descriptor of a person. Finally, the telephone number is not always made available to the other party because of limitations in the implementation of CLIP and COLP for a particular connection.

4.1.2 Trustworthiness

Requirement:

A person should be able to trust that, when an identifier purports to give an authentic description of a person or role, that description is correct. Any identifier that gives a name that is not a clear and correct name for that person or role should be clearly identified as an 'alias' and not an 'authentic' identifier.

Current practice:

The security features within most telephony networks mean that attempts to misrepresent the number that is calling, and that is shown in CLIP and COLP are rare. However, as telephone numbers fail the descriptiveness test, they cannot be said to be trustworthy identifiers of the person calling or being called. The only time that telephone numbers function as trustworthy identity descriptors is when they are stored in someone's phone book. Then it is possible for the phone to display the name associated with that number when an incoming call from that number arrives. If the number is always used by a single individual,

such as the number of someone's mobile phone, then the name displayed can be trusted by the user to quite a high degree.

With email, neither the email 'From' field, the originating email address, nor the return email address can be trusted. Those sending viruses, Trojan horses and spam know this very well and they find it is easy to falsify any of these, which means that none of them can be trusted.

4.2 Identity pointer requirements

There are a number of essential properties that an identifier needs to have to effectively fulfil the 'identity pointer' role. The most important properties, other than the obvious 'uniqueness', are listed in the following sections.

4.2.2 Ease of use

Requirement:

Any new identifier should be easy for people to use when communicating.

As indicated below, it may be impossible to ensure very high ease of use even where an identifier designer has complete freedom to choose the identifier format. As such, the ease of use may need to be enhanced by putting the identifier into an environment, which minimises the necessity for the user to recall and manually enter identifiers in order to establish communications.

Current practice:

Long numeric identifiers are something that people have difficulty in correctly remembering whilst dialling as has been shown in recent research [2, 3]. Long numbers are likely to be both difficult to dial and hard to memorise because of their lack of meaningfulness. For this reason, many people have seen potential benefits in moving to alphanumeric identifiers [4].

Alphanumeric identifiers might seem to have great advantages because of the potential meaningfulness of the text in an alphanumeric identifier compared to a numeric identifier that has no such inherent meaning. To function effectively, identifiers have to be unique within the context in which they are used which, for communications identifiers, is a global context. To achieve this uniqueness, the sort of simple short identifiers that would be most meaningful need to be augmented with additional information such as domain names. This additional information may be unknown to someone trying to guess or remember an identifier. In addition, many issuers of identifiers, such as Internet Service Providers (ISPs), put constraints on the form that the user name element of an email address can take, something that further divorces the name used from the obvious and easy to remember name that supporters of alphanumeric identifiers often cite. One final limitation of alphanumeric identifiers is that people often remember them phonetically. This can lead to errors when subsequently generating the identifier from recalled memory, as letter substitutions that are phonetically correct but practically wrong can occur. An example of where this can occur on a very well known company name is with Vodafone, where it is very easy for people to type 'vodaphone.com' rather than 'vodafone.com' when entering a Vodafone email address as 'phone' is the most natural spelling for the sound in Europe. Another example is with a small consultancy called 'W M Services' that has a domain name of 'wmserv.com'. This domain name can easily be miswritten as 'wmserve.com', as the 'serv' part has been phonetically encoded by the person trying to recall it as the more natural English word 'serve'.

It is clear that no current identifiers will be perfectly easy to use as both numeric or alphanumeric identifiers have genuine ease of use limitations. For this reason it may be impossible to design an identifier that has very high levels of ease of use built into it.

4.2.2 Service independence

Requirement:

An identifier will only be a good 'identity pointer' if it points to a person or role and it can be used to establish any kind of electronic communication.

Current practice:

Current identifiers are tied to a single communication service subscription, which is normally associated with one, or a very small number of, communication services. Having someone's email address will not allow a voice telephone call to be made to them, nor will a telephone number allow an email to be sent.

The only time a wide choice of electronic communication services will be available is if the person making the communication has a complete and up-to-date set of all of the other person's different service-specific identifiers.

4.2.3 Stability

Requirement:

An identifier will only be a good 'identity pointer' if it always enables communication to be established with the person or role referred to by the pointer.

Current practice:

Both telephony and Internet identifiers are issued to individuals by service providers or to employees by their companies. These identifiers are thus linked to active service subscriptions.

Where telephony number portability is available, and is utilised, it is possible to transfer a telephone number to a telephony service provided by a different service provider. No such arrangements exist to port email addresses between ISPs. Individuals could redirect mail to their own personal domain as a mechanism to detach email addresses from individual service subscriptions. However, this option is relatively expensive and only likely to be known by sophisticated Internet users who also have the understanding necessary to maintain such an arrangement over time.

4.3 What else should an identifier offer?

The above two sections list some of the most fundamental requirements that need to be met to enable the 'identity descriptor' and 'identity pointer' functions to be effectively performed. In practice there are very many other requirements that must be met for the identifier to be of any practical use. Two of these, which have been taken very seriously in the development of the UCI concept, are listed below.

4.3.1 Usable with legacy systems

Requirement:

It is important that it is easy to use any new identifier with existing communications networks and services as it is unrealistic to believe that the fundamental changes will be made to these long-established systems to accommodate the new identifier.

Current practice:

Current communications networks and services are all designed with some form of identifier to identify the end users of those networks and services. The UCI proposals do not attempt in any way to replace these identifiers. Instead, the UCI approach has been to propose an infrastructure to support UCI that overlays all current networks and services. As well as minimising any changes to existing networks and services (limiting changes to those necessary to interface to the UCI infrastructure) such an approach also means that UCI will not be limited to currently existing networks and services (as might be the case if UCI were too tightly integrated with legacy networks and services).

4.3.2 Future proofing

Requirement:

Any new identifier should be introduced on the basis that it will be suitable for its task at any time in the future. This requirement implies that neither the simple passage of time nor the emergence of new communications services or the redundancy of existing communications services should necessitate the replacement or major revision of the new identifier.

Current practice:

As there are no widely used universal identifiers at present, it is not possible to judge whether this requirement has been met. However, the sheer breadth of different types of currently used service specific communication identifier indicates that this requirement has not been met by several of these identifiers or they would have evolved into universal identifiers.

4.3.3 Simplifying a user's communications

Requirement:

The support environment for any new identifier should help users simplify the tasks needed to effectively exploit the benefits brought by the potentially wide range of available communication services. Without care being taken to look at the usability issues associated with exploiting many different means of communication, the multi-service possibilities that a single identifier enables might lead to unwanted complexity for the user.

Current practice:

At present, as a user starts to make use of a new method of communication, there are many management issues that have to be taken into account to allow that communication to be effective. A user currently may have to bar telephone calls from specific numbers, divert telephone calls to different telephones, filter unwanted emails, maintain an address book that lists multiple identifiers for each contact, etc. The mechanisms for exercising these controls are distributed across a number of different services and applications and there are no synergies between these services and applications that enable a control exercised in one place to have an effect in another service or application. What is needed to solve this latter issue is a means of exercising the required controls at one point and related to people and not service specific behaviour.

4.3.4 Other requirements

ETSI documents about UCI list many more requirements of an identifier and the support environment for that identifier than the subset mentioned above. ETSI EG 201 940 [5] lists 39 requirements for an effective identifier and EG 202 067 [6] refines that to 15 very essential requirements specifically related to the concept of a UCI. The requirements listed in these documents have been found to be a useful source by people who have been looking at the issue of identification, even those uninvolved in the UCI concept.

5 The UCI Solution

In [6], ETSI proposed an identifier solution that satisfies all the requirements detailed above as well as the majority of the other requirements identified in [5] and [6]. The solution, called the Universal Communications Identifier (UCI), is made up of three parts. These are:

- a globally unique numeric identifier;
- an alphanumeric label;
- an additional information field (coded information not for direct presentation to users).

Each of these parts is described in more detail in the following sections.

Some of the most notable characteristics of the UCI include:

- it is a unique identifier for a person, role or organization;
- it allows a label to be used as a "user-friendly" name that describes the originator and/or recipient of a communication;
- it allows the originator or recipient of a communication to claim authenticity for their identifier;
- where it is particularly important to claim authenticity, additional procedures can be invoked to make sure that it is not another person using the UCI owner's terminal and thus not the person it seems to be;
- it is independent of services and networks;
- it is independent of communication service provider;
- it can communicate a selection of important additional information such as:
  - flags that mark important properties of the UCI, e.g. that the name element is an "authentic name";
  - the user's preferred language;
  - the user's preferred alphabet;
  - whether the UCI is a business or personal one.

  These may, with the UCI owner's permission, be made available to other people.

5.1 The number

The numeric part of the UCI is globally unique and would be allocated by a trusted authority. In order that people can be sure that a UCI that they have stored will always allow the specified person to be contacted, the numeric element of the UCI must not change with time, even with a change of service provider. This characteristic has been called 'stability'. Thus, when a person changes his/her name (e.g. after marrying) or when a person cancels communications services and adopts new ones, the numeric part of the UCI will always allow the correct person to be reached. Currently, such changes are often the cause of people losing contact with the people stored in their address books.

The numeric part of the UCI is primarily intended for internal use within the infrastructure that supports UCI. People should only have to enter this numeric element into a system manually when they directly enter someone's details into an address book or when they wish to initiate a first contact with someone and they only have access to a written UCI. The system behind UCI should ensure that these manual entry activities form a very small proportion of the user's UCI-based communication activity.

There are a number of options for the range of numbers from which the UCI numeric element is chosen. Several of these options are discussed in detail in [7]. Each option has a number of benefits and disadvantages and the decision on which range(s) of numbers is chosen will be a future commercial and/or regulatory decision.

5.2 The label

It is the alphabetic label that is intended to be the primary benefit for the UCI user. Every UCI-based communication should allow each party to see (or hear) the UCI label of the other party. This label does not need to be unique as uniqueness is handled by the numeric part. For this reason, there are no constraints on the type of name that can appear in the label. The label will therefore permit a person to use their own name and present it in the way that they normally wish it to be seen.

The label field normally contains a person's name, or the name of a business role in a company. However, the label can have three major variants:

- the 'authentic name' – this is one of the UCI's most powerful new benefits as it ensures that the other person knows they are seeing your true name. The 'authentic name' is associated with a certificate, issued by a trusted authority, which certifies that the name used is one that the person is entitled to

use. This corresponds to how a passport or identity document allows a person to prove their identity in a trusted and recognised manner in the 'real world'. If the trusted authority in the 'virtual world' is the same as the organisation that issues identity documents in the 'real world', then it is reasonable to assume that the level of trust could be as high as that currently associated with the use of passports.

- the 'alias name' – any name appearing in the label field that is not backed by the 'authentic name' certificate is automatically classed as an alias name – even if it is a person's true name. The 'alias name' allows people to use friendly nicknames instead of their more formal 'authentic name'.

- anonymous – here the label field is blank. This variant is unlikely to be very frequently used and it is only provided to permit anonymous communication, which is something that is currently possible. As 'alias names' allow people to mask their true identity, there seems to be few cases where being anonymous would be beneficial to both parties in a communication.

The decision about when to use different versions of the label is of great importance in communicating with UCIs, as it can affect the likelihood of the communication being successful (i.e. bringing the result that the sender of the communication desires). When communicating with UCIs, the rules for choosing when to use different versions of the name can map very closely to the way that a person uses names and identities in the 'real world'. This is described further in section 6.5.

5.3 Additional information

The additional information field is the least precisely defined element of the UCI. This is the element that can be enhanced and expanded over time to allow UCI communication to respond to many different types of user requirement.

One element of the additional information field conveys the user's assertion that the UCI label field is authentic, is an alias or is anonymous. The value of this field will determine how the UCI is treated by the PUA (Personal User Agent) of the person receiving the UCI-based communication.

Other additional information field elements that may be required for some communications are:

- whether the UCI is a private or a corporate UCI;

- the order of presentation in the label (e.g. surname, forename, company);

- the preferred language for communications;

- the preferred alphabet for communications;

- the preferred communication mode – speech/text (this would be very useful for people with disabilities and for young children);

- the UCI registration authority.

The data in the additional information field can be encoded in a compact form as it is not intended for direct presentation to users. The information may be used by PUAs (Personal User Agents) when configuring a communication (see section 6 for a description of Personal User Agents). In addition, terminals and applications may interpret the data from the additional information field of a received or stored UCI and present it to a user in an appropriate way.

6 How the UCI is used

For a UCI to be of any use it needs to exist within a specially designed environment that assists UCI users to establish and receive communications and to manage their communications preferences. In the earliest phases of work on UCI, two options for providing such an environment were discounted. These were:

1. Making changes to all existing communications networks and services to enable them to directly support the use of UCIs. Such an option would require extensive modifications to or replacement of many of the protocols, APIs and hardware of these networks and services. Such an approach would be exceedingly costly, potentially disruptive to existing communications and would require every operator of a network or service to upgrade to the new specification simultaneously. It is clear that such an approach could never be achieved.

2. A solution that allows UCI to be supported across all networks and services, based on modifications to an existing communications network or service that has the features necessary to support UCI. Although this approach would remove the need to make wholesale changes to every network or service, it would still disrupt an existing network or service and it presupposes that the chosen network or service would be universally available and universally able to be modified both now and in the future (when a better network or service might cause the chosen network or service to become unpopular and uneconomic).

The approach that was chosen for the UCI support environment was to define abstract entities and capa-

bilities that could potentially be realised by a number of different technical solutions. The merit of this approach is that technical solutions that make use of existing entities could be built and that these might interoperate with other solutions built on different technologies, as long as both solutions fully adhere to the detailed abstract requirements. Such a solution aligns very well with the approach increasingly being taken in the specification of modern communications networks and services (e.g. ETSI TISPAN).

A simplified description of the proposed UCI support environment is shown in Figure 3.

Figure 3 shows the two entities, the Personal User Agent (PUA) and the Service Agent (SA), that are introduced to support communications that use UCIs. These two entities are described in more detail in the following sections.

6.1 The Personal User Agent

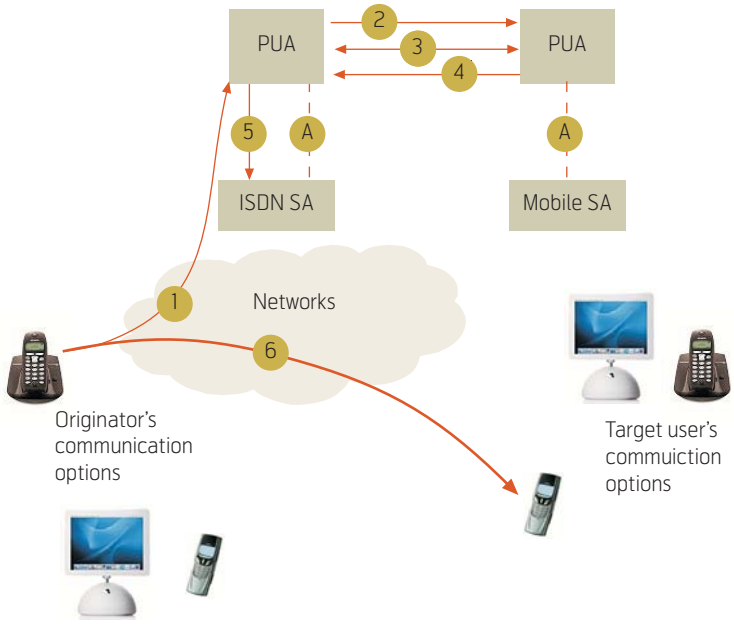
A Personal User Agent (PUA) is a functional entity that has a one-to-one relationship to a specific UCI. It stores, or has access to a user profile that contains information on all the UCI user's preferences about their communication services. The PUA also has access to information on the current status of these services (e.g. 'mobile phone switched on and reachable', 'unable to access home telephone', 'unable to read emails at this time', etc.) via Service Agents (see below).

A PUA only participates in communication with other PUAs, its own user and SAs associated with the user's service registrations. A PUA should never release personal information unless specifically authorized to do so by the UCI user.

6.2 The Service Agent

A Service Agent (SA) is a functional entity linked to a communication service (or network), and would typically be provided by a network or service provider. An SA is the link between the UCI and networks and services and it participates in communication with PUAs, its own network/service and, in a more restricted way, with other SAs. SAs related to a user's networks and services would be specially trusted by that user's PUA.

The SA provides a consistent interface to the PUA irrespective of the internal architecture of its network/service. Where the network/service already has entities that provide a common point of control over the network/service functionality, the SA merely provides an interface to these control functions. Where the network/service does not provide such a common point of control, the SA must also provide additional functionality that interfaces with the distributed control mechanisms within the network/service.



The originator requests a voice call to the target user:

A – Each PUA exchanges information with the SAs of its user's networks/services before, during and after communication attempts take place. The target user's PUA knows that the user's mobile phone is able to receive voice calls.

1 – The originating user enters the UCI of the target user
2 – The originating PUA makes a request to the PUA of the target user
3 – The PUAs negotiate communication options if necessary
4 – The target user's PUA takes account of its user's preferences and proposes the user's mobile phone to receive the call
5 – The originator's PUA instructs the originator's network to set-up the call
6 – A voice call between the originator's ISDN phone and the target user's mobile phone is established

Figure 3 UCI in operation

The SA should never release personal information (such as dialable terminal identifiers) unless specifically authorized by the owner, but it can use this information to expedite the set-up of a communication.

6.3 Entity relationships in a UCI environment

Figure 4 shows the relationships between PUAs, SAs, user roles and terminals. It shows how one user role can have a single PUA that helps the user to manage communication involving a number of terminals that are associated with a range of networks and services. It also shows how SAs are related to a communication service (or network) and that PUAs may be provided by a number of different PUA Provider organisations.

6.4 Privacy and security

Security of the information exchanged during the establishement of a communication based on UCIs was seen as a fundamental requirement of UCI. Similarly, the privacy of a user's personal information and communication preferences was also taken as one of the most fundamental design requirements in the design of the UCI support environment. Therefore the ETSI UCI work stressed the vital importance that:

- an entity claiming to be a UCI user, a PUA, an SA or a network/service can be reliably identified as legitimate and not another entity masquerading as a UCI user, a PUA, an SA or a network/service;
- communications taking place between UCI users, PUAs, SAs and networks/services should use mechanisms that ensure that the communication cannot be harmed or intercepted;
- information stored in PUAs and SAs should only be made available to other entities according to the rules that define the minimum information exchange needs of UCI-based communication or when explicitly permitted by the UCI user.

6.5 Communication power with the UCI

In the 'real world', when a person first meets someone, especially in a business transaction, the person needs to introduce themselves by stating their name and, in some door-to-door selling transactions, by showing an official identification document. In the circumstances of a first-time meeting or cold-call selling in the 'virtual world' of the UCI, the person would use their UCI with an 'authentic label' – their 'authentic UCI'. In most communications in the 'virtual world' it is not possible to look at the other person's eyes and body language to see if they are who they say they are or if they are lying and so the certification that goes with an 'authentic UCI' becomes much more important.

In a first time meeting or a cold-call selling situation in the 'real world', the person being communicated with might decide not to communicate with the originator if that person or organisation did not use a reliable form of identification. In communications using UCIs, many people might use the communications management facilities of UCI to refuse, or treat more cautiously, first-time communications from people or organisations not using 'authentic UCIs' – especially if the organisation was trying to sell something.

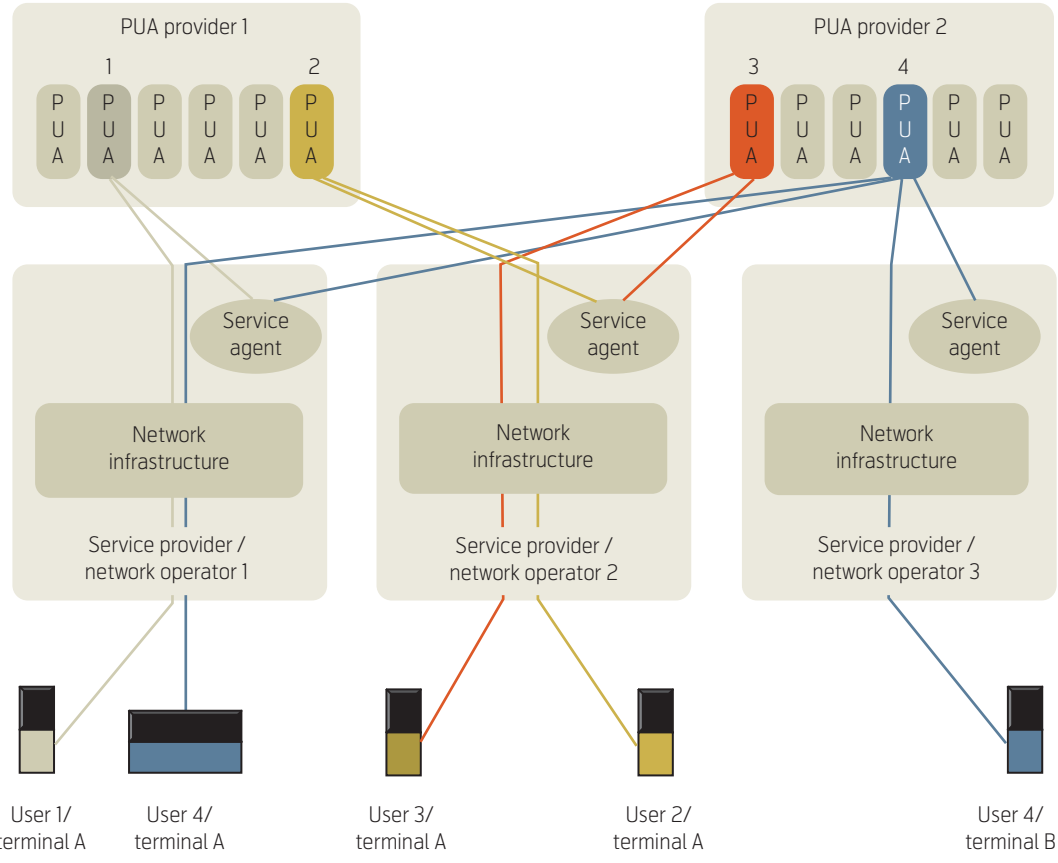


Figure 4 Typical UCI entity relationships

ISSN 0085-7130 © Telenor ASA 2004

In the 'real world', once people have communicated a few times they no longer need to address each other so formally and they do not need to see proof of identity. In these circumstances in the 'virtual world' people may wish to use the more friendly naming style of a UCI with an 'alias name' label field. In earlier communications the recipient may have received an 'authentic UCI' and this may have been stored in an address book. When the same person communicates again using a UCI with an 'alias name', the recipient's Personal User Agent (PUA) will be able to identify the 'authentic' and 'alias' variants as representing the same person or role, as the UCIs will both share the same numeric element. The communication will thus be recognised as a repeat one from a trusted source and not a first time communication that fails to use an 'authentic UCI'.

6.6 Using UCI to keep control of communications

People who subscribe to a communications service initially have almost full control of their communications privacy. Only the service supplier knows the communications identifier (e.g. the telephone number or the email address) associated with that service. At this stage, only the service supplier has the ability to send a communication to the subscriber – but they can do so whenever they choose.

With only the service supplier knowing the subscriber's communication address nobody else will be able to contact that subscriber. Subscribers will never receive the communications they want without giving their communications addresses to others. However, in conventional communications systems, once people have given their communication addresses to others, they have lost all control over how people contact them.

There are at least three ways in which most people enable others to communicate with them:

1. Giving their communications address to specific people who they want to be able to reach them (e.g. giving away a business card).

2. Attaching a communication address to individual communications (e.g. Calling Line Identity and 'From' addresses on emails).

3. The listing of communication addresses in public directories.

Anyone who has accessed a person's communication addresses via any of these means can then attempt to contact that person.

Currently the only way for people to manage receipt of communications from people who have their communication identifier is to use some form of management service (e.g. selective call barring or selective call diversion supplementary services or application specific services such as email filtering). A person trying to contact someone who uses these control mechanisms can easily overcome them by using an alternative identity (e.g. simply using another telephone or a different email address).

Throughout the evolution of UCI, giving users maximum control of their privacy has been a key aim. The elements of the UCI architecture that contribute to giving users fine-grain control over their privacy are:

- User identification – The UCI is expressly designed to provide reliable identification of a person (or role). This identification forms the basis for the UCI user to control who can contact them (communication privacy) and who can see selected personal information (information privacy). Most other identification systems only identify terminals or service subscriptions – not people.

- User control – The PUA contains rules that the user can modify in order to give very fine-grain control over their information and communication privacy.

- Information location – A principle that has been applied throughout the design of UCI is to try to store information where it is used rather than in entities that may have lower security than the original environment.

- In the design of UCI information flows, care has been taken to ensure that only the minimum information necessary to achieve an outcome is passed to the entity responsible for achieving that outcome.

- UCI search – Users would have the option of allowing their UCI to be found in any UCI search. There are various options of how such searches could be supported:

  - 'Classic' directory mechanism – Here the options available to the UCI owner are whether to allow their UCI to be listed in the directory or not. It is hoped that, as the UCI gives the user much greater control over their incoming communications, more people will allow their identifier to be listed than at present. Many people currently will not allow their telephone numbers to be listed in telephone directories for fear of receiving pestering calls they cannot stop, but UCI should help them avoid these difficulties.

- Some form of peer-to-peer mechanism between PUAs (where the search request is propagated to all PUAs or to PUA Providers) – In this case PUA rules can be applied to searches. These rules could allow UCIs to be unconditionally released to people in the UCI owner's address book, and prevent release to those people on a 'blacklist' within the PUA.

- A further level of subtlety for deciding when to release a UCI would be to ask the enquirer to leave a 'virtual calling card' which would reveal some information about them and their reasons for wanting to communicate. Such a mechanism, which in the 'real world' was normally only used to protect the privacy of a minority of 'important' people, could, in the 'virtual world', be available to everybody.

In practice there may be several variants and hybrids of these search options.

7 How the UCI can answer today's worrying problems

Modern communications services provide a rich environment, in which people communicate, learn and entertain themselves. However, as well as these significant benefits, there are a number of ways in which people can suffer from misuse of communications services. There are two abuses of communications services that are increasing rapidly, that cause people a great deal of concern and that have generated a very large amount of sensational publicity (e.g. 'the death of email' and 'children in danger on the internet'). These two problems are:

- the rapid increase in the nuisance caused by spam and viruses;

- the 'grooming' of children in chat rooms (and other means of communications) by paedophiles.

The way in which UCI can be used to help people avoid being affected by these two forms of communication service abuse are described below.

7.1 Spam and viruses

The senders of spam and viruses use a number of mechanisms to maximise the chance that the recipient will read the messages and activate the viruses. Most of the successful methods use falsified identities to persuade the recipient that the message is from a friend or from a reputable organisation. Also, the spam and viruses are sent from email accounts and Internet addresses that are frequently changed and that are made almost unidentifiable and untraceable. Because these techniques are used, even when the recipient realises that they have received a communication that they don't want, they are unable to block further communications from the same sender. The Internet was designed with very open and flexible naming strategies and this has made it easy for people sending spam and viruses to exploit these mechanisms to masquerade under falsified identities and to hide their true identity. In contrast, UCI is built on the need to provide accurate and trusted identities.

Even after the widespread uptake of UCI, existing Internet-based services such as email will continue to operate in the same way that they have previously. UCI will therefore do nothing to eliminate the sending of spam and viruses via email between people and organisations not using UCI. However, UCI offers the opportunity for the senders and receivers of communications that use UCI to adopt different rules of behaviour based on mutual trust of the identity of the sender and receiver.

The expectation when using UCIs is that all reputable UCI users will make initial contact with other UCI users using their authentic UCIs. All unsolicited communications that do not use an authentic UCI would automatically be treated with suspicion by UCI users, and unsolicited communications from commercial UCI users who don't use their authentic UCIs would be treated even less favourably and would probably be automatically deleted by the mail recipients (these senders are the 'virtual world' version of the suspicious characters shown in Figure 2). If UCI users adopted such a strategy, then legitimate commercial organisations that wanted to maximise the chance that their unsolicited mail would be read by the largest number of people would decide to use authentic UCIs with all of their communications.

If this communication behaviour became well established, UCI users would be happier to read commercial communications that used an authentic UCI in the knowledge that the person sending the communication was confident enough to reveal their true identity and that it would be possible to block all further communication from that sender if they wished to. In such an environment, the chaotic spam and virus ridden environment of unmodified internet emailing that currently exists might be seen as a poor substitute for communications using UCI and might begin to be less used.

The UCI-based solution would not need to replace existing Internet email, it could be implemented as an additional step, invisible to the end-user, that would occur prior to the despatch of an email message that the recipient's PUA had agreed that it is willing to

receive (e.g. an ID that represents the agreement between the PUAs could be part of the email header information).

7.2 Protecting children

The recommended behaviour in many chat rooms, particularly those for use by children, is not to reveal your true identity. For this reason, UCI users would not want to use their authentic UCI but would use an alias UCI instead. Using alias UCIs would ensure that the UCI users' true name is not known and the priority on keeping all personal information associated with a UCI totally secure will ensure that no other personal information is made available.

The most serious problem with children's chat rooms is to ensure that all of the people participating are genuinely children. The only way to guarantee that only children participate in the children's chat room is to check the age of each person before that person is allowed to join. Some people might object to having their age checked, but logic dictates that this is a necessary condition to ensure online child safety and would be a decision that would need to be made by anyone wishing to use a safe service.

Most forms of identification that can easily be used online, such as the possession of a credit card, can only help to determine if a person is older than a certain age and not younger. Therefore, currently, there are no easy methods to identify whether an online chat room subscriber is genuinely a child. UCI has the potential to provide such a method. The introduction of an 'authentic date of birth', certified in a similar manner to the 'authentic name', provides a UCI-based mechanism by which age can be checked. It would be straightforward for the 'authentic date of birth' to be certified at the same time and by the same authority as the 'authentic name'.

In joining a 'safe' child chat room; a child would have to agree that its 'authentic date of birth' could be checked by the chat room. Anyone not agreeing to the checking of their 'authentic date of birth' would be prevented from joining the chat room and anyone not meeting the age requirement associated with the chat room would also be excluded.

The legitimacy and security of the chat room's age checking process could be guaranteed by strict adherence to standard UCI processes and procedures. The chat room would have a UCI that indicated the 'authentic name' of the organisation running the chat room and the checking mechanism would be performed in a safe and secure manner by the PUAs of the child and the chat room. The concept of age verification of communications with children could be extended beyond chat rooms to other modes of communication once UCI-based communications and the 'authentic date of birth' concept have been accepted. For example, it might be possible to request an age-check on the sender of an email received by a child if the person's real age is in doubt. A refusal to permit the check (which all UCI users would have by right) or a negative check would be interpreted as a cause for concern.

No services would be allowed to check a person's 'authentic date of birth' without the UCI user's explicit permission. Once the concept of 'authentic date of birth' is established, it is possible to conceive of other services that could benefit from it. Various insurance services and services that provide discounts to elderly people might also be linked to 'authentic date of birth' checking processes. Again, with these services, no person would be forced to have their 'authentic date of birth' checked, but those that refused would be obliged to prove their age by some other means, most probably some form of time-consuming offline method.

8 Conclusion

When some people first encounter the UCI concept, and in particular the concept of the 'authentic UCI', they react very negatively to the concept of an official organisation certifying a person's identity. In some people's mind this gets associated with the misuse of national identity cards by some authoritarian governments. Usage of UCI is as different to such practices as it is possible to imagine.

The UCI is available for those people who wish to use them to enhance their communication experiences. Far from being a tool that is used to oppress people, the UCI offers the opportunity for people to protect themselves from oppression by spammers, virus writers and paedophiles. Unlike compulsory identity cards that people are obliged to use, it is hoped that people will wish to adopt UCIs for the benefits that they recognise that UCI-based communication will bring them.

If commercial organisations recognise how the use of authentic UCIs will differentiate them from the disreputable organisations that indulge in spamming, then they will be keen to adopt UCI as a badge of legitimacy. A trend could rapidly build whereby the only companies not using authentic UCIs will be spammers who would find that attempting to spam with authentic UCIs would lead to their communications being blocked by all UCI users. Businesses will also find that implementing UCI within their own organisations provides them with large benefits. In

particular it will help them to handle their communications and present a consistent view to the outside world in an environment where the organisation changes its internal organisation and personnel on a regular basis.

People should soon begin to realise that having a UCI allows them to identify legitimate companies, protect themselves from virus carrying emails that appear to come from their friends and business contacts and protect their children from online dangers (if the children also have UCIs). As this message becomes clear, it is hoped that the number of UCI users will rapidly increase, and when that happens the full benefits of UCI will be realised.

References

1 Maniatis, P et al. The Mobile People Architecture. *ACM Mobile Computing and Communications Review*, 1 (2), 1999.

2 Nordby, K, Raanaas, R K, Magnussen, S. The expanding telephone number I: Dialling briefly presented multi-digit numbers. *Behaviour & Information Technology*, 21 (1), 27–38, 2002.

3 Raanaas, R K, Nordby, K, Magnussen, S. The Expanding Telephone Number: Immediate Memory for Multiple-digit Numbers. *18th International Symposium on Human Factors in Telecommunication*, Bergen, Norway. November 5–7, 2001.

4 Nordby, K. The Expanding Telephone Number: Users' Needs for a Common Address Format in Future Converging Networks. *17th International Symposium on Human Factors in Telecommunication*, Copenhagen, Denmark. May 4–7, 1999.

5 ETSI. *Human Factors (HF); User identification solutions in converging networks*. Sophia Antipolis, 2001. ETSI EG 201 940.

6 ETSI. *Universal Communications Identifier (UCI); System framework*. Sophia Antipolis, 2002. ETSI EG 202 067.

7 ETSI. *Universal Communications Identifier (UCI); Results of a detailed study into the technical areas for identification harmonization; Recommendations on the UCI for NGN*. Sophia Antipolis, 2003. ETSI EG 203 072.

Mike Pluke founded Castle Consulting Ltd. in 1996 and, since that date, he has been leading and working in ETSI Specialist Task Forces (STFs). Much of this work was funded by the European Commission's eEurope *Programme and since 2001 has included development of the Universal Communications Identifier (UCI) concept.*

Prior to forming Castle Consulting Ltd., Mike worked for BT and, for several years, he was responsible for Human Factors standardisation and for development of user interface style guides.

Apart from his leading role in the evolution of UCI, Mike's current research interest is in cultural localisation of voice interfaces.

Mike.Pluke@castle-consult.com

Designing for all *e*World inhabitants using risk analysis as a design tool

ERIK DAGFINN WISLØFF

Erik Dagfinn Wisløff

To avoid unnecessary costs we need foresight, not hindsight. Operational risk analysis gives us the required foresight, while designing for all broadens our customer base. I will briefly illustrate how we can do both at the same time and at little added cost.

Designing good *e*Solutions for all *e*Citizens is an exciting, but challenging proposition. To an economist it must surely seem viable. The added value in reaching out to all *e*Citizen is obvious, isn't it? And the costs – they won't soar to new heights, will they? – Hah, the hard-core economist would say, – better *not* design for all. It's too risky! Let's do business as usual. Let's focus on our regular user group; the Caucasian Western male, 25–40 years old, highly paid, well educated (preferably holding a degree in computer sciences) individual who needs new *e*Products and *e*Services to simplify his *e*Life. Now that's a Really Good Business Decision. Our hard-core economist has just done away with 99 % of the planet's population. And in so doing, we have moved from designing for all towards designing for the minority.

Please note that I am speaking for myself in this article. I am not writing a popularised project report, I do not have any groundbreaking scientific progress to report, nor do I have any ideas that will turn the *e*World upside down in an *e*Revolution. Instead, I will give you my unsubstantiated thoughts on how to cut costs while improving the overall quality of a service or product by using operational risk management techniques to design for all *e*Citizens. The disclaimer done with, let us get down to *e*Business.

A hypothesis necessary to support this article

I have not met any project or business manager who deliberately wants to exclude profitable user groups by designing the product or service in such a way that it is impossible or cumbersome to use. Yet they manage to do that. The barriers they design are everywhere, and, depending on your viewpoint, those unwanted barriers are either a problem or an opportunity.

An example of a typical barrier is the *e*Commerce web site where one needs to hunt for the clickable link, the web page where the same concept has two different names or the web site where the contrast between the text and the text background is marginal. Oh, and let me mention the web site where the font has been reduced to 10 points or less. If you are really indifferent to your users, you will probably combine all these elements on the same page. Why is it so? I do not think it is done out of malice, I do not think it is a vile conspiracy to make me feel inadequate (though it sometimes does feel that way).

I hypothesize that it is a lack of knowledge or perhaps more precise; a lack of forethought, possibly combined with a lack of knowledge of the different user groups and maybe even lack of understanding of the product that is being designed or implemented. I further suggest that removing obstacle and failure modes will, in the long run, reduce costs. Finally, I suggest that we treat barriers to users as a cost since barriers prevent customers from spending their money in our *e*Shop.

Given this hypothesis, the question then is whether the hypothesis is correct. I postulate that it is.

The question then becomes, how do we rectify the web site's designers' lack of forethought and knowledge about potential problems, before the knowledgeable programmers start hammering out their code and before the costs of changing the design start rising?

My proposed solution

Consider a web site where I pay for my purchases using a credit card, and where I need to surrender private information in order to receive the goods I purchase. Let us assume an on-line *e*Pharmacy. The information I will give the site owner is my credit card details, my address and, of course, my medical status inasmuch as the doctor's prescription divulges this. Let us further assume that the *e*Pharmacy security is outsourced. Thus, I assume the cyber security issues are negligible in this case. I also assume the *e*Pharmacy fulfils national laws and EU-directives that require them to perform a security risk analysis before putting the web site into service[1)]. We should

[1)] *The IT industry is notorious for knowingly ignoring legal issues that require them to design secure and robust products. Assuming that the e*Pharmacy *will live up to the requirements is therefore not realistic.*

ISSN 0085-7130

view the obligation to perform a risk analysis as an opportunity to improve the profit margin over the operational life of the *e*Pharmacy.

My proposal is to include usability and design for all issues in the security risk analysis they are required to perform.

While security officers worry about things impacting on service availability, data integrity and confidentiality, they surprisingly often are interested in how a web site impacts on the user actions, and how a web site is perceived and used by the *e*Community. Our *e*Pharmacy security officer will not be overly concerned with firewalls, routers, anti-virus, denial of service attacks etc because his outsourcing partner is handling these issues. He will probably be much more concerned about information leakage caused by user inactions, web site design flaws, operational issues etc.

Therefore, during the obligatory security risk analysis it is actually a benefit for the security officer that we also address "design for all" and usability issues. There is, I believe, a strong synergy effect between the two subject areas. Designing for all will remove many security threats, while good security will improve the site's reliability and usability and user appeal.

Including usability and design for all issues in a security risk analysis is easy and the extra cost is small, strange as it may sound. All that is required is to expand the risk acceptance criteria, a slight extension to the threat assessment phase and a somewhat wider scope of analysis.

A few not very scientific words about risk

Before outlining how a security risk analysis can assist the *e*Pharmacy, let me briefly explain what operational risk analysis is.

Operational risk is usually understood as being a function of an event, its consequence and its probability. The event in question is often related to the business processes or the business logic, and the root causes are people's actions (or inaction), their (flawed) decisions and the "Acts of God" of the insurance industry. Further, operational risk is usually non-diversifiable. For the *e*Pharmacy it is no consolation that their cyber security has been hardened if a poor web site logic or bad design turns the customers away (the opposite is also true).

Operational risk analysis is a somewhat more clearly defined concept. It is just a structured way of answering the question set "what can fail in our business operation, how can it happen, what will be the consequences, how often do we expect the threat to materialize and what can we do about it?"

The actual work tasks are nothing more than a series of steps designed to answer the question sets. Depending on the requirement, the risk analysis may be purely qualitative or it may involve advanced modelling and serious number crunching.

The interested reader may wish to have a look at the Telenor group framework for operational risk analysis and risk management on the TeleRisk web site. Simplified Risk Analysis and Comprehensive Risk Analysis are part of the TeleRisk platform. They are concrete and easy to use methods that support the identification of design flaws before the design has been finalized and pinpoint some of the pitfalls one would want to avoid. The methods also support requirement analysis and the development of a robust test design, and they work equally well in the later stages of the business process, product or service's life cycle.

Unnecessary costs should not be unavoidable

I will boldly propose that performing a risk analysis – or more to the point – operational risk analysis is in fact rather easy. What is difficult is neither the analysis tasks nor the methods we use. No, the one thing that seems to be difficult is to take the time off from reactive work in order to exercise our minds, thus becoming proactive by choice.

Sometimes I feel the hard-core economists are to blame – costs never incurred due to foresight are not found in any general ledger. If you play the game wisely, you knowingly inflate the operating costs[2)] so that the economists can have a field day, cutting the costs that should never have been there in the first place. There is actually an added benefit to this. The project could finish before schedule, maybe even below target cost. Never mind that the costs incurred are avoidable, not necessary. The added costs are not borne by the design project, so why bother? Why not enter a win-win situation where the project can bask in the glory of using little resources while the economists can have their fun and games cutting the needless costs designed into the product? Never mind the fact that this reduces the firm's net profit.

---

2) *E.g. by hurried design, little testing and sloppy documentation.*

Over time, unnecessary costs will creep almost unnoticed into the operating costs. I am not going to elaborate this any further since I have decided to stick to the initial design phase of the *e*Pharmacy. This is a win-win situation; less writing for me – less reading for you. Moreover, you are not missing the action because the same principles and techniques apply equally in all stages of the life cycle.

Some authors believe that operational risk analysis is more black magic than science. Not so. I leave it to the reader to decide which part I disagree with. Is it the magic, the science or both?

Prerequisites for an efficient and effective operational risk analysis

Let us be clear about one thing. Perhaps the most important prerequisite to operational risk analysis is that you must be willing to shift valuable resources from the reactive and unnecessary work towards going after opportunities. There, I have said it. It is official. Operational risk analysis is not for those unwilling to be pro-active, however one goes about the task.

Understanding the mechanics of an operational risk analysis is beneficial. However, odd as it may seem, it is not really necessary, given that one has access to a framework such as TeleRisk, where tricks of the trade have been built in.

One of the tricks is to identify the intrinsic value of the target of analysis. All you have to do now is to find the things that will reduce this value and then do some post-processing of your findings. But be wary of the accountants and the economists' valuation of the *e*Pharmacy, because their valuation holds little or no meaning in the real world.

Consider the *e*Pharmacy once again. What is the value for the *p*Pharmacy (physical pharmacy) of setting up an *e*Pharmacy? Book value of the server, software and leased lines is probably not the real answer. I don't think you will find a pot of gold by going by the replacement cost of the equipment either. *e*Reputation is probably closer to the bull's eye than a number crunching discounted cash flow what-if-analysis is. Sadly, there is no fixed and correct answer to the question. Finding the True Value is difficult. In my experience, a good operational risk analysis should try to explore a different line of reasoning in the Quest for True Value.

We should consider what the user sees as value in the *e*Pharmacy. We should consider how the *e*Pharmacy adds value to the business of running a *p*Pharmacy. We should state the *e*Pharmacy's main and secondary functions. Following this line of reasoning will more often than not yield Important Insights.

Is it possible that the value of the *e*Pharmacy lies in the possibility of receiving orders around the clock? Maybe the value has something to do with providing customers with a lookup-function so that they can learn more about their prescription drugs? Could it be that the added business value lies in automating order handling, e.g. when the customer places an order, the inventory is checked and if the stock is low an automated order for replenishment is sent to the supplier?

Looking closer at the examples we find that interaction between the *e*Pharmacy and the *r*World (real world) is one of the core added values (in my example). The intelligent reader will by now have deducted that designing for all usually improves this interaction and therefore adds value. Designing for the minority may add *e*Reputation, but will deter *r*Users (real users) from shopping in the *e*Pharmacy.

The rest is comparatively easy. How do we reduce the value of interaction? Well, ah, we deny some user groups access, we use pop-ups and pop-behinds, we ... the list is rather long so I think I will stop now. I trust you see my point: designing for the minority is easy, whereas designing for all (in order to reap Big Dividends) requires us to think before we build the *e*Pharmacy.

Select methods for identifying design flaws

I have been led to believe that fixing problems with a 20/20 hindsight is seldom *e*Profitable. Is it possible to identify design flaws or problems with this design philosophy before the design is finalized? Yes, it is. What we want to look for is the mechanisms that will reduce the value of the target of analysis.

The two risk analysis techniques that spring to mind are HAZOP and FMECA. Detecting inconsistent or inappropriate design, non-standard interaction (both technical and human), timeout issues, and so on, are areas where both HAZOP and FMECA have proved themselves to work, and these issues are within the domain of usability engineering.

HAZOP is an acronym for HAZard and OPerability studies. It is a structured technique for focusing brainstorming sessions where the goal is to identify deviations from normal situations. Basically, the facilitator will present the group with a relevant description of what one knows about the *e*Pharmacy web site – e.g. its users, technology and legal issues, operations and

From http://www.humor911.com

maintenance philosophy etc. Following this, the facilitator will ask a series of pre-defined and highly structured questions that are designed to trigger discussions centered on "what can go wrong". This format has proven well suited to identify deviations during both the design phases and when the system is in production. In our *e*Pharmacy example I would expect usability issues such as ease of navigation, web browser (client side) issues, timeout requirements etc. to be important.

In the HAZOP workshop I would look for technical deviations, e.g. in web browser client set up vs. server requirements, and include user problems, errors and omissions. I might use "unintentional user logout" as one of my guiding questions, and if the ensuing discussion did not touch upon timeout issues, I would follow up with another question "unintentional timeout after logging on". Of course, this illustrates that the quality of a HAZOP depends on the facilitator and on the expertise and active cooperation of the group of people participating.

FMECA is an acronym for Failure Modes, failure Effects and Criticality Analysis. This is more of a tabletop exercise for situations where one wants to address technological issues. Unlike HAZOP, the analysis team usually consists of only a few people and the analysis is not performed in a brainstorming session. Before the FMECA-session, the target of analysis is broken down into its component parts. The goal of the FMECA-session is then to identify the possible failure modes, the failure causes, and how the failure will affect the target of analysis and to determine the consequences of the failure. FMECA is well suited for analysis of technical systems that are thoroughly documented, e.g. in our *e*Pharmacy example we would expect to use structured systems design documentation along with a precise description of the user target groups.

A well prepared and properly executed HAZOP or FMECA will yield interesting results. Even though both methods will identify deviations and both methods will yield robust results, I believe HAZOP to be a better-suited method for identifying usability issues. HAZOP is able to identify composite deviations; works well with redundant systems and it is easy to design a workshop that is forward looking user-focused instead of technically oriented towards a concrete systems design. The fact that most workshop participants find the process intuitive is no drawback. They quite often find the workshop rather interesting since it draws on the participants' combined creative skill set.

It is important to note that HAZOP utilizes our most valuable resource – viz. our brain and our ability to interact with other humans. There is no software that is anywhere near the brain when it comes to Thinking New Thoughts about Problems Not Yet Encountered.

The interested reader will find more information on FMECA and HAZOP in the TeleRisk web site.

"How to" and example results

Let us assume we would like to identify barriers to effective navigation on the *e*Pharmacy web site. During the HAZOP workshop, I would require participation by the web site designer, a marketing representative, a pharmacist, a user representative, (e.g. helpdesk) and systems maintenance. I would also like to add one or two external parties (e.g. competitor, customer, media) but this is usually not possible.

To give you the flavour of the HAZOP-experience, consider the HAZOP-sentence *"Incorrect item selection during browsing"*. The goal is to identify how the user could select the wrong item and briefly to discuss why this would happen and to identify possible remedies.

When I, as the facilitator, feel that the discussion is nearing conclusion, I will move on to my next HAZOP-sentence.

Examples of possible design flaws are *"too small font"* and *"ill chosen colour design"*. A good problem description from a HAZOP could answer journalistic questions such as who, where, what, when and why – e.g. *"blue/yellow colour blind user is not able to distinguish between out of stock and in stock items*

due to colour-only for coding stock status during item selection". Another example could be *"User is unable to locate the correct button due to low contrast between the button and the background"*, or perhaps *"User is unable to understand the web site navigation logic because the* e*Pharmacy design does not consider web browser usability configuration options"*.

Avoiding these Stupid Mistakes improves usability, reduces the need for urgent redesign due to user complaints – and will help keep your friendly security officer in the 'friendly' mode. Ah, I nearly forgot to say that the expected life cycle cost of the *e*Pharmacy will be reduced.

Having successfully failed to find any design flaw – the *e*Pharmacy turned out to be perfect (all IT projects are perfect, are they not?) – we can relax. The customers will enjoy a flawless site and our *e*Pharmacy will be a great success giving us Super Profit. No, I think not. I am hoping to see such a perfect product one day, but it has not happened yet.

The number of problems we identify varies. A successful brainstorming session will normally give us between 20 and 200 design flaws – in a well-designed site the majority of the design flaws will only have minor impacts and low probability. However, the typical web site designed according to the requirements of the hard-core economists by focusing on the minority (young Caucasian adult, computer science degree, money to spend) will have design issues with major impact and a high likelihood of occurring.

What do we do about the flaws wee have found? That's easy. First we identify a set of possible mitigating actions, then we let the business owner decide which actions are to be implemented.

In the previous, rather trivial example the actions could be to test the *e*Pharmacy on small and large screens, validate HTML code and avoid Active-X components.[3)] And we test the *e*Pharmacy with usability features of the web browser activated. That's more often than not an Enlightening Experience – and studying the face of your dedicated web designer during the exercise is rather fun.

As a side note, the *e*Pharmacy is also an interesting case because you should probably not induce the customer to buy more than planned. Consider what would happen if the functionality "other customers who bought your drug also bought these popular drugs ..." is implemented in the *e*Pharmacy. I sincerely hope you are able to see the potential for unwanted medical implications.

Determining consequence and estimating probability

I am not going to sketch out a process for identifying the consequences of the design flaws and the probability of their occurrence. This is described at the TeleRisk web site. I suggest you learn the process by performing an operational risk analysis. Go on, try it, it is not very difficult.

The heart of the matter – cutting costs the easy way

The enlightened business manager will no doubt know that an easy way to improve the Profit Margin is to avoid costs in the first place.

Identifying unnecessary costs before they are built into the *e*Pharmacy is easy if you use tools such as HAZOP. Prioritising design issues is easy once the flaws are identified and the problem potential defined by the flaw's consequence and probability of occurrence.

The hard part lies in convincing the hard-core economist that an operational risk analysis is beneficial even though the up-front costs increase marginally. Not incurring costs means there are no costs to be cut. If cost cutting is what the game is all about, you spoil the game when you avoid unnecessary costs. This, unfortunately, is the hard part – improving the profit margin means spoiling the cost-cutting game, a game we are so comfortably playing.

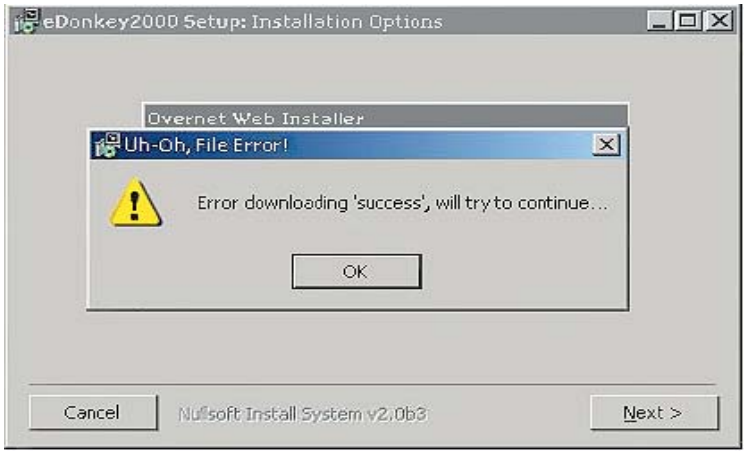
If you want to keep the costs down after the *e*Pharmacy has been rolled out, you could start a program of risk management. That is a bit harder, but still entirely possible. However, I need to warn you, there is a significant downside; you could become proactive, opportunity conscious and cost aversive through risk management – and that is something your colleagues may find disturbing.

We have identified a number of issues. Should we resolve them all? No, that is a bad choice. It is not a good idea to remove all *e*Issues. What we need to do is to select those *e*Issues we want to address and proceed to treat them.

Let me give you an example. While visual clues are important when a customer uses the *e*Pharmacy, we

3) Two good reasons to avoid Active-X: Many security officers are wary of them and, importantly, not all web browsers support them.

ISSN 0085-7130



From Harry Hurt / PC-World

will probably have to accept design flaws limiting blind users' access to the *e*Pharmacy. What I would like to see is higher levels of usability and a more comfortable and pleasing *e*Commerce experience through consciously selecting the risks and flaws one builds into the product.

I suggest that operational risk analysis is one of the best vehicles for identifying flaws and selecting treatment options, while balancing risk and rewards. This will remove barriers, streamline products and provide robust and secure services that benefit the firm, investors, users and ultimately the *e*Community.

Expected benefits

Successfully designing the *e*Pharmacy for the majority, not the minority, should be rewarding both on a personal level (the feel-good factor) and on a business level (the revenue factor). The benefit of avoiding unnecessary costs should be apparent.

However, according to the w3c Web Accessibility Initiative there are some secondary benefits by improving the ePharmacy's accessibility features:

- Increased market share and clientele reach
- Improved efficiency
- Demonstration of social responsibility
- Reduced legal liability.

These so-called secondary benefits are described at length on the w3c web site. Do you think the investors and board of directors of the *e*Pharmacy feel that increased operating efficiency and increased market share are goals worthwhile pursuing in their own right?

I hope you have seen how operational risk analysis can improve a group's understanding of potential problems with a business process or a product. If you have, I hope you also see how you could use this knowledge to avoid operating costs and to broaden your customer base.

I believe that everyone will benefit by removing the design flaws that create obstacles to users. I believe that web sites where one accepts that users have different reading skills have a competitive advantage. I believe that thinking about time-out issues, security issues and the general information design will benefit the *e*Pharmacy both in the short and in the long run. I believe that it is cheaper to invest 10 % extra time and money in the design phase, rather than forking out over the life span of the *e*Pharmacy. I must be mistaken. So many products and services exclude the majority by only caring for the minority.

Parting words

There is no real conclusion to this article. However, you may want to remember that the ultimate vehicle for exploring the frontier of unnecessary costs is the human mind – all we need to do is to unleash our mind's creative powers by using proven techniques for Thinking New Thoughts. HAZOP is such a tool – proven, efficient, effective and fun to use.

Life is an on-going process and we can move from privileged to non-privileged, from minority to majority, at the blink of an eye. Let us hope somebody has thought about the majority in case we suddenly find we are one of them.

You see, sites designed for the minority are surprisingly easy to spot in the *e*World. You may want to compare www.telenor.com with your favourite web site. Just for the fun of it, activate your browser's default accessibility layout[4)] and take part in the majority experience for a while.

Happy *e*Browsing.

[4)] *Internet Explorer users; select tools -> alternatives -> click Accessibility on the general tab lower right corner and check "ignore" in all boxes. Opera users; select view -> style -> user mode -> accessibility layout. Firefox users; give Opera a try.*

Interesting links

Web Accessibility Initiative (WAI):
http://www.w3.org/WAI/

Auxiliary Benefits of Accessible Web Design:
http://www.w3.org/WAI/bcase/benefits.html

A simulator to see how the visually impaired are challenged: http://www.blindeforbundet.no/Filer/simuleringsprogram/sbs_intro.html
http://www.absv.de/sbs/sbs_intro.html

TeleRisk web site:
http://tns-fbu-22-118/FoU/It/infotorg/Sikkerhet_mobilitet/Default.htm (Telenor internal)

Erik Dagfinn Wisløff (43) is a Certified Information Systems Auditor and Certified Information Security Manager, working in the Business Models and Disruptive Changes group of Telenor R&D within the field of operational risk management. Before joining Telenor in 1998 he worked in the Armed Services with communication technology.

erik-dagfinn.wisloff@telenor.com

Information security and human frailty

JAN A. AUDESTAD

Jan A Audestad

The computer industry is always rightfully blamed for not making the computers safe enough; the decision makers in industry and organisations often ignore security in order to make their ICT systems cheaper; the computer systems are so complex that the computer department is likely to make mistakes when the system is installed and later updated; the maintenance personnel is often not trained well enough to protect the systems against malicious attacks. However, the core problem of computer security is that most of the users of computer systems do not have a single clue that there is a problem. Most of us are also easy to fool, giving away the password for a ball pen.

The security of passwords is not the problem. The problem is the way in which people are forced to handle their password simply in order to memorise them. In *e*Commerce, commercial interests are likely to be more important than protecting the electronic payment method. People also tend to believe that they have nothing to hide so that leaving electronic traces everywhere is not a real concern.

E-mail is used more and more in professional and private exchange of information. This makes the interactions efficient but not necessarily secure since the e-mail system is flooded with spam and viruses. E-mail is at the same time good, bad and ugly!

1 The schism of the *e*Society: Ignorance versus knowledge

Society has become ICTated. Almost every activity of society depends on information technology or telecommunications, or both. This applies to our daily life; it applies to how we work, what we do and why we do it; it applies to the way society works and how it is operated; in short, it applies to all activities we associate with the notion of a society.

The ICTation has taken place during the last ten years. Before that, we also had a number of computer applications, and much of society depended on these applications. Thinking back, these applications were few and mostly local. Many of us had personal computers at home but they were not interconnected with other computers. However, it was not the existence of the personal computer that set off the evolution but the capability of simple interconnection of computing devices via the internet, in particular the Web, going commercial around 1995. The Global System for Mobile (GSM) communication commenced operation in 1991. This system only offered interconnection with the packet switched data network and not the internet, and none of us who developed the GSM had the slightest idea that this would change completely within a few years. The evolution was a surprise both to the professional communications engineer and to the layman. This applied also to the data communications engineer: the boards of all telecommunications operating companies had decided that X.25 packet switching was the road to the future. This was what the investments should be used for. Pretty soon this capital became sunk.

Ten years later we all depend on the capabilities offered by the internet. We cannot do without them. During the same time the first generation of people who really master the new technology has emerged. For them the computer is not a bigger mystery than the television set – they handle the computer with agility, without fear and with the disrespect of the juvenile. People my age treat the device with awe – hitting the wrong key may cause it to leap into your face, or even worse, laugh at you!

A friend of mine taught radio communications at an African university some 35 years ago. Several of the students came from the rural regions and had never seen a radio receiver before. For them, radio communications represented an almost insurmountable abstraction – they had no visual picture on which they could hang the ideas. Pure abstraction is hard to handle – it usually requires the brain of a hardcore mathematician to do so. Even theoretical elementary particle physics relates to something that we can depict and is thus less scaring and less abstract, though the mathematics of particle interactions is painstakingly difficult.

We are used to electronic gadgets of all kinds. Therefore, the computer is not a strange device. However, it requires courage to hit the first button on the keyboard!

ISSN 0085-7130 © Telenor ASA 2004

Although there is a new generation of people mastering the computer, there are still a large number of us who are illiterate when it comes to information technology but are forced to use the technology in order to carry out the daily duties we get paid for.

I am a professor of information security and distributed processing. Still I do not know how a data virus is written, what the antivirus program looks like, how I can configure my own computer, or what the different program files on the C disk of my computer do and how they interact in order to help me write this essay. I am half-good at Word and PowerPoint, and find Excel counterintuitive and awkward to use. Without heavy training and tediously studying the manuals I would most likely not pass the Microsoft examination to get a certificate of excellence in Office. Being a professional computer scientist able to prove the Turing stop theorem, I am still illiterate when it comes to the finer details of computing. And here we are at the heart of the problem: computer science has become so complex that none of us, even the most professional, understand all the details of the computer, the communications platform on which they reside and the interaction between computational processes. In such a complex world, anyone can make foolish mistakes that they may sincerely regret later. Also the professionals – I was one of them – swallowed the bait in the e-mail containing the "I love you" bug and activated it. I received afterwards several hundred copies of the bug, all of them from other professionals having swallowed the same bait. There is great comfort in not being alone – but it does not improve security.

Computer illiteracy is therefore much more than total lack of knowledge and training. The problem we are facing has to do with complexity so big that it is not possible for anyone to gain enough overview to avoid errors and faults. Even worse, the field is developing so fast that it is not possible to train the students of computer science in handling all the novelties.

However, there is another more basic illiteracy, and that has to do with the lack of some sort of fundamental knowledge. Such knowledge includes the fact that the industry, the infrastructure of the society, and the media are vulnerable to information attacks that may destroy them. Things become dangerous if the top management of the industry and the politicians ignore the threat and do not provide money and initiate actions to fight the abomination of viruses, worms, Trojan horses and spam that are flooding the network and invading our computers. The reason for this illiteracy is that these people do not realize how big the threat is and the vulnerability of the company or the society if it is hit, and also how weak defence security policies and firewalls are.

There has not been any serious damage so far and the decision maker is easily lulled into the comfort that it has not happened before, ignoring the fact that history is vastly different from the future: the dumbest thing you often do is to base the forecasts for the future on history. In statistical terms: the probability of the past is 1, the probability of the future is 0. Two quotations alluding to this fact are the following.

In her book *The Art of Decision Making* Helga Drummond says, "there is one thing more dangerous than success and that is repeated success". In his personal report to the Congress on the Challenger disaster in 1986, the Nobel laureate physicist Richard Feynman says, "when playing Russian roulette, the fact that the first shot got off safely is little comfort for the next".

With this in mind, let us move to a different arena that may cause havoc to our computer systems.

2 Keeping it simple but unsafe: Passwords and security

If you want to log on to a computer, you will first have to provide a password. Passwords must be memorised. This is the first problem associated with this simple technology. If you use the wrong password you are denied access, and if there is no way in which you can recover the password, you may never be able to access that computer anymore. However, there are companies excelling in accessing computers, in particular hard disks, and recovering information if the computer is destroyed or the password is lost.

If the password is used to encrypt the hard disk also, even these experts have to give up. Therefore, there must be some way in which you can avoid running into such problems. The simplest is that you write the password on a piece of paper you can carry in your wallet. Handling the password in this way is secure as long as no one gets hold of it by stealing your wallet. If you lose your wallet, you may again end up in a situation where you cannot access the computer. One way of avoiding this is to keep the password in several places, making it even easier for adversaries to get hold of it.

Computers connected to the local network of companies can be handled in a different way. The computer management centre can access your computer and alter the password so that you again can use the computer. If the hard disk is encrypted the situation is a little more complex. Then the computer management centre may keep you password or the encryption key

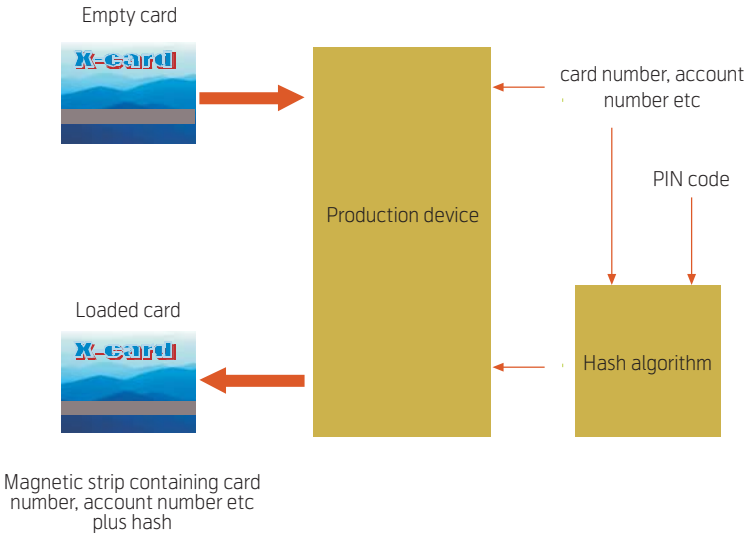


Figure 1 Producing the card

generated by the password in a safe database called a key escrow. If you lose the password, the computer management centre can retrieve the key from the escrow and open the computer.

The hard disk may also be opened by two keys: the key generated by the password you possess and a master key that can be used to open the hard disk of several or all of the computers of the company.

In both cases, your computer is accessed and manipulated remotely. This is, of course, a security threat because an intruder taking command of the computer management centre takes command over all computers of the company – and the company itself. The company thus possesses something that must be protected much more vigorously than any other part of the company.

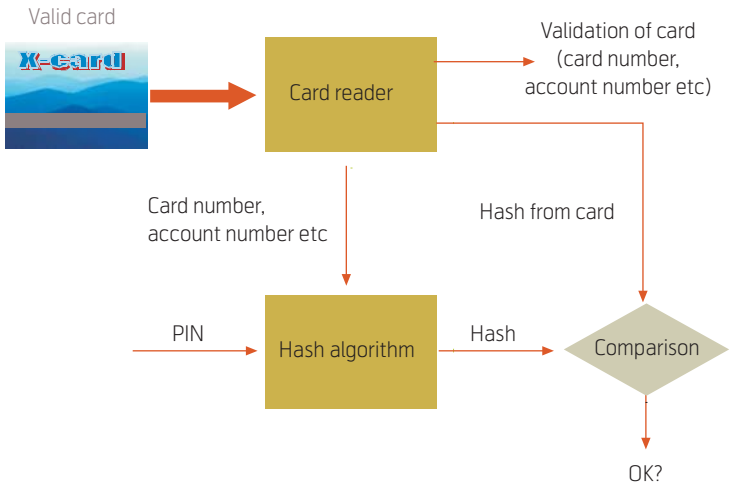


Figure 2 Using the card

This is then how passwords are managed. The next question is how they are made and how easy it is to guess them or steal them.

The simplest passwords are the PIN (Personal Identification Number) code used in the ATMs (Automatic Teller Machines) of the banks and the card readers of shops, hotels and so on. These PIN codes consist of four digits, making a total of 10,000 different codes (or slightly fewer if we subtract what the banks regard not to be random numbers (!): 0000, 1111, 2222 and so on, bringing the concept of randomness into the arena of psychology). The security is based on the fact that if a wrong PIN is used three times, the ATM keeps the card. The card readers in shops also accept only two wrong PINs. However, these machines do not keep the card. The personnel or the machine may raise an alarm.

Let me first explain the security of the PIN code and hopefully dispose of the claim that the technology is the weak security link. The method used for producing cards and verifying them is shown in Figures 1 and 2, respectively.

A card with an empty magnetic strip is inserted into the production device. The card number, the account number and possibly other information such as production time and expiry date are provided to the production device. This information identifies the user account and the issuer of the card that together with the expiry date is used to verify the validity of the card. The information provided to the production device is also provided to a hash algorithm together with the PIN code chosen by the issuer. The hash algorithm produces a bit string that is inserted on the magnetic strip together with the other information. The card number, account number and so on are then available in plaintext on the strip and can be read by anyone possessing a magnetic strip reader.

The card issuer keeps the card data and the PIN code in a safe place. The PIN code must be kept if the same code is to be used on later issues of the card; for example, if renewed after expiry. This is now common practice.

The verification of the card at the ATM is as shown in Figure 2.

The ATM separates the plaintext and the hash from the magnetic strip. When the user produces the PIN code, the ATM takes the plaintext and uses the PIN to produce the hash. The validity of the information on the magnetic strip is then ensured if this hash value is the same as the hash value contained on the magnetic strip.

Validation of the card – card number, account number, expiry data and so on – is then carried out by other local and remote procedures.

Smart cards containing a computer chip instead of the magnetic strip are no more secure in this respect. The PIN code must still be used in order to identify the user to the computer on the card. The only added security is that while the magnetic strip can be copied enabling several clones of the card to be made, the computer chip cannot be copied.

The hash algorithm is such that even if you know the hash value (that is, the string of bits produced by the hash algorithm), the hash algorithm and the plaintext used to produce the hash (card number etc), it is a computationally hard task to compute the PIN code. It is also a hard computational task to find another PIN code that produces the same hash value for the same plaintext. The security is therefore not broken because the algorithms used are weak. The weakness of the method depends on human frailty and the non-technical part of the verification procedure.

Still we may assume that no one possesses an algorithm that inverts the hash algorithm. This can be concluded from the observed patterns of credit card fraud. On the other hand, we do not know whether someone keeping the knowledge secret in order to exploit it under other circumstances knows such algorithms. This is the schizophrenia of information security and is an example of the eternal battle between the cryptographer and the cryptanalyst.

Since there are 10,000 possible PIN values, stealing a card and trying two random PIN codes leaves you with the probability of 1 in 5000 to succeed. Making one or two attempts at several different machines may increase this probability since these attempts are not correlated.

The PIN is only four digits long because it was deemed safe enough at the time the cards were introduced. Since then, it seems as if PIN = 4 digits has become the general formula for generating PIN codes. One reason is, of course, that ATMs and card readers are designed to handle only four digit numbers. The hash algorithm used to verify the card must be standard for all machines accepting this type of cards. The verification method used by all ATMs worldwide is therefore based on the same hash algorithm. In order to improve security the ATMs use several hash algorithms so that the verification check is more complex but still simple for a computer. The task to break the code is a little harder for the cryptanalyst.

In GSM it is specified that the PIN opening the mobile terminal may consist of four to six digits in order to offer higher security. In this case, there are one million possible PIN codes. However, the mobile operators still allocate only four digits, sticking to the old formula. Four digit PINs are also used for access cards for opening doors and for many other security applications.

The owner of the card may use several techniques in order to memorise the PIN codes. The simplest one is to use the same code for all cards. This is possible because the user can choose the PIN code on some types of cards when the card is produced, or the PIN code can be changed later. The user cannot alter the PIN codes on bankcards and credit cards, making these cards less vulnerable to attack. However, it does not preclude the user from using the PIN of the bankcard for all the other cards.

A thief stealing a wallet containing several cards for different purposes may therefore assume that all cards have the same PIN code and can try randomly chosen PIN codes against all cards twice. If there are ten cards in the wallet the thief can safely make 20 trials, increasing the probability of success to 1 in 500. However, PIN codes used for opening doors are not likely to be stored as safely as PIN codes issued by banks and credit card companies. The same may apply to customer loyalty cards, payment cards and identification cards issued by shops and transport companies. The thief may penetrate these databases and get hold of PIN codes that are then used to open the cards used for withdrawal of money and payment of large sums. Another problem is that juveniles are often borrowing GSM phones from one another. In order to do so, they must also disclose the PIN code of the phone. Even if the banks do not allow the customers to choose their own PIN codes, no one can prevent customers from using the code of their credit cards on their mobile phones, so long as the mobile operators allow their customers to select their own PIN codes[2)].

[2)] *It is not mandatory to use PIN codes on mobile phones in order to make the user interface simple, e.g. for older people. This fact and the habit of lending mobile phones to others have made the banks reluctant to accept the mobile phone as an electronic wallet for e*Commerce *and other banking services. However, the use of passwords is not mandatory for accessing personal computers and PDAs either, so where does this leave us? How many people know the password to the personal computer at home? If there is a juvenile member of the family, this is likely to include his or her friends. This may even be the same computer you use at work.*

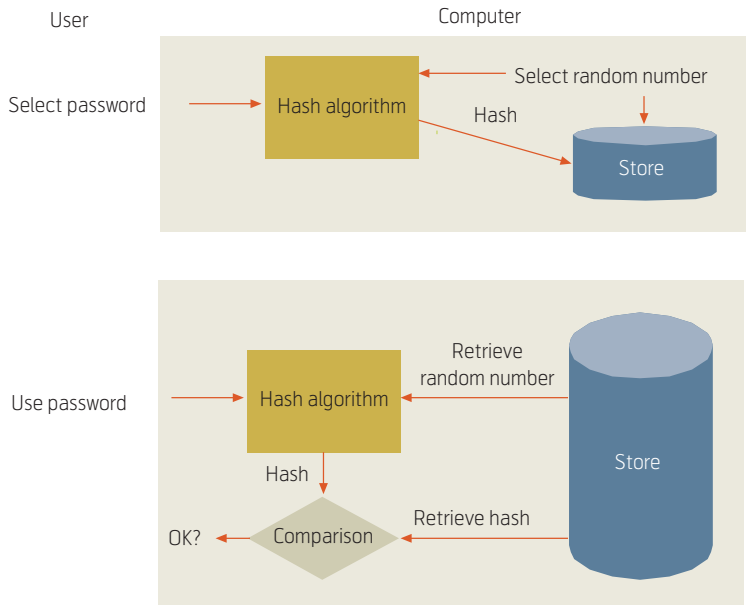


Figure 3 A secure password system

Most people are totally ignorant of these security risks. I am also convinced that companies issuing cards for various purposes are just as ignorant; otherwise they would not allow this practice in order to protect the users against their own ignorance.

Then there are many other ways in which the thief can fool you to give them your PIN code. One simple trick is to call the victim and claim that you are calling from the bank. Somebody has just handed in the card you have lost and the bank needs the PIN code in order to verify the card. The probability is high that the thief will get your PIN code! Most people have not a single clue as to how the bank operates, in particular with regard to stolen cards. Furthermore, most of us still believe that we live in a friendly world without foes.

At a railway station some time ago people were run through a quiz where they also were asked to disclose their computer password. In reward they got a cheap ballpoint pen. Astonishingly many people swallowed the bait and disclosed their password for the ballpoint pen.

As *e*Commerce increases, thieves masquerading as merchants may also fool people to disclose both their credit card number and the PIN code in order to pay for goods. This enables the thief to make several clones of the credit card, sell the clones to other people and also withdraw money himself. This may be a lucrative business. Among the spam I receive, there are several offers of products where they request my credit card number. This number alone is enough to make credit card purchases. Therefore, never respond to such e-mails!

Finally, the long arm of thieves may reach you by modifying the ATM, installing fake ATMs, and using video cameras with powerful lenses to intercept you PIN code. If you in addition throw the receipt in the wastebasket, it is simple play for the thief to produce a faked card.

Then there is the problem of the computer password. Passwords can be stored in three ways: in plaintext, as an un-keyed hash of the password or as a keyed hash of the password. The latter is shown in Figure 3. This method is secure because an intruder either has to steal the password itself, or bypass the hash function and present the stolen hash value to the comparator, or derive the password from the hash function. The latter is, as indicated above, computationally difficult; the second method depends on whether there is such a loophole in the operating system of the computer. The first one is the simplest approach, again based on the fact that though the human brain is a formidable computer when it comes to pattern recognition, managing semantics and other tasks that are meticulously difficult on electronic computers, the brain is a poor device when it comes to remembering random numbers and letter strings.

In order to gain future access to the computer the user selects a password (say, eight random symbols, see below) and the computer generates a random number. The password and the number are fed to a hash algorithm. The random number and the hash result are finally stored in the computer. When the user later accesses the computer by presenting the chosen password, the computer retrieves the random number and calculates the hash result and compares it with the hash result already stored.

This method is secure provided that no trace of the password is stored in plaintext anywhere in the computer. If so, an intruder may find it and then break through the computer security. This puts constraints on the operation of the computer.

Passwords are difficult to memorise and people prefer to choose simple passwords they think are real secrets such as the name of a relative, the name of the place where the family stayed last vacation, the name of well-known persons, the name of a Greek god, the number plate of the car, an uncommon word and so on. People have often approached me claiming that I am unable to guess *their* password. Of course, *I* cannot guess it, but a computer can. The number of passwords selected in this way may amount to, say 200,000 words. It is very hard for the human brain to

run through such a huge list; it is very easy for a computer to do so. Password detection software is based on this principle. It is also based on the curious fact that infinitely many password attempts can be made against a computer: the worst that can happen is that the computer after a number of unsuccessful attempts takes you back to start, which for the computer is "press Ctrl, Alt, Delete in order to continue", starting the procedure all over again. The forte of the hacker is patience, patience and more patience!

The procedure is as simple as this. First, select a list containing first names and run through this list. This list is tiny as viewed by the computer. The probability is high that you will find several passwords of this kind that takes you into a computer. Then you may search over names of gods, and so on. Then you attack the more sophisticated users by running through the words of a standard dictionary – also available on the net. In a short time you will have discovered a large number of passwords. You end up with a few cases where the password cannot be guessed. However, as an intruder you are not concerned with whose machine you penetrate; you are only concerned with finding a weak link that can take you further into the system.

In order to avoid this problem, most systems now require that you select a random password consisting of eight characters: digits, symbols and uppercase and lowercase letters based on certain rules such as the password must contain both uppercase and lowercase letters and at least one digit or symbol. The keyboard consists of 95 symbols. This gives $95^8 = 6.6 \times 10^{15}$ possible passwords. If we only allow letters and numbers, the number of passwords is $62^8 = 2 \times 10^{14}$. These are large numbers even for computers. However, if only lowercase letters are used, the number of possible passwords are $26^8 = 2 \times 10^{11}$. An estimate given in the Handbook (1997) is that it took 1.3 years to run through all passwords of length 8 consisting of only lowercase letters. Now, six years later, this figure is reduced to about 15 days, not because the algorithm is more ingenious but because the computers are this much faster by Moore's law. The corresponding figures for 62 and 95 characters are 44 years and 1300 years, respectively. In comparison, it takes only 1.3 seconds to run through a list consisting of 200,000 words.

Time is working in favour of the cryptanalyst; time is working against the cryptologist!

Moreover, the user must change password once a month and the same password may not occur twice within one year. This is increasing the security – it makes the task to penetrate the system impossible even for supercomputers.

However, the element that is forgotten is the memory of the user. This dynamic system of passwords will require that passwords are written down or certain *patterns* on the keyboard are exploited or both. Such patterns are 1q2W3e4R and >Zxc4321: easy to memorise, easy to test – at least for a computer. Another alternative is to choose Zeus1234 the first month, then change it to Zeus4321 the next month and Zeus2143 the month thereafter and so on. For a computer this sequence is random; for the user and the hacker it is not. How many words of this kind exist? Say that there are 100 names of gods. The numerals 1234 can be chosen in 10,000 ways. This means that 10^6, or a million, passwords of this kind can be chosen. For a human this is a huge number; for a computer this number is tiny: it takes the computer less than 10 seconds to run through the list of passwords of this kind.

Here again we observe a pattern: for a computer any sequence of symbols is random; for the human brain the same sequence may be strongly ordered. But then we may suggest that the password program of the computer is instructed to look for such patterns and not accept any password that appears to be written in this way. This is, of course, possible by simply running through the same lists as the hacker whenever a password is to be accepted. Even if we take this precaution, I am convinced that the human brain will find ways to circumvent even this precaution, making the passwords as insecure as before.

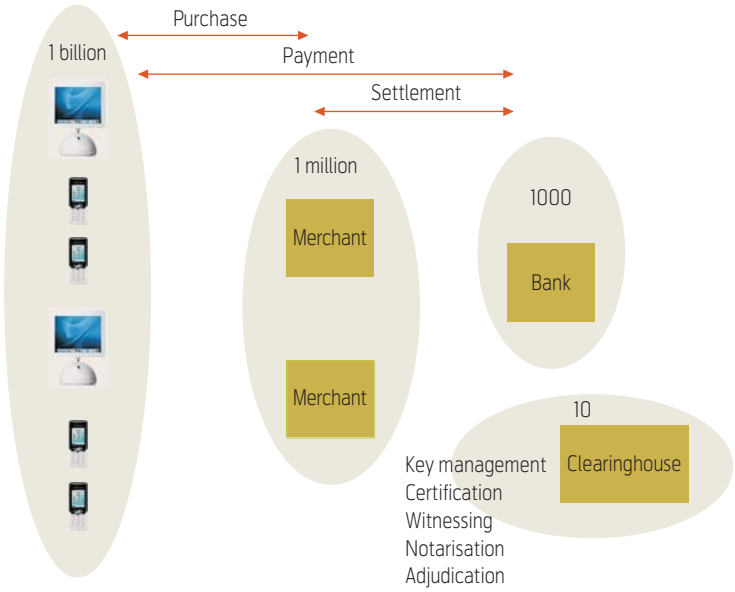


Figure 4 e*Commerce*

ISSN 0085-7130

However, there are still litter bins where people sometimes throw away secrets, there are people who give away secrets out of pure ignorance, and there are people who give away secrets because they get something in return. There are many possibilities open for the distorted minds of people who want to destroy us.

Then you can simply call the computer administration centre and claim that you have lost the password and ask them to give you a new one. In many cases, this will work!

3 Keeping secrets secret

Consumer-to-business (C2B) *e*Commerce is concerned with systems where, say, one billion worldwide customers can access one million *e*Shops worldwide, arrange payment via 1000 banks and assisted by a dozen or so clearinghouses or Trusted Third Parties (TTP). The principle is illustrated in Figure 4.

The purchase takes place between a customer and a merchant and the payment takes place by use of credit cards or electronic wallets. The bank of the customer settles the payment with the merchant and transfers money to the bank of the merchant. Finally, there are clearinghouses supporting secure management of public key systems, certification of the authenticity of customers, merchants and banks, witnessing the proceedings and producing evidence that goods are delivered and paid for, storing evidence of delivery and payment (notarisation), and, finally, resolving disputes (adjudication).

The interactions between customers and merchants are sporadic. These interactions must be secure; that is, there must be mechanisms that ensure that the goods are delivered to the customer and that the payment from the customer is made to the merchant's bank. Moreover, it must be impossible for adversaries to steal goods or money during or after the transaction; it must be impossible for adversaries to masquerade as customers or merchants; and it must be impossible for adversaries to launch denial of service attacks on merchants in order to weaken their brand or take customers away from them. It is a formidable complex task to achieve all this in a configuration as complex as that of Figure 4. The number of parties involved is enough to make the designer shiver.

Payment information, credit card numbers and other sensitive information is protected by encryption when it is transferred between the parties. Public key systems are used to provide digital signatures for authentication, certification and non-repudiation. Non-repudiation is concerned with gathering evidence that certain events have taken place. In *e*Commerce, this implies that the customer cannot deny having received the goods, that the customer can prove having paid for the goods, that the merchant can prove having sent the goods, and that the merchant cannot deny having received the payment.

Purchases must be anonymous in such a way that banks cannot build up databases containing details of all interactions a customer makes. We do not trust the banks that much: governments may easily include them as part of Big Brother if the need should arise.

On the other hand, we can trust that banks and clearinghouses can handle secrets such as encryption keys, certificates and information required for non-repudiation. We cannot trust that the systems used by customers or merchants can store secret information. This is the major problem of versatile *e*Commerce.

Moreover, the customer may do the purchase from any computing device: personal computer, set-top box, PDA, mobile phone and so on. Some of the devices can hide secret keys in tamper-free electronics; other devices cannot. The interaction may take place across any access network and technology: ISDN, digital modem, ADSL, GSM, GPRS, WLAN, Bluetooth, VSAT and so on. An adversary may eavesdrop on some of these accesses since information is exchanged in plaintext; other accesses are protected by encryption making it harder for the adversary to get access to secret information.

We are facing several problems here. I mentioned above that people are lending mobile phones to each other, and also that mobile phones, PDAs and personal computers can be accessed without passwords: this is your choice. When stolen, these devices can be used by anyone. In addition, all these devices, except the mobile phone, may not contain tamper-free electronics where secret keys can be stored safely. The common user of these systems is completely ignorant when it comes to complex types of security measures like this – never having heard of this type of risk.

Moreover, personal computers, PDAs and the latest generation mobile phones (smart phones) can communicate with other devices via infrared and Bluetooth. They can also be connected to your personal computers at home and at work via cable for synchronisation and updating of information and software. These accesses are not protected in the same way as accesses to the internet and can be explored by an adversary to discover a new way into the computer system; this may even take place when you pass someone in the street! We may call it "air-spread" virus in analogy with epidemiology. The virus entering the PDA or smart phone via Bluetooth can then

be spread to the whole system when the device is synchronised with the personal computer. Virus attacks against PDAs and smart phones have already been reported[3]. The owners of these devices do not know that this possibility exists.

Therefore, there is a long way to go until computing devices are designed in a secure manner.

It is not just *e*Commerce that requires secret information stored in tamper-free electronics. The medical record of all individuals in Norway is now being transformed into electronic form. There will then be 4.3 million electronic medical files in Norway. Everyone older than 18 years will have access to their own file and to the files of their children younger than 18 years – this the Government has decided. In addition, there are 40,000 practitioners who must be given access to the files of their patients. Then there are hospitals, specialists, ambulance personnel, psychologists, physiotherapists, pharmacies, social security authorities and so on that must have conditional access to this information. The complexity of *e*Commerce is simple compared to this system.

Handling of the medical files requires secure storage of them, stringent access control minutely defining the read/write access rights of everyone having admission to such files, and that the equipment used to access the file is tamper-free and cannot be taken over by an adversary. The first requirement is simple; the second requirement is complex but manageable. But the third requirement is a dream that can easily turn into a nightmare!

An adversary can take the information in your medical record and in your bank account and put together stories about you that can be used for extortion. An adversary may harm you by changing your medical records so that you receive wrong medication. And there is a lot more an adversary can do to you if security is not properly managed – he can even murder you.

In the hands of authors like Friedrich Dürrenmatt and Max Frisch, the medical record of your boss suddenly dying of a heart attack together with electronic traces you have left at petrol stations, the electronic door locks where you work, the ticket machine at the railway station and everywhere else, may be converted into the most incriminating evidence against you that could send you to jail for the rest of your life.

The problems discussed above are concerned with how secret information, such as encryption keys, is stored in computing devices and how easy it is for an adversary to get hold of this information. A satisfactory solution to these problems should be found *before* complex systems like the above are put into operation. This, unfortunately, is not the case because decision makers are not usually aware of these problems, or even that such problems exist at all. The expert in information security – and in any other complex technology for that matter – often feels like an idiot in the discussion of such things because the ignorants can state their points with absolute confidence while the expert is vague and not able to offer a unique solution. The point is that the wiseacre does not understand that there is a problem; the expert understands that there is a problem and that the problem has not yet been solved. The expert is likely to lose the argument leaving the discussion without honour and dignity. You need a strong psyche to be an expert!

4 Protecting the person: Personal integrity, untraceability and anonymity

Making medical records electronic is an issue that has to do with the protection of personal data. The issue discussed in the previous section had to do with hiding information and encryption keys in such a way that no one gets illegal access to personal information. Let us now look at what is meant by personal information.

Personal integrity means that no one can impersonate another person. This is, of course, closely related to the management of passwords and other information that can be used for secure identification of the person. One such method, which is also hidden under the three letter abbreviation PKI – Public Key Infrastructure – is the use of pairs of keys where one key is public, that is, available to everyone, and the other key is secret known only by the person possessing the pair of keys. The public and secret keys of one person are inverses of each another: a message encrypted by the secret key can be decrypted by applying the public key to the encrypted message; a message encrypted by the public key can be decrypted by applying the secret key. A message encrypted by the secret key can be decrypted by anyone but no one else than the holder of the secret key can have encrypted it. Anyone can encrypt a message using the public key, but the message can only be decrypted by the holder of the secret key.

[3] *New Scientist, 4 October 2003, pp 30–33.*

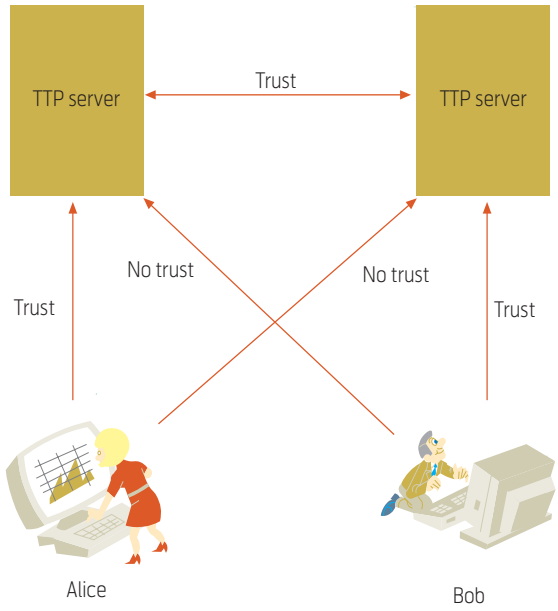


Figure 5 Trust relationships

This looks pretty safe. However, there is a problem. If Alice sends her identity (name), her public key and a message encrypted by her secret key to Bob, Bob can read the message by decrypting the message using the received public key. Bob trusts Alice and knows that it is only she who could have encrypted the message. Or does he? In this scheme, someone, say Cecil, can claim that he is Alice by sending her identity, his own public key and an encrypted message containing a virus destroying Bob's data files. What went wrong is that Bob has no way in which he can make sure that the claimed identity and the public key belong to the same person, in this case Alice. To make sure that impersonation cannot take place the assistance of a trusted third party or TTP is required. Trusted means that Bob trusts that all information Bob receives from the TTP is true, and that nobody can impersonate the TTP. What the TTP does is to provide Bob with information correlating the identity and the public key of Alice. Only Bob can read this information and neither Alice nor Cecil or anyone else can alter it. This is called a (secret key) certificate provided by the TTP.

The questions are then: can Bob really trust the TTP? What if the TTP producing the certificate resides in a different country and Bob does not even know that the TTP exists at all? How can trust then be established with an unknown TTP? *e*Commerce depends on the answer to these questions.

One way of doing this is shown in Figure 5. Alice and Bob may trust different TTPs. If these TTPs trust each other, Bob may trust the certificates of Alice if they are signed by both TTPs. If the certificates are signed only by Alice's TTP, Bob will not accept them. In this way, an international hierarchy of trust can be established. This in itself is difficult, not from a technical point of view, but from a political one. Trust is thus a fragile concept, but it must be the basis for all interactions involving persons where integrity is at risk and where cheating is possible.

Personal integrity also means that unauthorised parties do not get access to information a person does not want to disclose. This has to do with access control. We are not really conscious about how and where data about us is stored, accessed and used. As a matter of fact, we do not care. Data about us appears in a number of databases for a variety of purposes: social security, taxation, communications, banking, employment, shopping, subscriptions, health, medication and so on and so forth. The number of sites containing information about us increases as more electronic services are introduced. Whenever we use a credit card, a piece of information about us is stored. This information can then be combined with information contained in other databases, e.g. which movie we last saw at the hotel, and in this way a record containing our bad and good habits, our health, our economy, our parking and speeding fines, our interests, our sex life, and a lot of other things that we do not like to become public, can be built up.

So indeed we should care!

Almost no one is raising protests against this development. The reason is that we users neither understand that such information exists about us nor that it can be used against us. The claim you often hear is that "I have nothing to hide so it does not matter that someone puts together a record of my life". However, we all have secrets whether we know it or not, and we would become very embarrassed if someone disclosed the inner secrets of our personality. This is just what the psychologist can do if he or she knows enough about us. In this picture we would hardly recognise ourselves.

The archives of possible enemies of the state uncovered in East Germany after the fall of DDR should make us all tremble!

Two other aspects related to personal integrity are *traceability* and *anonymity*. Traceability means that someone can follow the electronic traces we are leaving everywhere and from them deduce all our movements. In GSM, a temporary identity and encryption is used in order to make it difficult for anyone to trace us by monitoring radio traffic. The feature was introduced so that politicians requiring the highest level of protection could safely use the GSM service, and the request for the anonymity service was written on

paper with the logo and watermark of the Palais de l'Élysée. Untraceability and anonymity are serious business!

If we are using WLAN and Bluetooth, we are no longer protected in the same way as in GSM. On the other hand, the protection mechanisms of GSM can also be switched off or bypassed so that Big Brother may still see us and hear us wherever we are.

Anonymity means that we can do things without disclosing our identity. Paying by cash is such an anonymous service. In the electronic world there are fewer and fewer places were we can be anonymous. And this does not worry us! The only ones who are still able to hide their identities are the crooks, the hackers, the extortionists and the terrorists. And this is not the direction in which we want society to develop – if we take the time to sit down and think about it.

5 The good, the bad and the ugly: e-mail, spam and viruses

We are moving towards a society where all interaction between people takes place via e-mail. The postal service is about to become history. We are not there yet, but the dismantling of the postal service has started.

E-mail increases the speed of communication by offering 'quasi-real-time' exchange of information. This means that the information can be sent when it suits the sender and be read and reacted upon when it suits the recipient. In this process there is no transfer delay. This is the way we want communication to take place because real-time communication between people has become more and more difficult: either the initiator of the communication or the recipient are in a meeting or on business travel. Furthermore, large amounts of information can be attached to the mail.

Even SMS (Short Message Services) has become accepted as a means of exchanging small pieces of information. SMS is thus a supplement to e-mail.

E-mail, SMS and web chatting have become the main vehicles for communications between people. E-mail, SMS and web pages are used for communication between the public and the authorities and between businesses and their customers. E-mail is also the major communication channel between companies for exchange of information and contracts.

The e-mail service is not protected by encryption. It is the responsibility of the sender of the information to encrypt the content of the message if secrets are not to be disclosed. Both the origin and the destination of the information are available in plaintext.

Eavesdropping on e-mails may thus be a lucrative business. Often the eavesdropper will find secret information sent in plaintext – it has never occurred to the sender that anyone can gain access to the mail system. This makes industrial espionage simple. If the information is encrypted, the eavesdropper may conclude that this is important information and record the addresses of sender and recipient. This is also industrial espionage because it is sometime enough to know that two parties are exchanging secret information. Making such knowledge public or using it for launching a denial of service attack, may destroy plans and cost much money. Since the address of the sender can be faked, the intruder may even use the knowledge to impersonate one of the parties and send bogus information to the other party.

Eavesdropping on e-mail is not difficult. The protection offered in the e-mail system is not much better than that of a postcard.

The industry and the authorities are then building their communications infrastructure on one of the weakest technologies we have. Money is saved, but what it may cost in the long run is impossible to guess.

Spam is making e-mail even more unattractive as communications infrastructure. About 60 % of all e-mails are spam. Spam causes congestion of the internet. This problem has not been too severe yet, but as the traffic on the network increases spam will, eventually, be the major factor causing overload and loss of information.

Spam also costs the recipient much money. One reason is the time required to remove the spam from the mailbox. Another reason is that in the process of cleaning up the mailbox important information may accidentally be deleted. Spam can be stopped by employing filters, but the filters scanning all e-mails looking for spam causes huge processor load on the mail server, making the e-mail service inefficient and slow. In addition, the spammer is clever to discover ways that circumvents the filter making the problem just as bad as it was. Finally, spam is a vehicle for other computer crime such as launching virus attacks.

Most viruses are spread via e-mail as executable (.exe) attachments. Even if the mail server (or firewall) deletes messages containing .exe files, the virus designer may call the file something else in order to fool the server and the user.

Altogether there are around 70,000 known viruses. These viruses can be stopped by antivirus software. However, the antivirus software must be able to destroy all these viruses because any one of them may still exist in remote corners of the internet coming to life now and then. Viruses are long-lived because of the scale-free structure of the internet. This means that the computer has to check against 70,000 viruses for every e-mail and other input received. This also causes much processor load that could have been used for better purposes.

The virus software cannot detect new viruses. This is hard mathematical fact and the efforts to make such a program (yes, there are people who try!) are similar to constructing the *perpetuum mobile*. Because of the scale-free structure of the internet, a new virus spreads rapidly and most of the network is infected in less than one hour. It takes the antivirus programmer much longer to develop the antivirus program, and thus counteractions can only be taken after almost every corner of the network has been contaminated.

Then we have the fake viruses. These are virus warnings often sent by the system administrator telling you that "your computer is infected by a virus and you have to follow a minutely detailed recipe to restore it". The end result is that you delete everything on the hard disk. Since the virus maker can fake the sender address, he can impersonate being the system administrator. Even if your company (or ISP) does not have any function carrying that name, it seems logical to you that this is a genuine position in the company and you are likely to swallow the bait.

E-mail is a Dr Jekyll and Mr Hyde technology: the benefits are enormous but the threats are even more frightening. Is it just pure ignorance that makes us build our society on this technology? Is it the same kind of psychology that causes us to build some of our biggest cities on geological fracture zones?

6 Where do we stand?

It has taken thousands of years to build society as we know it; that is, the society of the Western civilisation, but it may take only years or even days to destroy this civilisation. Humanity will survive because there is little to be destroyed in the Third World.

We had the nuclear threat during the Cold War, where the philosophy of safety was based on the MAD doctrine: Mutually Assured Destruction. This threat was against all of humanity and most other living creatures. The threat is still there but we do not think about it.

Now there is the information security threat. Society depends on ICT (Information and Communication Technology) – it cannot exist in the way we know it without this technology. The dependency is becoming deeper and deeper and soon we are at a point where destroying the ICT systems is the same as destroying society as we know it in the West. This evolution is like the arrow of time: it can only point in one direction; it can never be reversed.

This dependence on ICT has created new ways of running businesses and the prosperity of the developed countries is escalating. At the same time, society has become more fragile and susceptible to destructive attacks. Society can be attacked by anyone, for any reason and with almost no resources. This is what is coming up.

Implementation of countermeasures against the information war has gone on in the public domain for 25 years, starting with the development of DES (Data Encryption Standard) in 1977 and RSA (Rivets, Shamir and Adleman) in 1978. Before 1977 cryptography was a military pastime. That the information war is a serious threat against society became evident less than ten years ago, and the problem was put on the political agenda in Norway in 2000, and the first organs for keeping an eye on the evolution were established in 2002. Norway is not lagging behind in this area. The situation is much the same everywhere.

It has been reported from the USA, UK and other countries that after the dot.com boom, many companies started saving money by, amongst other things, reducing the information security staff and cutting down on research in this area. The reason is obviously that top managements are not taking the security threats seriously and do not realise that the losses caused by a virus attack may easily surpass the money saved. They are playing Russian roulette without realising that the probability that the chamber contains a bullet next time increases with the number of times the trigger has safely been pulled.

The ignorance concerning information security threats thus permeates all of society. People are sending their credit card numbers unprotected on the internet. Businesses base their commerce on e-mail though the network is flooded by spam and viruses. The industry simplifies the computer networks by reducing security in order to make them cheaper and introduces new flexible business processes allowing personal computers to be transported between protected and unprotected zones. The infrastructures of society cannot work without computers. This is the broad picture we are facing, and we do not know

exactly how to protect all this from being destroyed for reasons of politics, terrorism or self-assertion.

Bibliography

Bishop, M. *Computer Security : Art and Science*. Addison-Wesley, 2003.

Clark, D L. *Enterprise Security : The Manager's Defence Guide*. Addison-Wesley, 2003.

Drummond, H. *The Art of Decision Making : Mirrors of Imagination, Masks of Fate*. John Wiley, 2001.

Mr. Feynman Goes to Washington: Investigating the Space Shuttle Challenger Disaster. In: Feynman, R P. *What Do* You *Care What Other People Think : Further Adventures of a Curious Character*. Harper-Collins, 1992.

Kahn, D. *The Codebreakers*. Macmillan, 1967.

O'Mahony, D, Peircer, M, Tewari, H. *Electronic Payment Systems for E-commerce*. Artech House, 2002.

Menezes, A J, van Oorschot, P C, Vanstone, S A. *Handbook of Applied Cryptography*. CRC Press, 1997.

Pfleger, C P. *Security in Computing*, 2nd Edition. Prentice-Hall International, 1997.

Zhou, J. *Non-repudiation in Electronic Commerce*. Artech House, 2002.

Jan A Audestad (62) is Senior Advisor in Telenor. He is also Adjunct Professor of telematics at the Norwegian University of Science and Technology (NTNU), and he holds a professorship in informatiton security at Gjøvik University College, where his main task has been to build up a new master degree in information security, sponsored, amongst others, by Telenor. He has a Master degree in theoretical physics from NTNU in 1965. He joined Telenor in 1971 after four years in the electronics industry. 1971 – 1995 he did research primarily in satellite systems, mobile systems, intelligent networks and information security. Since 1995 he has worked in the area of business strategy. He has chaired a number of international working groups and research projects standardising and developing maritime satellite systems, GSM and intelligent networks.

jan-arild.audestad@telenor.com

On the mobile, its security issues and applicability potentials

TOR HJALMAR JOHANNESSEN

Tor Hjalmar Johannessen

The mobile, or the cellular phone has gone through a rapid development and deployment in recent years. Its capability to handle an increasing number of services, its mobility, and not least its capability as an identity carrier are elements for its technical success. Its commercial success is also due to the consideration to price policies and manufacturers' focus on design; e.g. youngsters finding it a cool symbol. This article will leave the sociological aspects alone and focus on its technology and the potentials derived thereof.

Introduction

Mobiles today are classified into 2nd generation (2G) and 3rd generation (3G) communication technologies. Typical candidates are the 2G Global System for Mobile Communications (GSM) and 3G Universal Mobile Telecommunication System (UMTS). However, a mobile may also be hybridised to contain other communication technologies like infrared, WLAN and Bluetooth®, so the clean definition is somewhat blurred, especially since these access methods can be combined. Communications, furthermore, can be split into two classes: voice and data.

A central part of the mobile today is the SIM (Subscriber Identification Module; USIM – Universal Subscriber Identity Module – in 3G), which is a smart card chip containing functions and data that are essential and sensitive. Basically, SIM contains the subscriber ID necessary for billing and functions for connecting to the network. As the available SIM capacities increase, also other sensitive functions can be employed like a security infrastructure, e.g. PKI (Public Key Infrastructure), as shown below. If the SIM is removed, replaced or locked by the operator so is all sensitive information for that mobile.

The very backbone of mobile systems is their arrangement into communication cells with a cell diameter varying from typically some few hundred metres up to several kilometres. When switched on, the phone is registered to one particular cell and thus there exists information in the system of the mobile's position down to the accurate cell or better. Since this information is kept in audit logs, it can be and has been used as legal proof in court in criminal cases.

Applications are of increasing importance able to exploit the data communication capabilities. SMS (Short Message Service) represents the basic data service, but with growing data rates and protocols, including Web browsing, *e*Commerce, and MMS (Multimedia Message System). New mobiles have cameras – send a snapshot to a friend or receive one from her. This year 45 million camera phones are reported sold world-wide, but only 30 million digital cameras. The MMS market has grown significantly. Telenor launched the service in 2003 and handled around 10 million MMS before the end of the year. This all makes the mobile an interesting target for the commercial industry. Due to the mobile's increasing multitude of capabilities an increasing interest is observed from the international R&D society; several EU-funded projects are being launched on mobile-related topics.

This article will shed some light on some of the potentials that are possible to implement and achieve using existing technology, and technology available in the very near future.

Some wireless technologies implemented in the mobile now or in the near future

Specific to the 2G GSM mobiles are the basic voice and data capabilities with a datarate of 9.6 kb/s. The data capabilities have been improved by coding technologies and efficient protocols like GPRS (General Packet Radio Service) and EDGE (Enhanced Data Rate for GSM Evolution). Additional wireless technologies like Bluetooth® and Infrared (IR) have been included, making the GSM phone a device with multiple access possibilities. The UMTS mobiles have been deployed recently. Plans also exist to integrate WLAN functionality with the SIM to improve authenciation in WLAN logon.

An indication of some of the available technologies is given in the list below [6,7]:

Mobile:

- 2G: GSM/GPRS/EDGE
- 3G WCDMA/UMTS, CDMA2000 [6]

ISSN 0085-7130 © Telenor ASA 2004

Type		Bandwidth / Data rate (kb/s)	Comment	Coverage range
Mobile 2G	GSM	9.6	Basic	~ Kilometres
	GPRS	8–2048	Peak throughput	As for GSM
	GPRS	22–58.2	Realistic	As for GSM
	EDGE (optimalisation of GSM)	473.6	Maximum	As for GSM
	EDGE	150	Realistic	As for GSM
Mobile 3G	UMTS	2048	Maximum	~ Kilometres
WLAN IEEE 802.11	11b	10000	Peak throughput	~ 100 m
	11b	~ 4000 ref [9]	Typical user data	
	11a/g	54000	Peak throughput	
	11a/g	~ 20,000 ref [9]	Typical user data	
IR		1.2 – 115.2		~ 1 – 5 m
Bluetooth®		1000		10/ 20 /100 m (*)
Modem Cable (RS 232 serial)		256		~ 1 m

Table 1 Some mobile access technologies, bandwidths and coverages. (*) = power class dependent

Wireless LAN:
- IEEE 802.11a, b, e, f, g, h, i ...
- ETSI: HIPERLAN2
- MMAC: HiSWANa/b (Japan)

Wireless Personal Area Networks (WPAN):
- IEEE 802.15x, BlueTooth®
- Home RF / DECT

Other:
- IR

WPAN is intended for the very close (~10 m) ranges, typically indoor, but not necessarily.

The data rates and ranges vary depending on many factors such as the number of users, the distance from connecting antenna and its transmit power, and also with the speed of the user if in motion. For example, if 11 users connect to the same WLAN hotspot, only 1000 kb/s is given to each. Except for IR, which only works within narrow angles and line of sight, the other technologies may operate through walls, although this will attenuate signals. Angle dependency, on the other hand, may reduce wiretapping and increase security.

The offered bandwidth and coverage will put constraints on which services that can be offered, so the figures are important for what the mobile can support, especially when browsing, multimedia or location based services are taken into account.

It also means that while basic 9.6 kb/s is sufficient for SMS, GPRS or EDGE must be employed to offer good quality for browsing and MMS. For live TV or other real-time applications, UMTS or WLAN will be needed.

Products that integrate different technologies have been on the market for some time. Today mobiles that offer several capabilities in the same unit overtake pure GSM phones. Modern mobiles and "Lightweight" PCs like PDAs (Personal Data Assistant) provide accessibilities to several network types making them very flexible, as indicated in Figure 1. In this article, however, "mobile" is used as a common notation for all variants. Note that GPS is only used for receiving positioning data, and not for data transfer in general.

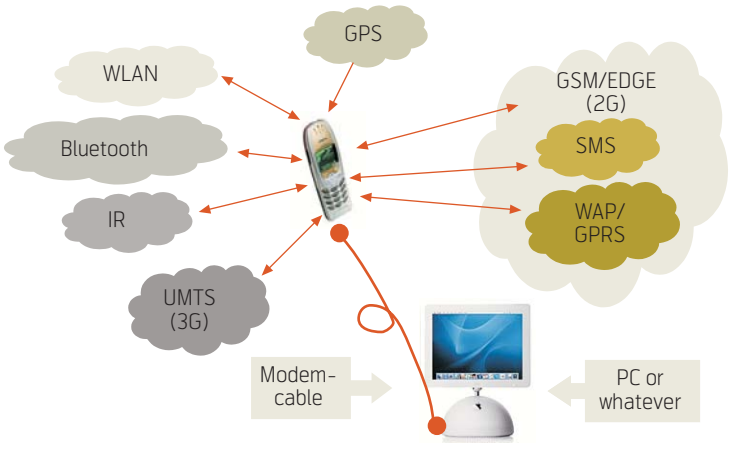


Figure 1 Data accesses for the mobile

ISSN 0085-7130

The (U)SIM as security nucleus of the mobile and identity carrier for the user

In the days of 1G, the old Nordic Mobile Telephone (NMT) times, the subscriber telephone number was hard-coded into the unit itself. With 2G came the SIM, which isolated the functionality for security and sensitive data into a removable chip, actually a smart card with reduced plastics. Today's technology defaults to 32 kbyte RAM, 64 kbyte RAM is available, and larger stores are soon to be expected. The basic sensitive information encompasses the mobile telephone number, the subscriber's identity number, IMSI (International Mobile Subscriber Identity) and the global chip ID, ICCID (Integrated Circuit Card Identifier), which are identifiers of 15 or 20 digits long numbers respectively which are associated with the physical mobile owner. The quality of how these numbers are associated with the owner is of course crucial to their trust. The mobile billing system, which depends on invoicing the correct person, depends on the IMSI, while Telenor's mobile PKI for *m*Commerce applications relies on the ICCID, a number also written on the rear of the SIM.

A person applying for a normal mobile subscription needs to provide proof of identity. In Norway this information is checked against the National Security Number system (Folkeregisteret) before the chip is delivered to the subscriber. Depending on the local control, this is baseline for assurance. Questions can be raised for personal check-up, especially to see whether the registration personnel do their jobs properly. For *m*Commerce necessary codes are delivered to the user through enforced registered mail postal services. One should not overlook that it is also possible to achieve anonymous non-subscription SIMs. Currently, these are quite flexible and can be loaded with a certain number of "cold ticks" (Norwegian 'Ring kontant') and are popular among young people. Anonymous SIMs are also attractive to criminals, but new laws may close this option.

Having the user ID on the SIM implies that a user may easily change his phone and reuse his subscription and other identity numbers simply by transferring his SIM.

Access to the SIM is secured by PIN and PUK. If erroneous attempts exceed a certain value (typically 3) the SIM will be blocked until the more complex PUK is entered on the keypad. This protects the owner if the mobile is stolen in a de-activated (off) state. If the mobile is on when stolen, the owner's only option to avoid misuse is to call the operator and require a blocking of the mobile number (including a revocation of the specific SIM). The thief may afterwards use the mobile, but only with another SIM.

Having an ID strongly associated with the mobile owner on the SIM opens for a large number of services, especially in the *m*Commerce world, where economic liabilities for purchases are important.

- Sensitive functions and data in hardware (SIM) such as PIN, PKI, ID and keys
- PIN: 4 – 8 digit (normally 4) and blocking after 3 erroneous attempts
- PUK: 8 digits
- Separate PINs for telephony and electronic signature
- Under R&D: Biometrics (fingerprint) to replace PIN

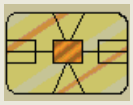


Table 2 Mobile, local security items

PKI on the SIM

The Telenor Mobile *m*Commerce department has enhanced the SIM functionality by implementing an explicit PKI functionality on the SIM. This implementation (in the SIM toolkit) allows for electronic signatures of received SMSs providing a non-repudiation service that connects the user who signs the SMS to the very content of the SMS. The signature can be regarded as legal proof according to EU directive 99/93 and Norwegian law on Electronic Signatures from 2001. Briefly, the PKI services include Qualified Certificates according to ETSI TS 101 456; furthermore, the PKI/SIM can be considered as an SSCD (Secure Signature Creation Device) and a candidate to fulfil the CEN CWAs 14167-14169 (class 3). These CEN standards define the requirements for how a Qualified Signature shall be generated.

Compared to a PC with a smart card reader the mobile itself offers better protection for a simple reason: the mobile's operating system (OS) and capacity have far fewer options for malignant software like Trojan horses than a PC; hence, the trust to a user that he really "signs what he sees" is higher for a mobile at the present time. But as smart phones enter the market the chances of phones with malignant intent is expected to increase.

CWA 14169 addresses Common Criteria [10] EAL 4+ (on a scale from 1–7) for SSCD. General-purpose OSs for workstations or PCs like Windows 2000 etc. achieve maximum EAL 3, while OS for a mobile may achieve 4 or 5 (thus fulfilling the saying that "the more stupid the more secure"). The user has to enter a dedicated PIN code to activate the signature. This Signature PIN is different from the normal PIN.

If the PKI is blocked or revoked, the user may still use the normal telephone functions

Having the mobile PKI, the association with the user ID and SIM is significantly increased. SIM alone can prove that a user was present in a conversation or in a geographical cell; PKI can bind the mobile user to a transaction with the strength of legal proof in court. The Sign-PIN application is basically implemented for *m*Commerce applications, and has been in use since May 2001 bundled with payment options, either by *e*Purse or a pre-registered credit card or bank account. The fine thing is that the connecting *e*Shop only addresses the user with his mobile number and will have no information of bank account etc. If the signature is OK (as verified by the PKI-platform of Telenor Mobile) the *e*Shop will receive a confirmation that the transaction is signed, confirmed and agreed payment is en route.

All new Telenor subscribers since 2001 have been equipped with a SIM card with potential PKI. Today more than two million subscribers in Norway are so equipped. However, in accordance with Norwegian law, a user must apply for activation of the service. More than 90 *m*Commerce services are currently available, ranging from purchase of cinema tickets to payment and warning of expiry of parking time. In the future we will see the function in use also in areas outside traditional *e*Commerce. For example, pilots for electronic signature of public forms for childrens' day care and building permit applications are tested this year. The Norwegian Government *e*2005-plans open for several projects in the public area. The technology can also be tailored for stock purchase, *e*Voting, and a number of applications and services where a personal signature is required. Using the mobile as an out-of-band login assistant was tested out in Telenor's broadband project FSN more than a year ago. The user could select various methods for login authentication, depending on what was available. One option was to sign an SMS challenge on the mobile. This method may also prove to be attractive for other purposes, e.g. every time you have to register for purchasing on the Web. Instead of entering your registration data by the keyboard and watching it go unencrypted over the Web, why not have a service for releasing standard registration info stored in the mobile terminal by signing an application in the mobile. The returned information is an encrypted hash-value and the ICCID of the actual SIM, only meaningful to the network provider that has an agreement with the actual subscriber.

Telenor's SIM-PKI-function comprises a 1024-bit private key, SHA-1 hashing and RSA encryption of a received SMS. For initialisation, the generation of keys is performed on the SIM under the user's control, and the public key is transferred for certification 3DES-encrypted over the air. The certificate is not stored on the SIM (today) but referenced through its ICCID. Signatures are validated in Telenor's *e*Commerce validation server. The certificates are maintained by the trusted third party (TTP) ZebSign (X.509 v3 format) and stored in a standard X.500 catalogue.

Figure 2 The Mobile as an electronic pen
Illustration by Ragnar Philip Rosenlund

The PKI-enabled mobile can truly be considered as an electronic pen, as the artist describes it in Figure 2. It is an obvious candidate for providing Qualified Signatures, according to EU directive 99/93. Minor adjustments may be necessary to pass evaluation, however.

All in all, the mobile has gone through a development revealing increasing capabilities as carrier of an electronic ID. A great leap has been taken every 10 years since the start of mobile in 1981, see Figure 3.

Further personal-ID-oriented issues — electronic passport

Electronic passports with biometric data have been a target for increasing interest recently.

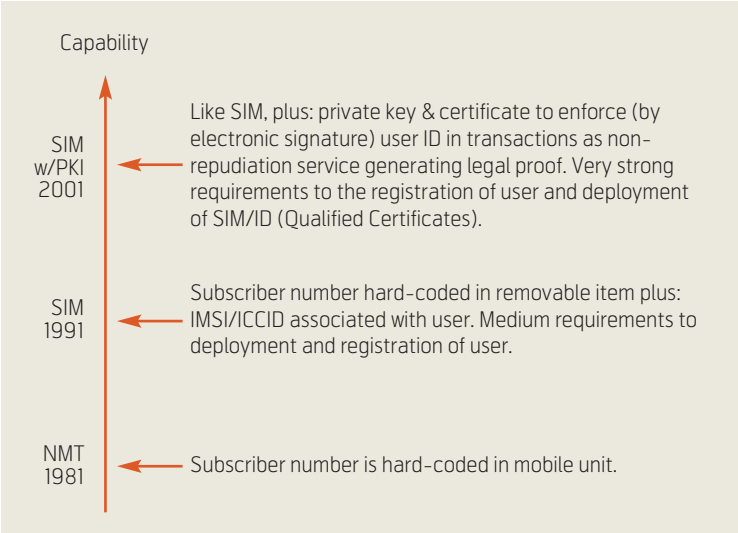


Figure 3 The Mobile as carrier of Personal ID

ISSN 0085-7130

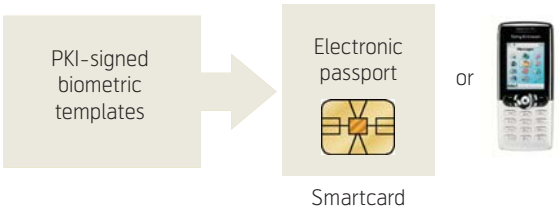


Figure 4 Smart card or the Mobile Phone as Electronic Passport

There are currently heavy efforts to select and standardise methods and means for this in the international society. ICAO (International Civil Aviation Organization) is in the lead in this work, and several methods have been tested at selected airfields around the world.

In principle, a biometric passport contains biometric data like recorded fingerprint, originally collected and stored in templates, e.g. on a chip. The template is electronically signed by the issuing authority and enclosed with a corresponding PKI certificate. At the gate, the user has to reproduce the biometric data by pressing a finger, scanning the iris, sampling the voice or whatever. This collected data is then compared with the data on the electronic passport, which first is verified through the certificate and the electronic signature. The mobile can easily be used, at least as a carrier for the signed biometric template. The data is transferred to the controlling machine through Bluetooth® or IR. R&D efforts should be launched to test such options. Biometric products claiming certification under [1] using PKI-signed voice templates have already been marketed [5].

Technically, both smart card and mobile phone solutions can do the job.

Other security issues

Securing the mobile and its traffic is of major concern both for the mobile user and the network provider.

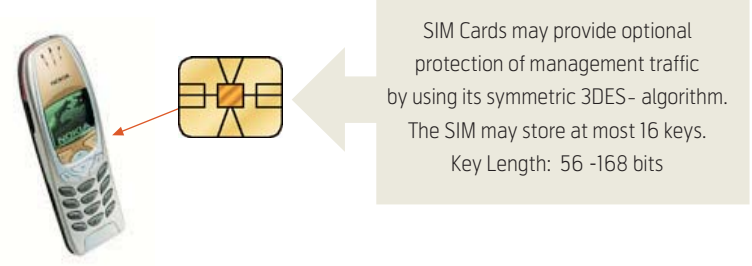


Figure 5 Addional Security Capabilities in the Mobile Phone

The most basic service is authentication to assure that the ID is correct. Traditionally, the network provider requires a mobile to authenticate itself in a unilateral manner. This is not least to assure that billing is sent to the correct addressee. Network access is given to subscribers only after a successful authentication. With the growing number of services, it will also be important that the network and service providers authenticate to the mobile user, to assure mutual authenticity. This is seen in UMTS, while 2G systems only provide unilateral authentication. PKI-based authentication is implemented as an option in WAP (WAP certificate class 1 and 2) for client-server authentication, WTLS, similar to the SSL method found in the Internet.

Another issue is how sensitive information is integrity- and confidentiality protected during transfer. Cryptographic methods are used to cover such services. In 2G and 3G systems the user traffic is encrypted over the air, but not in the fixed network. Neither of the addressed systems provides true end-to-end protection. Encryption keys are derived from secret information partly stored in the particular SIM or USIM. Bluetooth® and WLAN have optional encryption schemes. The hassle with these is that they have to be manually keyed by the user. Also, user traffic in WLAN is optionally protected by the WEP, a method thoroughly documented as easily broken by intruders. IEEE 802.1X is therefore increasingly taken into use for WLAN. 802.1X protocols provide authentication and key management. One problem for the WLAN is still that the management traffic is not protected, so it is proved that it can be easily jammed by DOS (Denial of Service) attacks.

Non-repudiation services are services directly offered to applications end-users, mostly in the form of electronic signature for selected transactions. PKI is almost mandatory for this service. Telenor Mobile's solution is one example (for signing SMS), and WAP Forum has also launched a certificate (class 3) for application signatures (SignText).

The embedded symmetric 3DES algorithm in the SIM can also protect management traffic for 2G and similar for 3G. In GSM, there is enough space for 16 different keys. In the basic system one key is shared with the network provider, so that specific messages and new software can be protected during the exchange. Telenor Mobile PKI exploits this facility when the public key is exported from the SIM to be certified. This service channel thus protects more in an "end-to-end" fashion, compared with the encryption of traffic, which is on the radio link only. This protection method could also be employed in wider scenarios, but this has to be done in agreement with

Type		Authentication Client – Server		Encryption (Non-repudiation)		Electronic Signature
		Client to server	Mutual	User traffic	Signalling	(PKI-based digital signature)
Mobile 2G	GSM	Yes	No	On radio link	No (in general)	Optional (SMS)
	WAP	Optional (PKI) Client–gateway	Optional (PKI) (Gateway-client)	No	No	Optional (SignText)
Mobile 3G	UMTS	Yes	Yes	On radio link	On radio link	Optional (PKI on USIM)
WLAN IEEE 802.11		Optional, mutual (WEP and/or IEEE 802.1X)		Optional radio link	Optional radio link	N/A
Bluetooth®		Optional, mutual		Optional Client-Server		N/A
IR		No data		No data		N/A
Modem Cable (RS 232 serial)		No data		No data	N/A	

Table 3 Security Services in mobiles

the network provider since the keys are installed in the SIM before distribution to the customer. One candidate can be an auxiliary to protect secrets like keys and passwords, e.g. for logging into a WLAN hot spot.

Table 3 gives an overview over the security capabilities in the mobiles today.

Electronic Signature is an application layer service, and is not applicable (N/A) for communication protocols alone. IR and modem cable may have proprietary solutions, but no obvious standards exist.

Applications and services

The basic data applications on a GSM are the editor and mailbox for SMS. This character-oriented function relies on a low bit rate, sufficient for one SMS that can be maximum 160 characters long. Note: a message may exceed 160 characters but will then be broken down and chained. EMS (Enhanced Messaging System) for 3G, which is based on MS protocols, breaks this limitation. The user will need an EMS-enabled phone, however.

One great leap for the 2G was the introduction of WAP (Wireless Application Protocol) and GPRS that allowed for browsing on the Web and higher bit rates. Now multimedia services (MMS) including transfer of pictures are possible. Camera on the mobile was introduced to utilize the improved communication channel.

With Bluetooth® and IR, which became available on 2G mobiles, it is also possible to send and receive pictures, files and messages, e.g. electronic business cards between terminals in close vicinity. Taking a photograph with the mobile and sending it by MMS or by IR or Bluetooth® to a nearby mobile or PC is today a piece of cake, and free of charge for the user as opposed to SMS or MMS. IR devices have already been in use for a long time as carport openers and TV remote controls. A mobile can in principle do the same. Telenor R&D has already implemented demonstrators that can control a slide show projector by using the mobile via Bluetooth®.

Utilisation of the RS323 cable interface is not new. Connecting the PC to the mobile provides Internet access with standard modem bit rates (9600 kb/s and upwards). Many services remain within this scope, e.g. applications where the PC is replaced by devices for a specialized purpose, like alarms, remote controls and even mobile credit card terminals, as increasingly seen in restaurants where the waiter no longer needs to take the credit card out of the owner's sight. Both DECT and GSM/GPRS are used.

Hence, the mobile plays an increasing role in remote operation. Commercial applications may encompass a bar code reader connected to the mobile. This may be a convenient reporting device for watchmen when inspecting a building, for social service people on home visits, for professional cleaning bureau people as they move from site to site. A convenient bar code is scanned and transmitted to the home base, reporting the starting or finishing of a job, etc. Such check-

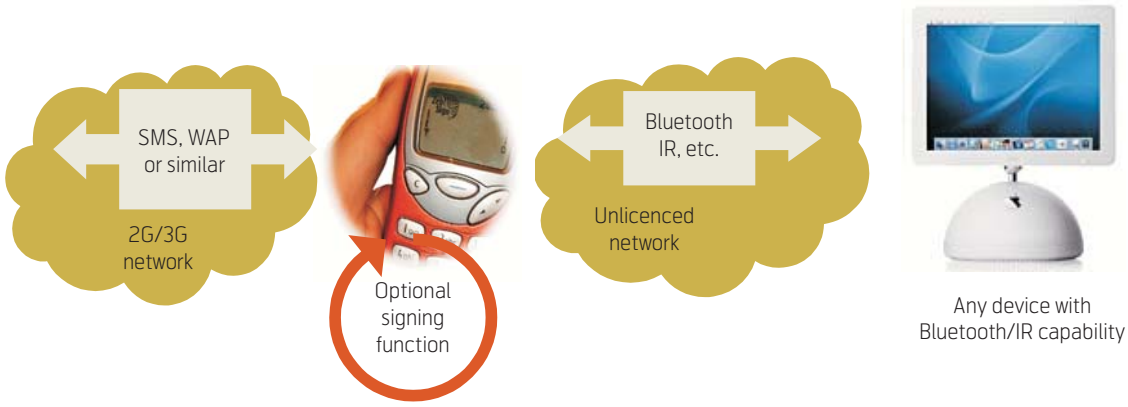


Figure 6 Mobile as Mediator between GSM /UMTS and WPAN Networks

ing-in and checking-out job segments are important to build workflow systems that can improve both management and even security in case of anomalous or missing reports. If additional confirmation is required, an SMS with a PKI-challenge to be signed by the phone user can be employed.

Hybridisation of communication channels is obviously a challenging task. In another implementation, the mobile is used to transfer passwords via SMS and Bluetooth® to a PC which is then automatically logged into a WLAN. A small application in the mobile captures the SMS containing the secret, and transfers it to the PC via Bluetooth®. It may just as well be to a physical door needing the "Open Sesame" code. Telenor R&D has also implemented a function for screen lock/unlock where the Bluetooth® application detects when the authorised mobile (actually SIM) is in close vicinity, i.e. less than 10 metres, and automatically opens or closes the screen depending on whether the mobile phone is within local range. A similar functionality can be implemented for printers that sometimes print classified material. A typical solution will mandate the printer to delay the printout of classified documents until the owner of the document is in physical vicinity. Today printer solutions are available where the document owner must supply the printer with a password for the same function. Users obviously experience this as a hassle, and it requires substantial effort from an organisation to make people use it. An automated solution with Bluetooth® would remove this hassle.

Bluetooth® can also be used to assist blind and visually impaired people when out in traffic. Position markers activated with Bluetooth® and placed along the streets communicate their position to any receiver passing by. An application in the mobile polls the signals continuously and transfers this information to audible voice messages like: "you are now on the corner of x- and y-street, half a kilometre from so and so place."

Bluetooth® will in such cases work superior to IR-based systems, since Bluetooth® is omni-directional and the device will be independent of placement within a certain radio signal beam. Due to the short-range coverage, the base stations can easily be placed outside each other's range to avoid overlaps and interfering signals. In cases where wiretapping is unwanted, IR is preferred, e.g. when authenticating to a bank automat or to the carport lock.

For handling the commercial transactions, the invention of *e*Purse makes it possible to associate payment with transactions. In its basic form the user has to fill up her *e*Purse by using a net-bank account or similar, but will later use the *e*Purse only by commands from the mobile. Telenor Mobile *m*Commerce solution has enforced these payment transactions also to include payment directly from a bank account or a credit card account. This solution (SmartPay) is established with the Norwegian bank DnB. Actually, the user may fill up her *e*Purse directly from her bank account only by commands from the mobile (MMS or WAP). This payment solution was security enforced by applying electronic signature to the transactions, meaning that the user had to explicitly sign it by PKI.

Of course, signatures add extra hassle and also cost to the service, so for low-value transactions the PKI-part is skipped and security relies on the authenticity of the personally registered SIM alone. An enforced identity control of the SIM is carried out in any case. For larger transactions electronic signature is mandatory, actually required by the bank. The solution also protects the customers against the *e*Shop that receives no information about bank account etc. which can be misused later. Only the mobile number is revealed. The *m*Commerce transactions include several phases:

1. The user must build an order by browsing (WAP) or SMS. These transactions only take place between the user and the *e*Shop. When the order is built, the *e*Shop contacts Telenor *m*Commerce PKI-server by the SOAP protocol with a requirement to confirm the payment for that user (referenced by his mobile).

2. The PKI-server generates an SMS request to the addressed user with two requirements: select payment method (if more than *e*Purse is registered), and then sign the transaction value.

3. When confirmed and signed, the server verifies the signature, activates the payment to the *e*Shop, sends a receipt back to the mobile user, and finally stores the signature for three months (configurable).

The PKI-server refers to the user's certificate by the ICCID-code that is returned together with the signature. Payment can in principle be combined with any service of relevance, whether physical items or music, lotto, or merely for transferring money from one account to another. Actually, the combination of payment and transfer of access code can also be attractive to Pay-TV or Video on demand. Combined with IR or Bluetooth® toll roads payment solutions can be developed where the car driver just directs his mobile to a detector, or the payment function is automatically detected when driving through the toll station.

The EU is especially concerned about money laundering. With requirements to PKI-based electronic signatures on financial transactions, the society will have a means to counter this growing market. This is because electronic signatures not only act as legal proof, but also because of their ability to participate in automatic handling and control in audit surveil-

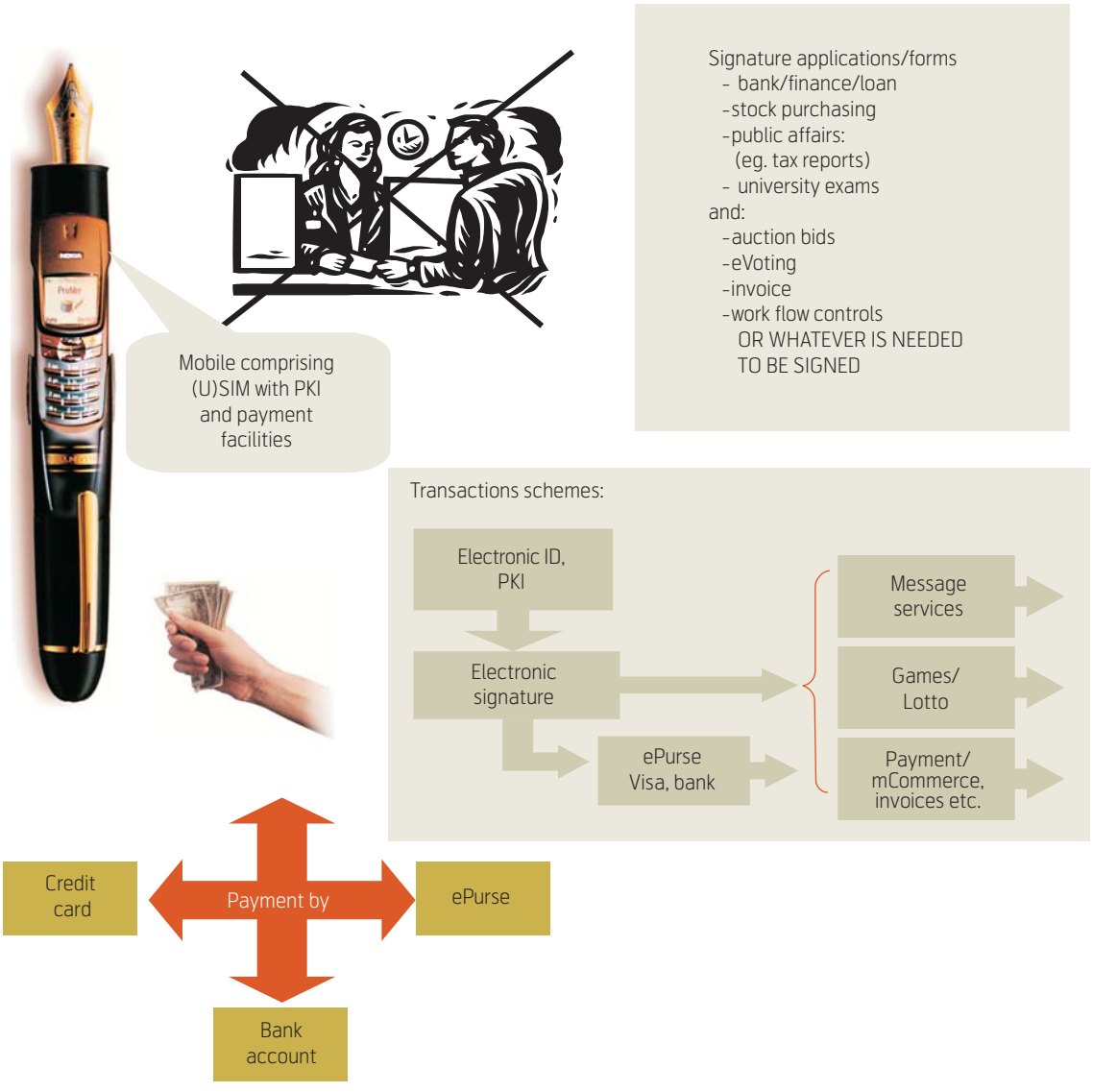


Figure 7 Mobile signature and payment in the e&m*Commerce world*

ISSN 0085-7130 © Telenor ASA 2004

lance. In the same field, automatic handling of invoices and their signatures can be incorporated in any transaction with or without payment. Analyses show huge potential savings by moving from manual to automatic methods. PKI is already prepared for automatic document handling, as the certificates have been standardised as XML format objects by ETSI.

In principle, the signature function is open for any transaction, also those not including payment. It can be used for distributed *e*Voting and confirmation in workflow situations, like signing receipt or delivery of patient journals in a hospital, and also for signing public forms. In a hybrid situation the user sits at her PC and fills in an application form or tax report. When finalised, the information must be reduced in size by a hashing method, e.g. SHA-1 that yields a 20 characters fingerprint of the original document. This hash-value is later transferred as an SMS to the mobile for signing, and the user has to compare the presented hash-value on the PC and the mobile before signing. The good thing is that this method transfers no sensitive information that can be wiretapped in transit. The question is whether the user is comfortable with signing a cryptic pattern or not.

However, the best solution would be to have the whole document appearing on the mobile, but for natural reasons (size of display, and also lack of encryption during transit) this may not be possible for still some time. On the other hand, whenever a really safe method is required, and the message is small and contains no or low sensitive information, mobile signature is very convenient.

Such situations can be foreseen in the remote control of hydroelectric dam ports, high value transactions like stock purchase (stock id, amount and timestamp are signed) and down to the more obvious control of the heating stove in a private mountain cabin (provided within mobile range). An expected visitor can be enabled to enter a certain gate and building by sending his business card by IR or Bluetooth® followed by a confirmation by signing an SMS challenge – Open Sesame (indeed). Systems requiring multiple-keys for opening are easily implemented: signatures from two or more different (and named) users are included in the solution.

Location based services

The ability to detect and use the actual geographical location of a mobile telephone opens for a portfolio of services. The cell location services of 2G (and 3G) have already been mentioned. Applications can use system information and detect whether two (or more) particular mobiles and thus their owners are in the same or a neighbouring cell, and can enable them easily to meet each other. Otherwise, it can launch a map of that area to the mobile to assist in finding the way to some place of interest. Higher precision than the size of a GSM cell (diameter from several hundred metres to some kilometres) is also possible. Since the cells partly overlap, higher accuracy, i.e. down to ~50 metres can be obtained by measuring time differences in signal arrival times from three adjacent GSM cells. The lower part of Figure 8 indicates this (a base station (BS) represents each cell). 3G-systems are similar.

Bluetooth®, on the other hand, can detect if a unit is within a short range (a few metres). Bluetooth® coverage for a larger area thus requires a large number of deployed base stations, however. A real enhancement would be to include the satellite based GPS (Global Positioning System) for measuring the exact location (~1 metre) almost anywhere in the world. However, GPS is not always reliable in cities (the "canyon" effect of tall buildings).

Location based services can be used to control anything from senile people gone astray, to prisoners serving open jail sentences, and of course, to keep track of your car, boat, children or pets.

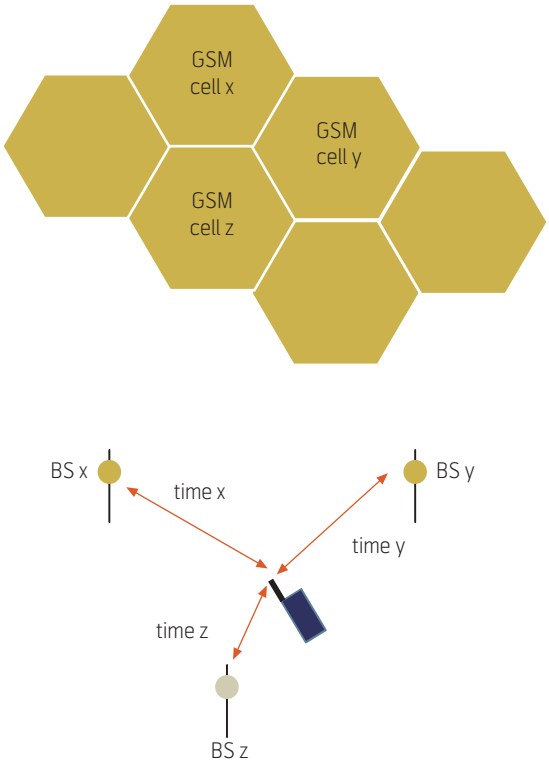


Figure 8 Idealized GSM cells and base stations

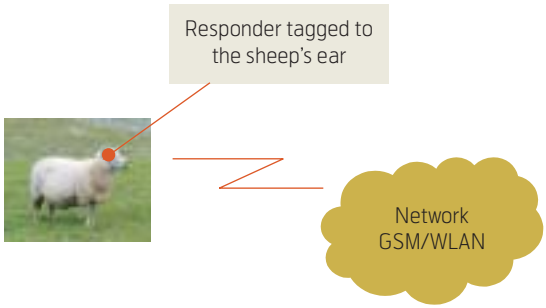


Figure 9 A wireless sheep and network

In the "smart" home WPAN Bluetooth® sensors can detect when father (carrying his mobile and SIM) is home or entering a room. Applications can then automatically turn down the too loud music (teenager is home before father) and change it to his taste given by a profile, turn light on or off to save energy etc. etc., and display any messages on some screen intended just for him, e.g. "no more beer in the fridge"; and certainly switch the TV to his preferred channel or set to preferred temperature when taking a shower. Otherwise, WPAN/Bluetooth® systems may be the target for many applications, including reporting from fire or burglar alarms, baby cry sensors or emergency alarms for disabled or senior people.

Mastercard and Nokia cooperate on a design of a mobile phone for card reader applications, typically payment in shops. Instead of sweeping the credit card through the card reader, you simply sweep your mobile phone over the card reader. Realisation is done through a low-effect radio unit in the mobile, which sends the credit card information to the card reader. Other developments indicating short range identification are the usage of radio frequency identification tags (RFID), which are already now in use in exclusive goods to prevent theft [8].

An interesting project comprised the herding of sheep in the wilderness (sheep normally roam free in the Norwegian mountains and rural areas). Here GSM and WLAN technologies were employed with base stations and a detector physically strapped to each sheep. By frequently reporting its position an application could detect anomalous situations like "no movement" – dead (?), "no signal" – possibly eaten by a wolf or bear who also crushed the radio while eating (?), "rapid movement over larger areas" – the wolf ate the sheep and also the radio without crushing it (?), or more obvious: illegally slaughtered and now inside a car in transfer to a refrigerator etc. The project *Cordless Sheep* ("Trådløse Sauer") has been thoroughly reported by Telenor R&D and was successfully carried out in cooperation with MIT [4].

Other deployed mobile location services, often GPS-based, are found in the transport industry, where position information of taxis and long haul trucks are used partly for business reasons: find and schedule the taxi closest to a passenger, or nearest available truck to pick up goods for express delivery, otherwise for emergency like urgent assistance in an accident or robbery, or to trace a stolen vehicle. Such services have been on the market for some years now. The technology and network are partly proprietary and may even be satellite based to cover rural areas or areas with e.g. low GSM coverage. Satellite based systems are mandatory for meteorological balloons or transponders drifting in the ocean, or units connected to a whale or a wolf reporting migration data for scientific research.

Security level based on location data

Having knowledge of position, it is possible to build policies with adjusted security requirements. When a terminal is located inside an area that is otherwise physically secured, some security issues may be relaxed, e.g. during logon to a network. This may imply a more efficient network – and for the user – potentially less of a hassle in several cases. Location data may be utilised to select strength, e.g. among the services authentication, authorization, access control, and communication confidentiality. Policies could be generated that sometimes require passwords and sometimes smart card-enforced login. Other policy elements could mandate encryption or explicit electronic signature for transactions performed in certain areas. Up to now, such policies are normally associated with the line-based networks; today it is also possible to deploy them in the wireless zone.

Seamless roaming

In the growing patchwork of different networks, the mobile phone users may find themselves in a world of access-options depending on where they may go or drive, and also the variety of access technologies implemented in each mobile unit. A simple situation is depicted in Figure 10, showing that different technologies may offer a "best effort" (as a combination of quality and cost) as the position changes.

Roaming in GSM is a well-established means, but inhomogeneous networks represent a challenge. Focusing on the technological part only, the different access-points like WLAN hotspots, Bluetooth® WPANs etc. may require authentication before con-

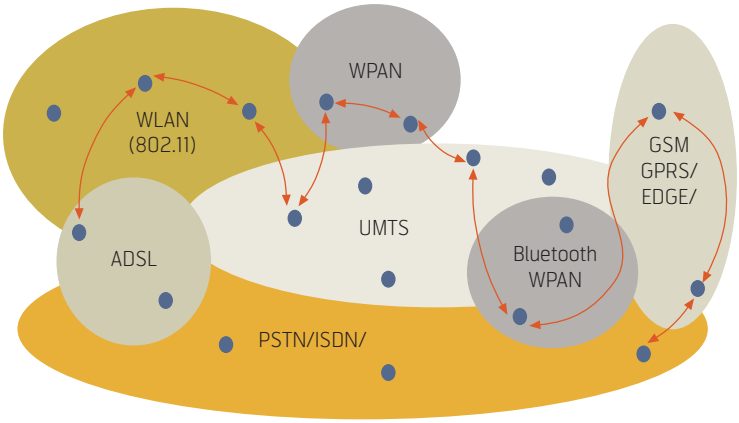


Figure 10 Seamless roaming – a vision. The blue dots represent access points

necting. All kinds of authentication methods depend on "shared secrets" that have to be distributed to the authorized community and the actual devices. By utilising message facilities like SMS, possibly enhanced with service channel protection (3DES), and also combined with electronic payment, the necessary tools are already present but need to be assembled. By mediating messaging over GSM/SMS and Bluetooth®, Telenor R&D has demonstrated that all transactions for WLAN log-on may be automated, compared to today's manual methods of entering passwords or initialisation keys etc. into the terminals. New projects should be launched to develop such methodologies and products aiming to connect to networks more easily.

Summing up

By combining the mobile's already partly existing communication capabilities, various applications, location based facilities, payment systems and security features, it is possible to implement a vast number of services. Actual demonstrators have been implemented demonstrating several possibilities, and show that it is possible to carry it out with rather small effort.

Figure 11 indicates some of the elements and combinations that may be exploited to generate new services.

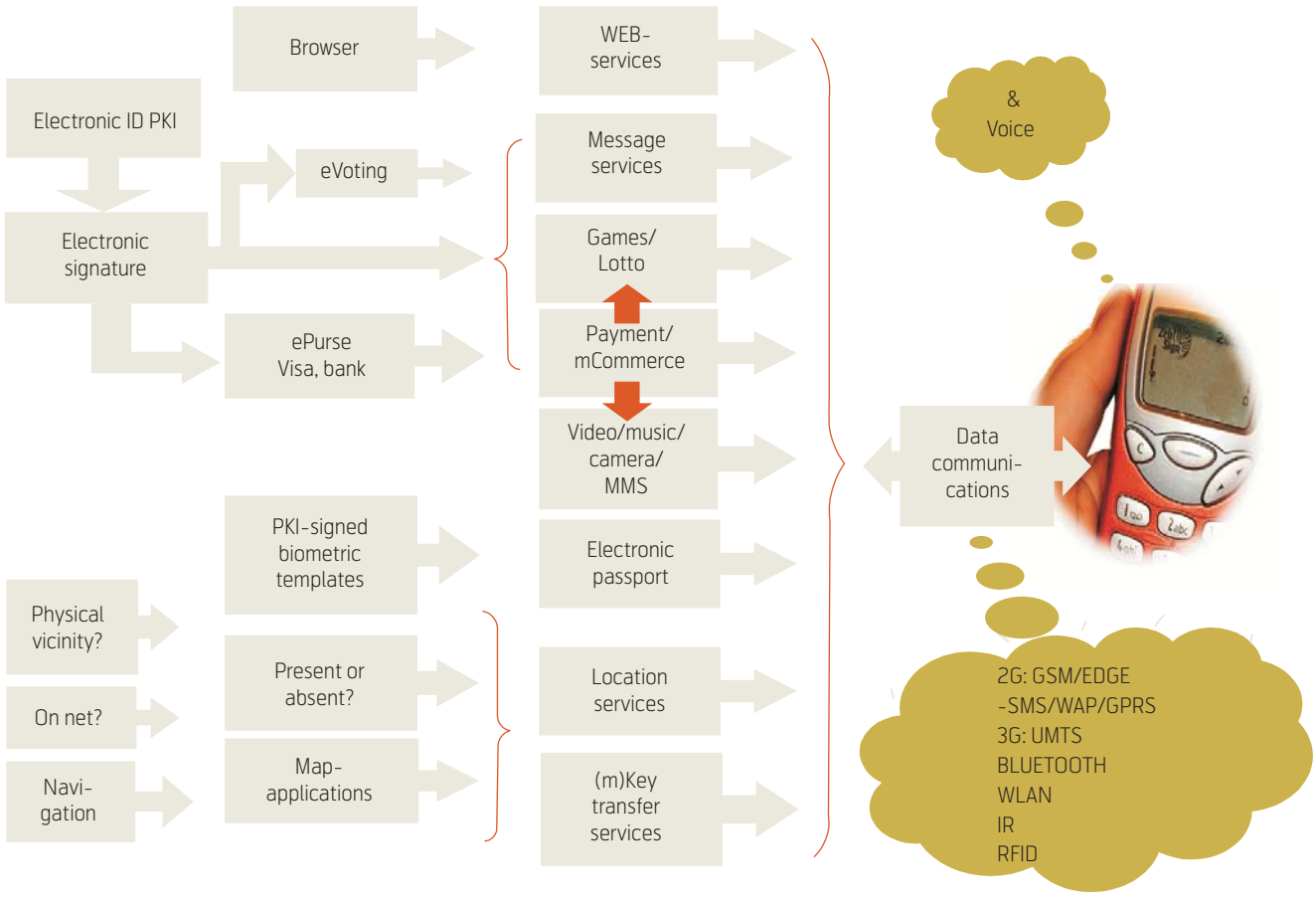


Figure 11 Baselines for existing and new mobile services

Service on Mobile	Status
Biometric Passport	Under consideration by several governments. Norwegian and EU solution planned launched 2005 at latest. US in 2004. Global activities lead by ICAO. Pilot solutions ongoing in several airports (Heathrow, Schipol etc.) Mobile's role = to be determined.
Authentication, Login assistant, Internet portals	Test implementations, FSN. Also see Signature applications below.
*e*Purse, *e*Payment terminal (*m*Commerce)	Deployed and in production, including PKI-signature. More than 90 services available Telenor *m*Commerce (mid 2003). Also competitor solutions in Norway, but without PKI.
*e*Purse, *e*Payment terminal, hybridized with PC-browsing or non-mobile applications	Implementation plans exist and easily activated (order on PC – confirm and pay by phone). *e*Purse & password transfer for WLAN login: found in market.
*e*Voting	Still in the thinking box, R&D projects in EU countries exist.
Location based services	Many systems already on market, such as taxi, lorry and wild game tracking. New R&D projects and plans frequently launched.
PC Lock/ Unlock	R&D test implementations (via Bluetooth®).
Remote controlled access (via IR) and electronic signature	R&D test implementations.
Signature applications for e.g. financial market (stock purchase, contracts, invoicing etc.	Telenor Mobile with system integrators have deployed signature and authentication solutions since Q4 2002. Tests carried out by and in the municipality of Oslo (signature of application forms). Financial market: Nothing yet.
Signature applications for public forms (signed hash)	Ongoing pilot for signing applications for house building and kindergarten. (Telenor Mobile)
Seamless roaming / Automatic transport of passwords, crypto keys etc. to devices	Products available for seamless roaming between WLAN, GPRS and UMTS. R&D test implementations (password via Bluetooth®). *e*Purse & password transfer for WLAN login: found on the market.

Table 4 Status for some of the ideas presented above

Deployments and plans

As indicated above, many of the ideas shown have already been implemented. Some of them only in demonstrators, others have been in production for some time. Still, many of the combinations should be studied further, implemented and tested for potential new products and businesses.

A total status is impossible to give, since the work takes place in laboratories and sites worldwide, and new services are launched frequently. From my own experience, I have seen new products marketed by various vendors only few weeks after just starting to think of the same idea. Table 4 indicates the status of some of the ideas today, as seen from Telenor R&D (in alphabetical order).

Final words

One important question now is how many of these services, accesses and mobility aspects can be compressed into manageable devices that are easy to use

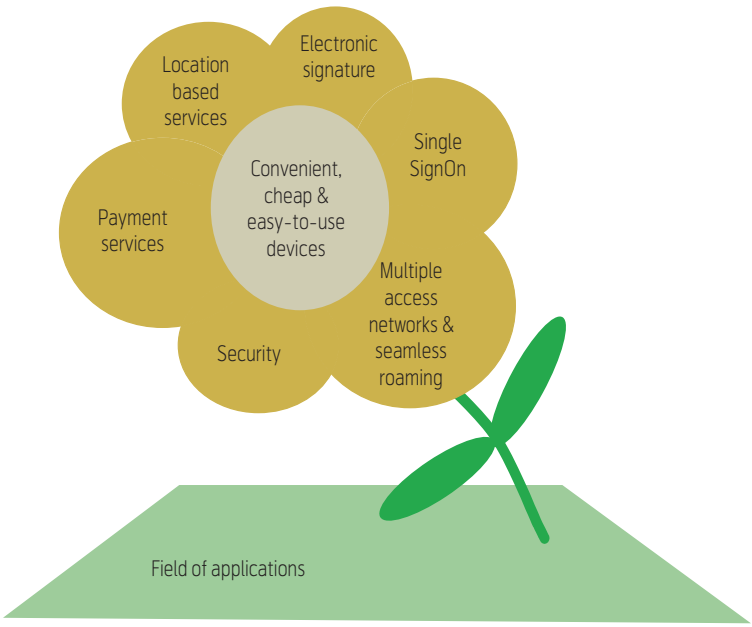


Figure 12 The flower of services and capabilities

ISSN 0085-7130

by the consumer, reasonable priced, and also secured and personalised to avoid misuse if lost or stolen. Biometric control, PKI-based electronic ID and single sign-on facilities represent some of the clues. Another important question is based on the fact that the technology increasingly facilitates "big brother" supervision of private activities and movements. How will legislation in society react to control and possibly block deployment of features?

Peeping into the crystal bowl of the future, the flower of success is assembled by petals of many flavours, as indicated in Figure 12. Correct and adequate nourishment at the right times will also be important in order to produce good fruit. Standardised trimming of the plant, and also good business plans are needed. Wise navigation concerning issues like personal privacy principles to deter public attacks and obstacles are also required. However, although at the infantile stage, buds have already started to grow and are observed with an increasing frequency. It is likely that some ideas, but not all, end up as useless weeds in the field of future applications.

There are still some miles to go before we know all the answers, but the movement has started.

Acknowledgements

I would like to thank Juan Carlos Lopez Calvet for providing ideas and demonstrators. I am also grateful for inspiring comments from other colleagues at Telenor R&D, especially Judith Rossebø, Dr. Josef Noll, and also Ingrid Aabø and Dag Thomassen at Telenor Mobile for sharing information and technology with respect to Telenor *m*Commerce PKI production platform.

Abbreviations

1/2/3G	1st, 2nd, 3rd Generation Mobile systems
3DES	Triple Data Encryption Standard (an encryption algorithm)
BS	Base Station
CDMA	Wideband Code Division Multiple Access
CEN	Comité Européen de Normalisation (European Committee for Standardization)
CWA	CEN Working Agreement
DECT	Digital European Cordless Telephone
DOS	Denial of Service
EAL	Evaluation Assurance Level (in Common Criteria)
EDGE	Enhanced Data Rate for GSM Evolution
EMS	Enhanced Messaging System
ETSI	European Telecommunications Standards Institute
FSN	Full Service Network (a Telenor piloted broadband project)
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications
ICCID	Integrated Circuit Card Identifier (typically 20 characters)
ID	Identity
kb/s	Kilobits per second
IMSI	International Mobile Subscriber Identity
IR	Infrared
LAN	Local Area Network
MIT	Massachusetts Institute of Technology
MMAC	Multimedia Mobile Access Communication
MMS	Multimedia Message System
NMT	Nordic Mobile Telephone (1G)
N/A	Not Applicable
PDA	Personal Digital Assistant (electronic handheld information device)
PIN	Personal Identification Code (4–8 digits)
PKI	Public Key Infrastructure
PUK	PIN Unblocking Code (8 digits)
RFID	Radio Frequency IDentification
RSA	Public Key algorithm (RSA) (in Telenor SIM-PKI: using 1024 bits key length)
SIM	Subscriber Identification Module
SMS	Short Message Service (up to 160 characters)
SSL	Secure Socket Layer (Internet protocol)
UMTS	Universal Mobile Telecommunication System
USIM	Universal Subscriber Identity Module (SIM for UMTS)
WAP	Wireless Application Protocol
WCDMA	Wideband Code Division Multiple Access
WEP	Wireless Equivalent Protocol
WLAN	Wireless LAN (IEEE 802.11)
WPAN	Wireless Personal Area Network
WTLS	Wireless Transport Secure Layer (for WAP)
XML	Extensible Markup Language

References

1 The European Parliament and the Council of the European Union. *Directive 1999/93/EC on a Community framework for electronic signature*. Brussels, 2000.

2 CEN European Committee for Standardization, CEN Workshop Agreement. *Secure Signature-Creation Devices, version 'EAL 4+'*. Brussels, 2002. (CWA 14169 EAL 4+)

3 ETSI. *Policy requirements for certification authorities issuing qualified certificates*. Sophia Antipolis, 2002. (ETSI TS 101 456)

4 Thorstensen, B et al. *"Trådløse dyr" – Teknologi og evaluering av feltforsøk*. Fornebu, Telenor Research and Development, 2002. (R&D report R48/2002)

5 *VoiceVault*. May 10, 2004. [online] – URL: http://www.voicevault.com/.

6 Eurescom Project P921. *UMTS radio access*. Heidelberg, 2000. (P921)

7 Eurescom Project P1203. *Operator's vision on systems beyond 3G*. Heidelberg, 2003/2004. (P1203)

8 Eurescom Project P1346. *Potential of the RFID technology for Telecom Operators*. Heidelberg, 2003/2004. (P1346)

9 Xylomenos, G et al. TCP Performance Issues over Wireless Links. *IEEE Comm Mag*, 39 (4), 52–58, 2001.

10 BS ISO/IEC. *Common Criteria for Information Technology Security Evaluation (CC)*. Geneva, 2000. (ISO/IEC 15408)

Tor Hjalmar Johannessen (56) is Senior Adviser at Telenor R&D. He graduated from the University of Oslo in 1975 as Cand.Real. After working with military crypto systems at Alcatel Telecom since 1989, he joined Telenor R&D Security Group in 2000. His main interest and occupation has been security in general and deployment of PKI systems in particular, which includes several engagements for ZebSign and Telenor Mobil /mCommerce. He participates regularly in ETSI ESI and CEN/ISSS WS on Electronic Signatures. He has been co-writer of EURESCOM P1001 deliverables, and also given several lectures on PKI and security topics.

Tor-Hjalmar.Johannessen@telenor.com

Status

Introduction

PER HJALMAR LEHNE

Per Hjalmar Lehne

The objective of *Telektronikk*'s Status section is to keep our readers up-to-date about organisation, working procedures, latest results, contributors, future activities, etc. from international standardisation bodies and research organisations.

In this issue of *Telektronikk*'s Status section, we present work and results from the *International Telecommunication Union (ITU)*. First, *Astrid Solem* and *Evi Zouganeli* report from the work on ITU-T's vision of *Next Generation Network (NGN)*. The concept was introduced as a result of the work on the *Global Information Infrastructure (GII)*, which was started in 1995. The GII has previously been presented in the Status section of *Telektronikk* 1.1996. The GII allows a heterogeneous mix of technological and operational domains. The concept of NGN was introduced to address the implementation issues and to facilitate a convergence of networks and services. The article explains the concept in terms of the basic capabilities and presents the NGN reference model with its basic functionality divisions. QoS and mobility are discussed specifically and the paper concludes with some remarks on further work and thoughts about the possibility for success.

In two papers, *Arve Meisingset* presents short technical notes on work in ITU-T. The first paper addresses the topic of *data exchange between operators*. In Europe, the current situation is that different operators use different data definitions, so that the exchange of data during an interoperation agreement often must be done by use of telefax or e-mail. The paper discusses this problem and presents the work on recommendation M.1401 from Study Group 4 (Telecommunication management, including TMN). In his second paper Arve Meisingset presents the work on *notations* in the *Universal Markup Language (UML)*. The article presents the different specification languages from ITU, like SDL, eODL, MSC, URN, TTCN and ASN.1. Study Group 17 (Data Networks and Telecommunication Software) is trying to provide profiles in UML for all the notations used in the aforementioned languages. Mr. Meisingset is currently the Vice Chairman of Study Group 17.

Ole Grøndalen also presents a technical note on *Telecommunication for disaster relief (TDR)*. This is an application addressed by Study Group 16 (Multimedia services, systems and terminals). Specifically, the capabilities for authorities to use public telecommunications services and networks for emergency operations are addressed. Study Group 16 has taken the role of co-ordinating the different players involved and the work done in other parts of the ITU. The article presents a possible TDR core topology and which features are expected to be included.

The final article is a presentation of the *Wireless World Research Forum (WWRF)* from *Erik Lillevold*. The WWRF sprang out of a joint activity in EU's 5th Framework Programme, called the *Wireless Strategic Initiative (WSI)* that started in 2000. WWRF, created early 2001, is probably one of the most important fora in which important visionary work for tomorrow's wireless telecommunications is taking place. The article explains the background and objectives of WWRF and briefly explains the organisation and the reference model. The results from WG1 on Human Perspective and WG2 on Service Architecture are explained in more detail.

Per Hjalmar Lehne (46) is Research Scientist at Telenor R&D and Editor in Chief of Telektronikk. He obtained his M.Sc. 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. His work since 1993 has been in the area of radio propagation and access technolgy, especially on smart antennas for GSM and UMTS. He has participated in several RACE, ACTS and IST projects as well as COST actions in the field. His current interests are in the use of MIMO technology in terrestrial mobile and wireless networks and on access network convergence, where he participates in the IST project FLOWS.

per-hjalmar.lehne@telenor.com

 ISSN 0085-7130

Next Generation Network – an ITU-T vision

ASTRID SOLEM AND EVI ZOUGANELI

Astrid Solem

The Next Generation Network (NGN) is a network concept that aims at providing a framework to encompass the large variety of existing and emerging protocols and services, facilitate a further evolution of these, decouple this evolution from the underlying network infrastructure, and facilitate the interfacing of a plethora of available media. The rationale behind NGN lies well founded in paradigm shifts that have been taking place within the technological solutions and the business models in the telecom industry as a whole. And whereas the concept itself is all but new, the attempt by ITU-T to standardise the complete concept based on IP technology is in fact a rather new notion.

Evi Zouganeli

Introduction

Telecommunications has undergone a radical transformation in the past decade driven by new business circumstances, technological factors, and evolved user requirements. Prominent among these driving forces are the total deregulation of markets and open competition between operators, the explosion of digital traffic, the dominance of data over other types of traffic, a steadily increasing user demand for new multimedia services and general mobility – to name a few. In the face of this change, a new network concept had to emerge – a more flexible one that could encompass a variety of protocols, provide a range of services, and facilitate the interfacing of many media.

The ITU-T paved the way with the introduction of the Global Information Infrastructure (GII), on which work was started already in 1995 [Y.100, Y.110]. GII allows a heterogeneous mix of technological and operational domains, where the choices of core technologies can vary at the same time that a full set of services is being provided by e.g. a multi-service network. Implementation issues were, however, not a part of GII and therefore additional Recommendations were required – i.e. additional specifications and implementation guidelines. Naturally enough, the concept of Next Generation Network (NGN) was introduced with the mission to facilitate a convergence of networks and services [Y.NGN-overview].

Besides its purely technical scope, NGN aims at promoting fair competition and encouraging private investment. It shall define a framework that facilitates the fulfilment of a range of regulatory requirements and assist in providing open access to networks via well-defined open interfaces. There is a clear underlined socio-political and economic base here that aims at promoting broadband and diversity of content, and advocating equality for citizens worldwide.

The primary role of NGN was to constitute a concrete realisation of GII. However, in view of the rapid changes that took place and not least the challenges of the implementation process, NGN has grown beyond the original GII. In addition, a clear demand from the market for preliminary NGN standards has been quite evident in recent years. Hence 2004 was proposed as the target date to prepare first Recommendations on NGN. The basic ideas behind NGN are by no means new – the main building blocks it consists of have been around for a while. And yet NGN presents in fact a great challenge for the ITU-T. Indeed, the attempt to standardise and recommend the complete concept based on IP technology is an accomplishment in itself – especially so when it is initiated and carried out by ITU-T.

In the following, we present an overview of the rationale for the creation of NGN and the current status of the work, including the main characteristics of NGN and the capabilities it provides. We then proceed to a presentation of a selection of aspects related to the QoS and mobility architectures and, finally, conclude with a brief evaluation of the future prospects of NGN.

Basic characteristics and capabilities of NGN

A Next Generation Network is a packet-based network able to provide a range of services – including telecommunications services; able to make use of multiple broadband, QoS-enabled transport technologies; and in which service-related functions are independent of the underlying transport-related technologies. It offers unrestricted access for users to different service providers and supports generalised mobility, allowing consistent and ubiquitous provisioning of services to users [Y.NGN-overview].

ISSN 0085-7130 © Telenor ASA 2004

In addition to the above, the NGN is characterised by [Y.NGN-GRM]:

- a separation of control functions into bearer capabilities call/session and application/service
- de-coupling of service provision from the network
- provision of open interfaces
- fixed/mobile converged services
- generalised mobility
- unified service characteristics for the same service as perceived by the user
- compliant with all Regulatory requirements.

NGN will have to provide the required resources (infrastructure, protocols, etc.) in order to make possible the creation, deployment and management of all kinds of services – both known and yet unknown. This encompasses services over all kinds of media, with all kinds of encoding schemes and data services – conversational, uni-cast, multicast and broadcast, messaging, simple data transfer services, real time and non-real time, delay sensitive and delay tolerant services – with bandwidth demands ranging from a few kbit/s to hundreds of Mbit/s, guaranteed or not. NGN will provide service related Application Programming Interfaces (APIs) in order to support an efficient creation, provisioning and management of services and thus allow a service customisation and personalisation.

One of the main characteristics of NGN is the decoupling between services and networks, allowing them to be offered separately and to evolve independently. Therefore in the NGN architectures that are proposed, there is a clear separation between the functions for the services and the functions for the transport. NGN allows the provisioning of both existing and new services independently of the network and the access type used.

In NGN the functional entities that control policy, sessions, media, resources, service delivery, security, etc, may be distributed over the infrastructure, including both existing and new networks. When they are physically distributed they communicate over open interfaces. Consequently, the identification of reference points is an important aspect of NGN. New protocols are being standardized to provide the communication between the functional entities and interworking between NGN and existing networks such as PSTN, ISDN and GSM is provided by means of Gateways. In addition, NGN will support both existing and 'NGN aware' end-terminal devices. Hence, terminals connected to NGN will include both old (e.g. analogue telephone sets) and new types of terminals (e.g. Ethernet phones through PCs). Finally, a major feature of NGN will be generalized mobility, i.e. a consistent provision of services seamlessly across access technologies where the users will be identified uniquely – independent of access technology – as discussed later in this paper.

NGI Reference Model

The need for a generic model that is applicable to all open systems standards has been evident for a while. Indeed, most of the viable – existing and emerging – network technologies do not fit in the OSI Reference Model – a well known example among these being IP. The OSI model specifies a hierarchical seven-layer architecture and has in practice become a rather rigid reflection of the seven layers it comprises. A set of characteristics have been specifically defined for these layers, and OSI tailored protocols have been developed to match these characteristics.

On the other hand, the concepts of a layered architecture that is used in the OSI Reference Model [X.200] apply to all layered architectures. Difficulties arise when the more specific seven-layered model is considered. NGN systems (non-OSI systems) may encounter a number of situations, such as that the number of the layers is not seven, or that the functions of individual layers or the protocols involved do not correspond to those of OSI, or that the compliance requirements of the OSI model are not applicable. The functionality described in OSI may be present in most systems, it may however be distributed in a different way, e.g. within fewer layers or not layered in the rigid fashion defined by the OSI model. NGN, therefore, defines a much more flexible architecture in its reference model, aspects of which are described in the following [Y.NGN-GRM].

Ordering of protocol layers in NGN

Contrary to OSI where the ordering of protocol layers is hierarchical in nature, in NGN there may be no natural order to certain instances of protocol layering – this may be rather arbitrarily imposed as a result of differences in core technologies and services offered.

Peer semantics of layers

In today's environment there are many cases where the transport protocol, e.g. TCP, does not operate between the ultimate source or the ultimate destination of data – for example when firewalls are being used, or in the case where network based devices offer simulated services. In NGN no absolute assumption is made as to the nature or the location of the end-point at which particular protocols are generated or terminated.

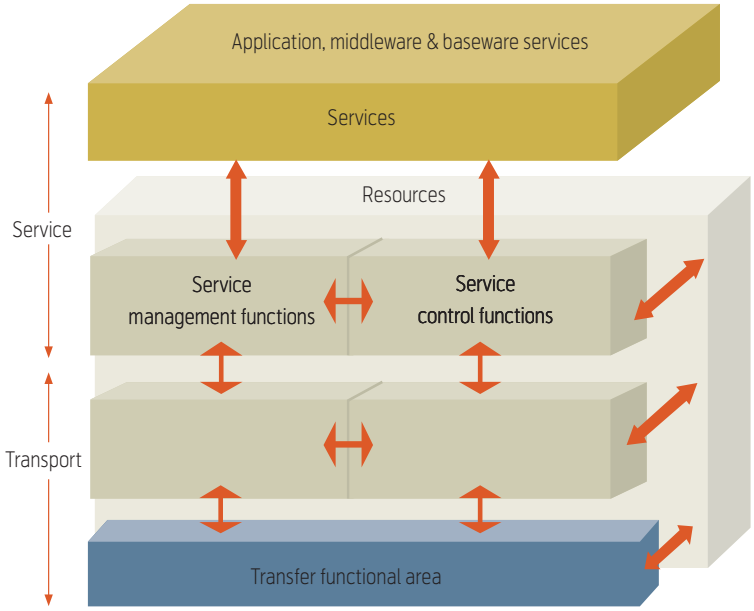


Figure 1 General Functional Reference Model. NGN provides horizontal separation of services from the network (transport) and the vertical division of control and management

Mode of transmission

Traditionally if there is a connection-oriented transport service there must be a connection-oriented network service. In many environments, however, this rule is clearly violated and a connection-mode transport service operates over a connectionless-mode network service (TCP over IP is a notable example).

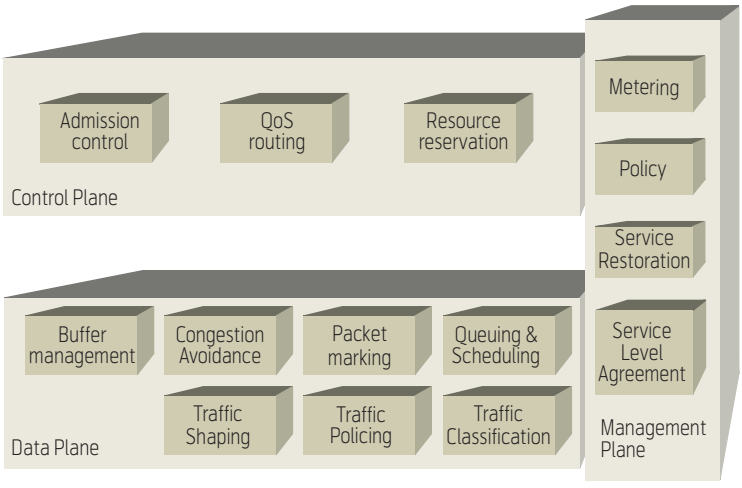


Figure 2 Architectural Framework for QoS Support. The QoS framework provides a complete set of QoS mechanisms needed to deliver exhaustive QoS. The Data Plane contains mechanisms dealing with the user traffic directly. The Control Plane contains mechanisms dealing with admission to and control of the pathways through which user traffic travels, and the Management Plane contains mechanisms dealing with the operation, administration, and management aspects of the network

NGN disregards this requirement for harmonized connection-orientation between network and transport services.

Basic functionality divisions in NGN

A key cornerstone of the NGN is the separation of services from the network. This is represented by two distinct blocks of functionality: the service layer and the transport layer (Figure 1) [Y.NGN.GRM].

The service layer, or group of layers, is concerned with the application and the services that are to be operated between peer entities. Services may be related to voice, data or video applications, arranged separately or in some kind of combination in the case of multi-media applications.

The transport layer, or group of layers, is concerned with transfer of information between peer entities. For the purposes of such transfers dynamic or static associations may be established to control the information transfer between such entities. Associations may be of extremely short duration, medium term (minutes), or long term (hours, days – or longer).

QoS – an important aspect of NGN

Controlled service provisioning for different types of services through the use of QoS is an important aspect of NGN. Work in this field has been carried out both in ITU and IETF in the past years. Several ITU Recommendations deal with service performance, e.g. Recommendation E.800, which defines QoS as the degree of satisfaction a user of a service will have. Recommendation G.1000 provides a framework for service performance (or service quality) and breaks down the service performance into functional elements and links these functional elements to network performance as defined in ITU-T Recommendations I.350, Y.1540 and Y.1541. Recommendation G.1010 provides end-user-centric application requirements in terms of broad categories (such as interactive and error tolerant). Concerning specific applications or performance parameters, among related standards, ITU-T Recommendation M.1079 defines the end-to-end speech and data quality and performance requirements for IMT2000 access networks, while ITU-T Recommendation G.114 specifies the bounds for transmission time for connections across a digital network.

In order to deliver the required network performance a number of mechanisms need to be in place within the network, namely mechanisms that are important in order to control and deliver various network services. For instance, fair resource allocation schemes

are needed in the network. The required mechanisms as well as a signaling method to indicate the desired level of network performance are the focus of the architectural framework for support of quality of service (QoS) in packet networks [Y.qosar]. This recommendation provides a basis for further Recommendations related to QoS and manageability in NGN [Y.e2eqos], [NGN-MAN].

A set of generic QoS network mechanisms and the frameworks structure for these are shown in Figure 2, which illustrates the architectural framework for support of QoS.

The different network mechanisms shall be combined in order to deliver the overall performance that is satisfactory for a wide range of applications. In terms of QoS, the Control Plane mechanisms include admission control, QoS routing, and resource reservation. The Data Plane mechanisms include buffer management, congestion avoidance, packet marking, queuing and scheduling, traffic classification, traffic policing and traffic shaping. Finally, the Management Plane mechanisms include Service Level Agreement (SLA), traffic restoration, metering and recording. The recommended Control Plane mechanisms related to admission control as well as some Management Plane mechanisms are described is some more detail in the following.

Whether traffic is admitted to the network depends on whether there are resources available in the network and on the applicable SLA. It is in the interest of the service provider that maximum new traffic is admitted to the network while at the same time the same level of QoS is maintained for the existing traffic. One approach to admission control is to ensure satisfactory compliance for a set of metrics (e.g. packet loss, delay and jitter) and this approach is appropriate for providing hard QoS for real time services. This approach implies the use of a resource reservation request for securing the necessary resources. The use of such a type of admission control is illustrated in Figure 3.

The bearer resource manager (BRM) is an independent resource control function that manages all the bearer resources in an administrative domain. The BRM records and maintains the network topology and resources in a database as well as using this database in order to make intra-domain path selection, resource allocation and admission control for a service flow. Resource control for inter-domain application flows is achieved through signaling. The BRM may also perform functions like policy management, SLA management, LSP traffic metering, and interface with AAA servers.

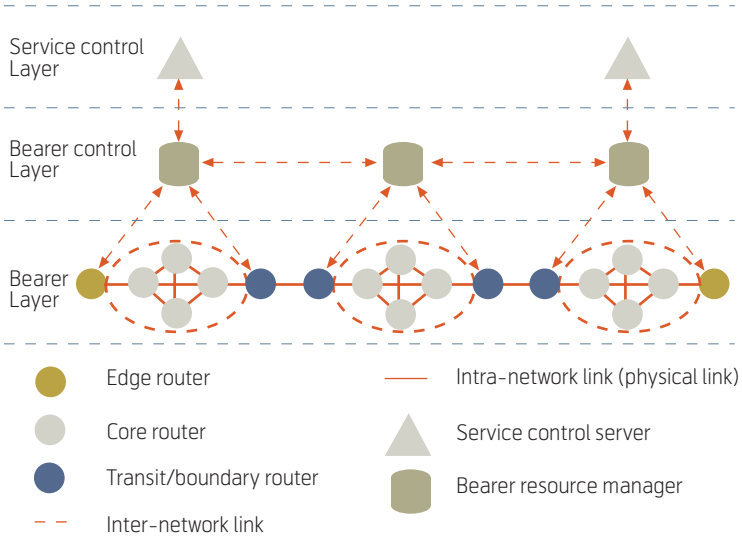


Figure 3 Comprehensive QoS approach based on independent resource control. Iteration between the different layers provides hard QoS guarantees using Service Control Servers (SCS) which interact with the Bearer Resource Managers (BRMs), that control resources in each administrative domain, through the use of clear signaling interface between control plane and data plane

A variety of service control servers (SCS) are responsible for controlling various service requests (e.g. signaling voice calls), identifying the originating and terminating point of each service request, translating number (or name) into IP address, and then sending the resource requests to the BRM of the originating domain. For services that are not provided by use of service control servers, hosts can initiate a QoS service request through RSVP or other QoS signaling protocols. The BRMs only manage the intra-network link resources, whereas application gateways or boundary routers manage the inter-network link resources by the specified inter-network SLAs and an application gateway or boundary router acts as the ingress or egress edge router. The building blocks interact primarily through signaling at a per-flow level and on the basis of resource management per logical bearer network (LBN)[1)]. In this model there is a clear signaling interface between control plane and data plane and this approach is further detailed in conjunction with networks using DiffServ-aware MPLS and packet networks without MPLS support.

[1)] *A Logical Bearer Network consists of edge routers, intermediate transit routers and LSPs (in case of MPLS) between nodes.*

Levels	Descriptions	Features	Remarks
0	No management	No monitoring, No resource control	• No mechanism to detect network fault and congestion • No mechanism to control network resources
1	Overall network resource management	Overall monitoring, No resource control	• Notify overall network fault and resource status by network provider • Manage the group level resources by the customer
2	Group level resource management	Group level resource monitoring and control	• Notify group level network fault and resource status by network provider • Manage the group level resources by the customer
3	Individual resource management	Individual level resource monitoring and control	• Notify individual network fault and resource status for end-to-end connectivity • Manage the end-to-end resources by the customer

Table 1 Levels of Manageability. Management may be performed on overall network resources, on group level or on an individual resource. The different levels of manageability require different levels of functionality for the network operator and different access to the management mechanisms for the customers, where the customers may be other network providers, service providers, or end users

By contrast, the measurement-based approach uses measurements of existing traffic for making an admission decision. It does not guarantee throughput or hard bounds on packet loss, delay or jitter and it is therefore appropriate for providing soft or relative QoS. In a measurement-based QoS approach based on the Priority Promotion Scheme (PPS) a form of admission control is achieved as follows: prior to establishing a session the source measures or probes the availability of network resources by sending out packets with a priority level that lies one level below that of normal packets. In order to establish the session, the priority of succeeding packets is raised, or the packets are promoted, by increasing their DiffServ Code Point (DSCP) value of the succeeding packets. The DSCP may be lowered in order to leave resources with existing sessions or otherwise adjusted so that the number of packets does not exceed the available capacity. By demanding that all end systems follow the above behavior, end-to-end QoS is achieved without the maintenance of per-flow states in network nodes. This approach has in general higher network resource utilization than the parameter-based one.

In the Management plane a Service Level Agreement (SLA) between a customer and a provider specifies the level of availability, serviceability, performance, operation or other attributes of the service. It may include aspects such as pricing that are of a business nature. ITU-T Recommendation E.860 defines a general SLA framework for a multi-vendor environment. The technical part of the agreement, the Service Level Specification (SLS), will specifically include a set of parameters with given values to define the service offered by the network. SLS parameters may either be general such as those defined in the Y-series Recommendations, or they may be rather technology specific as in IntServ or DiffServ. Based on traffic metering, recording and monitoring of traffic stream against the agreed traffic profile necessary actions (e.g. dropping or shaping) can be performed.

In an NGN service environment providing SLAs the IP network should be reliable and manageable. The end-to-end connectivity should meet the negotiated SLAs according to the various application types.

Different levels of manageability for manageable IP services are described in Table 1. The different levels will depend both on the customer and the network operator's point of view.

SLAs are (among other things) used for making assurances about performance and availability of the network to the customer. It is expected that an NGN must satisfy the requirements to support the business model for differentiated service concepts, including the usage-based billing and charging model. In addition, the network should be stable and secure with reliability performance of 99.999 %. This implies that advanced capabilities and protocols have to be introduced in order to make the IP network reliable and manageable as described in [Y.MAN-NGN], and manageability of the network will increase the complexity of network mechanisms and protocols. For instance, management with feedback control and access control will be determining features in order

to meet the differentiating service and business needs of NGN. The work on manageable IP networks has among other things resulted in a reference architecture, as defined in [Y.MAN-NGN], that illustrates the functions required to provide a Manageable NGN Network.

Mobility in NGN

One of the most crucial requirements for NGN is to provide mobility management for users and terminals in order to ensure nomadicity within the network, roaming and nomadicity across different networks, as well as seamless mobility for on-going sessions in the networks. The Mobility Management Requirements and Architecture for NGN are described in [Y.NGN-MOB]. Mobility can be divided into Intra-Access Network Mobility, Intra-Network Mobility, and Inter-Network Mobility and further separated into personal mobility and terminal mobility.

Different Mobility Management (MM) techniques have been proposed and are deployed in different types of networks to effectively manage the identification, registration, authentication, and movement of mobile users. Some of these techniques have been unique to the respective system and hence manage only the movement of users within a specific homogeneous mobile system. The provision of seamless service and mobility across different heterogeneous systems is currently not possible in many cases, for instance due to differences in the (wired/wireless) access technologies used, differences in the available services and their non-portability, or differences in the MM techniques deployed. Users have up to now been considered as different customers if they are on different access networks with independent service configurations and no bridging between these.

An important aim of the NGN is to overcome these restrictions of mobility. Another requirement put forward is that the NGN Mobility Management shall avoid the need to implement as many authentication mechanisms (with different identities, logins and passwords) as there are access technologies. In other words, by using Mobility Management functions, the network should be able to address efficiently identification, authentication, and mediation between different access technologies to access the same services seamlessly. It is expected that there will co-exist a variety of the existing and new wired/wireless access network technologies, for example the WLAN, xDSL and the 2G/3G/4G mobile networks, as illustrated in Figure 4. Each of the access networks would be connected to the IP-based core network and be interworking with other core networks as well as with other access networks. This way it will be possible to provide the same set of services for users, preferably independent of the access type.

Further work

The NGN is a network concept that aims at providing a framework to encompass the large variety of existing and emerging protocols and services, facilitate a further evolution of these as well as decouple it from the underlying network infrastructure, and facilitate the interfacing of a plethora of available media. The rationale behind NGN lies well founded in paradigm shifts that have been taking place within the technological solutions and the business models in the telecom industry as a whole and is certainly a concept that was badly needed. The basic ideas behind NGN are, however, by no means new. The different building blocks that comprise the concept – such as separation of control functions into bearer capabilities call/session and application/service, de-coupling of service provision from the network, all services provided by one network also including fixed/mobile converged services as well as generalised mobility – have been around for a while. On the other hand, the attempt to standardise and recommend the complete concept based on IP technology, and in particular within ITU, is a rather new notion.

The work so far has focused on the definition and development of the overall framework of the concept, with the addition of some more detailed recommen-

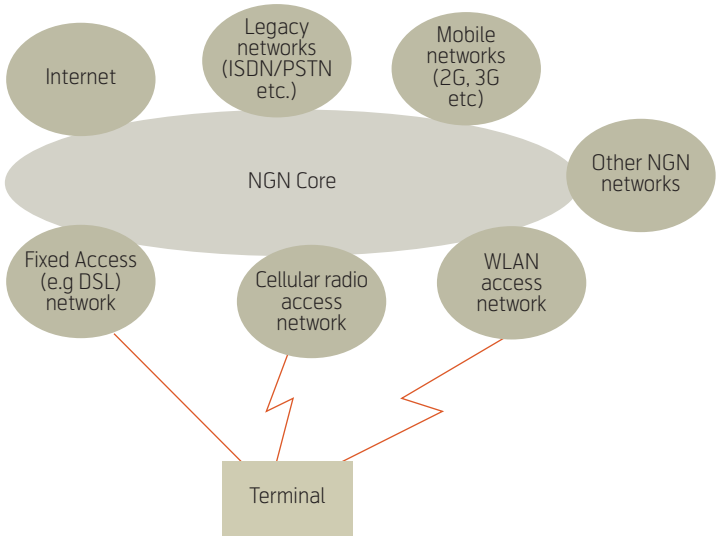


Figure 4 NGN core with several types of access networks. The aim of NGN mobility is to provide mobility mechanisms in order to attain seamless service provision and mobility across different heterogeneous systems

dations in certain areas. It is important to note that NGN work is followed up by all relevant Study Groups and that the complete set of more detailed Recommendations are produced in order to give guidance and standards to vendors, network operators and service providers. Some specific issues that need to be further addressed include the migration of voice services to the NGN infrastructure, QoS related to real time voice services (bandwidth guarantees, delay guarantees, packet loss guarantees etc.) as well as security. NGN should provide the security mechanisms to protect the exchange of sensitive information over its infrastructure, to protect against the fraudulent use of the services provided by the Service Providers and to protect its own infrastructure from outside attacks.

NGN is an ambitious goal, and success should not be taken for granted. On the one hand, it may seem like the days of thoroughly standardised concepts are over. On the other hand, however, it appears that the lack of a complete set of standards has led to a rather slow evolution of advanced functionality in IP-based networks, and this despite the fact that de-facto standards and proprietary solutions have been around for at least some parts of the concept. It remains to be seen whether ITU will succeed in its ambitious task regarding NGN – in any case operators should hope so!

Abbreviations

AAA	Administration, Authorization, and Authentication
BRM	Bearer Resource Manager
DSCP	DiffServ Code Point
GII	Global Information Infrastructure
GSM	Global System for Mobile Communications
IP	Internet Protocol
ISDN	Integrated Services Digital Network
LBN	Logical Bearer Network
LSP	Label Switched Path
PSTN	Public Switched Telephone Network
MPLS	Multi Protocol Label Switching
MM	Mobility Management
NGN	Next Generation Network
OSI	Open System Interconnection
PPS	Priority Promotion Scheme
SCS	Service Control Servers
SLA	Service Level Agreement
SLS	Service Level Specification
TCP	Transport Control Protocol
xDSL	X Digital Subscriber Line
WLAN	Wireless Local Area Network

References

[E.800] ITU. *Terms and definitions related to quality of service and network performance including dependability*. Geneva, 1994. ITU-T Recommendation E.800.

[E.860] ITU. *Framework of a Service Level Agreement*. Geneva, 2002. ITU-T Recommendation E.860.

[G.114] ITU. *One-way transmission time*. Geneva, 2000. ITU-T Recommendation G.114.

[G.1000] ITU. *Communications Quality of Service: A framework and definitions*. Geneva, 2001. ITU-T Recommendation G.1000.

[G.1010] ITU. *End-user multimedia QoS categories*. Geneva, 2001. ITU-T Recommendation G.1010.

[I.350] ITU. *General aspects of quality of service and network performance in digital networks, including ISDNs*. Geneva, 1993. ITU-T Recommendation I.350.

[M.1079] ITU. *Performance and quality of service requirements for International Mobile Telecommunications-2000 (IMT-2000) access networks*. Geneva, 2003. ITU-T Recommendation M.1079.

[X.200] ITU. *Open Systems Interconnection – Basic Reference Model*. Geneva, 1994. ITU-T Recommendation X.200.

[Y.e2eqos] ITU. *End-to-end QoS Architecture for IP Networks evolving into NGN COM 13-D537-E*. Geneva, 2004. ITU-T Draft Recommendation Y.e2eqos.

[Y.NGN-FRA] ITU. *Functional Requirements and Architecture of the NGN TD 38 (WP2)*. Geneva, 2004. ITU-T Draft Recommendation. Y.NGN-FRA.

[Y.NGN-GRM] ITU. *General Reference Model of the NGN TD 45 (WP2)*. Geneva, 2004. ITU-T Draft Recommendation Y.NGN-GRM.

[Y.NGN-Overview] ITU. *Overview, General Reference Model of the NGN TD 59 (WP2)*. Geneva, 2004. ITU-T Draft Recommendation Y.NGN.

[Y.NGN-MAN] ITU. *Framework for Manageable IP Networks TD 48 (WP2)*. Geneva, 2004. ITU-T Draft Recommendation Y.NGN-MAN.

[Y.NGN-MIG] ITU. *Migration of networks (including TDM networks) to NGNTD 39 (WP2)*. Geneva, 2004. ITU-T Draft Recommendation Y.NGN-MIG.

[Y.NGN-MOB] ITU. *Mobility management and its Architecture of the NGN TD 19 (WP3)*. Geneva, 2004. ITU-T Draft Recommendation Y.NGN-MOB.

[Y.100] ITU. *GII Overview*. Geneva, 1998. ITU-T Recommendation Y.100.

[Y.110] ITU. *GII Principles and Framework Architecture*. Geneva, 1998. ITU-T Recommendation Y.110.

[Y.1291] ITU. *An architectual Framework for Support of QoS in Packet Networks*. Geneva, 2004. ITU-T Recommendation Y.1291(Y.qosar).

[Y.1540] ITU. *IP Packet Transfer and Availability Performance Parameters*. Geneva, 1999. ITU-T Recommendation Y.1540.

[Y.1541] ITU. *IP Packet Transfer Performance Objectives*. Geneva, 2002. ITU-T Recommendation Y.1541.

Astrid Solem (40) graduated from the Norwegian University of Science and Technology in 1988. She has since then been working in Telenor, the first four years in the Network division and then in Telenor R&D where she has been working with technology and business aspects of network architectures for fixed and mobile networks on contract from Telenor business units and European Research projects. She is currently in charge of monitoring the work in Study Group 13 on behalf of Telenor.

astrid.solem@telenor.com

Evi Zouganeli (41) holds a Master of Management (2001, BI, Norway), a PhD in Opto-electronics (1992, U. College London), an MSc in Telecommunications (1987, U. College London), and a BSc in Applied Physics (1985, U. of Patras, Greece). After postdoctoral work at the Swiss Federal Institute of Technology (ETH, Zurich), she joined Telenor R&D in 1994, and has since worked on high capacity optical networks, network migration strategies and technology evaluations on contract from Telenor business units as well as in a number of European Research projects. She is currently a senior research scientist at Telenor R&D.

evi.zouganeli@telenor.com

Data exchange between operators

ARVE MEISINGSET

Arve Meisingset

Data may only be exchanged automatically between computer systems if the systems share a common understanding of the data. Therefore, ITU-T is providing Recommendation M.1401 Formalisation of designations for interconnections among operators' networks.

Telecom operators use electronic data interchange extensively with their customers, retailers and vendors, but they are not good at using the same medicine in their own interoperations. The reason for this situation is simple; incumbent operators have old proprietary data definitions that are not compatible with those of other operators. Therefore, in Europe an operator ordering a leased line via another operator is typically sending an informal facsimile or e-mail, the data are converted and registered manually in a computer system by the other operator, and so on. The US is in a better position, as after the divestiture of AT&T, Telcordia established a business on selling the company standard Common Language, much of this has been standardised by ANSI, and the trade organization NECA is managing common codes between operators. But without harmonized data definitions, Europe and others cannot do *e*Business between operators on an equal basis. However, if one operator is accepting to act as a customer to the other operator, he may use the other operator's proprietary standards and manually use his web interface.

For international manual interoperation, all operators use Recommendation M.1400 [1] Designations for interconnections among operators' networks. This Recommendation was in 2001 extended by ITU-T Study Group 4 to also apply to domestic use, and operators are during 2004 undertaking conversion of software and data in order to comply to this Recommendation. While originally, M.1400 applied only to a handful of international transmission centres by each operator, now it may apply to every junction box within its network, as this junction box may become an access point for another operator and become a site for telehosting of his equipment.

However, M.1400 is not as formal as required for computer systems to understand each other. Therefore, Telenor has contributed to the development of Recommendation M.1401 [2] Formalisation of designations for interconnections among operators' networks.

Work in ITU Study Groups is organised into Questions, and the relevant Questions here are

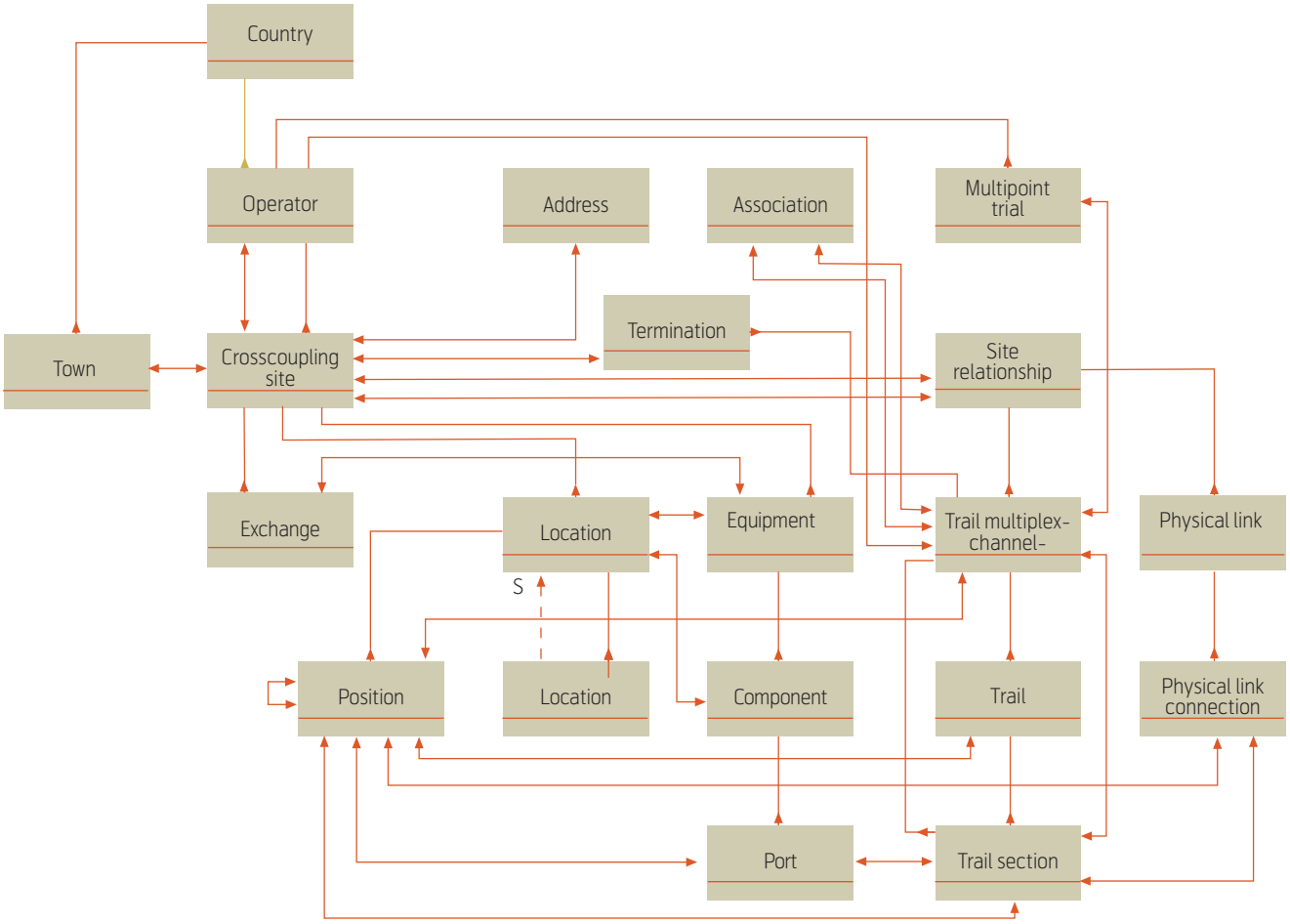
- Question 2/4 Designation for interconnections among network operators
- Question 9/4 Requirements for the TMN X-Interface.

Telcordia wants to define Common Language (CL) as a specialisation of M.1401. Hence, we may achieve a kind of global standardisation allowing for global *e*Business between operators. However, many details remain to be put into place to achieve this goal. Therefore, Telenor has defined data structures for the physical network resources (PNR) [3], as well. Telcordia will define Common Language by using the same specification technique and then try to map CL to M.1401 and PNR.

To define data is not straightforward. Some define information or concepts using class diagrams in the Unified Modelling Language (UML) from the Object Management Group (OMG). However, data may appear in many formats for different usage. There may be different formats for databases, like relational, networked or object based databases. There may be different formats for data communication, IDL for Corba, Coarsegrained IDL, XML and GDMO for CMIP. And there may be graphical and alphanumeric formats for end users of the same data. UML class diagrams do not provide sufficient richness to map between all these formats. Therefore, M.1401 and PNR use the HMI specification technique from Study Group 17 Z.350 series Recommendations [4], [5].

The HMI specification technique is used to specify the External Terminology Schema [6] for end users. This technique allows for using the end user's own words as class labels, and does not require use of 'computer speak'. This allows for company specific codes and mappings between national language ter-

ISSN 0085-7130 © Telenor ASA 2004



Recommendation M.1401 Formalisation of designations for interconnections among operators' networks. This Recommendation defines network resources that may be addressed in communication between network operators. The Recommendation defines object classes (boxes), name bindings (lines with reversed arrowheads), references (two-way arrows), identifiers (not depicted) and additional information (not depicted)

minologies, like French, Japanese etc [7]. The External Terminology Schema may be mapped into Internal Terminology Schemata for databases and data communication, e.g. in XML. UML class diagrams may not be needed or useful for this mapping.

As depicted in the figure above, cross-coupling sites are identified local to operators. Operators are assigned unique ITU Carrier Codes (ICC) within each country. These are assigned by the administration in each country. ITU-T has established a distributed web containing all authorised ICCs [8]. This includes the company codes from NECA, and countries are assigned three letter codes according to ISO 3166 [9].

References

1 ITU. *Designations for interconnections among operators' networks*. Geneva, 2001. ITU-T Recommendation 1400.

2 ITU. *Formalisation of designations for interconnections among operators' networks*. Geneva, 2004. ITU-T Recommendation 1401.

3 Meisingset, A. *Identification of Physical Network Resources*. Geneva, 26 April – 7 May 2004. ITU-T SG4 TD-7 (WP 1/4). URL: http://www.itu.int/md/meetingdoc.asp?type=mitems&lang=e&parent=T01-SG04-040426-TD-WP1-0007

4 ITU. *Data oriented human-machine interface specification technique – Introduction*. Geneva, 1993. Recommendation Z.351.

5 ITU. *Data oriented human-machine interface specification technique – Scope, approach and reference model*. Geneva 1993. Recommendation Z.352.

6 Meisingset, A. *Data Architecture*. ITU-T COM 17-C 84-E. URL: http://www.itu.int/md/meeting-doc.asp?type=sitems&lang=e&parent=T01-SG17-C-0084. Geneva, 09.01.2004.

7 Meisingset, A. *The HMI specification technique*. Kjeller, Telenor R&D, 1996. Telenor R&D note N 54/96.

8 ITU-T. *ITU Carrier Codes*. May 4, 2004. [online] – URL: http://www.itu.int/ITU-T/inr/icc/index.html.

9 ITU-T. *Guidelines for Data Conversion to Revised Recommendation M.1400. TSB Circular 183*. May 4, 2004. [online] – URL: http://www.itu.int/md/meetingdoc.asp?type=sitems&lang=e&parent=T01-TSB-CIR-0183.

Arve Meisingset (56) is senior research scientist at Telenor R&D. He is editor of Recommendation M.1401 Formalisation of designations for interconnections among operators' networks.

arve.meisingset@telenor.com

Where notations come from

ARVE MEISINGSET

Arve Meisingset

Many of the new notations in UML 2.0 come from the 2000 version of ITU languages, like SDL 2000 and MSC 2000.

The Unified Modelling Language (UML) family has been improved by many new notations in its version UML 2.0 from the Object Management Group (OMG). The improvements have been made by Thomas Weigert, Motorola rapporteur for SDL, Birger Möller Pedersen, Ericsson previous rapporteur for UML for SDL, and Øystein Haugen, Ericssson previous rapporteur for MSC carrying language notations from ITU-T Study Group 17 into OMG. We expect that UML 2.0 will be stable for some time and that tool vendors will need some time to deliver implementations. Meanwhile, Study Group 17 is providing UML profiles of all its language notations. Also, Study Group 17 provides a more stable environment for formal definition and maintenance of the languages than provided by OMG.

Parts of UML have grown popular in requirement and early design specifications of most kinds of software. However, UML is typically too informal for implementations and even for mapping to implementations. Note, however, that Study Group 17 has provided Extended ODL, ie. eODL, as an attempt to map from computational to engineering specifications. Those who want notations for more than just a means of creating illustrations may find much of interest in the ITU-T language family.

The following presentations of the ITU languages are taken from the SG17 web [1].

SDL (Specification and Description Language) is used to specify the behaviour of complex systems. SDL can be applied at various levels: from simple diagrams that illustrate the prose description of a protocol, through models that can simulate behaviour at normative interfaces, to actual implementation descriptions [2]. SDL is defined in the Z.100 series Recommendations [3] and is addressed in Question N/17 Specification and Implementation Languages [4].

eODL (extended Object Definition Language) is used to map between computational and engineering languages. This allows for component implementations and deployment in various implementation languages [5]. eODL is defined in the Z.130 series Recommendations [3] and is addressed in Question N/17 Specification and Implementation Languages [4].

MSC (Message Sequence Chart). MSCs are used to show interactions between system components. MSC diagrams provide a clear description of system communication in the form of message flows [6]. MSC is defined in ITU-T Recommendation Z.120 [3] and is addressed in Question O/17 Requirement Languages [4].

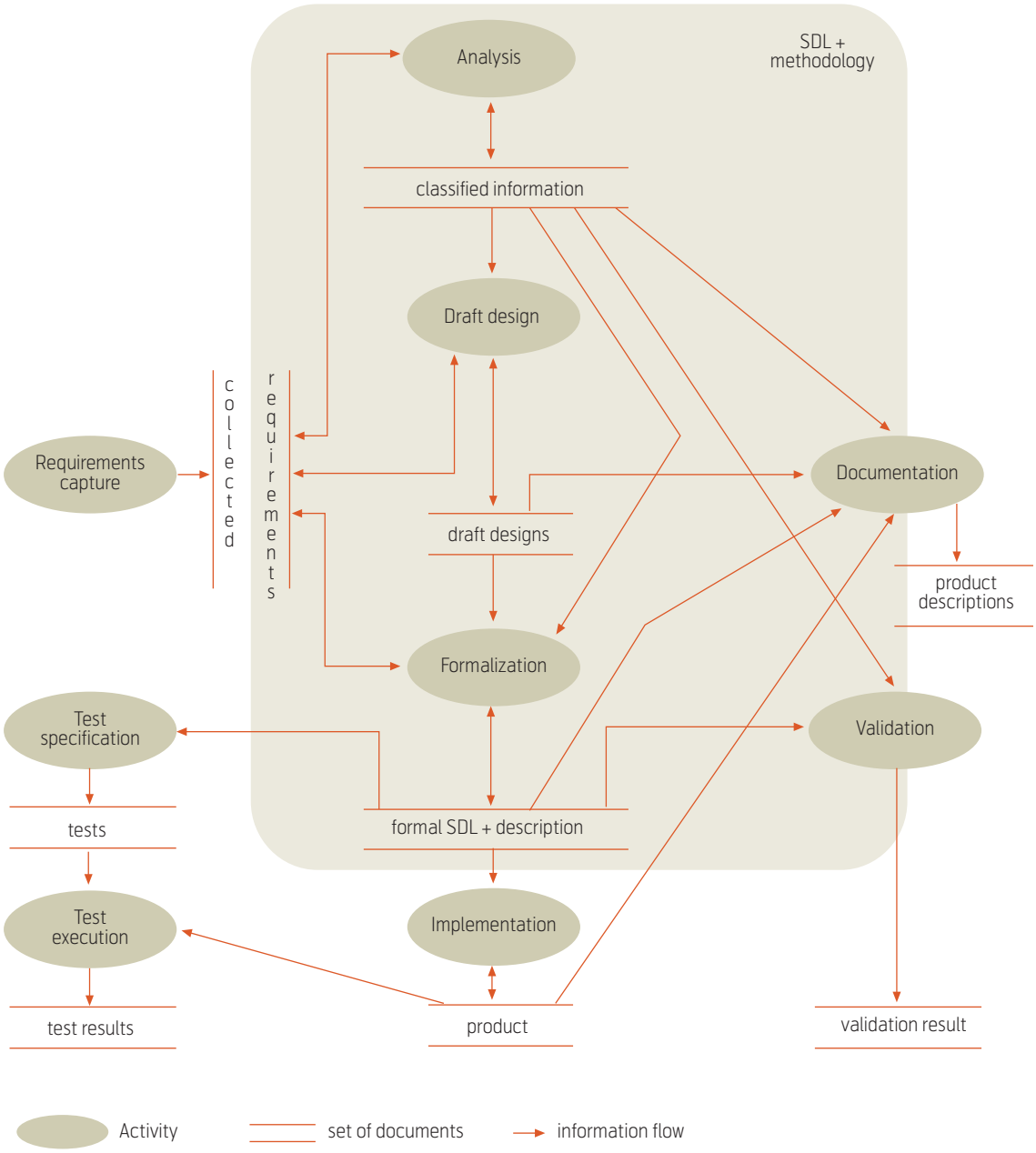
URN (User Requirement Notation) provides both a Goal-oriented Requirement Language (GRL) and a Use Case Map Notation (UCM) [7]. URN is defined in the Z.150 series Recommendations [3] and is addressed in Question O/17 Requirements Languages [4].

TTCN (Testing and Test Control Notation). The latest version of the language is TTCN version 3 (TTCN-3). The work of drafting this version was done within ETSI and the results were contributed to ITU-T. Typical areas of application for TTCN-3 are protocols, services, APIs, software modules etc. TTCN-3 is not restricted to conformance testing. It can be used in many areas, for example interoperability testing, robustness testing, performance testing, regression testing, system testing and integration testing [8]. TTCN is defined in the Z.140 series Recommendations [3] and is addressed in Question Q/17 Testing languages, methodologies and framework [4].

ASN.1 (Abstract Syntax Notation One) is an international standard for describing the structure of data exchanged between communicating systems. ASN.1 has been extensively used in telecommunication standards. Notable examples are protocol standards for intelligent networks, UMTS (3G), Voice over IP, Interactive television and HiperLAN/2. The use of ASN.1 goes well beyond telecom standards and products [9]. ASN.1 is defined in the X.690 series Recommendations [3] and is addressed in Question M/17 Abstract Syntax Notation One (ASN.1) and other Data Languages [4].

In particular for ASN.1, Study Group 17 has been reviewing use of the notation in any standard and

ISSN 0085-7130



Recommendation Z.100 Supplement 1: Methodology for use of MSC and SDL (with ASN.1). This supplement provides guidelines for how to develop and test real-time systems and recommend notations to be used in each phase (box). Extensions to other kinds of applications are e.g. provided through the work on UML for ODP. The Open Distributed Processing framework is already incorporated into the methodology

provided the definitions in the ITU Formal Language Database [10].

Study Group 17 is developing Recommendations jointly with ISO/IEC/JTC 1. As an example of this collaboration, they are developing 'UML for ODP', i.e. guidelines on how to use UML-based notation in the various viewpoints of the Open Distributed Processing framework [11]. Also Study Group 17 is investigating whether to standardize Model Driven Architecture for telecommunications based on results from EU projects.

Throughout the last four year Study Period, Study Group 17 has been organising Workshops at each study group meeting, addressing aspects of language co-ordination. Documentation from these can be found at the Study Group 17 web site [1]. See under Workshops and Seminars.

Also, Study Group 17 has been organising Tutorials on hot issues during every Study Group meeting [12].

And Study Group 17 organise and collaborate with several fora.

Several of the Study Group 17 Recommendations are available online [13] free of charge.

References

1 ITU-T. *ITU-T Study Group 17 (Study Period 2001–2004)*. May 4, 2004. [online] – URL: http://www.itu.int/ITU-T/studygroups/com17/index.asp.

2 Reed, R. *SDL-2000 for new millennium systems*. May 4, 2004. [online] – URL: http://www.itu.int/itudoc/itu-t/com17/tutorial/ 78255.pdf.

3 ITU. *ITU-T Recommendations. Series Z*. May 4, 2004. [online] – URL: http://www.itu.int/rec/recommendation.asp?type=products&lang=e&parent=T-REC-Z.

4 ITU-T. *Report of the plenary of Study Group 17*. Geneva, 10–19 September 2003. COM 17 – R 13 – E. May 4, 2004. [online] – URL: http://www.itu.int/md/meetingdoc.asp?type=sitems&lang=e&parent=T01-SG17-R-0013.

5 Fischer, J. *Integrated Application of eODL*. May 4, 2004. [online] – URL: http://www.itu.int/ITU-T/worksem/iafl/documents/iafl_006.ppt.

6 Jervis, C. *Message Sequence Charts*. May 4, 2004. [online] – URL: http://www.itu.int/itudoc/itu-t/workshop/joint/ s4p1.html.

7 Amyot, D. *Integrated Application of URN*. May 4, 2004. [online] – URL: http://www.itu.int/ITU-T/worksem/iafl/documents/iafl_005.ppt.

8 Hogrefe, D. *Integrated application of TTCN-3*. May 4, 2004. [online] – URL: http://www.itu.int/ITU-T/worksem/iafl/documents/iafl_004.ppt.

9 Larmouth, J. *ASN.1 Today and Tomorrow*. May 4, 2004. [online] – URL: http://www.itu.int/itudoc/itu-t/com17/tutorial/78247.html.

10 ITU-T. *ITU Formal Language Database*. May 4, 2004. [online] – URL: http://www.itu.int/ITU-T/formal-language/index.html.

11 Wood, B. *UML for ODP viewpoint specifications*. May 4, 2004. [online] – URL: http://www.itu.int/itudoc/itu-t/com17/tutorial/81998.html.

12 ITU-T. *Tutorials*. May 4, 2004. [online] – URL: http://www.itu.int/itudoc/itu-t/com17/tutorial/index.html.

13 ITU-T. *ITU-T Study Group 17 – Languages for telecommunication systems*. May 4, 2004. [online] – URL: http://www.itu.int/ITU-T/studygroups/com17/languages/index.html.

Arve Meisingset (56) is senior research scientist at Telenor R&D. He is Vice Chairman of Study Group 17, organiser of many of the Study Group Workshops and Tutorials, and rapporteur of Question R/17 Open Distributed Processing (ODP).

arve.meisingset@telenor.com

Telecommunication for disaster relief

OLE GRØNDALEN

Ole Grøndalen

ITU-T Study Group 16 Multimedia is addressing a set of special applications, one of these being Telecommunication for disaster relief.

A new trend in standardization is to consider special applications. An example of this is ITU-T Study Group 16's question I/16, which has been established to address public telecommunication services that authorities can use to communicate during emergency and disaster operations. This capability, referred to as the emergency telecommunication service (ETS) or telecommunications for disaster relief (TDR), will enable communications from authorized users to have preferential treatment for organizing and coordinating disaster relief operations.

TDR addresses the need of authorized users in terms of facilities established in the public network infrastructure, including the inter-working aspects with dedicated/private networks. TDR work does not specifically address systems for the use of the public in general (e.g. emergency numbers 112/911 and broadcasting network to forward emergency relevant information to the public). Since ETS is more generic, TDR is the preferred term in order to avoid the confusion with the systems described above.

There is a variety of telecommunication capabilities for use in emergency situations. The scope of question I/16 addresses the capabilities specifically for authorities to use public telecommunication services for emergency and disaster relief operations. However, it is recognized that some of the technical solutions emerging from this work could be applicable to other emergency telecommunication capabilities. In addition, consideration could be given to possible interfacing between public networks providing TDR capabilities and dedicated systems used by authorities during disaster relief operations. Standards that emerge from this work are intended to apply to international TDR traffic.

The development of TDR capabilities is being addressed by many standards development and disaster relief organizations. Therefore, cooperation and liaison (coordination) between the many organizations representing different interest areas is essential to ensure consistency and completeness in the provisioning of effective telecommunication capabilities to support emergency and disaster relief operations. To support cooperation between the players involved in TDR, ITU has created an information exchange platform named Telecommunication for Disaster Relief and Mitigation Partnership Coordination Panel (TDR-PCP). Among its aims are to monitor the progress of technical standardization work and map the requirements of different users. The platform will enable an invaluable dialogue between standards-development organizations and, most importantly, the users of the equipment that will enable efficient preparation and reaction to disaster events.

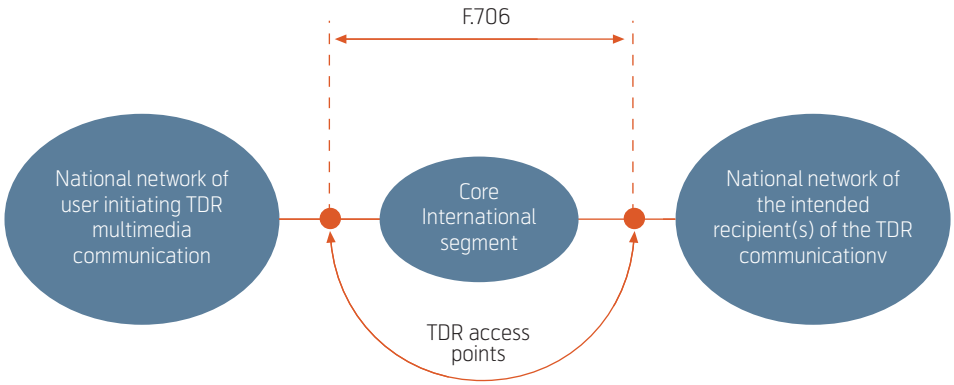
The work on TDR in SG 16 will be coordinated with the work in other parts of ITU, in particular with the work performed under the responsibility of ITU-T Study Group 2. Two new recommendations are expected to be developed in SG16. The first will be a "TDR Requirements" Recommendation that will define TDR aspects of multimedia applications and services. The second will be a "TDR System Framework" Recommendation that will explain how network, service, and operational capabilities can be used to allow TDR communications originating from authorized users to have preferential access to network resources.

Figure 1 illustrates a possible multimedia TDR core topology. The core TDR functions operate between the designated TDR access points of the communicating national networks in accordance with the operational agreements established between the administrations involved. Only authorized users can initiate a TDR communication, the recipient(s) can be any user. TDR traffic passing through any transit country should be treated transparently and may not receive preferential treatment, but the session management information should be preserved whenever possible.

The core TDR multimedia service is expected to include features like:

- Identification of TDR communications from any (or predetermined) access point;
- Authentication of TDR users from any access point;

ISSN 0085-7130 © Telenor ASA 2004



Figur 1 Possible multimedia TDR core topology

- Provision of preferential processing;
- Provision of priority treatment;
- Implementation of ancillary service features and, for example, the protocol mechanisms and routing procedures required by the service provider to offer a particular TDR feature.

Ole Grøndalen (42) is a research scientist at Telenor R&D. He follows the work in SG16, but is himself not involved in the TDR work. This paper is provided to give a simple introduction and overview on TDR standardization work in Study Group 16.

ole.grondalen@telenor.com

ISSN 0085-7130 © Telenor ASA 2004

Wireless World Research Forum (WWRF)

ERIK LILLEVOLD

Erik Lillevold

In May 2000 the joint activity Wireless Strategic Initiative (WSI) was started by Alcatel, Ericsson, Nokia and Siemens. The main objective of the four largest wireless equipment manufacturers in Europe was to develop a vision of the total wireless communication system that may follow third generation, i.e. beyond the specifications made by 3GPP. As an indication of a time frame for such a system it is planned to be fully operational by 2012. The system is termed the Wireless World and its Reference Model and Architectural Framework models are expected to become important guidelines for the future research work in this field. As the necessary enabling technologies become mature, it is expected that the Wireless World will be realised and commercialised in steps.

Abstract

The organisation and the results obtained so far by WWRF are described. The main result is the analytic work of a large amount of projects done within the EU IST Framework Program in later years. These analyses are used as the foundation for the *"Book of Visions"* which is a living document that will be updated in new and better releases as time goes by.

The results of WWRF described below are based on *"Book of Vision 2001"* and working documents in the WGs. The most important work so far is done by WG1 and WG2 and the outcome has been a "WWRF Common Reference Model", a "User Centric Reference Model", a "Service Architecture Reference Model" and an evaluation of several "Enabling Technologies". The results of WWRF described below are based on *"Book of Vision 2001"* and working documents in the WGs. Of the enabling technologies agent technology seems to be the most valuable due to its maturity and flexibility.

Introduction

The WSI project, which was partly funded by the EU IST 5th Framework Program, decided to make the process to develop the vision totally open and invited everybody who wanted to contribute. The open approach was a major success and the Wireless World Research Forum (WWRF – http://www.wireless-world-research.org/) was created in early 2001.

It soon became obvious that some structuring principles were needed to guide the work of the forum and its working groups. It was therefore decided to define a reference model; i.e. a structural framework for the definitions and research on the Wireless World. The idea is to achieve the same benefits of the WWRF reference model to the WWRF as the Open System Interconnections (OSI) reference model had for the development of data communications from the mid-1970s. OSI was by its time probably the most important structuring scheme used in communication engineering as a research tool, but rarely used for direct implementations of products.

This paper describes the WWRF reference model and the models derived from this overall model in the respective working groups, i.e. the WWRF current proposals of a structuring scheme for wireless communication, which is expected to be able to support the definition of the complex mobile communication concepts and the structuring of the research work on the Wireless World.

The objective of the forum is to formulate visions on strategic future research directions in the wireless field, involving industry and academia, and to generate, identify, and promote research areas and technical trends for mobile and wireless system technologies.

The Book of Visions 2001 is the main outcome of the work of WWRF together with a range of working documents in several working groups and Special Interest Groups (see Figure 1). The work done in WG1 and WG2 represents the more overall view of the system and appeals to the application and software developers. WG3 – WG7 are for network and radio specialists and have lately been reorganised from originally two groups called "New Communication Environment and Heterogeneous Networks" and "Spectrum, New Air Interfaces and Ad-hoc Networking". Below is the work done by WG1 and WG2 described in more details since they have the greatest impact on the WWRF vision. This does not mean that the work done by the rest of the groups is more important, but rather not so important to get an overall view of the Wireless World.

ISSN 0085-7130 © Telenor ASA 2004

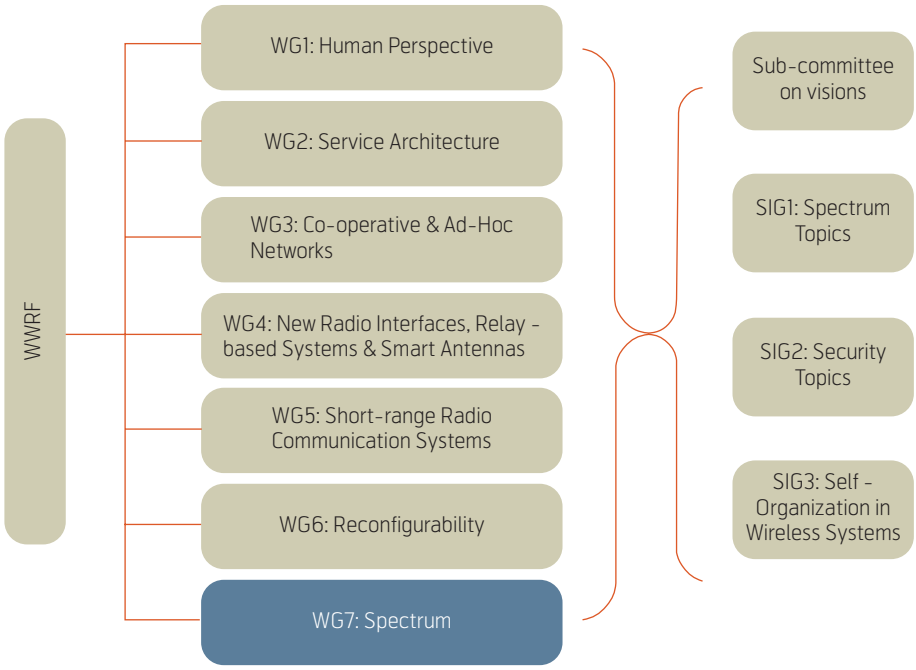


Figure 1 Organisation of the current work within WWRF

The WWRF common reference model

During the discussions to prepare the visions of a Wireless World it became apparent that a reference model was needed. As a result the MultiSphere model was sketched out. It is a common model that should assist in putting the issues and ideas into a common context.

Driven by the "horizontalisation" of system functionality introduced by 3G's mobile Internet, it is believed that future vertical applications and services will be composed in an ad-hoc manner by a multitude of wireless technologies. Those elements will be around us like a number of spheres in which we live.

The common reference model of WWRF therefore puts the human being in the centre surrounded by six spheres of equipments, networks and functions (see Figure 2).

They are from the inner to the outmost sphere:

1. *PAN (Personal Area Network):* This sphere contains the data components that are closest to the user and that he/she carries all the time, e.g. mobile phones, watches, cameras, glasses, and other body-near equipment.

2. *The immediate environment:* This is the immediate surroundings consisting of fixed or little movable equipments, e.g. TV, PC Workstation, refrigerator, etc.

3. *Instant partners:* This spherical level is reached when equipment is added, which we need to communicate and interact with people around us.

4. *Radio access:* The sphere of equipment and networks needed to obtain ubiquitous coverage for distributed systems. It is a fundamental requirement that all radio interfaces shall offer access for PAN and "Instant Partners", i.e. spheres 1 and 3.

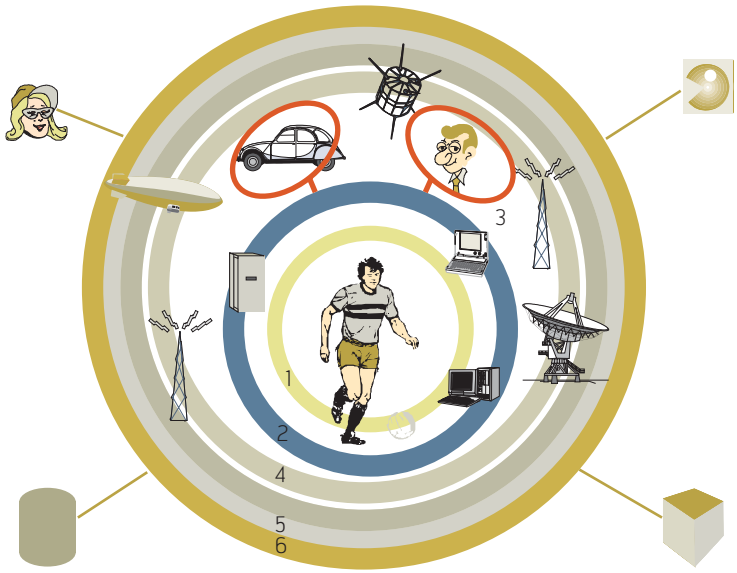


Figure 2 WWRF MultiSphere Reference model

ISSN 0085-7130 © Telenor ASA 2004

5. *Interconnectivity:* This sphere is a functional layer that also gives interconnectivity to mobile systems. The user is offered a real mobile Internet independent of radio access network and terminal.

6. *Cyberworld:* This is the outermost sphere that makes up an enhanced reality (Cyberworld) created by all the applications offered in next generation mobile system. Today's Internet and Web is just a start of what will come. Step-by-step we will be inhabitants of the Cyberworld, where our interests, needs and wishes will be maintained by individual agents, which are autonomous data programs in the complex system represented by the Wireless World.

The building blocks of the Wireless World

The previous chapter describes a user-centric view of the Wireless World, using the sphere model. A system view emerged when the visions and issues were further analysed in the different working groups.

WWRF has so far identified nine major system elements or functions that characterise and make up the Wireless World (see Figure 3):

1. *Augmented Reality/Cyberworld:* The reality seen by the user will be enhanced with useful and powerful information and other computerised services.

2. *Semantic Aware Services:* The information and services offered should be adapted to the user's needs and wishes. The adaptation is based on personal profiles and the contextual conditions of the user, e.g. location and day.

3. *Peer Discovery:* The Wireless World system must have mechanisms to localise and register user services, and connect other users. Addressing is a very important functionality.

4. *End-to-end Security & Privacy:* If vital services as e.g. payment and health services shall be commonly used, security and personal information protection must be better than today.

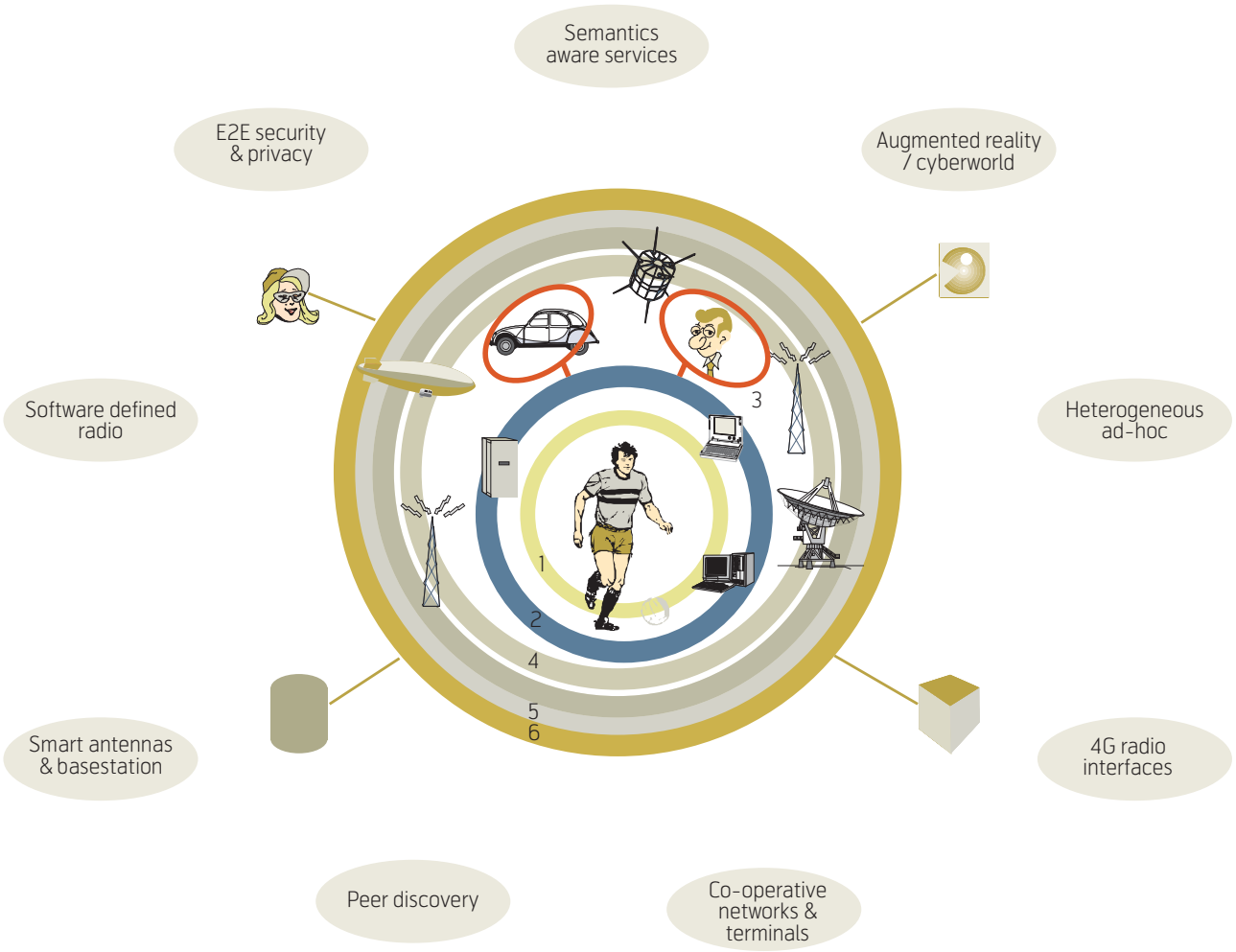


Figure 3 The building blocks of the Wireless World

5. *Co-operative Networks & Terminals:* The mobile system will contain a number of heterogeneous network and computer technologies that have to be co-ordinated to enable seamless user services.

6. *Heterogeneous Ad-Hoc Networking:* The communication network of the Wireless World will include ad-hoc elements that collaborate to construct network islands of increased direct communication needs, e.g. such configurations are supposed to appear at hot-spots like airports, shopping malls, etc.

7. *Radio Interfaces:* Different radio technologies, which are tailored to certain environments need to be defined for application in the Wireless World network. Their spectral co-existence needs to be guaranteed by defining appropriate rules for frequency etiquette.

8. *Smart Antennas & Base stations:* R&D in antenna and base station technology should be evaluated and promoted to contribute significantly to the future radio access technology. Among potential technologies are Smart Antennas, High Altitude Platforms, Beam forming, Multiple Input Multiple Output (MIMO) Channel, Space Time Coding, Radio heads and optical fibre

9. *Software Defined Radio:* This technology is a key enabler for flexible network architecture, allowing an easy adaptation to the application's demands. Thus, it ensures a future proof network architecture, which can keep pace with the application innovation process by changing the mobile station's protocol stacks remotely.

The WWRF reference model with building blocks is used by the different working groups as a common overall model in their work, which is a more detailed work within specific areas of the Wireless World.

Results from Working Groups 1 and 2

In the same way as the WWRF forum felt they needed a reference model to structure their discussion, some of the specific Working Groups also have their own reference models, which are in line with the common WWRF reference model. These models are briefly described below.

WG1: Human Perspective

This group has made a User Centric Reference Model for the Cyberworld and its services shown in Figure 4. This sees the individual in different perspectives, the Value Plane, the System Capability Plane and the Cyberworld.

Meeting the needs described in the value plane demands that the system provides a set of basic functionalities to the user. These are described as components within the capability plane. Our initial analysis led us to group the capabilities into six focus areas: ubiquitous communications, presence awareness, personalisation, natural interaction, ubiquitous information, and context adaptation. The details of each of these components and their inter-relationships are described in this section.

- *The Value Plane* describes the core human needs that wireless systems must satisfy in order to be successful. Research on technology and changes in the user's behaviour has established a conviction that successful applications and services must address and support the human values to be successful. WG1's initial analysis has indicated that the core needs may be grouped into six focus areas: safety, belonging, privacy, control, self-actualisation and human capability augmentation. WG1 will continue to work on the details of each of these components and their inter-relationships.

- *The System Capability Plane* provides a set of basic functionalities to the user in order to meet the needs described in the value plane. These capabilities are described as components within the capability plane. Our initial analysis led us to group the capabilities into six focus areas: ubiquitous communications, presence awareness, personalisation, natural interaction, ubiquitous information, and context adaptation. The details of each of these components are also for further studies in WG1.

- *CyberWorld* is the augmented reality made available to us by the future mobile and global information and communication system. It is likely that our presence in CyberWorld will soon be almost as important to us as presence in the real world. A deep understanding of this "world" is necessary to develop Wireless World technologies that satisfy our fundamental needs. CyberWorld is made up of five functional components:

  - *Presence*, which offer information about services and people who are available to the user.

  - *Identity*. This identifies users precisely, and is the foundation for security and privacy.

  - *Interaction*. This is about user interfaces (e.g. multi-modality), localisation (e.g. position and orientation) and other things that are needed to enable valuable augmented reality for the user.

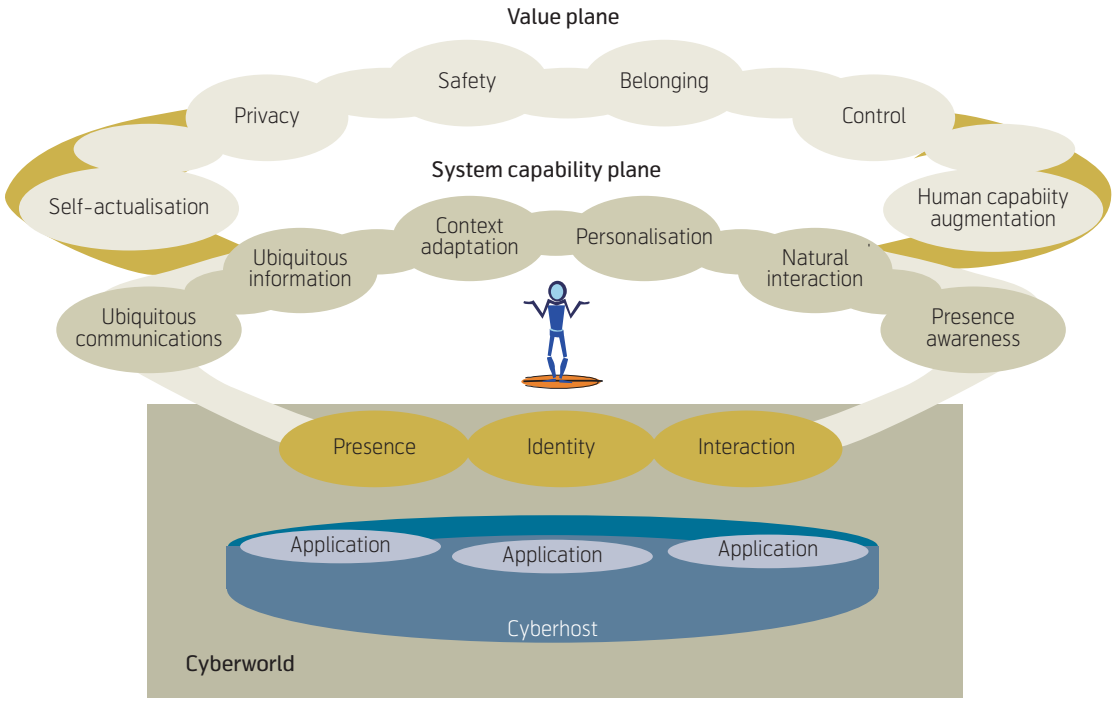


Figure 4 WG 1 User Centric Reference Model

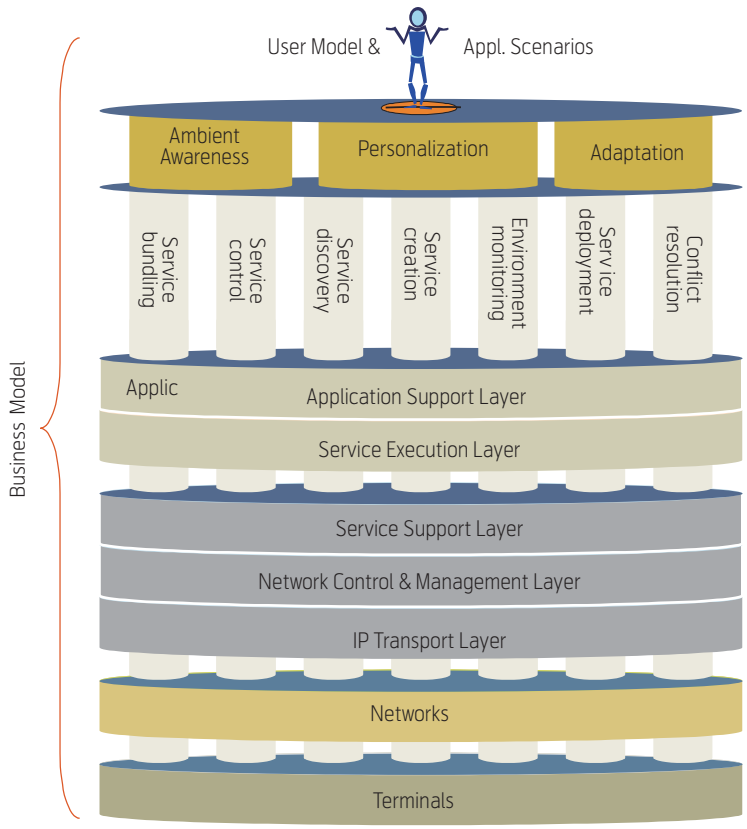


Figure 5 WG2 Service Architecture Reference Model

- *Applications*. They are built to make an instantaneous realisation of services, which help the user achieve his/her tasks and goals.

- *Cyberhost*. This is the part of the system that offers an apparently "local" run-time environment for the mobile applications and services.

WG 1 uses this reference model to structure the discussion about the human perspective of the visions of the Wireless World.

WG2 – Service Architecture

This working group has made their own reference model for the service architecture of the Wireless World based on the common reference model and anything else envisioned in the *"Book of Visions"*.

The WG2 Service Architecture reference model

The model is simple, but still seems to cover most of the architectural features required by the Wireless World vision.

The model is both horizontally and vertically layered. The horizontal functional layering is much the same as the classical Open Systems Interconnection (OSI) made by the International Standardisation Organisa-

tion (ISO) in the 1980s. This means that functions/services created in one layer of the WG2 reference model is based on the services/functions from the layer beneath and offered to the layer above. The vertical layers are made up by entities from several horizontal layers to build functional components like Service Discovery, Service Control, Service Creation, etc. These maybe viewed as system applications and services used by developers, not by end users.

The model also emphasises that development of applications should be made through application scenarios and user models. The tools to be used could be those recommended from the WWRF WG1.

The horizontal layers are separated by interfaces that may be standardised and open. Such interfaces are crucial to which business models can be developed.

Enabling technologies

To realise systems based on the WG2 reference model, which is an open service architecture with ubiquitous personalised and context dependent services, it will be necessary to have advanced software technologies. Working Group 2 also has the responsibility to analyse and recommend the development of existing and new technologies of relevance. WG2 is watching closely the following technologies:

- Open Interfaces (Parlay, OSA, JAIN)
- XML and beyond (RDF, ebXML, SOAP, Web Services, Semantic Web, etc)
- Agent technology
- Peer-to-peer (p2p)
- Plug and Play Technologies (Jini, UPnP)
- Aspect-oriented Programming (AOP)
- Event-based programming

Parlay/OSA and XML based Web Services are already state of the art for realising telecom and global Internet services respectively. These will obviously be fundamental to next generation services. However, as outlined above new software technology is needed to give support for a certain minimum of "intelligent" functionality, i.e. adapt services to the current situation and environment of the user. Semantic Web, Agent Technology, p2p, Adaptive Grid and maybe more other technologies are examples of technologies trying to meet the requirements of the next generation services.

Of these, agent technology seems to be central due to its maturity and flexibility, i.e. easy integration with one or more of the other technologies. It is called Agent Based System (ABS) technology and may be an important enabler for highly distributed, complex, self-organising, and collaborative systems. ABS technology has the ability to communicate, co-ordinate and associate the usage of learning, do scheduling and other advanced techniques like matchmaking and even move agents to other platforms.

Conclusions

The WWRF is definitely not the only forum or consortium trying to look into the future of information and communication technologies (ICT) and systems. However, it is one of the few that gather so many of the most important telecommunication vendors (Nokia, Ericsson, Motorola and Siemens) paired with a long range of universities, operators and R&D institutions in a long term vision of the ICT systems, an evolution in the time span until 2010 and beyond including 3G and next generations mobile systems.

It is expected that WWRF will be a major contributor to critically analysing and coordinating the work on future telecom and data communication. It is also expected that the result of the work may be stepwise commercialised in the time frame of their visions.

Acronyms

3GPP	3rd Generation Partnership Project
ebXML	electronic business and XML – stands for a global standard for business-to-business communication
JAIN	Java in Advanced Intelligent Networks – an initiative to provide open solutions based on Java technology to the Telecom Intelligent Network environment
Jini	Is not an acronym. Jini is a network architecture for the construction of distributed systems where scale, rate of change and complexity of interactions within and between networks are extremely important and cannot be satisfactorily addressed by existing technologies. Jini technology provides a flexible infrastructure for delivering services in a network and for creating spontaneous interactions between clients that use these services regardless of their hardware or software implementations.
MIMO	Multiple Input Multiple Output
OSA	Open Service Access

Parlay	The Parlay Group is an open multi-vendor consortium formed to develop open technology-independent application programming interfaces
RDF	Resource Description Framework – a general purpose language for representing information in the Web
SIG	Special Interest Group
SOAP	Simple Object Access Protocol
UPnP	Universal Plug and Play
XML	XML – Extensible Markup Language – a document format for the Web that is more flexible than HTML because it allows tags to be defined by the developer of the page

Sources

Wireless Strategic Initiative. May 5, 2004. [online] URL: http://www.ist-wsi.org/

Wireless World Research Forum. May 5, 2004. [online] – URL: http://www.wireless-world-research.org/

Book of Visions 2001. May 5, 2004. [online] – URL: http://www.wireless-world-research.org/general_info/default.htm

---

Erik Lillevold (61) obtained his MSc in Physics from the Norwegian University of Science and Technology (NTNU) in 1970 and joined the Norwegian Defense Research Institute (NDRE) in 1971. He worked there as a research scientist until 1986 when he joined Telenor R&D. He has most of his time worked in the field of messaging and application services, e.g. X.400, Internet Mail, WWW and WAP. In the last few years he has devoted his work to Open Service Platforms (Parlay, OSA, Web Services, etc.).

erik.lillevold@telenor.com