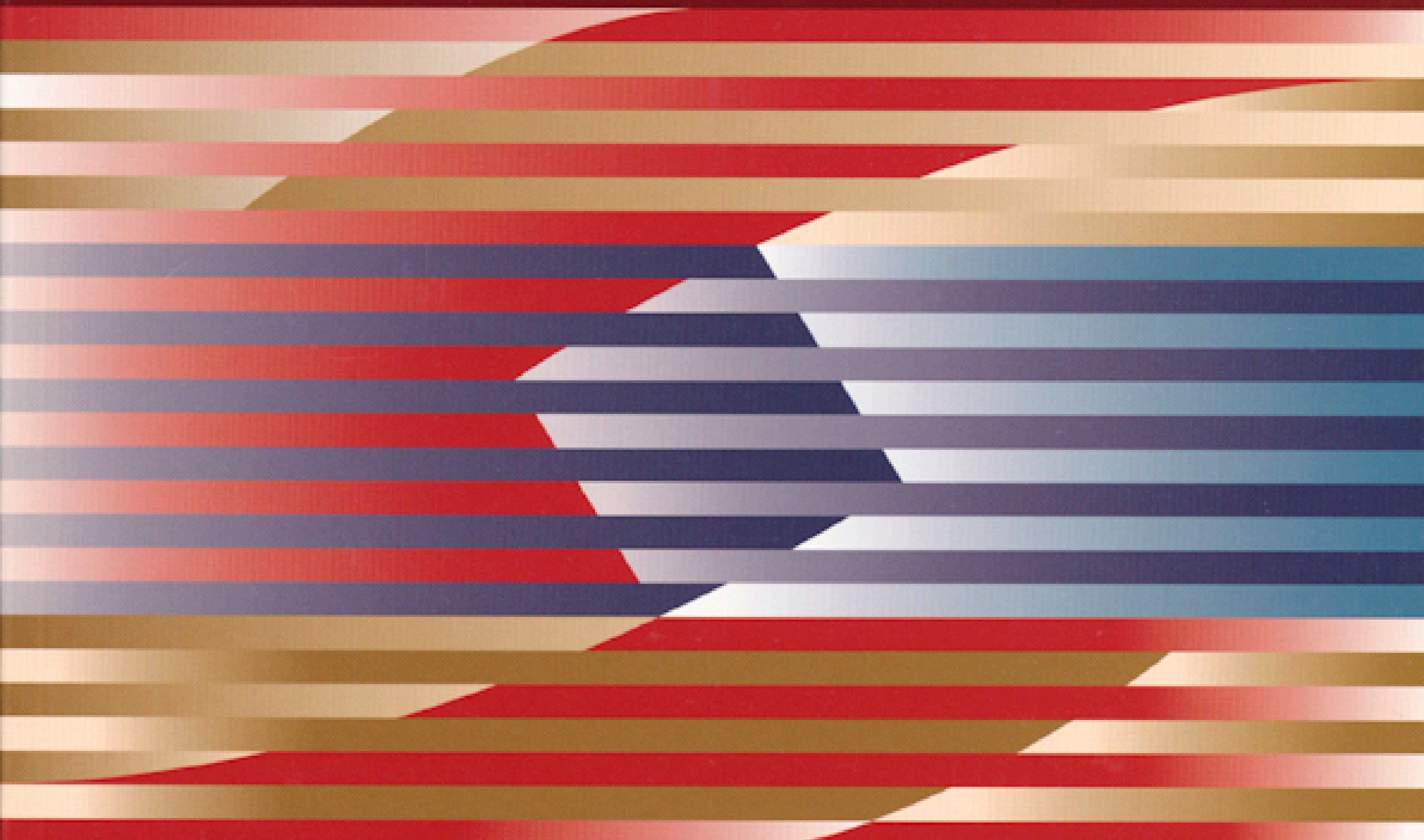
Telektronikk 2.98

Personal Communications



Contents

Feature

Status

Kaleidoscope

Telektronikk

Volume 94 No. 2 – 1998
ISSN 0085-7130

Editor:
Ola Espvik
Tel: (+ 47) 63 84 88 83

Status section editor:
Per Hjalmar Lehne
Tel: (+ 47) 63 84 88 26

Editorial assistant:
Gunhild Luke
Tel: (+ 47) 63 84 86 52

Editorial office:
Telektronikk
Telenor AS, Telenor Research & Development
P.O. Box 83
N-2007 Kjeller, Norway
Tel: (+ 47) 63 84 84 00
Fax: (+ 47) 63 81 00 76
e-mail: telektronikk@fou.telenor.no

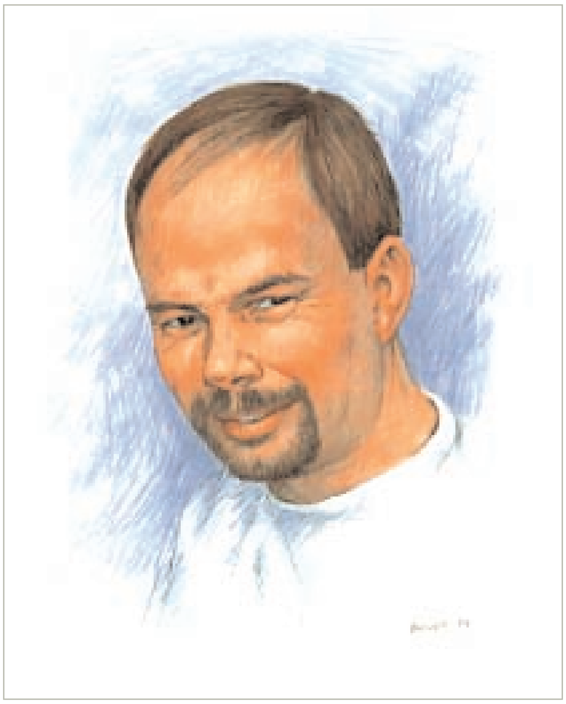
Graphic design:
Design Consult AS

Layout and illustrations:
Gunhild Luke, Britt Kjus, Åse Aardal
Telenor Research & Development

Telenor

Guest editorial

RUNE HARALD RÆKKEN



Tailoring of services, service mobility, personal mobility and terminal mobility are essential elements in *personal communications*. With the integration of fixed and mobile services, and interworking between different technologies, personal communications will soon be available. The transition from plain old telephony to true personal terminals will then become a major step in the history of telecommunications. The mobile operators may within few years change from being mobile network operators to being providers of value added services to personal terminals.

The European Commission's Green Paper on mobile and personal communications (1994) said: "... personal communications has the ultimate potential to reach near 80 % of the population (ie. up to one connection per adult)." This is far beyond today's penetration of wireline phones, even in industrialised countries. With the introduction of personal multimedia communications the vision of 'the wireless' as the customer's first choice will become true.

This issue focuses on the transition from mobile telephony to personal communications. We are used to seeing people communicating via true pocket cellular phones in almost any situation. Mobile communications has, during the last decades, evolved from being the businessman's tool to becoming the consumer's means of receiving and making calls when on the move. This highly affects the social life of the mobile users, giving them the possibility of being available whenever wanted – avoiding 'being held prisoner' by the wireline phone.

The world-wide success of mobile communications, based on the GSM standard, has led to a tremendous interest for mobile systems. The growth of cellular services has been boosted by mobile communications being exposed to competition. New operators are entering the scene, and their best means to get a piece of the still growing telecommunications market is by use of radio solutions.

Until now the main cellular service has been speech, even though data services have been present in the operational networks for several years. However, over the last years the focus of the mobile society has changed from speech to data services due to the tremendous growth in the demand for information services and Internet access.

The next step will be integration of the different fixed, cellular and Internet services. Mobility functions in both wireless and fixed networks, and flexible service creation and management makes the differences between fixed services, cordless telephony and cellular diminish. Introduction of packet switched data into the GSM network, like the General Packet Radio Service (GPRS) is turning the original circuit switched GSM network into a hybrid network.

The standardisation of a third generation cellular system – Universal Mobile Telecommunications System (UMTS) has been going on for several years. In addition, real broadband mobile systems offering bit rates of up to 155 Mbit/s are being developed.

UMTS is aimed at giving access to multimedia applications. It predominantly evolves from GSM, and will interwork with GSM, possibly also picking up elements from the DECT standard. In the longer term, integration between UMTS and IP networks offering differentiated quality of services will be the next evolutionary step. UMTS may then give global access to multimedia services across platforms like mobile, fixed and satellite-based networks.

Mobility across terminals, locations and infrastructure together with tailoring of high quality services will make the vision of *communicating anytime, with anyone, anywhere* come through. The broadening of mobile communications into general personal communications will substantially affect the lives of individual citizens as well as the functioning of the society.

From mobile telephony to personal communications

JOAR LØVSLETTEN AND RUNE HARALD RÆKKEN

Mobile telephony – a retrospective glance

From land coverage to cellular services

Mobile communications in the form of communications between a base station and a number of mobile stations has been available for several decades. Communications between land and ships using short wave radio is an example of such an application that has been in use in the last century. Early communications took place using Morse telegraphy; later, it was changed to voice telephony.

Land mobile applications have been available since the 1930s, in the form of closed user groups connected to a base station. The police were among the pilot users of such services.

Since then, several public land mobile services have been launched. In Norway a service called OLT (Offentlig Landmobil Tjeneste = public land mobile service) was launched in the late 1960s. All subscribers to the OLT service listened to the call channel and were requested by an operator to change to a traffic channel when a call was received. To set up a call to a mobile subscriber, one needed to know the approximate position of the mobile subscriber, in order to direct the call to the appropriate paging area.

It soon became evident that this was not the way to offer mobile communications to the mass market. Capacity problems were envisaged due to limited available radio spectrum. Hence, it was regarded an advantage to be able to use a cellular layout of the base stations to allow reuse of the radio frequency spectrum to raise the total system capacity. During the same period, the 1970s, switchboard operators were removed, and calls in the wireline networks were set up automatically. Hence, the idea of forming an automatic international mobile network emerged in the Nordic telecommunications administrations. Among the objectives for the system were [1]:

- The system should automatically set up and charge a call
- It should be possible to communicate between any mobile subscriber and wireline subscriber, or between any two mobile subscribers
- The system should operate in all the Nordic countries
- The use of the mobile phone should be as similar as possible to the use of the wireline phone
- The mobile subscriber should have access to as many as possible of the wireline services
- Calls to the mobile should be automatically routed to the position of the mobile, the mobile should be roaming in the network and data bases should keep track of its position
- The system should be able to automatically switch from one base station to another during conversation.

Such characteristics of a mobile communications system seem inevitable today, but particularly the two last bullets were quite revolutionary in the late 1970s.

The first steps towards the Nordic Mobile Telephone (NMT) system had been taken.

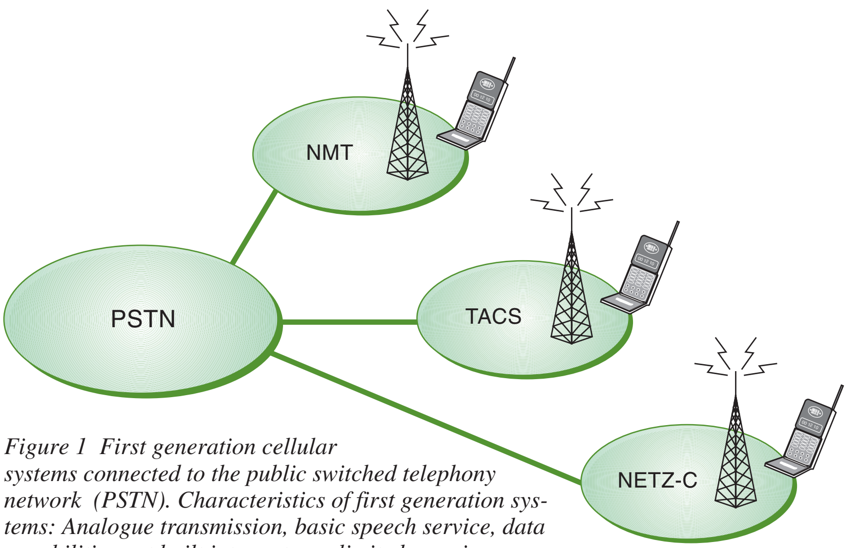


Figure 1 First generation cellular systems connected to the public switched telephony network (PSTN). Characteristics of first generation systems: Analogue transmission, basic speech service, data capabilities not built into systems, limited roaming capabilities

The analogue mobile telephone systems – first generation cellular

In 1981 the world's first automatic cellular international mobile telephone system was put into commercial operation in the countries of Norway, Sweden, Denmark and Finland. When the system was initiated, it operated in the 450 MHz region and was therefore named NMT 450.

The NMT system was mainly designed for speech communications, and the system relies on analogue frequency modulation (FM). However, the signalling is digital, with a transfer rate of 1200 bit/s.

Data communications is also possible using special modems. However, since the mobile radio channel is rapidly changing, it may be difficult to transfer error-free data in a system with no data error correction methods present.

The first versions of NMT terminals weighed more than 10 kg and were mainly manufactured for installation in cars. With today's measures the equipment would be regarded as highly impractical, expensive to buy and expensive to use. Despite this, the launch of the NMT system was a great enhancement in making public telephony available outside the wireline network. The growth of traffic in the NMT system was thus much higher than the most optimistic prognoses, and capacity problems were soon experienced. The NMT system was therefore extended to the 900 MHz band in 1986. The system is known as NMT 900.

In the late 1980s hand portable mobile phones started to emerge. The first ones weighed about 750 grammes, at that time regarded as remarkably small for a cellular phone. Miniaturisation has continued, and today true pocket phones are available for the NMT system.

Both NMT 450 and NMT 900 operate in parallel with GSM today, and there are still some 3,500,000 subscribers to the NMT service throughout Europe [2].

In addition to NMT there are several other systems belonging to what we call first generation cellular mobile phones, like the AMPS (Advanced Mobile Phone Services), TACS (Total Access Communication Systems), NETZ-C, MATS-E and others, totally holding some 93 million subscribers world-wide [3].

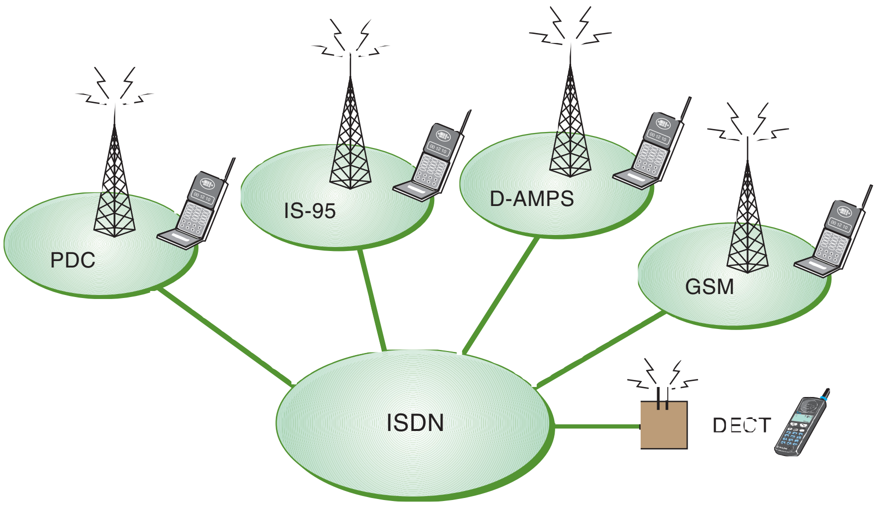


Figure 2 Second generation systems connected to the integrated services digital network. Characteristics of second generation cellular systems: Digital transmission, basic speech service, supplementary services, limited data capabilities, continental roaming

Next step – pan-European cellular

During the 1980s several automatic mobile telephony systems had grown up throughout Europe, mostly designed and set up to meet national needs for mobile telephony. Parallel to this, standardisation work was going on to design a harmonised mobile communications system for use throughout Western Europe.

Reservation of a common frequency band throughout Western Europe, with the European Post and Telecommunications Union (CEPT) as a co-ordinator of this work, made it possible to set time pressures and national industry policies aside and agree upon a common standard for mobile communications for a unified Western Europe.

Some objectives for the system design performed by the Group Spécial Mobile (GSM) were [4]:

- Mobile stations should be used throughout Europe. The system should route calls automatically to a mobile in any position inside the coverage area
- In addition to plain telephony, adapted ISDN services should also be made available to the mobile subscribers
- The system should offer services to car mounted as well as hand portable terminals
- Speech quality should be at least as good as for NMT 900 under good conditions
- The system should allow for encryption of user data over the air interface
- The system should utilise radio spectrum in an optimum way
- The system should not demand large upgrades of the fixed network
- It should be possible to have different billing structures in the different GSM systems
- It should be possible to have several GSM systems in one country.

The goal was that the GSM system should be in operation by 1991. In Norway a trial service was launched in 1992, and both Norwegian operators put GSM into commercial operation in 1993. Second generation cellular was born.

There are some 82 million users of GSM world-wide by mid-1998 [5].

Digital communications – tailored for data communications?

GSM and its sister systems D-AMPS and IS-95 in USA and PDC (Personal Digital Cellular) in Japan are digital by nature, hence they are better suited to data communications than the first generation cellular systems. Also, error correcting mechanisms are present, making error free data communication more easily available to the customers.

Available services in the GSM networks are for instance:

- Speech
- Emergency calls
- Facsimile
- Data services (2.4, 4.8, 9.6, 14.4 kbit/s)[1]
- Short message service, point-to-point as well as point-to-multipoint
- Supplementary services like call forwarding, calling line identification presentation and calling line identification rejection
- Mobile information services.

In addition, development of cordless systems like the DECT (Digital Enhanced Cordless Telecommunications) system has taken place. DECT is designed as a business and residential cordless system with higher bit rates available than GSM, but with a much less robust air interface. Hence, DECT is mainly intended for indoor wireless applications.

Being digital, the second generation mobile systems are better tailored to data communications than first generation cellular services. Still, a vast majority of the connections made using GSM is for speech telephony. For several reasons, only the very advanced users have started using their GSM for messaging and data communications:

- Terminals are not optimised for data communications because of the man-machine interface (too small keys and displays)
- High user threshold to get started using data services
- Users are unfamiliar with messaging services.

[1] *Enhanced data services like HSCSD (High Speed Circuit Switched Data services) and GPRS (Generalised Packet Radio Service) are under specification in what has been named GSM phase 2+.*

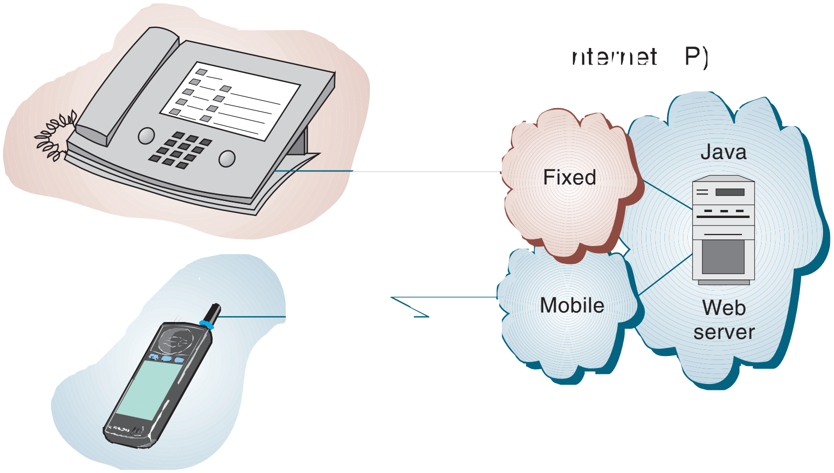


Figure 3 Terminals will increasingly communicate with servers. It is expected that within few years 50 % of the mobile traffic will be data communications

Popular – but still not the first choice

The mobile market has experienced a tremendous growth in the last years, boosted by the mobile operators subsidising terminals to attract customers to their network. At the moment (mid-1998), the mobile penetration in Norway has surpassed 40 %[2] of the population.

We do however note that the cellular is still not the customers' first choice. Most cellular users still have both a telephone at the office and a wireline phone at home. However, it is interesting to note that the mobile penetration is approaching the fixed phone penetration (including both residential and business connections) in the pioneering mobile markets.

---

2 *By the end of 1997, the mobile penetration in Norway was 38.5 %, only exceeded by Finland with a penetration of 42.1 % [6]. In Finland they soon expect to reach a penetration of 50 %.*

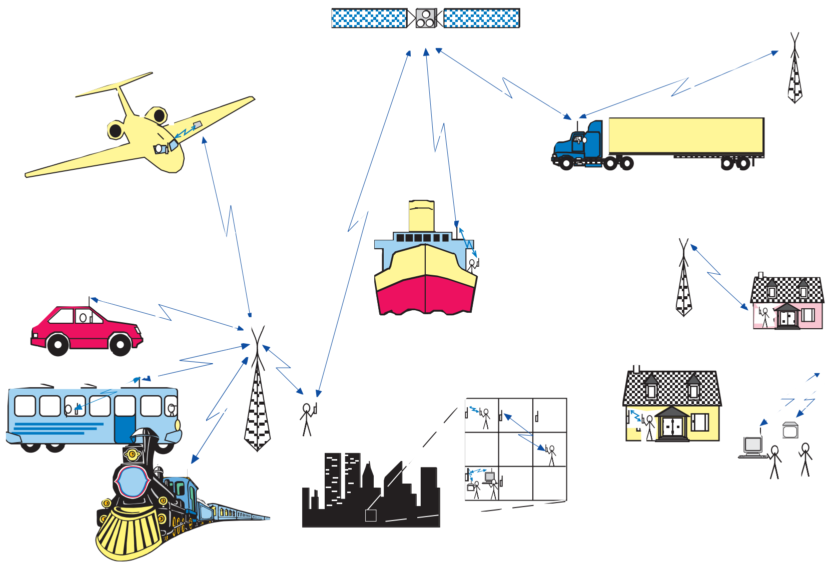


Figure 4 UMTS is recognised as an opportunity to provide mass market wireless multimedia services at any location. A fully developed UMTS may give global access to multimedia services across platforms as mobile, fixed and satellite-based networks

There is, however, one major difference between the wireline phone and the mobile phone. Experience shows that people are unwilling to share mobile phone, hence mobile phones are often personal, as opposed to fixed phones where a location and not a person is addressed.

The UMTS vision

In Europe third generation cellular systems have been called UMTS – Universal Mobile Telecommunications System. Research on UMTS started – typically enough – in the late 1980s, several years before the GSM standardisation was finalised.

The UMTS key objectives are summarised as follows [7]:

- Integration of services and application areas. Within one system UMTS will support services and applications which today are provided by dedicated systems. Target service areas are paging, mobility functions, mobile data communications, mobile telephony, public cellular applications, private business applications, residential cordless applications, wireless PABX applications and mobile satellite access.
- Integration of fixed and mobile networks through integration of fixed and mobile system technologies
- High quality of service, comparable with the fixed network
- Services requiring a variety of bit rates (up to 2 Mbit/s in phase 1) and variable bit rate services, including multimedia services
- Global terminal roaming capability, enabling the user to access UMTS services in all regions of the world
- Satellite and terrestrial based coverage, also extending UMTS coverage to areas where it is not techno-economically feasible to provide terrestrial coverage.

This leads to the UMTS vision of communicating anytime, with anyone, anywhere.

When research on UMTS started in the late 1980s, UMTS was regarded as revolutionary compared to second generation systems like GSM and D-AMPS. Later, several of the UMTS great thoughts have been implemented by enhancement of second generation systems or by interworking between second generation systems.

- *Global coverage:* GSM and its twins (DCS 1800, PCS 1900) are heading towards global operation. 293 administrations, network and satellite operators from 120 countries have signed the MoU stating that they want to provide access for GSM users
- *Service integration:* GSM's point-to-point short message service is an advanced form of radio paging (with store and forward and receipt to sender when message is received) integrated into a cellular system
- *Interworking between systems:* standardisation of interworking between cellular and cordless systems (DECT and GSM) is going on within ETSI
- *Integration of satellite components:* systems for handheld satellite based communications are being planned or implemented, like ICO, Iridium and Globalstar. All these system concepts rely on combination terminals communicating with the land based infrastructure when available, otherwise relying on satellite based communications
- *Personal mobility:* GSM is based on the use of SIM cards containing the subscriber's identity (allowing for 'plastic roaming' which for instance was offered to GSM customers during the Nagano Winter Olympic games)
- *Tailoring of services:* Intelligent network functionality is being implemented in the fixed as well as in the GSM networks, giving room for tailoring of services and service mobility.

Hence, it seems that services and network aspects of UMTS is predominantly going to be an evolution rather than a revolution from the second generation systems. With the ongoing upgrade of GSM to EDGE (Enhanced Data rates for GSM evolution) to offer bit rates up to 384 kbit/s, it is also a question of definition if UMTS Phase 1 is going to be an evolution or a revolution compared to the enhanced GSM radio subsystem.

In later stages of the UMTS implementation, integration with IP networks offering differentiated quality of service will take place. In addition, UMTS may support access networks, providing higher bit rates than 2 Mbit/s. UMTS may then give global access to multimedia services across platforms as mobile, fixed and satellite-based networks, turning UMTS into the real universal telecommunication system.

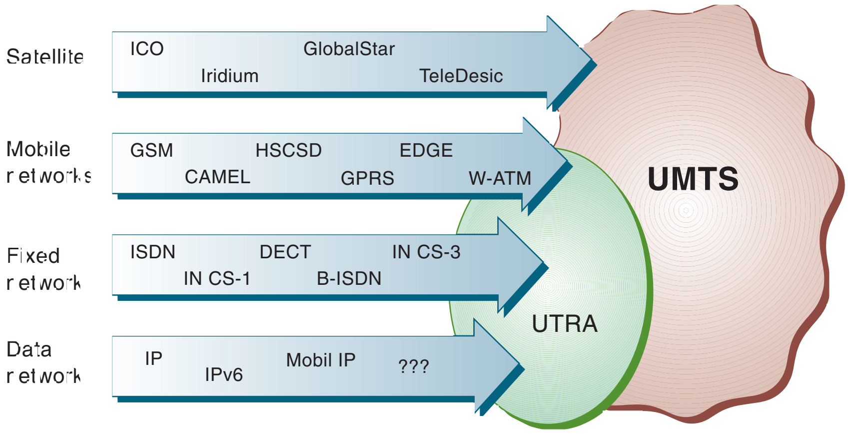


Figure 5 Possible evolutionary paths from today's telecommunications solutions to the UMTS vision. UTRA stands for UMTS Terrestrial Radio Access

The plan is to implement UMTS in 2002.

Mobile broadband systems

In addition to the UMTS work, real broadband mobile systems are also being developed, like in the ACTS project SAMBA (System for Advanced Mobile Broadband Applications).

At the moment work is focusing on a trial platform providing transparent ATM (Asynchronous Transfer Mode) connections and supporting bearer services at up to 34 Mbit/s in a cellular radio environment operating at 40 GHz.

The ultimate aim for SAMBA is to develop mobile broadband systems offering bit rates of up to 155 Mbit/s to mobile users.

Looking into the crystal ball

A general trend is that when new services have been introduced in the fixed network, a demand has arisen for the same services also being available in the mobile network.

Trends – mobility

One trend that has been present since mobile telephony was introduced is the aim of increased mobility. With the introduction of the S-PCN (Satellite Personal Communications Networks) based on Low Earth Orbit satellites the coverage will be ubiquitous.

Four types of mobility are being referred to:

- *Personal mobility:* one access number is connected to a person, regardless of terminal and access point. One sort of personal mobility is the UPT (Universal Personal Telecommunications) service in the fixed network, another form of personal mobility is SIM card roaming ('plastic roaming') within GSM
- *Service mobility:* uniform access to the same set of services across terminal and service platforms
- *Terminal mobility:* continuous mobility across locations, relying on radio based mobile terminals that can be used within the radio coverage area
- *Session mobility:* during a communication session the user can move between service platforms. One example is a person being alerted on her radio pager that someone wants a video conference with her. Then the user has the possibility to transfer the session to a service platform offering the requested service.

Trends – FMC

FMC has been a buzz word for some time, an acronym with different interpretations from person to person. FMC could be bundling of mobile and fixed services by the service provider, or even

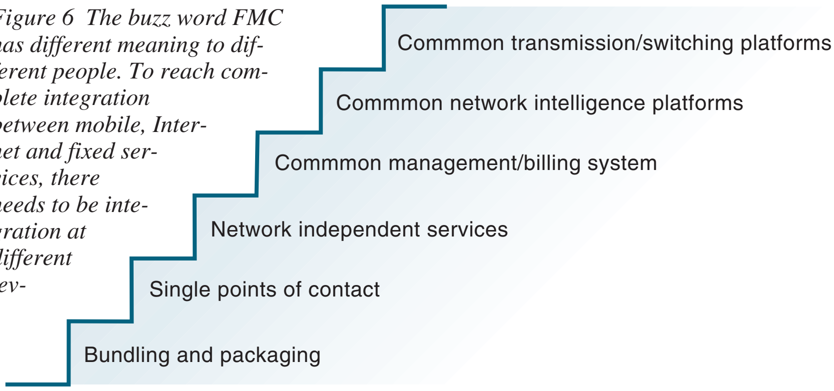


Figure 6 The buzz word FMC has different meaning to different people. To reach complete integration between mobile, Internet and fixed services, there needs to be integration at different levels

issuing one bill for mobile and fixed subscriptions. The service provider's bundling of services is, however, dependent on the regulatory situation.

One-number services which are offered by several service providers can be realised in different ways. One solution is using DECT/GSM dual mode terminals with support of network functionality to keep track of the terminals both fixed and mobile networks are used.

By using GSM home base stations, which are being standardised, a one-number service could be offered based entirely on the mobile platform if the home base station is connected to the MSC.

In the same way, DECT as access to the fixed network can be used, eg. the Fido service offered by Telecom Italia, where a DECT terminal can be reached within the area of a city. For coverage beyond the city border, a combined DECT/GSM terminal is needed.

There is a general trend that similar services can be based on a mobile or fixed network platform. The next generation of IP also provides for some mobility, eg. suited for nomadic computing, and Internet is an interesting platform for personal telecommunications services.

Trends – multimedia services

Another trend is that the users are requesting a variety of services requiring different service quality and different bit rates. Examples of such services are:

- 7 kHz audio (AM quality)
- 20 kHz audio (CD quality)
- video telephony
- videoconferencing (128 – 768 kbit/s)
- messaging services
- telefax group 4
- data base access
- broadcasting services
- Internet access.

These services are made available to mobile users, either by designing systems for higher bit rates or by accommodating protocols to tailor the services to the bit rates being offered by the various radio systems.

Another consequence of multimedia services is that the demanded bandwidth might be different in the different link directions.

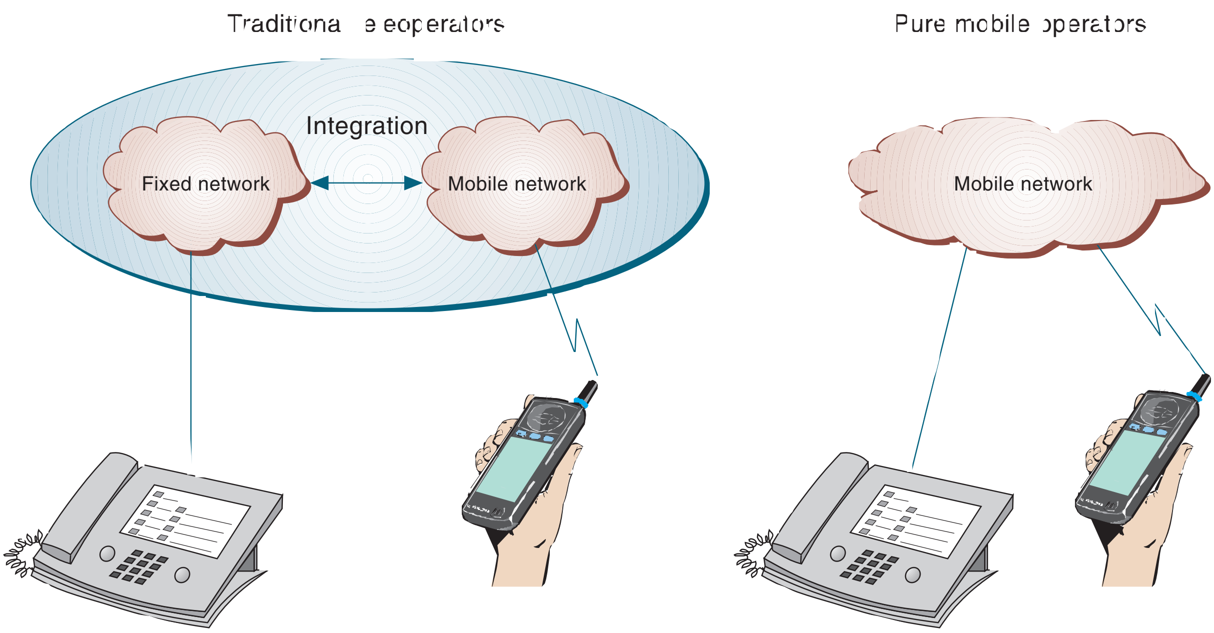


Figure 7 Traditional telecom operators have a different view of FMC from the pure mobile operators

Trends – wireless application protocols

The design of advanced data applications and services based on Internet Web technology in mobile environments has recently started. Up to now Internet technology has been aimed at powerful PCs utilising access networks like ISDN or Ethernet. To fulfil the demand for mobility without limiting the communication capabilities, future application platforms must handle communication platforms with lower transmission capabilities, using less powerful terminals like smartphones, microbrowsers, PDAs (Personal Digital Assistants), palm top computers, etc.

Within year 2000 the prognosis is for 22 million users world-wide using other types of terminals than PCs to access the Web [8].

The first joint initiative to standardising a protocol for wireless Web access is the work on the Wireless Application Protocol (WAP) in the WAP Forum, driven by Nokia, Unwired Planet, Motorola and Ericsson. WAP Forum has recently published the draft WAP specification, identifying a set of protocols and programming languages, which will allow further development of mobile phones into microbrowsers.

Up till now the mobile terminals have been much less standardised than the traditional PCs using MS Windows as operating system. For mobile terminals there is an ETSI standard for applications on SIM (Subscriber Identity Module) cards named SIM-ToolKit. The idea is that program code stored on the chip inside the SIM can be used to run an application dependent code, allowing for remote applications downloaded by the operator and accessed via the terminal. Hence, for instance user interface can become a competitive factor.

The introduction of the WAP is based on the assumption that there is a demand for Internet access from portable terminals. WAP has paved the way for making a variety of terminals, including the mass-market handsets, true information appliances.

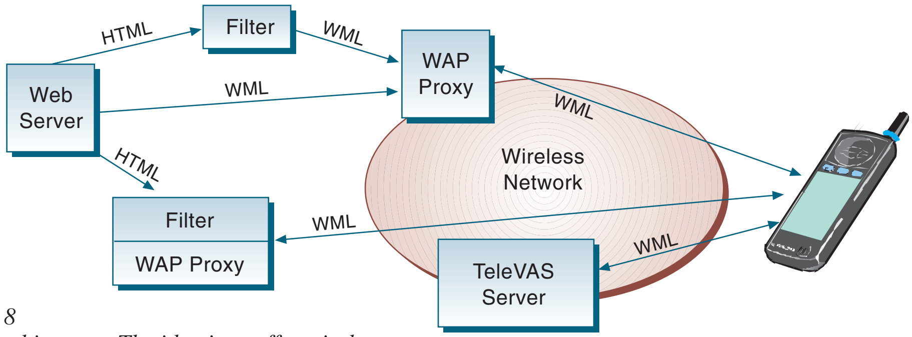


Figure 8
WAP architecture. The idea is to offer wireless Web access by filtering of HTML (HyperText Markup Language) to WML (Wireless Markup Language) and thus adapt the web interface to the capabilities of the mobile terminals

Trends – the telecommunication terminal as a module in a multifunctional device

Bringing the cellular phone and the personal organiser more or less wherever they go is quite common among people today. There is a trend towards using electronic personal organisers. Products containing both an organiser and a mobile phone in one package are on the market. As electronic commerce is becoming more and more widespread the next step of integration could be that even the wallet turns electronic and becomes part of a communication enhanced PDA. Keys and physical access control could also be made electronic and integrated in a multifunctional device.

A development in the direction of integrating what people usually carry in their pockets and which can have an electronic version into one device, will depend on the level of security that can be obtained. Satisfactory security solutions to prevent fraud, eavesdropping or unauthorised use are essential. Acceptable ways of handling lost or stolen devises also have to be found.

Personal communication – what will it be?

Personal communications is an ambiguous term, eg. in USA, Personal Communications Service (PCS) is based on second generation mobile technology.
In Europe the term Personal Communication Networks (PCN) was used when the first DCS 1800 networks were established.

The basic understanding of personal communications implies that a person can be reached on one number at any time wherever they are and that there is no limitation on outgoing calls.

Using the fixed network personal communications can be achieved if the incoming call is always routed to a terminal located close to the wanted person – personal mobility is mobility across terminals. Utilising personal mobility the network must know where to route the incoming call. This means that some kind of action is needed from the user to inform the network of which terminal they can be reached at. When leaving the location the network must also be informed. To register and deregister seems too much to ask from users.

Figure 9 When personal communications is present the basic telecommunications device is a true pocket terminal offering speech services, Internet access and various multimedia services. All services made available via the personal terminal are accessed using one personal number, which can be the same during the customer's life cycle

The alternative to personal mobility is terminal mobility, ie. users bring a personal terminal with them. Terminal mobility offers contiguous mobility across locations. The penetration of cellulars is surpassing 40 % in the pioneer markets, indicating that terminal mobility will be the basic element in personal communications.

In Europe potential technologies for personal communications based on terminal mobility are GSM, DECT and UMTS. Standards are made for both GSM and DECT to handle high bit rates and also to handle packet data. Multimode terminals capable of accessing different networks might be an important part of personal communications and users will have a choice from limited local mobility to full global mobility.

In the European Commission's Green Paper on Personal Communications [9] it says: *"With the broadening of mobile communications into general personal communications, the sector will, before the end of the decade, substantially affect the lives of every citizen of the European Union"* and *"– personal communications has the ultimate potential to reach near 80 % of the population (ie. up to one connection per adult.)"*

There are already trends in some countries that mobile subscriptions are starting to replace fixed line subscriptions.

The first phase of personal communications seems to be a change from POTS (plain old telephony system) to personal terminals.

Personal communications is present when the basic telecommunications device has become a personal pocket terminal instead of a wireline terminal and the standard subscription implies using a personal pocket terminal instead of POTS.

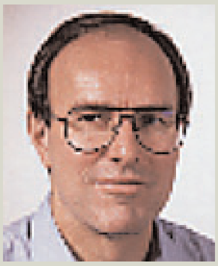
A next phase, following the transition from POTS to personal terminals, is when services beyond speech, like Internet access and multimedia services, will be available over pocket terminals.

The technology will be able to handle the development outlined. The user demand for personal communications is of course uncertain, but it seems that using personal terminals will be a requirement.

The introduction of personal pocket terminals creates new problems. For example, nobody wants to be available at all times. The situation and the surroundings have to be taken into account before making or receiving calls. In aircraft, for example, the use of cellular phones is prohibited and few will use the phone during a dinner party. Silence modes for monitoring incoming calls already exist, but users have to learn to behave when using pocket terminals. In general, flexible ways of handling received calls when not available will be a requirement when personal pocket terminals are used.

Will users want to use only one terminal, their personal pocket terminal, both at work and private? Will the differentiation between working and non-working hours decrease? These, among others, are questions that might be raised regarding the introduction of personal communications.

The transition from POTS to personal phones carried at all times, will be a major step in the history of telecommunication and will of course influence people's everyday life and also have an impact on the functioning of society.



Joar Løvsletten has been with Telenor since graduating from the Norwegian University of Science and Technology in 1976. He worked in the Network Department, dealing with radio relay systems, until 1990, when he started a group working on radio access systems. Løvsletten came to Telenor R&D in 1995 and is now Senior Engineer working with personal communications.

e-mail: joar.lovsletten@fou.telenor.no

Rune Harald Rækken is Senior Research Scientist at Telenor R&D, where he has been employed since 1987, working with standardisation of the GSM system, with radio propagation aspects, FMC, and mobile communications in general. He is currently in charge of the mobile and personal communications group at Telenor R&D.

e-mail: rune.rakken@fou.telenor.no

References

1 Løken, B et al. NMT-systemet : et fellesnordisk automatisk mobiltelefonsystem. *Telektronikk,* 75, (2), 1979.

2 European analogue & digital cellular subscriber up-date. *Mobile Communications International,* 50, April 1998.

3 Helme, S. There's still life in the old dog yet. *Mobile Communications International*, 50, April 1998.

4 Audestad, J A. *Det felleseuropeiske landmobile system : status og framdrift.* (The Pan European land mobile system. Status and drive.) Guest lecture in the course "Radio systems" at the Norwegian University of Technology, 1986. Paper in Norwegian.

5 The GSM MoU Association's web site (http://www.gsmworld.com/)

6 *Mobile Europe*, February 1998.

7 Bjørnland, D F, Lauritzen, G O. UMTS : the Universal Mobile Telecommunications System. *Telektronikk,* 91, (4), 127–132, 1995.

8 *ComputerWorld Norge,* 37, 1997.

9 Commission of the European Communities. *Towards the Personal Communications Environment.* Green Paper on a common approach in the field of mobile and personal communications in the European Union, 1994.

Third Generation Mobile Communications

JAN ERIKSEN, GEIR OLAV LAURITZEN, BJØRN HARALD PEDERSEN AND STEIN SVAET

This article considers the next generation mobile communications systems. The characteristics of these systems are described by outlining the system requirements and the services envisaged. Further, a short overview of European research activities is given, followed by an overview of standardisation activities within ETSI and ITU. The work on next generation mobile communications systems is now in an intense stage, aiming at finalising the first sets of specifications by the end of the century. Details of the system concept are thus subject to continuous change, and in order to avoid publishing outdated results, this article therefore takes a rather general approach.

1 Introduction

In Europe and in the Nordic countries in specific, the history of automatic land mobile systems [1] starts in 1969 with the development of NMT (Nordic Mobile Telephone). The analogue NMT system became a success, partly because of its robustness and simple mobile stations. Basic telephony with roaming possibility was the service offered. NMT hit the market at the right time, and is considered to be a first generation cellular system. Other first generation systems emerged, notably TACS and Netz-C.

Second generation systems, based on digital technology all the way from the interconnection point to the mobile station, emerged with the introduction of GSM. The development of GSM started already in 1982 within CEPT (European Post and Telecommunication Conference), and was carried on in ETSI from 1988. As for the analogue systems, the dominant service of GSM has until now been basic telephony. GSM introduced the separation of terminal and user identity, which enabled enhanced security and user roaming across terminals.

GSM was launched for the first time for commercial operation in 1991; in 1997 more than 100 countries had at least one GSM network. GSM is a success, now deployed in almost every corner of the world.

The research society moved onto third generation studies many years before GSM was launched for commercial operation. The idea that a third generation system should consist of different access components, including a dominating terrestrial access component providing at most 2 Mbit/s, was established relatively early. Regarding terminals and services, however, the focus of attention in the early days centred on the concept of a personalised lightweight hand portable. This is now commonplace also in second generation systems, and the focus for third generation has been widened to include a wide range of terminals to support various services, including multimedia.

On the standardisation arena, an early step towards the standardisation of a third generation mobile system was taken in 1985 by the establishment of CCITT Interim Working Party 8/13 (IWP 8/13), now known as ITU-R Task Group 8/1. The group was responsible for the standardisation of Future Public Land Mobile Telecommunications Systems (FPLMTS), now known as IMT-2000. ETSI followed the ITU in 1991 with the formation of a Sub Technical Committee called Special Mobile Group 5 (SMG 5) responsible for the standardisation of a third generation mobile system called Universal Mobile Telecommunications System (UMTS).[1]

Perhaps the most important result from the first years was the contribution to WARC 92 [2], resulting in an allocation of 230 GHz for future mobile communications in the 2 GHz band. Being dominated by operators and academia, there were strong forces within standardisation to repeat the GSM success by standardising a new system from scratch. Later, the view has changed to take more account of existing technologies. This is particularly pronounced in ETSI, where the legacy from GSM is strong. The development takes into account the fact that the UMTS technology should come about from an evolution of GSM, or at least the interoperation with GSM should be eased. This ensure that the vast amount already invested in GSM technology pays off also by the introduction of UMTS. Further, this enables the access to UMTS or the provision of UMTS-like services based on the GSM access, which is an important goal for GSM operators.

The development of third generation standards is gradually becoming more market driven. Large markets are still uncovered by mobile services, at the same time as more advanced markets demand new and better services. Two mega trends which are important enablers are the 'Internet explosion' and the huge growth in cordless and mobile terminals. Indeed, due to this market prospect, large resources are already spent on development, both from industry and operators/service providers. It is easy to forecast that the rate of investment in third generation development will increase even further in the nearest coming years.

However, the development is not solely market driven. Today's extremely high competence and know-how within mobile communications compared to previous years represents in itself a driving force. On the one hand, operators and service providers have developed advanced skills for network deployment and operation as well as for service design and provision. On the other hand, equipment manufacturers make continuing technological improvements supporting the possibility of increasing system complexity and capability, as well as reducing the equipment size and price. Developments contributing to this are found within microelectronics, modulation and coding, battery technology, antenna design, MMI, etc.

As opposed to this good basis for development, there is, however, a possibility that the development is slowed down or threatened by diverse interests in new technology. This may be experienced from industry or operators aiming to exploit beneficial positions as producers and owners of existing technology or systems. Having said that, a situation representing a serious threat to the development is not anticipated; the ideas to evolve the development from existing technology have rather contributed to cost-effective and viable solutions. Another situation that may seem more threatening to the growth of third generation systems is the availability of spectrum. The bands identified by WARC-92 are shared with other radio communication systems and services, and are in many cases already in use, and surveys show that additional spectrum is required [2, 17].

2 Services and System Requirements

2.1 Overview

First generation systems offer a switched telephony service to users who may

[1] *The UMTS concept was strongly influenced by the RACE Mobile (R-1043) project at the end of the 1980s.*

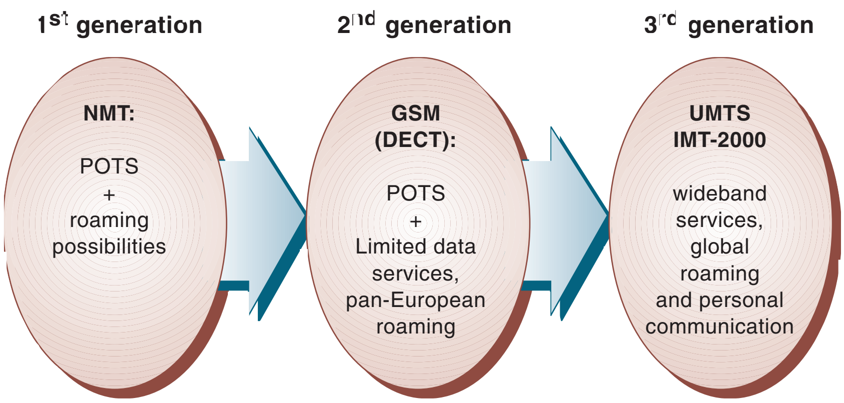


Figure 1 The different generations of mobile communications systems

move freely within the service area. This may be referred to as a plain old telephony system (POTS) with roaming capabilities. In second generation systems, the service repertoire has been extended to include not only POTS, but also a limited set of data services. Further, the mobility management functions are more powerful with more intelligent signalling. The concept of mobility has also been widened introducing mobility of users across terminals by separating the user and terminal identity. Seen from the user, however, there is not much difference in the functionality offered by first and second generation systems. Both the bandwidth limitation and the rigidly standardised set of services are obstacles for introducing advanced data services. The radio interface is designed for narrowband transmission, and the systems rely on switched narrowband connections in the fixed infrastructure.

Major goals of third generation systems are to overcome the limitations in bandwidth and to provide flexibility for service provision. This paves the way for a wide range of services, including multimedia, and supports personification of the services. A goal of third generation systems is also to provide a uniform access to personal communication based on different access networks, both wired and wireless, in different operating environments. The aim is to provide personal communication anytime and anywhere.

2.2 Service Provision in Different Environments

Third generation systems concentrate currently on new radio access networks that can provide services up to 2 Mbit/s. Although this type of access network will be a major component, the third generation systems also intend to provide access through enhancements of existing radio access networks (eg. GSM), as well as through wired access networks. At a later stage, microwave radio access networks may be introduced which provide bearer capabilities far beyond 2 Mbit/s. As will be discussed below, the services provided will not only depend on the type of access network, but also on the environment in which the network operates.

It is foreseen that wireless will gradually become the access of choice for all personal communications. Third generation systems aim to provide a quality of service for wireless access that is comparable to that of wired access. Service mobility is supported, so that the service is available to the user in a uniform way regardless of the environment. Although the services are similar both in quality and appearance in different environments, the bit rate of the bearer services provided will depend on the type of access network and the operating environment. Services requiring the highest bit rates will therefore not be supported in certain access environments. However, many services are expected to be designed in a flexible way, so that they can be carried on bearers with different bit rates depending on the capabilities of the access network.

It is required that third generation systems accommodate different types of terminals which may have different capabilities for supporting the various services. The actual services provided will thus also be dependent on the capabilities of the terminal.

As discussed above, the interdependence between terminal capabilities, environments and the services provided have implications on the system requirements. This is illustrated in Figure 2.

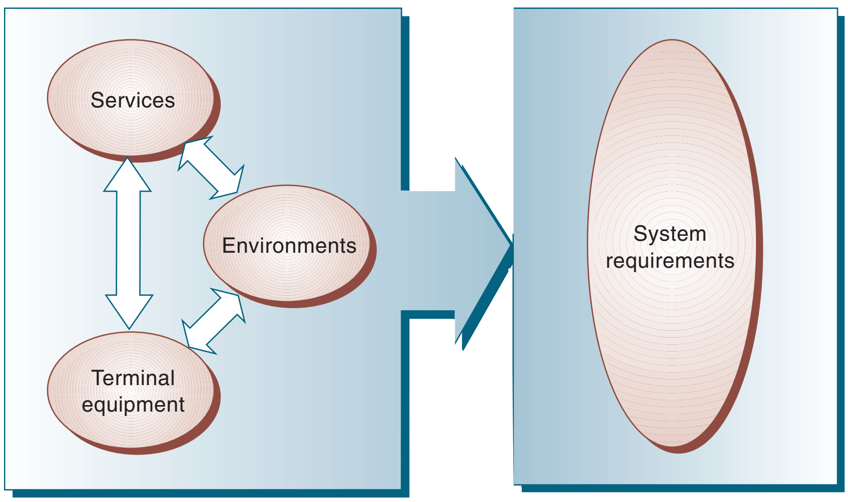


Figure 2 Impact on system requirements by services, environments and terminal equipment

2.3 Flexibility in Services and Applications

The flexibility in the design of services in third generation systems does not only facilitate the personification of services. It also enables service providers to distinguish their services and it eases the service creation by minimising the re-engineering required to provide new services. In order to provide this flexibility, the service components (service capabilities) are standardised rather than the services themselves, an approach already known from Intelligent Networks (INs) [3].

Service capabilities consist of bearers defined by QoS (Quality of Service) parameters and mechanisms to enable service providers and operators to create their own supplementary services, teleservices and end-user applications. Thus, to give a list of services for third generation would just be based on speculation. However, some services may be foreseen, such as telephony, different data transmission, video and multimedia applications.

Third generation systems will be designed to be flexible not only for service provision, but also for deployment in different environments, and easily expandable. This is dealt with in the chapters considering radio and network requirements.

The versatility of third generation systems makes them suitable for many applications, such as office applications to access services normally provided by PBXs and LANs, for cordless applications in the domestic domain, and for cellular operation for access to public telecommunications. Further, the third generation systems are candidates for radio in the local loop, which is a viable solution for rapid deployment of fixed telecommunications, particularly applicable in developing countries. The market is believed to be at least as large as the market for mobile communications.

2.4 Requirements for the Radio Interface

The requirements to the radio interface for third generation systems will obviously be quite different from earlier generation systems. In NMT the offered services were speech and low rate data. Even in GSM and other digital second generation systems the services are limited to speech and low rate data, though there has been some enhancement within GSM to offer rates in the order of magnitudes of tenths of kbit/s, and recently even packet switched services.

For third generation systems the requirements to services will be very different. These systems are required to offer a wide range of services from traditional speech to packet or circuit switched data having transfer rates in the order of a few million bits per second. However, all services will not necessarily be offered in all kinds of environments. As discussed previously, there is an interdependence between terminal capabilities, environments and the services provided.

The environments may be classified into the outdoor vehicular macro cell environment, the outdoor vehicular suburban environment, the urban environment with low mobility pedestrian users, and the indoor very low mobility environment [4,5]. It is obvious that the requirements to the radio interface will be quite different in these different environments. At one end we have the vehicular macro cell environment where the terminals will have high velocity and rather low density. In this environment deployment will be focused on coverage, ie. covering a large area, and giving some minimum service to all users within the area of coverage. At the other end we have the indoor pico cell environment where the terminals will have very low or no velocity, but some mobility. In this environment deployment will be focused on capacity, and giving high data rate services to the users in the area of coverage.

In the remainder of this section requirements for UTRA (UMTS Terrestrial Radio Access) are considered [6], though similar requirements within ITU-R are referenced [5]. Thus the term third generation system in most cases means UMTS in this context.

The requirements on the radio interface are focused on

- Bearer capabilities
- Operational capabilities
- Spectrum usage
- Complexity and cost.

2.4.1 Bearer Capabilities

The radio interface should support a range of maximum user bit rates that depends upon the environment the user is currently located in as follows:

- Rural Outdoor (Vehicular Macro cell): at least 144 kbit/s (goal to achieve 384 kbit/s) at speed of 120 km/h (maximum speed 500 km/h)
- Suburban Outdoor (Vehicular suburban): at least 384 kbit/s (goal to achieve 512 kbit/s) at speed of 120 km/h
- Indoor / Short range outdoor (Micro / Pico cell): at least 2 Mbit/s with mobility.

It is desirable that the radio interface should allow evolution to higher bit rates.

The radio interface should be flexible in allowing a range of user bit rates and types of services in different user environments:

- Parallel bearer services (service mix), real-time / non-real-time
- Circuit and packet switched bearers
- Scheduling of bearers according to priority
- Adoption of link to quality, traffic and network load, and radio conditions
- Wide range of bit rates with sufficiently low granularity
- Variable bit rate real time capabilities.

The radio interface should support user mobility within and between radio operating environments:

- Provide handover between cells of one operator
- Provide handover between different deployment environments
- Allow handover between different operators or access networks
- Allow efficient handover to second generation systems, eg. GSM.

2.4.2 Operational Capabilities

The radio interface should provide compatibility with services that are provided by present core networks:

- ATM bearer services
- GSM services
- Internet Protocol based services
- ISDN services.

On a global basis numerous core networks exist. The modes of operation and the range of services and bearers that third generation is foreseen to provide are very different. Therefore, the radio interface of these systems must exhibit great flexibility in order to be able to meet requirements given by these different aspects. This implies that the radio interface should be capable of meeting requirements given by various radio operating environments, operation across multiple operators, multiple types of operators, operation across different regulatory regimes, a variety of terminal types, and a variety of services and bit rates.

It is seen as a mandatory requirement that the operator should be excepted from exhausting frequency planning after reconfiguration of his network due to service or traffic requirements. Whether such a requirement is feasible without losing in other ways, however, is yet to be investigated. Therefore, the radio interface should allow for flexible use of the allocated frequency bands with minimum required spectrum planning after re-configuration of the network due to changing traffic or service requirements.

It is envisaged that public, private and residential operators will operate systems within the third generation concept. Public operators will operate systems covering a wide range of coverage area and services with minimum guaranteed grade of service in wide coverage areas and with very high capacity and user data rates in environments with limited coverage. Private operators will operate in limited areas and give service to a limited group of users, possibly with some limitation on services and mobility. Residential operators will operate within very small cells (pico cells), probably with a very limited range of service capabilities. The radio interface should allow for these three types of operators, but with predetermined and guaranteed levels on quality of service for the public operator when in the presence of other operators.

2.4.3 Spectrum Usage

The radio interface should have high spectrum efficiency for typical mixtures of different bearer services, and for low rate speech services the efficiency should be equally good as that for GSM. Of course it is difficult, or more likely impossible, to tell what the typical mixtures of bearer services will be. However, it is requested that the chosen technology is likely to have high spectrum efficiency in order to exploit the given frequency band in the most efficient way.

It is likely that some of the provided services, for instance Internet connection will require higher capabilities at the downlink compared to the uplink. This implies that the radio interface should provide for asymmetric use of the radio resources on up- and downlink.

Within the bands that are allocated to UMTS multiple operators should be allowed without co-ordination. It is assumed that different operators will be given different frequency bands to deploy their services within. Given such an assumption, any specific operator is free to deploy his third generation network without mandatory co-ordination with operators using an adjacent band.

Some of the frequency bands that are currently used by first and second generation systems will undoubtedly be suitable for services which first will be provided by third generation. To allow for such services to be deployed even further in an efficient manner it is preferable that the third generation technology can migrate into these frequency bands in an easy and non-prohibited manner. Therefore, it is necessary that the interface is capable of this migration.

2.4.4 Complexity and Cost

It is likely that third generation terminals will come in several types. This will allow for a wide range of features to be included within the different types, giving different terminal classes. It is required that terminals are available within the range of classes, both simple and inexpensive ones for low end users and more complex and thereby expensive terminals for high end users.

For the network it is desirable to have one radio interface which can adapt to different modes of operation in order to meet the predefined needs of the radio operating environment and service environment.

The overall cost of third generation is influenced by the technology chosen and the desired quality and reliability. The main elements that contribute to the overall cost of UMTS are:

- Development cost
- Equipment cost
- Deployment cost
- Operational cost.

2.4.5 Other Requirements

It is desirable that the radio interface technology that is chosen within ETSI meets at least the technical requirements as a candidate technology for IMT-2000 within ITU.

It should be possible to deploy and operate a third generation network within a limited bandwidth in order to meet regional regulatory requirements.

A third generation system should be capable of co-existing with other systems within the same or neighbouring frequency bands depending on systems and regulations.

2.5 Network Functions and Requirements

This chapter deals with the requirements imposed on third generation systems regarding the functionality and design of the networks.

2.5.1 Network Modularity

One of the basic requirements on third generation network architecture is the modular system concept where the access networks are clearly separated from core networks at specified interfaces. Further, the standards will open for a many-to-many relation between access and core networks.

This concept is supported by the need for different access networks due to different environments (eg. satellite, wide-area cellular, cordless, fixed) and different core networks. The concept also enables competition between different standards and implementations of access networks and core networks.

Within ITU, these principles are referred to as the IMT-2000 Family of Systems Concept (IFS). Within ETSI, similar ideas have been developed, based on the work on Global Multimedia Mobility (GMM) [7]. The GMM is a framework that allows different access networks to interact with different core networks based on a common architectural concept. This is illustrated in Figure 3 (taken from [7]).

This modular approach invokes the need for standardised interfaces between different system components.

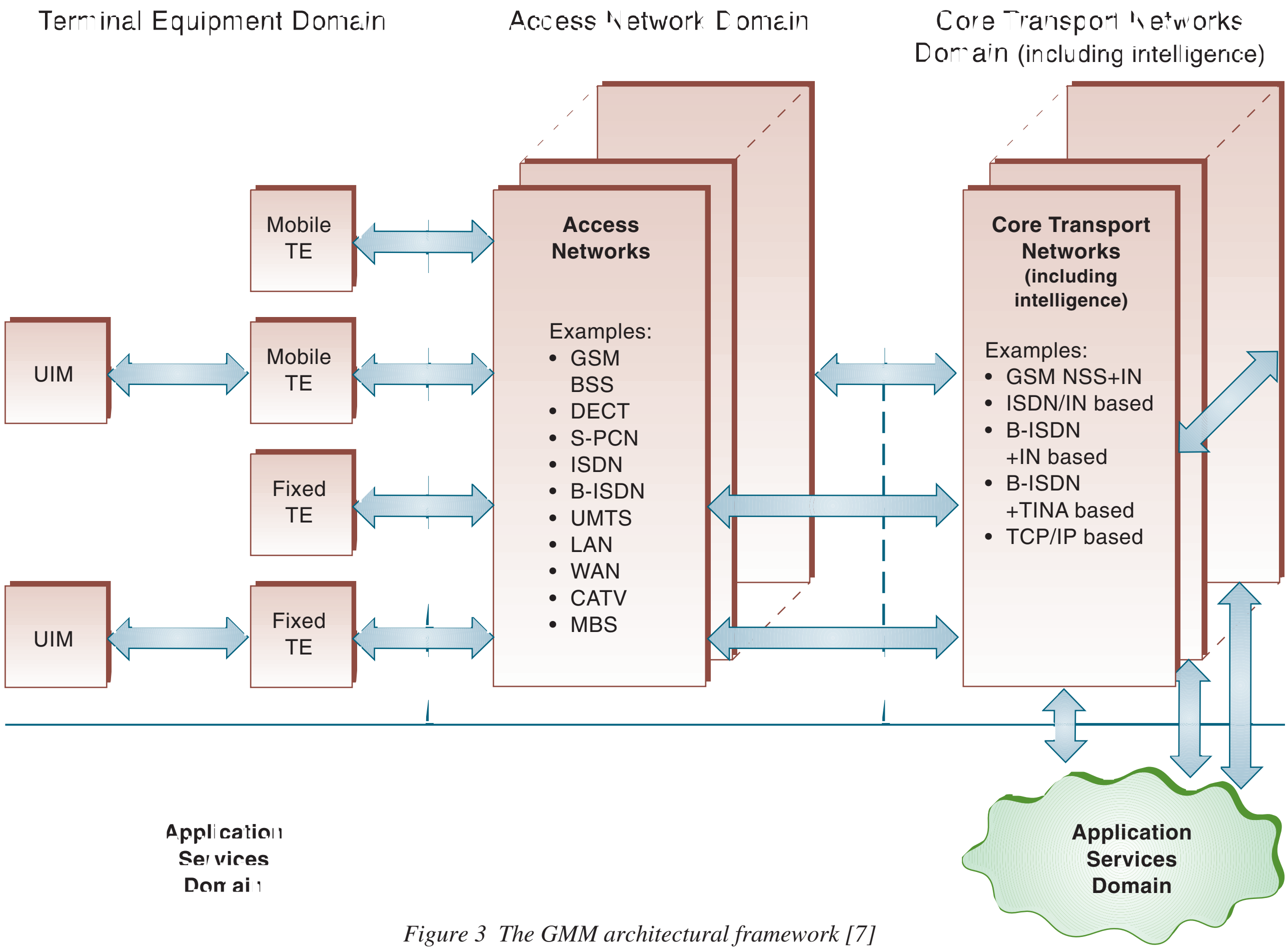


Figure 3 The GMM architectural framework [7]

2.5.2 Network Functionality

The required network functions are deduced from considerations of operational requirements. The network functions are grouped into functional entities, and functional architectures are created by identifying the functional entities and their interactions. Functional architectures are under continuous development both for IMT-2000 and UMTS [8,9]. These architectures include functional entities for (the list is not exhaustive):

- Radio resource management
- Mobility management
- Call-, connection- and bearer control
- Interworking
- Access- and service control
- Security control
- Network management.

Functions for radio resource management are necessary for allocation and control of radio communication resources. This includes allocation of radio resources, set-up and release of radio channels, power control, modulation, coding and compression. In third generation systems, the radio resources must be shared between circuit switched and packet switched communication. Additional functions for packet data transfer capability are required, including access control, multiplexing, error and flow control.

Functions for mobility management include location registration, paging, functions for supplying routing information, and functions for handover. Regarding the handover functions, it is necessary to have functions for radio channel quality estimation, handover decision and execution. Macro diversity may be provided in the access network, and special functions are required in order to control the recombination of information during a handover.

Call-, connection- and bearer control functions are required to perform the set-up and release of calls. For connection oriented services, the functionality is well developed and described. For connectionless services, the notion of a call is absent, and work is progressing in order to specify necessary functionality for end-to-end control.

The call-, connection- and bearer control functions may act in a hierarchical manner, supervised by the call control, which performs end-to-end QoS negotiations and set-up or release of the required number of connections associated with the call. Connections are end-to-end associations of bearers and the separation between call and connection gives extra freedom in allocation of the functionality to different network elements, which may be useful eg. for multimedia calls. The connection control performs the set-up and release of the connections, based on performance requirement such as BER and time delay. Finally, the bearer control functions allocate the required number of bearers to each connection, and triggers the establishment of the actual physical channels. Bearer control functions are required both for the air interface and the interface between the access- and core networks, and will be different according to the characteristics of the physical channels they control.

The interworking functions are necessary in order to connect different access networks to different core networks. Functions will be specified for interworking with ISDN, B-ISDN, X.25 PDN (public data network) and IP data traffic. Interworking with existing second generation mobile systems is also important.

Access- and service control is concerned with the control and access to services, and handles the service related processing. In an IN approach, as eg. taken for IMT-2000 [8,10], the service control functionality is triggered by the call control, and interacts with functions for service access and switching, as well as specialised resource functions. Service access requires the provision of terminal capability information, user- and subscription information and triggering of authentication. The specialised resource functions are responsible for various kinds of user interaction, eg. receiving digits and playing announcements, both for IN-provided services, multimedia services and packet data transfer.

Security control is becoming increasingly important in communications. Mobile and personal telecommunications services impose additional security requirements, since mobile communication systems inherently involves wireless access, and third generation systems must allow users to roam across network and national boundaries. The security control includes functions for [11,12]:

- Authentication of user, mobile terminal and service provider
- Privacy, anonymity and confidentiality of user location, user identity and user data
- Authorisation and access control to limit the access to user and service profile data
- Denial of access to particular services.

Functions for network management are necessary to ease and support tasks such as planning, installation, provisioning, operation, maintenance, administration, and customer service. Efficient network management functions are important for the ease of operation and thus the cost-effectiveness of the system. IMT-2000 bases the network management on the concept of a Telecommunications Management Network (TMN), which is being studied by ITU-T.

2.5.3 Flexibility and Evolution

For the mapping of functional entities to network elements, it is essential that a high degree of flexibility be supported. In other words, the mapping should give freedom to create a number of different reference configurations, so that the system may be optimised regarding performance and cost to different applications and environments. Both stand alone configurations with gateway to the fixed networks and integrated solutions with shared functionality should be possible.

The modular structure and the freedom to add extra functionality also support another aspect requiring flexibility, namely that the system should be easily expandable to allow growth in size and complexity.

3 Research and Standardisation Activities

This chapter considers the research and standardisation activities on third generation mobile communications systems. Research on third generation technology is performed at universities, within industry, and by telecommunications operators. This chapter considers some research activities performed in European organisations, and gives by no means a complete or representative picture of the research carried out in a global scale.

Standardisation of the next generation cellular system is done in all regions of the world, the Far East, America and Europe, by amongst others, the International Telecommunications Union (ITU) and the European Telecommunications Standardisation Institute (ETSI).

3.1 Research within ACTS

ACTS (Advanced Communications Technology and Services) is one of the specific programmes of the 'Fourth Framework Programme of European Community activities in the field of research and technological development and demonstration (1994 to 1998)'. ACTS builds on the work of the earlier RACE programmes (Research and Development in Advanced Communications Technologies for Europe). Within ACTS a large number of projects study issues related to third generation mobile communication. One of the most important goals is to develop and demonstrate the viability of technology and solutions. Many results will be forwarded to standardisation bodies.

Some of the issues studied in ACTS that are applicable for third generation mobile communication are:

- *Wireless ATM*
  In particular low range high bit rate systems, typically 25 – 155 Mbit/s, normally with limited mobility.
- *Satellite component of UMTS*
  Addressing performance parameters, economic and technical feasibility.
- *Migration paths towards third generation*
  Network architectures, interworking, mobility management, service platform technology.
- *Radio transmission technology*
  Addressing adaptive antennas, multi-mode operation and software radio, amongst others
- *Mobile terminals and applications.*

3.2 Research within COST

COST (European Co-operation in the field of Science and Technical Research) is a framework for R&D co-operation in Europe, and involves participants from 25 European countries.

COST has research activities in many different domains, the largest of these being the telecommunications domain. Within this domain, around five ongoing projects address issues within mobile communications. Topics addressed include:

- *Antenna technology*
- *Mobile satellite communications*
- *Radio technology*
- *Radio wave propagation*
- *Network aspects.*

3.3 Research within EURESCOM

EURESCOM (European Institute for Research and Strategic Studies within Telecommunications) has several projects on third generation mobile communications issues. This includes projects on:

- *Mobility management for cordless terminals realised by IN*
- *Support of terminal mobility based on TINA*
- *Radio technology for broadband access*
- *Wireless ATM including mobility management.*

3.4 Standardisation within ITU

As mentioned previously, ITU is responsible for the work on IMT-2000. The standardisation of IMT-2000 (at that time called FPLMTS) was initiated in 1985.

3.4.1 The IMT Vision and IMT-2000 System

The standardisation of future mobile communications within ITU is based on the International Mobile Telecommunications (IMT) vision, which aims to provide seamless operation of mobile terminals throughout the world, where Anywhere – Anytime communications require coverage by both terrestrial and satellite networks. The satellite component, denoted Global Mobile Personal Communications by Satellite (GMPCS), comprises a wide range of global satellite systems providing fixed and mobile services in different frequency bands.

The current work concentrates primarily on IMT capabilities that will be available around the year 2000, with the target system designated IMT-2000. IMT-2000 represents the satellite and terrestrial component of IMT that will be available around the year 2000 and operate in the 2 GHz frequency band. Standardisation of IMT-2000 is one of the strategic priorities of the ITU.

One of the important early results was the contribution to the allocation of frequencies in the 2 GHz band, which was resolved by WARC 92.

3.4.2 Organisation

The work on IMT-2000 within ITU is split between ITU-R and ITU-T [10]. A special group (Intersector Co-ordination Group – ICG) takes care of the co-ordination of the work. ITU-R Task Group 8/1 does all the radio related work on IMT-2000/FPLMTS. Within ITU-T, the work is split in several study groups. Study Group 11, responsible for the signalling and protocols, is the lead Study Group.

3.4.3 Time Schedule

A phased approach is adopted for the definition of IMT-2000. According to the time schedule the specifications for Phase 1 will be ready by the end of this century [13]. The finalisation of the specifications will be preceded by a period of proposals for standardisation, evaluation and consensus building, as illustrated in Figure 4.

3.4.4 Services and Applications

It is the intention that IMT-2000 shall meet the service requirements as given in chapter 2 based on bit rates up to 2 Mbit/s. A list of capabilities is defined in [14] for the initial phase (called capability set 1), and a short summary is given in Table 1. Phase 2 will include bit rates higher than 2 Mbit/s, and provide services such as support of high data rate needs of portable computing users and support of enhanced multimedia communications requirements.

Services will be available virtually everywhere, including land, maritime, and aeronautical situations. The user of the personal terminal will be able to access at least a minimum set of services in any situation. This set of services includes voice telephony, a selection of data services, access to UPT, and an indication of other services available [15].

The applications in developing countries are considered important in the work on IMT-2000. The wireless local loop appli-

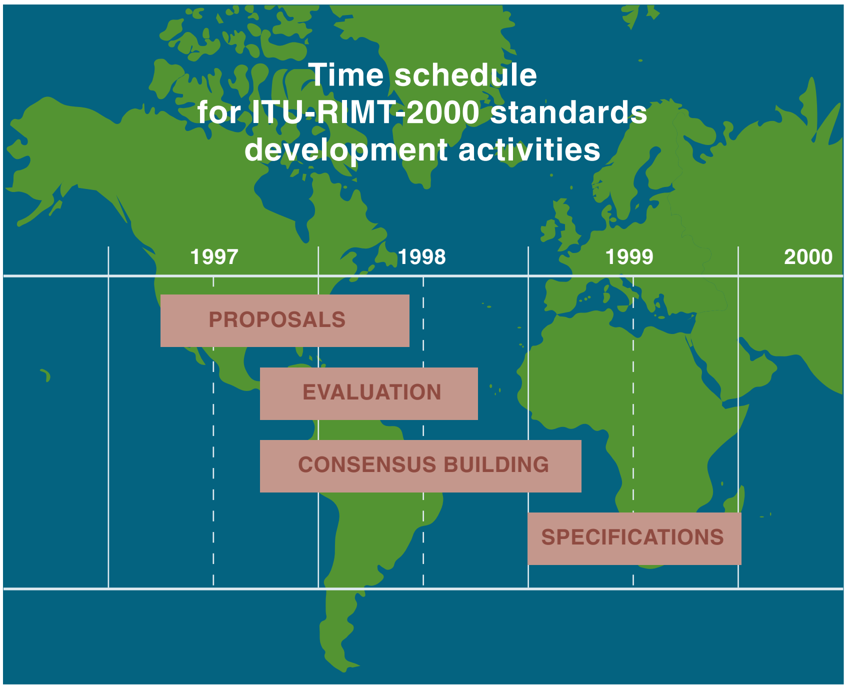


Figure 4 Time schedule for IMT-2000 standards development, Phase 1 (from ITU homepage, May 98, ref: http:// www.itu.int/imt/2-radio-dev/time/time.html)

Table 1 Key features in IMT-2000 Phase 1

Services
• At least 144 kbit/s in vehicular environment, at least 384 kbit/s for pedestrian environments and at least 2 Mbit/s for office environments
• High quality speech
• Support of advanced and existing addressing mechanisms
• Virtual home environment for service portability between IMT-2000 networks
• Intra-Family handover, seamless handover between cells of one IMT-2000 network
• Connection oriented and connectionless services
• Roaming among IMT-2000 family members
Terminals
• Range of terminals with different capabilities, from voice to multimedia, multiband and multimode terminals
Access networks
• New BSSs in the UMTS/IMT-2000 spectrum
• Support of packet services
Core networks
• Possible evolution from existing core networks
• Support of packet switched and circuit switched operation, based on Broadband Network Architecture (PDH/SDH/ATM)
• Interworking with ISDN, B-ISDN, X.25 PDN and IP data traffic

cation is important in this context, and the market is likely to be at least as big as the mobile market. Likewise, the satellite component of IMT-2000, together with other global satellite systems, will likely provide the first telephone in many rural villages. The terrestrial infrastructure will then follow as demand increases. Thus, integration of future terrestrial and satellite systems is highly desirable to facilitate seamless evolution.

3.4.5 System Design Considerations

IMT-2000 will meet the requirements discussed in chapter 2. It is stressed that it is important to have a high degree of commonality in the IMT-2000 family of systems. Commonality is important to reduce development costs, obtain cheaper equipment through economy of scale and limit the need for interworking functions. Commonality of radio interfaces is stressed in particular, to simplify the task of building multimode mobile terminals covering more than one operating environment. Within ITU, and especially from Japan and Europe, it is seen as mandatory that the radio interface should be able to operate in two different modes, FDD (frequency division duplex) and TDD (time division duplex), or rather, there will be different types of terminals for each of these modes. Depending on the specific technological solution that is chosen the deployment of the systems with these two modes will differ somewhat. At this point it is not possible to tell how the two modes should work together and be compatible with each other.

On a global basis there are some differences on how spectrum is used in different countries. This will obviously have impact on the specific implementation of any system and technology. In addition to this there are quite different existing technological platforms and regulatory regimes, and it is obvious that these differences will have an impact on the choice of system technology, system design and deployment. Because specific solutions have not yet been chosen, it is rather difficult to tell what the considerations and impact on design will be.

The IMT-2000 functional architecture is based on the IN concept, and the evolution of its functional capabilities is therefore closely linked to the evolution of IN. It must therefore be anticipated that IMT-2000 will take on board later versions of IN, switching and signalling standards, where radio and mobility management is a natural part of the protocols. A description of the functional model is given in Appendix 1.

3.5 Standardisation in ETSI

The ETSI General Assembly resolved a strategy for UMTS standardisation in 1996, stating that the work should proceed based on GMM recommendations. The work responsibilities are allocated to the Technical Committees SMG (Special Mobile Group), NA (Network Aspects) and SES (Satellite Earth Stations and Systems).

3.5.1 Organisation and Responsibilities

SMG is responsible for the UMTS Generic Radio Access Network, and for the GSM Core Network Evolution. NA6 UMTS is responsible for the ISDN Core Network Evolution. The UMTS satellite component is developed in a co-operation between SMG and SES, under the leadership of SMG.

Within SMG, the following Sub-Technical Committees are involved:

- SMG1 – responsible for specification of services
- SMG2 (UMTS Working Party) – responsible for the physical layer of the radio interface and the study of all radio engineering aspects
- SMG4 – responsible for UMTS data service requirements

- SMG 6 (WPA UMTS) – responsible for the network management functions and interfaces
- SMG 9 (UMTS Working Party) – responsible for the specification of the UMTS SIM-card / Mobile Equipment interface
- SMG10 (Working Party C) – responsible for UMTS security
- SMG11 – responsible for Speech Coding
- SMG12 – responsible for the specification of the network architecture, functions and signalling.

3.5.2 Time Schedule

The time schedule for UMTS development is subdivided into the following phases [4]:

- UMTS Phase 0 is the implementation of UMTS services in evolved GSM networks (using the evolved GSM radio access).
- UMTS Phase 1 will use the UMTS radio access. Operation is planned for 2002.
- After year 2002 UMTS will evolve in annual phases.

To reach the goals in Phase 1, the more detailed milestones have been identified. The UMTS concept is by now relatively clear, and the specifications for Phase 1 should be finalised before the year 2000. After a period of trials and industrial development, it is hoped that UMTS can be launched in 2002. This is illustrated in Figure 5.

3.5.3 Content of UMTS Phase 1

The overall content of UMTS has not yet been decided because the future development of UMTS will be dependent on the overall technology development and cannot be forecast in detail.

The content of the first phase of UMTS, however, has been agreed upon [16] and is described in Table 2 in terms of services, terminals, access networks and core transport.

3.5.4 Services and Applications

UMTS, like IMT-2000, will meet the service requirements as given in chapter 2. Although the services are similar to those of IMT-2000, there are some natural differences in the importance placed on different applications. In particular, it is seen that UMTS places less importance on wireless local loop applications and the satellite access component, which are particularly interesting applications in developing countries.

	1998	1999	2000	2001	2002
UMTS Phase 1 standards	■	■			
Pre-operational trials				■	
UMTS Phase 1 commercial operation					■

Figure 5 Time schedule for UMTS development, Phase 1

Table 2 Key features of UMTS Phase 1

Services
• Up to 144 kbit/s with high mobility, 144–512 kbit/s for wide area mobility, 2 Mbit/s for restricted mobility
• High quality speech using low bit rates
• Advanced addressing mechanisms
• Virtual home environment service portability
• Seamless indoor, outdoor and far outdoor
• Dual mode / band of operation of GSM/UMTS in one network
• Roaming between GSM and UMTS networks
Terminals
• Dual mode/band GSM/UMTS
• Multimedia terminals
• Adaptive terminals
Access networks
• New BSS in the UMTS/IMT-2000 spectrum
• Flexible, adaptive radio interface
Core transport
• Evolution from GSM and ISDN
• Support of variable bit rates and mixed traffic types
• Support of packet data by Internet protocols
• Mobile/fixed convergence elements

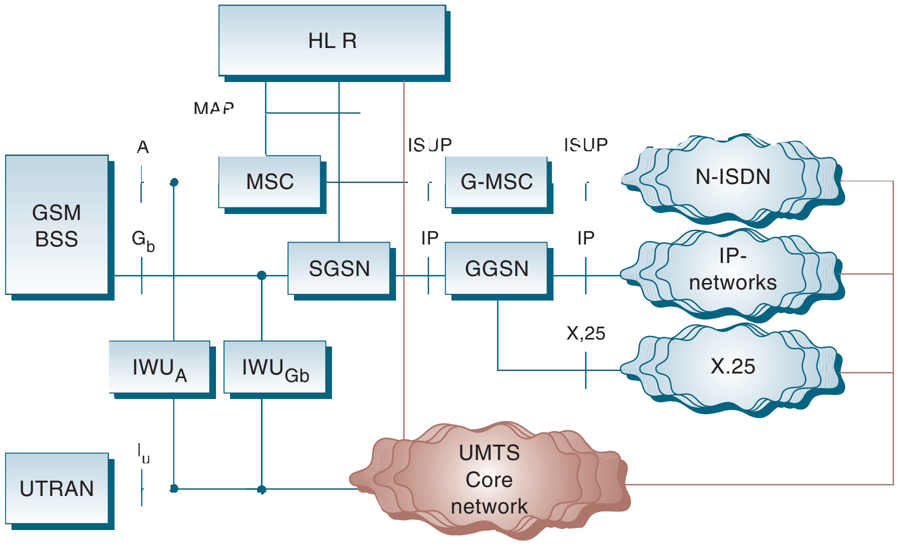


Figure 6 Evolution of UMTS. Blue part refers to UMTS Phase 1, red part to later phases

3.5.5 System Design Considerations

UMTS will meet the system design requirements given in chapter 2. The importance of modularity, flexibility and commonality is equally important as for IMT-2000.

Even within Europe it is regarded mandatory to have both TDD and FDD mode terminals. The specific solutions at a technical level are discussed at this moment. However, within Europe and countries that choose the UMTS system, the basis from second generation is probably not very different. These countries probably have GSM-900 and/or GSM-1800, and some of them even DECT. In addition, they have some Wireless Local Loop (WLL) systems based on DECT or other technology. System design will therefore very much depend on the specific need of different services in different deployment scenarios. In addition and very much in connection with this, the regulatory framework within the region (countries) will put some framework on operation of second and third generation systems. This framework will even create new requirements and initiate new development within second generation systems.

The original idea of UMTS was a new network that integrated all existing and new services in one universal network. This view has changed, and the current idea is to support services in different networks based on interworking, in particular interworking with GSM. This close link to GSM is one of the differences between UMTS and IMT-2000. It is the intention that UMTS will evolve from the GSM as illustrated in Figure 6.

UMTS Phase 1 will consist of a UMTS Terrestrial Radio Access Network (UTRAN). The UMTS Core Network will be introduced in later phases. Thus in Phase 1, UTRAN will use the GSM technology as the core network. As seen, this includes both the circuit and packet switched infrastructure. In order to interact with the GSM infrastructure, interworking units are required at the UTRAN interface.

The functional architecture proposed for UMTS is not so closely linked to the IN architecture as is seen for IMT-2000. One particular case is the fact that UMTS intends to employ the protocol developed for GSM for interrogation with the HLR, ie. the Mobile Application Part – MAP. The corresponding protocol in IMT-2000 is defined to be the IN Application Part – INAP. This difference is more politically than technically motivated, and moreover, it is not crucial, since the conversion between the protocols is easily performed.

Whereas ATM is specified as an important transport mechanism in IMT-2000, there is not yet full consensus on this issue in UMTS. It is anticipated, however, that ATM will become one of several optional transport mechanisms, as for IMT-2000. Originally, the idea was that UMTS should be a mobile extension of B-ISDN services. However, the fast growth of the Internet has made this an equally important partner network.

Among the system design considerations it is important to mention that an agreement on the UMTS Terrestrial Radio Access interface is reached. This agreement is a compromise between the concepts W-CDMA and CD-TDMA, both techniques being supported and allocated to different parts of the frequency band.

Appendix 1 – IMT-2000 functional architecture

A suitable functional model based on the service and feature requirements for IMT-2000 is essential for developing the necessary signalling protocols and for specifying the appropriate physical interfaces. The functional model specified for IMT-2000 is implementation-independent and captures the basic functional entities (FEs), and specifies relationships between the FEs. The physical interfaces to be specified are identified by using suitable mapping of various FEs in the functional model to physical entities (eg. base station (BS), mobile switching centre (MSC), home location register (HLR), visited location register (VLR), and service control point (SCP). Figure A1 shows an example of a mapping into physical entities based on the FEs described in the following.

The functional architecture for IMT-2000 as described in [10] is based on the modelling principles used for developing the signalling protocols for IN (Intelligent Networks) and is shown in Figure A2.

The management and control of radio resources will be carried out by procedures that operate in parallel with and rather independent of the overall communication control procedures. Two planes have therefore been defined:

- Radio Resource Control (RRC) plane
- Communication Control (CC) plane.

The basic management of radio resources resides in the RRC plane, while the overall bearer and connection control resides in the CC plane. Special plane interaction processes perform the synchronisation between the two planes.

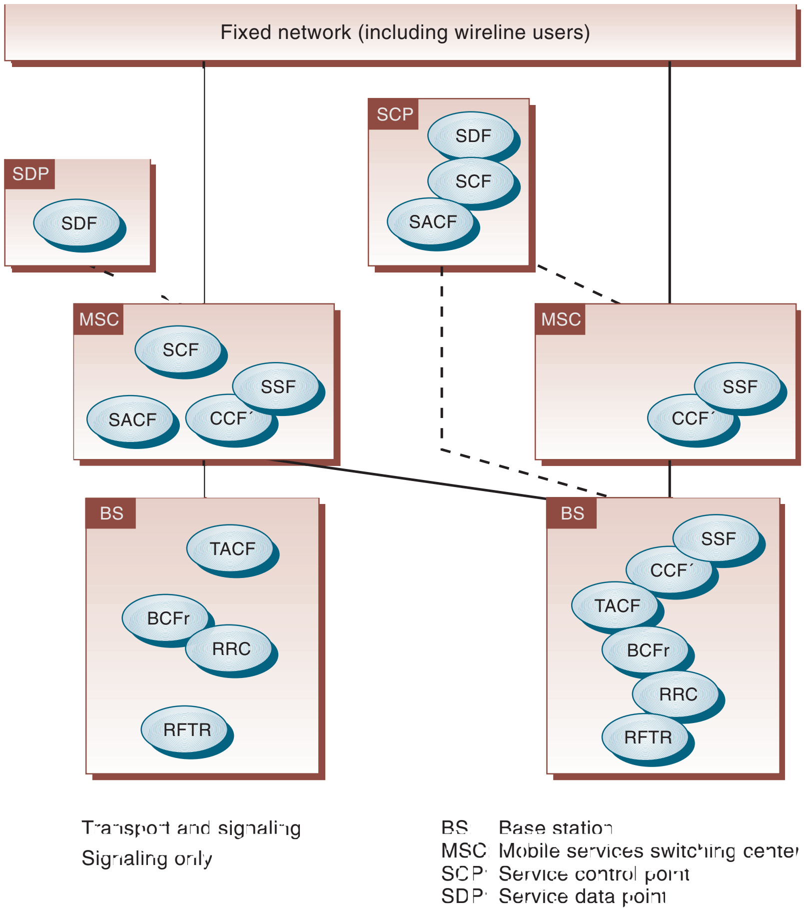


Figure A1 A standalone IMT-2000 network and allocation of functional entities to physical nodes. The functional entities are described in Table A1

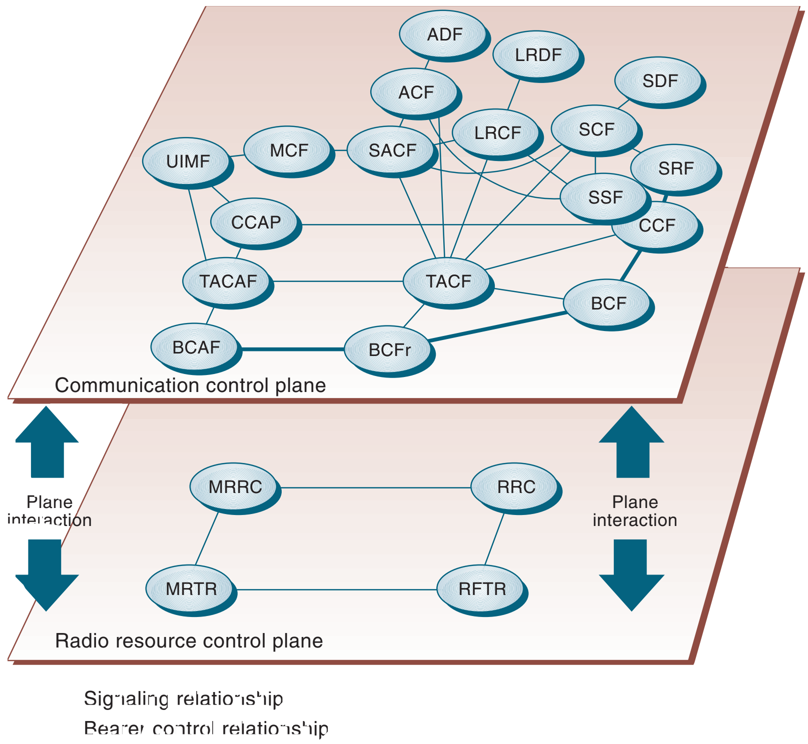


Figure A2 The IMT-2000 functional architecture

Table A1

RRC plane functions

These functions are in charge of assigning and supervising radio resources as required in the radio access subsystems.

- **RRC – Radio Resource Control**
  Handles overall control of the radio resources, such as selection and reservation of radio resources, handover decision, RF power control, and system information broadcasting.
- **MRRC – Mobile Radio Resource Control**
  Handles the mobile side of RRC.
- **RFTR – Radio Frequency Transmission and Reception**
  Handles the radio transmission and reception of control and user information at the fixed side of the radio interface, including radio channel and error protection coding and decoding.
- **MRTR – Mobile Radio Transmission and Reception**
  Handles the radio transmission and reception of control and user information at the mobile side of the radio interface. This includes radio channel and error protection coding and decoding.

CC plane functions

These functions are in charge of overall access, service, call, bearer and connection control.

- **SDF – Service Data Function**
  FE that handles storage and access to service- and network-related data, and provides consistency checks on data (eg. service profile and mobile multimedia attributes)
- **SCF – Service Control Function**
  FE that contains the overall service control logic and handles service related processing activity
- **LRDF – Location Registration Data Function**
  Handles storage and access to subscriber identity and location data, and provides consistency checks on data (eg. location information, identity data, and active/inactive status)
- **LRCF – Location Registration Control Function**
  Contains the mobility control logic for location registration. The functionalities include location management, identity management, user verification, paging control, and providing routing data.
- **ADF – Authentication Data Function**
  Handles storage and access to authentication data (eg. security-related parameters) and provides consistency checks on data.
- **ACF – Authentication Control Function**
  Contains the mobility control logic for authentication. The functionalities include user authentication, authentication processing, and confidentiality control (eg. ciphering management)
- **SSF – Service Switching Function**
  FE associated with the CCF'. It provides the set of functions required for the interaction between CCF and SCF.
- **CCF' – Call Control Function (enhanced)**
  FE that provides call/connection processing control (eg. establishes, maintains and releases call instances); provides trigger mechanisms to access IN functionalities; establishes, maintains, and releases bearer connections in the network, and so on.
- **SRF – Specialised Resource Function**
  FE that provides the specialised resources required for the execution of IN-provided services (eg. user signalling receivers, announcements, conference bridges), mobile multimedia services and packet data transfer services.
- **SACF – Service Access Control Function**
  Provides non-call- and non-bearer-associated processing and control (eg. in relation to mobility management).
- **TACF – Terminal Access Control Function**
  Provides the overall control of the association and connection(s) between a mobile terminal and the network. Examples of functionalities include terminal paging execution, paging response detection and handling, handover decision and handover execution.
- **BCF – Bearer Control Function**
  Controls the interconnection of bearers.
- **BCFr – Bearer Control Function (radio bearer associated)**
  Controls the interconnection and adaptation of a radio bearer to the corresponding fixed line bearer.
- **MCF – Mobile Control Function**
  Provides the overall service access control logic and processing at the mobile side of the radio interface. Specifically, it interacts with the network for non-call/non-bearer-associated services (eg. mobility management).
- **CCAF' – Call Control Agent Function (enhanced)**
  FE that provides service access for users. It is the interface between user and network CCFs.
- **TACAF – Terminal Access Control Agent Function**
  Provides access for the mobile terminal. Examples of functionalities include terminal paging detection and paging response initiation.
- **BCAF – Bearer Control Agent Function**
  Controls the interconnection and adaptation of a radio bearer to the rest of the mobile terminal.
- **UIMF – User Identification Management Function**
  Provides the means to identify both the user and the mobile terminal (or termination) to the network and/or to the service provider. It includes functionalities to store user-related information such as user identity and security- and privacy-related information, provide for user authentication, ciphering key generation, and so on.

References

1 Grimstveit, L, Hans Myhre, H. The history of mobile communications in Norway. *Telektronikk,* 91, (4), 1995, 15–20.

2 WARC'92. *The world administrative radio conference for dealing with frequency allocations in certain parts of the spectrum.* Malaga-Torremolinos, 1992.

3 TS 22.05 ver. 3.0.0 1998-03. Universal Mobile Telecommunications System (UMTS). *Services and Service Capabilities,* UMTS 22.05 ver. 3.0.0.

4 Universal Mobile Telecommunications System (UMTS). *Selection procedures for the choice of radio transmission technologies of the UMTS.* (UMTS 30.03).

6 UMTS Baseline document. *Positions on UMTS agreed by SMG.* (UMTS 30.01).

5 ITU-R. *Guidelines for evaluation of radio transmission : Technologies for IMT-2000.* (Recommendation ITU-R M.1225.)

7 ETSI programme Advisory Committee (PAC). *Global Multimedia Mobility (GMM) : a standardisation framework.* ETSI/PAC16(96)16, 1996.

8 ITU-T SG11. *Draft new recommendation Q.FNA, network functional model for FPLMTS, Version 6.1.0.* Geneva, January 1997.

9 ETSI. *Special Mobile Group : UMTS. Network Principles.* Sophia Antipolis, 1997. (Draft Technical Specification, UMTS 23.05 v. 0.10.0.)

10 Pandya, R et al. IMT-2000 Standards : Network Aspects. *IEEE Personal Communications,* August 1997.

11 ITU-R M.1078.

12 TS 33.20 ver. 3.0.1, 1997-07, Universal Mobile Telecommunications System (UMTS). *Security Principles for UMTS.* UMTS 33.20 ver. 3.0.1.

13 ITU home page, time schedule http://www.itu.int/imt/2-radio-dev/time/time.html

14 ITU-T Study Group 2. *Liason Statement to Task Group 8/1 on IMT-2000 Family Concept for Members of the IMT-2000 Family Phase 1.* Document 8-1/6-E, 27 November 1997.

15 ITU-R. *Framework for services supported on International Mobile Telecommunications-1000 (IMT-2000).* Rec. M.816-1.

16 UMTS Baseline document, UMTS 30.01 V3.1.0 (1997-10). *Positions on UMTS agreed by SMG.*

17 *Spectrum for IMT-2000,* UMTS Forum Report, October 1997.

Jan Eriksen is Research Scientist at Telenor R&D, Kjeller, where he is working with issues related to cellular systems and personal communications. His main involvement is within the physical layer of the radio interface and wave propagation. He is currently also involved in the ongoing standardisation work within ETSI SMG2 and ITU-R RG 8/1 on 3rd generation systems.

e-mail: jan.eriksen@fou.telenor.no

Geir Olav Lauritzen is Research Scientist at Telenor R&D, Kjeller. He is working in the field of mobile communication with special interests in network functionality and capacity aspects for land mobile systems.

e-mail: geir.lauritzen@fou.telenor.no

Bjørn Harald Pedersen is Research Scientist at Telenor R&D, Kjeller, in the Mobile Communication group. He is currently working with network and service aspects, functional architecture and requirements for third generation mobile communication systems. He has also been participating in the standardisation of UMTS in ETSI.

e-mail: bjorn.pedersen@fou.telenor.no

Stein Svaet is Research Scientist at Telenor R&D, Kjeller, where he has been employed since 1988. He is working in the field of mobile and personal communications, with a special interest in network architectures for UMTS and mobile broadband systems.

e-mail: stein.svaet@fou.telenor.no

Global Mobile Personal Communications by Satellite

ARILD FLYSTVEIT AND ARVID BERTHEAU JOHANNESSEN

With the forthcoming introduction of systems for Global Mobile Personal Communications by Satellite (GMPCS), mobile satellite services will finally become truly personal. Already available is Mobiq, a notebook PC sized terminal based on the Inmarsat mini-M standard, heating up the marketplace for Iridium, Globalstar and ICO announcing introduction of their handheld services from late 1998 onwards.

Iridium will undoubtedly be first to market, emphasized by fifteen successful launches in twelve months completing the venture's 66 satellite constellation. However, the challenge of providing a continuous global service is high due to the fact that these systems rely on large constellations of medium- or low earth orbiting satellites.

The GMPCS service concept is almost identical to GSM, including voice, fax and low speed data services. The systems even integrate with GSM and other cellular standards by the provision of multi-mode terminals.

The global pocket phone represents a milestone in telecommunications and it is expected that future evolution will focus on increasing the transmission speed rather than shrinking the terminal size. Along this evolutionary path, satellite platforms represented by Teledesic and Inmarsat Horizons, may very well take the lead in the race where cellular systems so far have been the front runners.

1 Introduction

For some years several systems have been under development for provision of voice- and data services to multi-mode handsets world-wide using low or medium earth orbiting satellites. These systems have been referred to under acronyms such as S-PCS (Satellite Personal Communication Services) and Big LEOs (Low Earth Orbits). In 1996, the International Telecommunications Union (ITU) World Telecommunications Policy Forum labelled them GMPCS – Global Mobile Personal Communications by Satellite.

The adventure began back in 1985, when Motorola engineer Bary Bertiger conceived of the Iridium system, as a result of seeking a means to enhance cellular[1] coverage. Bertiger shared his idea with colleagues Ken Peterson and Ray Leopold, and the trio considered various options – such as high-flying balloons – before settling on LEO satellites. Two years later, research and development got under way and in 1990 the Iridium system was announced at simultaneous press conferences in Beijing, London, Melbourne and New York [1].

Shortly after Iridium's announcements, several other concepts and players entered the arena, and today Iridium, Globalstar and ICO are seen as the main contenders. The expectations are high and so are the risk factors, both technologically and economically. In addition, there are complex issues of political and regulatory issues to be solved before these multi-billion dollar systems can offer their services as promised from late 1998.

In this article we will discuss some of the technical aspects and parameters of the GMPCS systems and illustrate the planned constellations. The market opportunities will also be briefly touched upon as will recent developments within the cellular and satellite environment, including ongoing regulatory activities. Finally, some thoughts on the expected future evolution of personal satellite communications are presented.

2 GMPCS – technical description

The planned GMPCS systems have much in common. They all target similar markets with similar services and they intend to complement the coverage of cellular systems. However, their technical characteristics such as choice of orbit, multiple access method and frequency band, differ.

In this chapter we will discuss services and technical aspects for GMPCS systems.

2.1 Service portfolio

The GMPCS service portfolio is almost identical to what we are familiar with in GSM and typically includes:

- voice – quality comparable to GSM
- 2.4 – 7.2 kbps data and fax[2]
- paging
- positioning[3]
- GSM compatible SMS (Short Message Service)
- ISDN (Integrated Services Digital Network) supplementary services.

The end-user charges are not settled, but the potential operators have indicated that they will be in the $1–3 range. Landline charges may be added for long distance calls, making the end-user charges even higher.

2.2 Space segment

Orbit alternatives

There is no such thing as an optimum orbit. This is clearly illustrated by the fact that none of the system operators have landed on the same orbital characteristics, resulting in four definitions of the perfect orbit for delivery of communications to pocket-sized terminals.

The high altitude (36,000 km) of traditional communication satellites in the geostationary orbit (GEO) results in time delays of the signal that could make them disadvantageous for GMPCS. Additionally, powerful transmitters and receivers, difficult to incorporate into a pocket-sized unit, are required to compensate for the large signal attenuation. Also, users in higher latitudes can experience service problems due to the low elevation angle, which can result in signals being blocked by buildings, trees and mountains.

To overcome these problems (but possibly introducing others), the GMPCS proposals are based on highly inclined LEO and MEO (Medium Earth Orbit) orbits made possible by advances in satellite technology, typically operating around 1,000 km and 10,000 km above the earth surface respectively.

As a consequence of the lower satellite altitude, the number of satellites needs to

[1] *In this article, the term 'cellular' means terrestrial Public Land Mobile Networks (PLMNs), although GMPCS systems also apply cell structured coverage.*

[2] *Higher data rates are offered for fixed installations.*

[3] *The GMPCS positioning services are not designed to compete with the far more accurate GPS (Global Positioning System).*

be increased to achieve continuous global coverage. Typically, a global LEO constellation consists of 40–80 satellites, whereas a MEO constellation requires approximately 10–15 satellites to cover the earth.

Inter-satellite links

The concept of Inter Satellite communication Links (ISLs) makes it possible to switch traffic between satellites in the same or adjacent planes. Traffic may be carried over long distances independently of ground networks, thereby minimising the required number of gateways for global coverage.

ISLs are seen as one of the major technological challenges for new satellite systems, such as GMPCS.

Handovers

While a GEO satellite covers a fixed area of the earth, LEO and MEO satellites move relative to the earth with the coverage area of a specific satellite varying over time as its beams move across the earth. Because of this satellite motion, GMPCS systems require a more complex ground segment, tracking satellites and performing handovers to avoid losing calls as satellites move over the horizon.

Two types of handovers exist, comparable to intra-MSC and inter-MSC handovers in GSM. Intra-satellite handover, which occurs in systems where a satellite's coverage area moves continuously as the satellite moves, means that the call is handed over to another beam (cell) on the same satellite. In systems where the satellites redirect their antenna to maintain a fixed coverage area on the ground as the satellites move, intra-satellite handovers are rare. Handovers between satellites are called inter-satellite handovers.

For systems with ISLs handovers are more complicated, since the whole chain of crosslinks must be handed over to the next chain of crosslinks simultaneously. Even though interworking with a cellular system will be implemented, functionality for handling of inter-system handovers between satellite systems and cellular systems for multi-mode terminals will not be present.

The average time between inter-satellite handovers is approximately 4–8 minutes for LEO constellations depending on altitude, but in special situations handovers can occur as frequently as every 10 seconds. For MEO systems, inter-satellite handovers occur every hour on average. Of course, intra-satellite handovers occur more frequently and depend on the number of spot beams provided by the satellite [2].

End-to-end communication delay

As the transmission path between user and satellite is much shorter for LEO and MEO systems than for GEO systems, the propagation delay is reduced dramatically. Isolated, this affects voice quality, which will be perceived as better. For data communications, the reduced delay has positive impacts on protocol compatibility.

Table 1 illustrates typical minimum and maximum one-way propagation delays between user and satellite for a LEO (Iridium) and a MEO system (ICO) [2][3]. As a reference, the minimum one-way propagation delay for a GEO satellite is approximately 120 ms.

Table 1 One-way propagation delay user-satellite

	LEO	MEO
Minimum [ms]	2.60	34.5
Maximum [ms]	8.22	48.0

In addition to propagation delay, the end-to-end communication delay is influenced by several other parameters. Significant contributors are voice codecs, switching/buffering operations and assembling/disassembling processes. [2] indicates that the delay in a typical TDMA system like Iridium is in the range of 100 ms, excluding propagation delay. Similar delays are expected also in CDMA based systems.

Coverage – truly global?

The GMPCS operators claim to offer global coverage. For provision of services to the global *population,* this is virtually true. However, regarding global *geographic* coverage, the systems are slightly different. These differences are mainly due to the space segment configuration chosen and in particular the inclination of the orbital planes:

- In LEO systems with polar orbits, the highest concentration of satellites is above the North and South poles, resulting in excellent coverage of the polar areas. At the equator, the distance between the orbits is at its maximum. Decreasing the inclination results in reduced coverage at high latitudes and increased 'satellite density' at equator.
- Due to the large coverage area of a MEO satellite, decreasing the inclination does not affect the coverage of the polar area in the same way as it does for LEO systems.

Unlike GSM, the GMPCS handheld terminals cannot be used indoors except at windows with line of sight to at least one satellite. Obstacles like trees, mountains and buildings may cause blockage, and the higher the minimum elevation angle for which a system can guarantee the availability of at least one satellite at any time, the better the quality of the system. None of the systems have been designed to provide communication in cities.

Multiple access methods

The two most frequently used methods for shared access on the user link in modern cellular systems are CDMA (Code Division Multiple Access) and TDMA (Time Division Multiple Access). As an example, the GSM cellular standard uses TDMA, whereas the IS-95 standard in the US is based on CDMA. GMPCS systems will be based on either of these two principles.

Frequency bands for MSS

In 1992 WARC (World Administrative Radio Conference) allocated frequency bands in L-band (16.5 MHz) and S-band (2 x 30 MHz plus 16.5 MHz) to MSS (Mobile Satellite Systems) systems. Figure 1 illustrates the frequencies to be used on the mobile link for the various GMPCS operators.

The sharing of the 16.5 MHz wide L-band spectrum to be used by the Iridium TDMA system and the two CDMA systems Globalstar and Odyssey has been an issue for negotiations and discussions in various forums over the last years. Currently, the agreement is that Iridium will use the band 1610–1615.15 MHz, whereas the CDMA systems will share the remaining 11.35 MHz of the band [4]. However, if some of the systems never go operational, the band will be

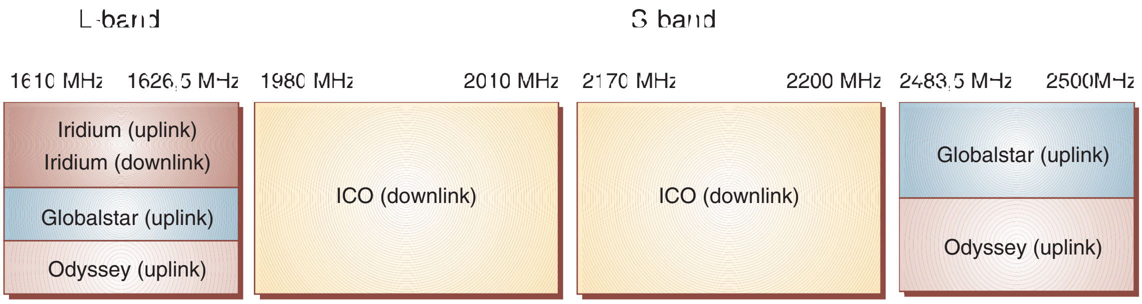


Figure 1 Mobile link frequency assignments

shared equally among the remaining operators.

C-band and Ka-band frequencies will be used for feeder links.

2.3 Ground and user segment

Cellular systems integration

The GMPCS systems are designed to be compatible with cellular systems and complement their coverage. GSM network elements and functionality like mobility management will be reused with minor modifications, and multi-mode handsets are offered for various cellular standards. Typically, the terminal will operate in its cellular mode when it is within cellular coverage and only be switched to satellite mode (manually or automatically) when the user is out of cellular coverage or in an area only offering an incompatible standard. Figure 2 illustrates this concept.

Gateways

All systems use a set of gateways to interconnect with terrestrial networks, fixed and mobile. The exact number of gateways depends on the orbital height of the satellites, use of ISLs and the operational strategy.

The operation of a gateway could, from a revenue perspective, be desirable or be a means to avoid bypassing of the national fixed networks. Therefore, the number of gateways in a system could be higher than strictly required.

GMPCS gateways are interconnected through a global network used for signalling, exchange of administrative data and in some cases even for routing of traffic.

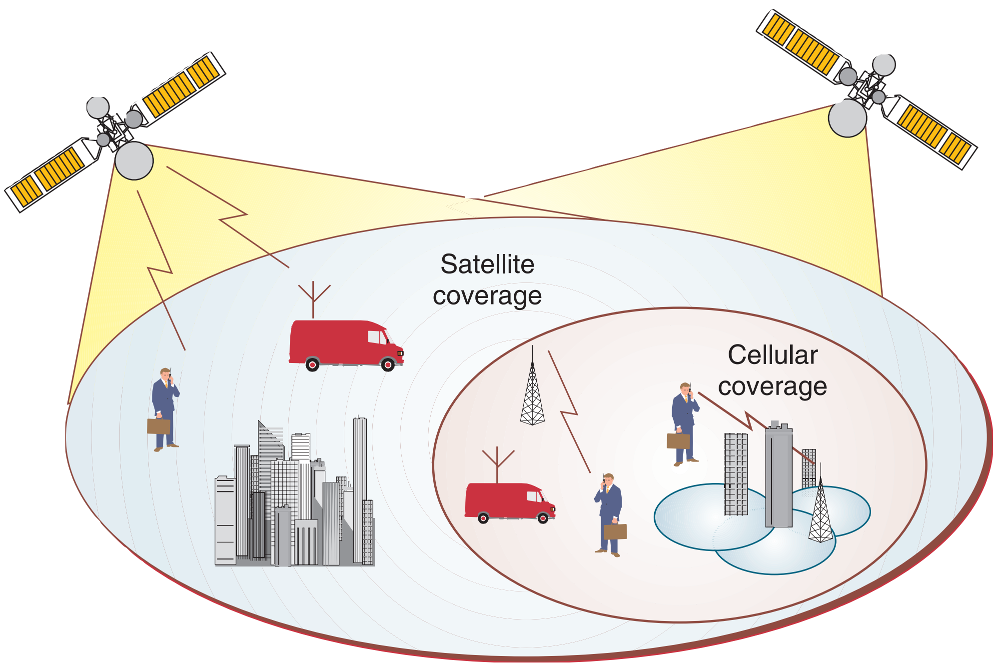


Figure 2 GMPCS complements cellular coverage

Figure 3 Prototype of the Kyocera and Motorola dual-mode handsets for Iridium

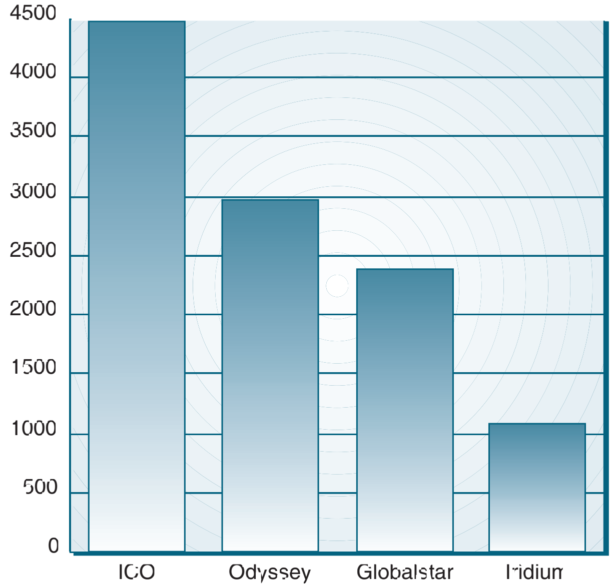


Figure 4 Capacity per satellite in number of voice channels

All systems aim to reuse existing GSM technology and interfaces in their gateways.

Terminal equipment

The handheld, portable phones used for GMPCS are slightly larger and heavier than ordinary single mode cellular handsets. Terminals will be available as single mode satellite only and as dual (triple) mode cellular/satellite. Available prototypes of such dual mode handsets weigh approximately 400 g. The average transmit power of the terminals is in the same range as for cellular only handsets, which is a necessity in order to comply with the type approval conditions for handheld terminal equipment.

Terminals are expected to be more expensive than existing cellular handsets. Information available indicates prices varying between $500 and $3,000 per unit.

Pagers and fixed terminal types, mainly designed for rural and remote areas, will also be available. These include telephone booths and satellite radio-antenna units for standard PABX equipment.

2.4 Numbering and addressing

One number

The ITU has assigned the shared country code 881 to GMPCS operators on a shared basis. One digit Network Identification Codes following the country code identifies each GMPCS operator as one interconnected global entity and enables routing of GMPCS calls to the nearest gateway. Single mode GMPCS terminals will be assigned numbers from this numbering plan.

Dual-mode GMPCS subscribers will also be addressed on one number, independent of whether she/he is communicating in cellular or satellite mode. This single number could either be a national cellular number or an 881 satellite number depending on the profile and needs of each subscriber.

Apart from the user-convenience of having one address, the choice of numbering plan impacts routing and charging, especially in the to-mobile direction, as discussed below.

National cellular numbers

A call *to* a dual-mode GMPCS subscriber with a national cellular number will always be routed via his/her home cellular network (country), irrespective of the communication mode used, because cellular numbers are part of the national numbering plan and not recognised as GMPCS/cellular numbers until they are analysed in the home network.

If the call originates outside the subscriber's home country, it will always be charged as any other international call to that country. If the call originates within the subscriber's home country, it will be charged as a cellular call. When the dual-mode GMPCS subscriber is operating in satellite-mode, this charge is not expected to be sufficient to cover the expenses. As a result, the called GMPCS subscriber will need to be charged for incoming calls (split billing).

The question then is whether a dual-mode GMPCS subscriber finds it attractive to pay for incoming calls when operating in satellite mode.

Dedicated 881 country code

A call to a dual-mode GMPCS subscriber with an 881 number will always be routed to a satellite gateway. If the terminal is in cellular mode, the call will be re-routed to the destination cellular network.

Independent on where in the world the call originates such calls will always be charged at the correct GMPCS charge, irrespective of whether the terminal is operating in cellular or satellite mode. This charge is expected to be higher than for cellular calls.

The question in this case is whether the calling party finds it attractive always to pay GMPCS charge to reach a dual mode GMPCS subscriber, even when this subscriber is within GSM coverage and communicating in GSM mode.

2.5 System capacity

The GMPCS systems promote themselves as an extension to cellular systems rather than direct competitors. The ability to compete is in fact rather low, as the total number of voice channels for all four systems combined is below 200,000.

The total frequency allocations for MSS in L- and S-band are 93 MHz. For comparison, one channel for analogue satellite TV broadcast requires approximately 27 MHz. This illustrates the limitation in the total capacity these systems can offer, and the limitation in number of systems. A high degree of frequency reuse is required to offer an acceptable capacity.

Satellite capacity is for Iridium and Globalstar satellites limited by power constraints. For replacement satellites the capacity may be higher, as this constraint is expected to be overcome.

The number of voice channels per satellite is higher for the MEO systems, which is natural due to the lower number of satellites and the instant area to be covered by one satellite. Since satellite capacity influences satellite cost, using MEO type satellites in a LEO configuration to increase capacity would raise the total system cost to an unacceptable level.

3 Systems overview

3.1 Iridium

Motorola announced the ambitious Iridium plan eight years ago, as the first of nearly a dozen quite similar proposals from various vendors and operators. The system was named after the element with 77 electrons, since the original proposal assumed a 77 satellite constellation. In 1992 Motorola announced system enhancements enabling a reduction in size of the constellation to 66 satellites. The 66 satellites are located 780 km above the earth's surface in a LEO orbit [5]. 11 satellites are equally spaced in each of the 6 near polar orbits (inclination 86.4°) together providing global coverage. Each satellite completes a full orbit cycle every 100 minutes. An illustration of the Iridium satellite constellation is given in Figure 5.

Each of the 66 satellites weighs approximately 700 kg and provides coverage in an area of 600 km in diameter. The 48 spot beams provide a cellular footprint structure. The satellite lifetime is expected to be 5–8 years. Iridium satellites have been launched using Delta II (Boeing, USA), Long March 2C/2D (China Great Wall Industry, China) and Proton (Khrunichev State Research and Production Space Centre, Russia) rockets, each carrying five, two and seven satellites respectively. Fifteen successful launches since May 1997 have placed 67 operational satellites in orbit. The last five were added on 17 May 1998 and marked the technical completion of the venture's 66 satellite constellation. Five of the 72 satellites which have been launched are suffering from mechanical or communication problems, reducing the number of spare satellites to one. Table 2 lists the Iridium launches.

Each satellite is linked with two satellites in the same plane (one in front and one behind) and with one in the adjacent plane on both sides via ISLs. None of

Table 2 The Iridium satellite launch history

Launch Date	Launch Vehicle	Launch Provider	# of Satellites	Launch Site
5 May 1997	Delta II	Boeing	5	Vandenberg, California
18 June 1997	Proton	Krunichev	7	Baikonur, Kazakhstan
9 July 1997	Delta II	Boeing	5	Vandenberg, California
20 August 1997	Delta II	Boeing	5	Vandenberg, California
14 September 1997	Proton	Krunichev	7	Baikonur, Kazakhstan
26 September 1997	Delta II	Boeing	5	Vandenberg, California
8 November 1997	Delta II	Boeing	5	Vandenberg, California
8 December 1997	Long March 2C/2D	China Great Wall Industry	2	Taiyuan, China
20 December 1997	Delta II	Boeing	5	Vandenberg, California
18 February 1998	Delta II	Boeing	5	Vandenberg, California
25 March 1998	Long March 2C/2D	China Great Wall Industry	2	Taiyuan, China
30 March 1998	Proton	Krunichev	7	Baikonur, Kazakhstan
6 April 1998	Delta II	Boeing	5	Vandenberg, California
2 May 1998	Long March 2C/2D	China Great Wall Industry	2	Taiyuan, China
17 May 1998	Delta II	Boeing	5	Vandenberg, California

Iridium's competitors use this technology.

The Iridium ground network consists of 11 gateways interconnecting the Iridium satellite network with the PSTN (Public Switched Telephone Network) and other terrestrial cellular networks. The small number of gateways is possible since ISLs are used.

Iridium has signed a $3,450 million contract with Motorola for building, launching and operating the space segment, excluding gateways. In addition, Iridium has executed a five year $2,880 million contract with Motorola for operation and maintenance of the satellite network, including satellite launches for replenishment. Raising $4,715 to fund the ongoing infrastructure build-out and commercialisation, Iridium has completed its anticipated funding needs.

The system is owned by Iridium LLC, an international consortium of 19 equity partners from 15 countries. Motorola is the largest investor (19 %), but the list of investors also includes other large manufacturers as Lockheed Martin and strong telecom operators as Korea Mobile Telecom.

Iridium will launch their service 23 September 1998.

Figure 5 The Iridium satellite configuration

3.2 Globalstar

The Globalstar system [6] was originally proposed in 1986 by Ford Aerospace to provide mobile satellite communications services to automobiles. Globalstar LP, which was founded by Loral Space & Communications and Qualcomm in 1991, has now evolved into a world-wide consortium of 12 strategic partners including Alcatel, France Telecom, Air Touch and Vodafone.

48 LEO satellites organised in eight circular orbits inclined 52° relative to the equatorial plane with six equally spaced satellites per orbit constitute the Globalstar space segment. Each satellite completes an orbit cycle every 113 minutes at an orbital height of 1,414 km. The constellation covers the area between 70°S and 70°N latitude with the 'satellite density' on each hemisphere at its highest in a band centred approximately at ±40° latitude. The non-coverage area over the poles can be recognised in Figure 7.

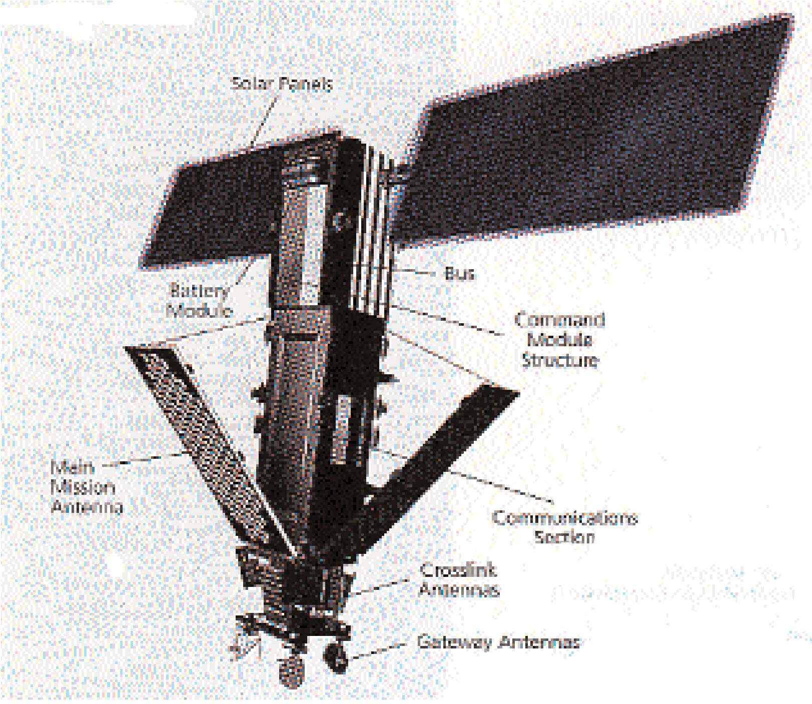


Figure 6 The Iridium satellite

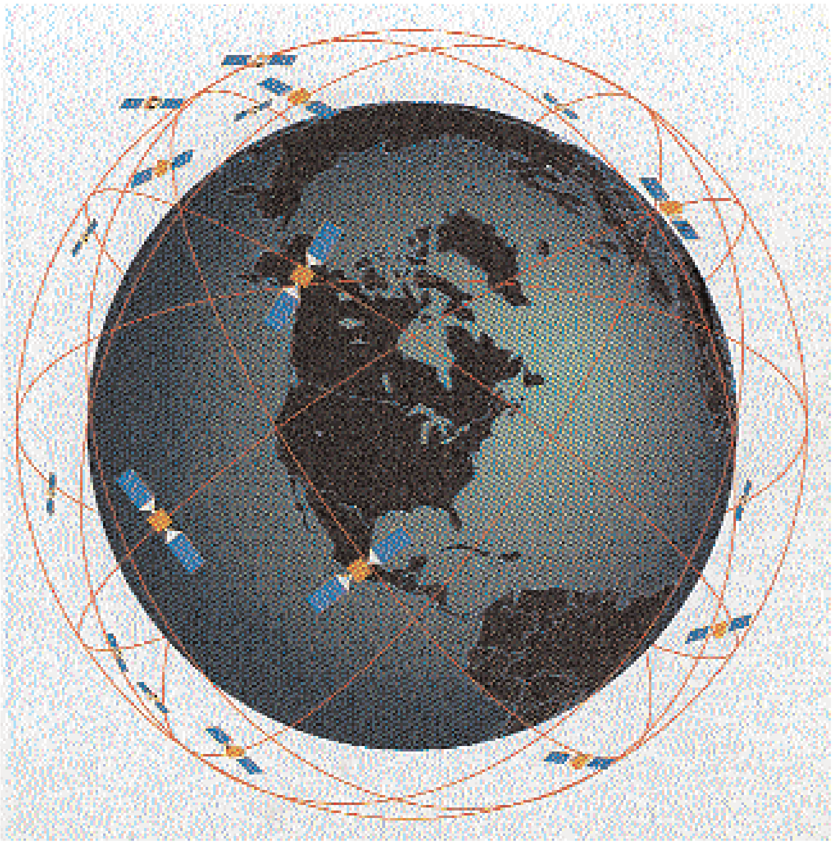


Figure 7 The Globalstar satellite configuration

Table 3 Globalstar calendar of completed and scheduled launches

Launch Date	Launch Vehicle	Launch Provider	# of Satellites	Launch Site
14 February 1998	Delta II	Boeing	4	Cape Canaveral, Florida
24 April 1998	Delta II	Boeing	4	Cape Canaveral, Florida
3Q98	Zenit	NPO Yuzhnoya	12	Baikonur, Kazakhstan
4Q98	Zenit	NPO Yuzhnoya	12	Baikonur, Kazakhstan
4Q98	Zenit	NPO Yuzhnoya	12	Baikonur, Kazakhstan
1Q99	Soyuz	Starsem	4	Baikonur, Kazakhstan
1Q99	Soyuz	Starsem	4 (1)	Baikonur, Kazakhstan
2Q99	Soyuz	Starsem	4 (1)	Baikonur, Kazakhstan

The 450 kg Globalstar satellites are unlike the Iridium satellites of the bent-pipe type. Each satellite has 16 beams and is designed for a minimum lifetime of 7.5 years. The Globalstar constellation will be lifted into space using Delta II (Boeing, USA), Zenit-2 (NPO Yuzhnoye, Ukraine) and Soyuz (Starsem, France) launch vehicles carrying eight, twelve and four satellites respectively. In February and April 1998 eight Globalstar satellites were placed in orbit using two Delta II launch vehicles. The Globalstar launch schedule is presented in Table 3.

Initially, Globalstar planned to build up to 210 gateways to minimise land-based long-distance charges and to respect national boundaries and thereby overcome potential problems with the national regulatory authorities. However, Globalstar recently announced that it will rely on only 50–75 gateways since some countries have decided to use gateways in neighbouring countries rather than go through the expense of building their own.

The total accumulation of capital is approximately $2,875 million including equity, bank financing and resale of gateways. Loral Space & Communications is the largest investor with approximately 33 % ownership. The initial Globalstar system cost is estimated to be $2,600 million.

Globalstar plans to launch a service for 'premium customers' with 44 satellites early 1999 [6] [7].

3.3 ICO Global Communications

ICO Global Communications [8] was established in 1995. The original proposal came from Inmarsat, which still is a large investor. The ICO owners comprises 60 investors from 51 countries, including Hughes, TRW[4], BT and T-Mobil. In addition, all Inmarsat members are indirect investors through Inmarsat's collective contribution to ICO.

The 10 ICO satellites are arranged in two circular MEO orbits 10,355 km above the earth surface. Each of the orbital planes are inclined 45° relative to the equatorial plane as Figure 8 illustrates. A satellite

[4] *TRW joined forces with ICO in December 1997 and terminated the Odyssey project due to lach of financing.*

completes one cycle around the earth in six hours. Each satellite will cover approximately 30 per cent of the earth's surface at a given time ensuring a significant coverage overlap.

MEO satellites are heavier than LEO satellites as a consequence of the higher power requirements. The 2,600 kg heavy satellites are manufactured by Hughes, a design based on the HS601 geostationary satellite bus. Communication through the 163 beam satellite is transparent. The life span of ICO satellites is expected to be approximately 10–12 years. The satellites will be launched by a variety of vehicles: Atlas IIAS (Lockheed Martin, USA), Delta III (McDonnell Douglas, USA), Proton (Khrunichev State Research and Production Space Centre, Russia) and Zenit/Sea Launch (NPO Yuzhnoye, Ukraine and a Boeing Consortium) will launch one, five, three and three satellites respectively. The first launch is scheduled for late 1998.

The 12 ICO gateways, called SANs (Satellite Access Nodes), are interconnected via the ICONET ground network. The SANs will be the primary interface between the satellites and the terrestrial networks.

ICO has awarded Hughes Space and Communications two contracts worth $2,200 million to build and launch the 12 satellites. In addition, an NEC led consortium is building the ground network in a contract initially worth $616 million. Including other minor contracts, ICO is investing more than $3,000 in their global satellite system. ICO has managed to raise approximately $1,500 million in equity from investors.

Current plans assume that the ICO system will be operational third quarter 2000.

3.4 Odyssey

Odyssey was proposed by TRW Inc. [9] with Teleglobe Canada as the only additional investor. In January 1997, Odyssey signed a service provider agreement with ChinaSat and on 11 March 1997, Daewoo Corp. of South Korea and Kumho Group announced that they had in principle agreed with TRW to join the project.

17 December 1997 TRW announced that it was acquiring an equity interest in ICO

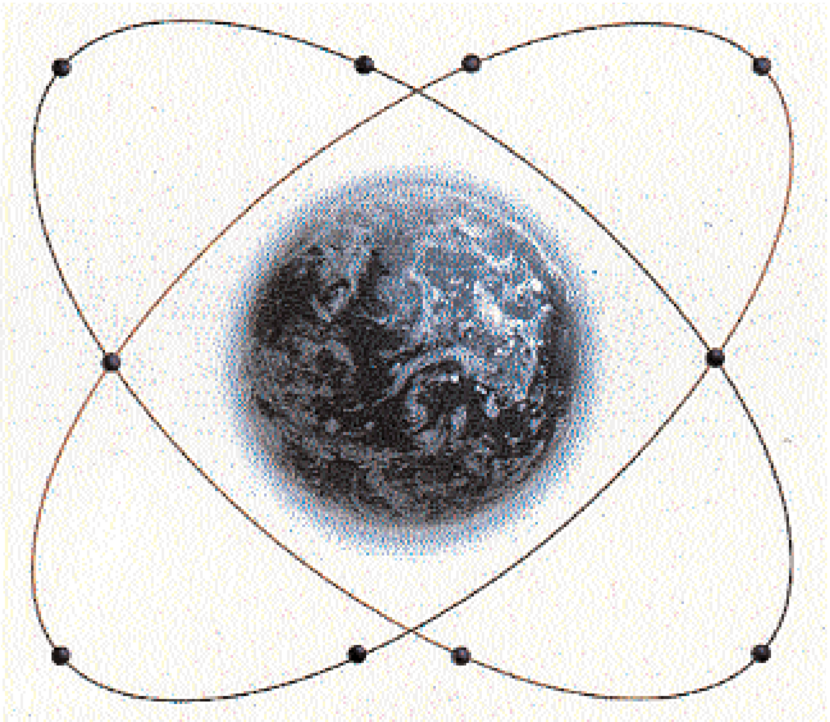


Figure 8 The ICO satellite configuration

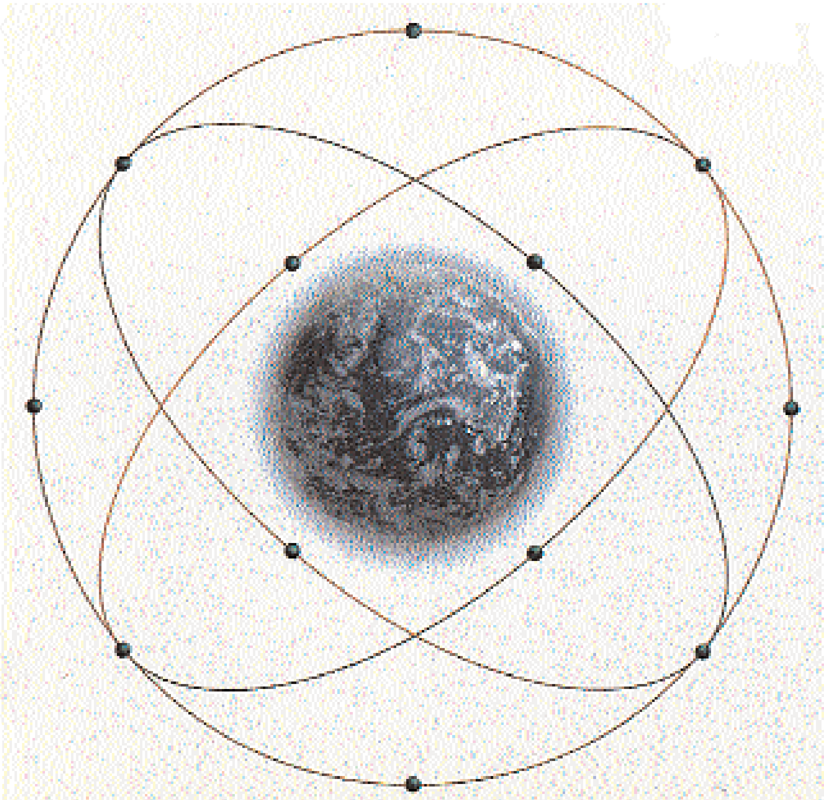


Figure 9 The Odyssey satellite configuration

and that the Odyssey project thereby was terminated. Odyssey as envisaged is described below.

The Odyssey satellite network was similar to ICO. 12 satellites circumferenced the earth in three circular MEO orbits at 10,354 km altitude. The orbits were inclined 50° relative to the equatorial plane. It took a satellite six hours to complete a full cycle. The planned constellation is illustrated in Figure 9.

The Odyssey satellite footprint constituted a 61 cell structure. The expected satellite lifetime was 15 years. Only seven gateways were used to connect with terrestrial networks.

The cost figure of $3,200 million included gateway costs. Odyssey had serious problems with the financing of the system with only $150 million committed to the project at the time of termination.

The operational target date for Odyssey was 2001.

Table 4 Summary of GMPCS systems characteristics

	Iridium	Globalstar	ICO	Odyssey
Orbit	LEO	LEO	MEO	MEO
Altitude [km]	780	1,414	10,355	10,354
Number of satellites	66	48	10	12
Number of planes	6	8	2	3
Inclination	86.4 °	52 °	45 °	50 °
Orbital period	100 minutes	113 minutes	6 hours	6 hours
Number of gateways	11	50–75	12	7
System cost [billion $]	3.4	2.6	3.0	3.2
Commercial operation	23 Sep 1998	Early 1999	3Q 2000	–

4 Recent developments

It is almost ten years since the first GMPCS systems were announced, and none of them are yet in service. During those ten years there has been a significant change of scenery, affecting those assumptions initially made by the founders of the GMPCS systems. These developments will inevitably affect the size and characteristics of the GMPCS market as well as the conditions for operating such services globally. We will discuss some of these developments, namely the rapid penetration of terrestrial cellular services, the introduction of regional satellite-based systems, the demand for broadband services and the regulatory barriers and possible measures.

4.1 The cellular explosion

AT&T projected early on that the total world market for cellular phones would be about 900,000. With more than 165 million cellphones in use world-wide today, we know that estimate was off by at least a factor of 180, and still growing. With such working assumptions, the outlooks for a global handheld satellite-based phone, nearly independent of local infrastructure, were of course different then than they are today. GSM and other cellular standards are present in more than 140 countries [10], and this number is rapidly increasing. Market forecasts[5] indicate that the global market for GMPCS services will reach 12 million subscribers in 2005. This represents 2–3 % of the expected global cellular market. However, large areas remain unserved by cellular and fixed infrastructure and a patchwork of cellular network standards around the world makes travelling with a cellular phone difficult. This development has caused the GMPCS operators to seek new business opportunities in 1997. Iridium announced the concept 'Iridium Cellular Roaming Service'. By the use of multi-mode (dual-/triple-mode) terminals, and translating differing cellular protocols, Iridium will act as an interworking system across incompatible cellular standards. In principle, a dual mode terminal supporting two cellular standards (without satellite mode) may be used, and Iridium, as a third party, will only provide the interworking and billing for all calls [11].

4.2 Regional satellite systems

Another category of systems expected to offer the GMPCS systems strong competition is the growing number of GEO-based regional systems. These systems are established with significantly lower investments than the GMPCS systems.

[5] *Performed for Telenor by Gemini C[4] Labs.*

One such system is the Asia Cellular Satellite System (ACeS). This system will provide voice, facsimile and paging services via GEO satellites to hand-held mobile and fixed terminals throughout Southeast Asia, India and China. Launch of the first ACeS satellite is planned for 1998, with service beginning a few months later.

4.3 The demand for more bandwidth

The GMPCS systems, like the cellular systems, provide data communications at low bitrates. There is an increasing demand in the fixed networks for more bandwidth, partly fuelled by the rapid growth of Internet services. Mobile users will also require higher bitrates for more efficient data communication and Internet access. The GSM operators are already planning the implementation of both circuit switched and packet switched higher rate enhancements to their GSM data services. HSCSD (High Speed Circuit Switched Data) will provide data rates up to 57.6 kb/s and GPRS (General Packet Radio Service) will offer data rates up to 115 kb/s to GSM users. Even before the service introduction of the GMPCS systems, one might claim that GMPCS, in certain market segments, will provide less satisfactory data service capabilities. In light of these developments, both Iridium and Globalstar, among others, have announced next generation systems targeting broadband and multimedia service distribution. This will be further discussed in chapter 5.

4.4 Regulatory restrictions, measures and deliberation initiatives

Regulatory restrictions are recognised as a key problem for global and regional satellite systems seeking to provide trans-border communications. These problems are often less predictable and may easily be underestimated when forecasting the global market potential. There may be several reasons for maintaining regulatory barriers:

- Concern about bypass of local networks, of prime importance for developing countries with poor infrastructures
- Security (especially a concern for countries in conflict areas and in countries with drug dealer problems)
- Lack of adequate legislation for the handling of a more liberalised environment
- Protection of national telecom operators and service providers
- Radio frequency interference considerations
- Lack of qualified staff and in handling licence requests etc.

The regulatory measures may take different forms, and among such measures are:

- Prohibitions on use of satellite telecom equipment
- High custom duties and licence fees. Custom duties exceeding 50–60 % of the equipment cost, and annual licence fees of several thousand US dollars are not uncommon, especially in some developing regions outside Europe and North America. This may apply to system and gateway operators as well as service providers
- Incomplete and badly co-ordinated procedures, creating bureaucratic delays in handling applications and commissioning of terminals
- Neglect of a country to respect type approvals of another country, even in cases where both countries have signed agreements of mutual recognition of type approvals
- Jamming of satellite systems not complying with national requirements. This is an extreme measure that has not so far been applied in practice, but which is talked about by representatives of certain African countries as a possible ultimate reaction to enforce national policies.

However, several initiatives have recently been made to progress the liberalisation of the regulatory regimes for satellite communications, and mobile satellite communications in particular. The introduction of the GMPCS systems in the near future has been one of the driving factors for this progress. On the global scale we should first of all take notice of the work done by the two United Nations (UN) organs, ITU and WTO (World Trade Organisation), in particular of the following initiatives:

GMPCS MoU

A Memorandum of Understanding (MoU) has been worked out by the ITU World Telecommunications Policy Forum in order to facilitate GMPCS. The document was opened for signing in March 1997.

The signatories to the MoU are administrations, GMPCS operators, service providers and manufacturers. Within their respective roles they agree to co-operate to stimulate free circulation of equipment, by addressing topics such as type approval, licensing and marketing of terminals, custom arrangements, and access to traffic data. Both Telenor and PT (Norwegian Post and Telecommunications Authority) have signed this ITU document.

Activities within the WTO

Ministers from 28 countries (including Norway), representing 84 % of the trade in information technology, issued a Declaration in November 1996 on Trade in Information Technology (ITA) that will come into effect once the subscription represents 90 % of the trade. A principal goal is to have the customs duties on equipment reduced to zero by year 2000.

The work of WTO's Group on Basic Telecommunications (GBT) to negotiate commitments to liberalise basic telecom markets was successfully concluded on 15 February 1997. The commitments relate to issues such as licensing, interconnection, transparency, independence of regulatory bodies, spectrum and universal service obligations. 69 countries made commitments by the February deadline. The formal entry into force of the commitments is scheduled to be 1 January 1998. But where an individual participant's commitment for a special service is to be phased in, the actual implementation will take place at the date specified in the schedule.

Additional activities to alter regulations in Europe

The EU (European Union) Commission and the European Radio Communications Committee (ERC), an organ within CEPT (Conférence Européenne des Postes et Télécommunications), have both taken steps to expedite the development towards a more liberalised telecom market in Europe, through a series of Commission Directives, Council resolutions, recommendations and action plans, addressing all the essential issues involved in removing exclusivity rights and opening the market for free competition.

5 Future Evolution of Personal Satellite Communications

As already mentioned, a major trend in communications is the demand for more bandwidth to cater for higher bitrate requirements for the provision of Internet and multimedia type services. The trend is well established in the fixed networks, the terrestrial cellular systems, like GSM, are already planning enhancements for higher bitrate data services, and the satellite community is of course also well awake.

The next generation mobile communication has to a large extent been synonymous to the standardisation work going on in ETSI (UMTS) and ITU (IMT-2000). Also in UMTS (Universal Mobile Telecommunications System), focus is put on the bandwidth trends, and in a recent report from the UMTS Forum [12], multimedia services are noted as a key driver for UMTS. UMTS will need to provide both narrow and wideband services. Part of the UMTS concept is a satellite component, providing satellite based services, both narrowband and wideband. The UMTS Forum states that these will have a place in the market in their own right, and in providing both early and temporary coverage before and as terrestrial networks are rolled out. Integration of the terrestrial and satellite components of UMTS is highly desirable to provide actual global coverage for at least a subset of UMTS services. Such commonalities are also expected to reduce the costs of the satellite terminals, and thereby make them more attractive.

The work within ETSI (European Telecommunications Standards Institute) of specifying a UMTS-specific satellite system to take the role as satellite component has not been successful and is not likely to happen. We regard it as more likely that proprietary systems, such as the GMPCS systems and other future systems under development, will be adopted as part of the UMTS 'family' in the long term. Among wide- and broadband satellite system concepts under development are Microsoft's Teledesic, Motorola's Celestri and Inmarsat's Horizons.

These initiatives represent new challenges in the history of satellite communications. Teledesic's original proposal included a network in the sky of no less than 840 LEO satellites providing Internet services to companies and end-users globally [13]. The concept has now been scaled down to 288 satellites. The Teledesic project is backed by initiator Microsoft along with the Boeing corporation, Matra Marconi Space, and since May 1998, also Motorola. After Motorola joined Teledesic the faith of Celestri [14], a hybrid LEO/GEO system (63 LEO satellites) for broadband services is announced. Inmarsat's Project Horizons plans a GEO system providing wideband services to portable terminal units the size of a notebook PC. These are only a few of several plans and concepts revealed. The use of higher frequency bands, such as the 20/30 GHz Ka-band is a common feature for many of these plans opening up for much more capacity than what is available in the currently used L- and S-bands.

A certain degree of commonality across services provided by the fixed networks, terrestrial cellular networks (current and enhanced versions), UMTS/IMT-2000 (International Mobile Telecommunications System) and the satellite systems is most desirable, especially from the end-user point of view. Commonality in user interfaces, consistent user profiles and simple numbering and addressing schemes are among the aspects expected to be strong future user requirements in the potential jungle of systems and terminal types. This fits well into the trend of Fixed Mobile Convergence (FMC) and should be kept in mind when developing and introducing new systems and services, both terrestrial- and satellite-based.

In light of this, personal communications takes on a broader meaning. Multimedia services may in nature not be well suited for pocket sized terminals and the aspect of Terminal Mobility isolated, well known from GSM, will not be the main focus. Personal Mobility and Service Mobility, the ability for a user to be independent of a specific terminal when accessing his/hers services, and the ability to access those personalised services independent of terminal type and serving network, will be important aspects. In developing the future satellite systems for multimedia services, these aspects must be taken into consideration to secure the position of satellite communications into the exciting next century of telecommunications.

Acknowledgements

The authors would like to thank Mr. Kristen Folkestad, Telenor Satellite Services, who has contributed to this article with his experience and knowledge on regulatory matters. Our appreciation also goes to Mr. Håkon Lundamo, formerly Telenor Satellite Services, and Mr. Scott Jacobs, Gemini Consulting, who worked with us on the GMPCS issue for nearly six months.

References

1 *Iridium Today,* 3 (4), July 1997.

2 Werner et al. Analysis of System Parameters for LEO/ICO-Satellite Communication Networks. *IEEE J. Selected Areas Commun.,* 13 (2), 371–381, 1995.

3 Ramesh, R. Availability Calculations for Mobile Satellite Communications Systems. *IEEE Proc. of Vehicular technology Conference,* April 1995.

4 Ananasso, F, Priscoli, F D. The Role of Satellite in Personal Communications Services. *IEEE Journal on Selected Areas in Communications,* 13 (2), 1995.

5 Iridium official Internet web-site: http://www.iridium.com.

6 Globalstar official Internet web-site: http://www.globalstar.com.

7 Kiernan, V. Globalstar at the starting gate. *Satellite Communications,* 36–39, July 1997.

8 ICO official Internet web-site: http://www.i-co.co.uk.

9 TRW official Internet web-site: http:// www.trw.com/seg/sats/ODY.html.

10 *Mobile Communications International*, 43, 67, 1997.

11 *Iridium Today,* 3 (3), 1997.

12 UMTS Forum. *A Regulatory Framework for UMTS.* 25 June 1997. (Report no. 1.)

13 Teledesic official Internet web-site: http://www.teledesic.com.

14 Celestri official Internet web-site: http://www.mot.com/GSS/SSTG/projects/celestri.

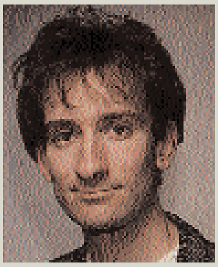
Abbreviations

ACeS	Asia Cellular Satellite System
CDMA	Code Division Multiple Access
CEPT	Conférence Européenne des Postes et Télécommunications
ERC	European Radio Communications Commission
ETSI	European Telecommunications Standards Institute
EU	European Union
FMC	Fixed Mobile Convergence
GBT	Group on Basic Telecom
GEO	Geostationary Orbit
GMPCS	Global Mobile Personal Communication Services
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communication
HSCSD	High Speed Circuit Switched Data
ICO	Intermediate Circular Orbit

IMT-2000	International Mobile Telecommunications System
Inmarsat	International Mobile Satellite Organization
ISDN	Integrated Services Digital Network
ISL	Inter Satellite Link
ITU	International Telecommunications Union
kbps	kilobit per second
LEO	Low Earth Orbit
LLC	Limited Liability Company
LP	Limited Partnership
MEO	Medium Earth Orbit
MoU	Memorandum of Understanding
MSC	Mobile Switching Centre
MSS	Mobile Satellite Systems
PABX	Private Automatic Branch Exchange
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
PT	Norwegian Post and Telecommunications Authority
SAN	Satellite Access Node
SMS	Short Message Service
S-PCS	Satellite Personal Communication Services
TDMA	Time Division Multiple Access
UMTS	Universal Mobile Telecommunications System
UN	United Nations
WARC	World Administrative Radio Conference
WTO	World Trade Organization

Arild Flystveit was, at the time of writing this article, Research Scientist at Telenor R&D, Kjeller, Personal Communications Group. He has been employed by Telenor since 1995, most of the time working with system- and network aspects of mobile satellite communication systems. Since March 1998 he has been employed by Telenor Mobil, Network Division, working with product development in general and implementation of Intelligen Networks in particular.

e-mail: arild.flystveit@oslo.mobil.telenor.no



Arvid Bertheau Johannessen was, at the time of writing this article, Research Scientist at Telenor R&D, Kjeller. He has been with Telenor since 1990, working with system- and network aspects of mobile communication systems in general and satellite based systems in particular. Since February 1998 he has been employed by Telenor Satellite Services AS, Mobile Division, with focus on development and operation of the Inmarsat services as well as future strategies for mobile satelite services within Telenor.

e-mail: arvid.johannessen@oslo.satellite.telenor.no

Cordless Terminal Mobility (CTM)

– Support of cordless mobility within the fixed network

IVAR OLDERVIK, KNUT BALTZERSEN, TORIL NATVIG AND GISLE PEDERSEN

This article presents the development of CTM within ETSI (European Telecommunications Standards Institute) and EURESCOM (European Institute for Research and Strategic Studies in Telecommunications), with a focus on the work done in EURESCOM Project 606 CTM (Cordless Terminal Mobility) Phase 2. Views and statements presented in this article are Telenor's only, and do not necessarily represent the views of other EURESCOM P606 partners. Partners in the EURESCOM P606 are British Telecom, CSELT, FINNET Group, France Telecom, Swiss Telecom PTT, Telefonica, Telenor and Telia.

1 Introduction

The mobile phone seems today to be considered by most customers as an addition to their fixed line subscription, and not yet as a substitute. A reason for this could be that users prefer the higher speech quality and quality of service of fixed telephony, but more important is probably that the cost of using a mobile phone is considerably higher than using a fixed line. One should, however, expect to see a rapid decrease in the price difference over the next few years.

Thus, in light of the rapid growth in the mobile communications market, one might argue that introducing CTM will help the fixed network operators to prevent a massive loss of customers to mobile operators. At the same time, CTM could be looked at as an important value-added service in the fixed network, representing a potential source of income for the fixed network operator. Another aspect that could prove to be an advantage for CTM is the higher capacity of DECT compared to GSM, for example. The increased capacity could be used to offer higher data rates and higher traffic capacity in certain areas. As the GSM operators are expecting an increase in the data traffic in the years to come, higher data rates could prove to be a powerful feature for CTM. Also, the development of packed data will be an important feature in order to support efficient solutions for Internet access, e-mail, file transfer, etc. Finally, with the development of CTM/GSM roaming, the end-users will gain a higher quality of service. Such a combination of technologies may represent a first step towards a third generation mobile system.

2 History and background

2.1 ETSI work on CTM

The mobile communication market has shown a great increase and has become a big success in later years. The Scandinavian countries have a mobile penetration around 40 %, and there are still no signs of saturation in the market. In order to be able to offer mobility even in the fixed network, the work on CTM has been carried out within both ETSI and EURESCOM since 1994. The concept of CTM is to offer IN supported terminal and service mobility, the cordless access part being based on DECT (Digital Enhanced Cordless Telecommunications) or CT2 (Cordless Telecommunications generation 2). ETSI have divided the work on CTM into a phased approach: CTM phase 1, CTM phase 2 and CTM phase 2+. The work is co-ordinated by the ETSI Project CTM (EP CTM), and involves several bodies within ETSI. NA1/2 is responsible for the service description, NA6 is responsible for architecture and functionality, the DECT project is responsible for the radio access part and SPS is responsible for the network protocols. The CTM phase 1 standard was finalised in 1997. CTM phase 2 will probably be finalised in 1998, and CTM phase 2+ is divided into different feature packages (eg. CTM Phase 2+ FP 1). So far, no estimate for finalisation of the FPs is given. A short description of the different phases will be contained in the following.

ETSI CTM Phase 1 will support the following core requirements [2]:

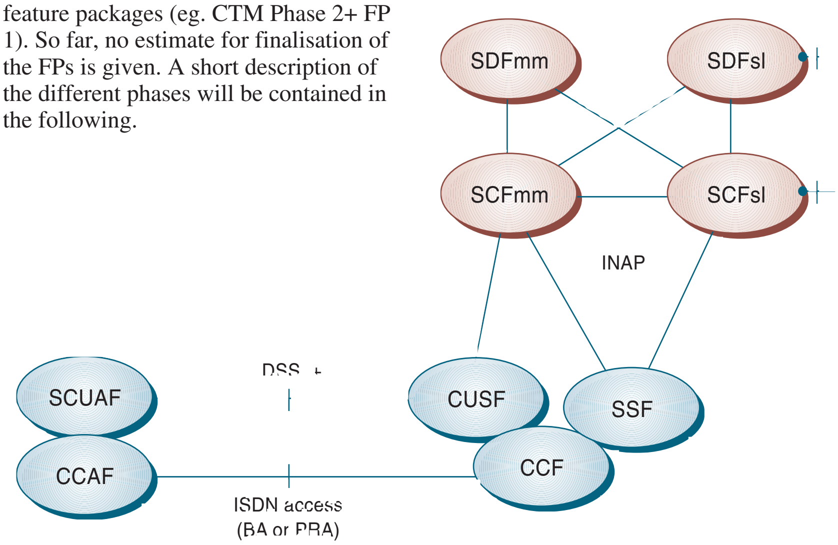
- outgoing calls
- incoming calls
- location handling
- authentication and ciphering
- emergency calls.

In addition, the optional requirements defined for CTM phase 1 are [2]:

- service profile modification
- service profile interrogation
- call forwarding not reachable
- subscription registration / deregistration
- location registration suggest
- network authentication.

CTM phase 1 is defined for the single public network case, for roaming between public IN based networks, and for roaming between private networks and public IN networks. For the single public network, the functional architecture shown in Figure 1 has been developed.

In phase 1, the radio access part will be supported by DECT GAP (Generic



mm = mobility management
sl = service logic
SDF = Service Data Function
SCF = Service Control Function
SSF = Service Switching Function
CCF = Call Control Function
CUSF = Call Unrelated Service Function
SCUAF = Service Control User Agent Function
CCAF = Call Control Agent Function

Figure 1 CTM phase 1 architecture

Access Profile), or alternatively by CT-2. The access will be based on ISDN access and DSS1+. The signalling within the IN network will be based on ETSI Core-INAP CS1 and CS2.

To further develop the concept of CTM, ETSI is currently also undertaking work on CTM Phase 2. This phase is adding some new aspects to the phase 1 definition of CTM, like several CTM supplementary services and the possibility to indicate for a CTM user his or her mailbox status (message waiting indication) [6]. Another important aspect of phase 2 is the work on external handover (see clause 3.5). The radio access part will in phase 2 be supported by DECT CAP (CTM Access Part).

In order to be able to introduce new features and services in a flexible manner, the ETSI CTM project has agreed that CTM Phase 2+ should be divided into feature packages (FPs). So far, the CTM project has initiated work on four new FPs:

- CTM Phase 2+ FP 1: Circuit Switched data Services; 32 and 64 kbps Unrestricted Digital Information
- CTM Phase 2+ FP 2: CTM Short Message Service (SMS)
- CTM Phase 2+ FP 3: CTM/GSM Roaming (feasibility study)
- CTM Phase 2+ FP 4: Point-to-Point Protocol (PPP) interworking for Internet access and general multi-protocol datagram transport (suggested).

New FPs could be added in the further development of CTM.

2.2 EURESCOM work on CTM

Eurescom P606 CTM Phase 2 work was initiated in June 96, and was completed in January 98. The main focus of the work has been to develop the CTM feature further, contributing to the ETSI standardisation work on CTM with focus on ETSI CTM Phase 2+. P606 has also undertaken work on the migration towards UMTS (Universal Mobile Telecommunications System) by studying what in EURESCOM has been called CTM pre-UMTS.

In EURESCOM, earlier projects have also had a focus on personal mobility / terminal mobility within IN. The first EURESCOM project addressing mobility aspects within IN was P230 ‘Enabling pan-European services by co-operation between PNO’s IN platforms’. P230 mainly focused on interworking between IN platforms, but also aspects of DECT used as radio access and IN mobility control were treated. To continue the work on mobility in IN, EURESCOM P507 ‘Mobility Applications Integration in IN’ was initiated, focusing on the support of mobile services on IN architectures, considering the CTM service as a basis for short-term scenarios [1]. The EURESCOM P606 was a follow up of earlier projects, but with a stronger focus on developing features and services for CTM. The pre-UMTS part of P606 was seeking an evolution path for CTM towards the UMTS, contributing to UMTS work in ETSI, among others.

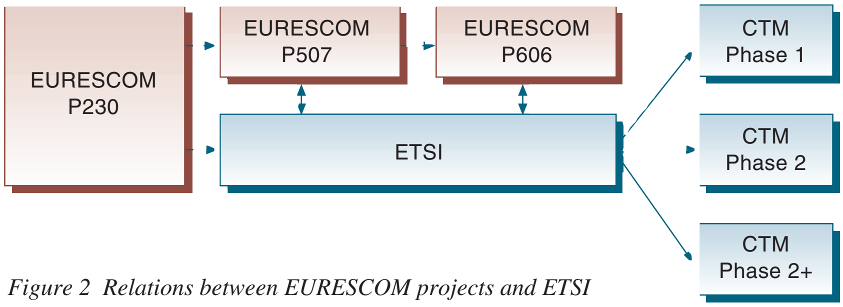


Figure 2 Relations between EURESCOM projects and ETSI

3 EURESCOM P606

EURESCOM P606 CTM Phase 2 was divided into three different tasks, more or less following the three stage method:

- Task 1 focused on service features and detailed service description both for CTM and for CTM pre-UMTS
- Task 2 focused on functional description, architecture and information flows for CTM and CTM pre-UMTS
- Task 3 focused on the additions necessary on the protocol level to support the new services / features developed for CTM and CTM pre-UMTS.

Work done in P606 has also been used as a basis for contributions towards ETSI CTM standardisation, with focus on ETSI CTM Phase 2 and ETSI CTM Phase 2+. It should be mentioned that the phasing of CTM within EURESCOM was not equal to the phasing of CTM within ETSI. Main work areas in P606 have dealt with features like CTM/GSM roaming, SMS, handover, on-air subscription, data-services, Virtual Home Environment, etc. EURESCOM P606 was also looking into the development within standardisation towards UMTS.

This chapter will describe some of the work done in EURESCOM P606.

3.1 64 kbps Unrestricted Digital Information (UDI)

An important set of applications for CTM data services in the consumer market will probably be several forms of Internet access (eg. www browsers and ftp). For the business user, teleworking may represent a group of important applications. File transfer (eg. ftp) and e-mail should be supported, but www browsers would also be interesting to these users.

These applications need to be supported in an optimised manner. One possible way to meet these demands is the 64 kbps UDI. The 64 kbps UDI is a fixed bandwidth symmetric ISDN bearer service functioning more or less as a ‘bit pipe’, which makes it very flexible. It is already being utilised in fixed networks for Internet access today.

The major advantage to the users is that prices may be comparable with standard analogue services, while transmission speed will be well over that of standard modems. The set-up time will also be much shorter than for modems, which take some 20 seconds to do their negotiation process.

3.1.1 DECT/ISDN Interworking Profiles

Within the DECT community there is already considerable progress made in developing standards for ISDN over DECT. Two different DECT/ISDN interworking profiles described below might help cater for the user needs mentioned above:

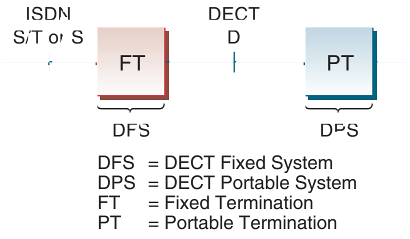


Figure 3 ISDN Access Profile

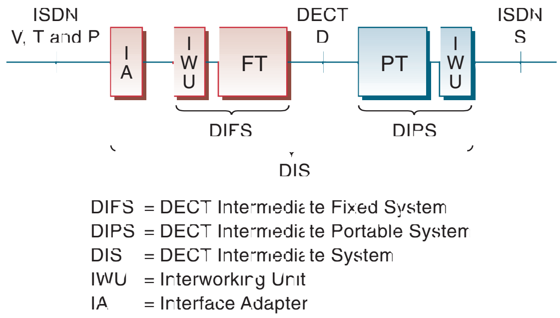


Figure 4 Intermediate ISDN Access Profile

In April 1996 the first edition of 'DECT/ ISDN Interworking for end system configuration', ETS 300 434 (also called ISDN Access Profile – IAP), was published.

The IAP applies when FP (Fixed Part) and PP (Portable Part) together constitute an ISDN terminal. The ISDN applications are located in the PP. There is no IWU in the PP. The FP maps the received layer 3 messages at the ISDN interface to DECT layer 3 messages and vice versa.

The bearer services being covered are [ETS 300 434-1]:

- circuit mode speech
- circuit mode 3.1 kHz audio
- circuit-mode 64 kbps unrestricted digital information.

A work program has been started within ETSI Project DECT on 'DECT/ISDN Interworking for Intermediate System Configuration' DE/RES 03039 (also called Intermediate ISDN access Profile – IIP) [DE/RES-03039].

The IIP applies when FP and PP together constitute a gateway between an ISDN network and an ISDN terminal. The FP and the PP have an IWU each, that maps the messages between the ISDN interface and the DECT air interface. This implies that an ISDN terminal can be connected to a DECT PP supporting IIP via the ISDN S interface.

The bearer services being supported are the same as for the IAP[1].

The profile best supporting the CTM user needs is the IAP, since the IIP is mostly tailor made for Radio in the Local Loop (RLL) solutions. Nevertheless, the IIP can be useful in supporting a standard ISDN terminal connected to a DECT PP supporting ISDN as described above.

3.1.2 Impact on existing standards

One important aspect of this service is that the implementation is only foreseen to impact on the access network. This means that only the DECT FP and PP need to be enhanced to support 64 kbps UDI. In particular, there is a need to support 'double slot' (two 32 kbps time slots concatenated) to be able to reach the demanded data rate of 64 kbps. This will probably mean that a hardware upgrade of all FPs in a voice only DECT network is necessary. It is also important to notice that the DECT/ISDN interworking profile adds extra error correction and ARQ mechanisms to the 64 kbps bit stream, to the cost of a need for new buffers.

The FP will also have to be able to interwork with ISDN networks.

With regard to standardisation, the upgrade can probably be made by incorporating parts of the DECT/ISDN profile(s) mentioned earlier with the CAP (CTM Access Profile). The only special demand 64 kbps UDI will have on mobility aspects is for handover. For voice connections bit errors during handover might be acceptable to the user. This might not be the case for a data connection as bit errors will lower effective throughput and in the worst case even lead to loss of connection. This is highly dependent on the type of application. For example, Internet applications will benefit from strong error recovery procedures in the TCP/IP protocols.

Another issue is the type of user equipment and user interface. Most of today's data terminals expect the user to be stationary, which limits the need for support of handover.

3.2 Short Message Service

3.2.1 SMS description

The Short Message Service (SMS) investigated in P606 has been modelled on the basis of the GSM Short Message Service, preferably to be perceived by the user as virtually equivalent to the SMS service in GSM.

This implies that we may define three different SMS services, as in GSM:

1. Mobile (portable part) originated point-to-point SMS
2. Mobile (portable part) terminated point-to-point SMS
3. Cell broadcast SMS.

The point-to-point services enable CTM users to send and receive short messages to and from other CTM users or in principle any other suitable kind of parties (eg. GSM users). The cell broadcast short message service enables the CTM operator to broadcast unaddressed and unenciphered messages at regular intervals to all CTM users located within a given geographical area. The cell broadcast SMS was not investigated by P606.

3.2.2 A possible implementation method for the point-to-point SMS

In EURESCOM P606, CTM SMS is described as a call unrelated service, making use of the SCUAF (Service Control User Agent Function) and CUSF (Call-unrelated User Service Function) as defined in IN CS-2.

When transmitting a point-to-point short message, the message will be relayed by a short message service centre (SM-SC). Unlike GSM, the information flows between this service centre and the rest of the CTM network are considered to be within the scope of the CTM specifications, ie. making use of a standardised protocol.

As in GSM, the CTM network should contain mechanisms to ensure that a message is stored in the short message centre until the message has been successfully delivered to the CTM user (or is timed

---

[1] *In addition, packet-mode (X-31 case B) D-channel (packet data) and packet-mode (X-31 case B) B-channel (packet data) may be supported.*

out), as well as to report back to the originating CTM user when the message has been received at the SM-SC.

A key driver in the method described below is to implement SMS with a minimum of impact on existing protocols. Hence, the most suitable method for interworking against the SM-SC is considered to be the one described in GSM 03.47, basing the interworking upon ITU-T Signalling System No. 7.

3.2.3 Implementation of SMS in CTM network

Figure 5 shows the architecture for CTM SMS, where the user is located in his or her home network, or in a visited network. The interface between the IWF and the SMSC is based on GSM 03.47 description of SMS-MAP. The interface between the SCFsl and the IWF is based on INAP. Communication across network boundaries is supported by a SCF-SCF relationship.

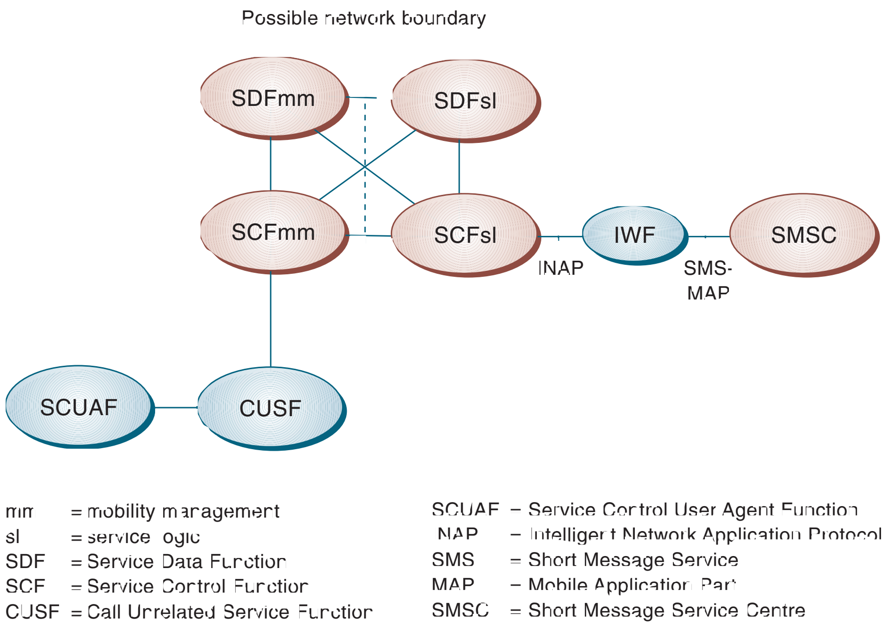


Figure 5 Architecture for SMS in a CTM network

3.3 CTM/GSM roaming

CTM and GSM are technologies with different characteristics which could be combined to provide a more powerful solution. CTM focuses on being a fixed network solution for the support of cordless mobility, which gives the user good speech quality, advanced services and high capacity. However, CTM could experience difficulties providing efficient coverage in rural areas, as well as continuing coverage in suburban areas. On the other hand, GSM is attractive because of its wide area coverage, both in Europe and in other parts of the world, as well as supporting a radio interface adapted to mobile environments.

In a scenario consisting of both CTM and GSM in combination, the two systems would behave complementary to each other and the scenario would appear very attractive in the evolutionary path towards UMTS utilising the best of two blessings.

The co-operation between GSM operators and CTM operators represents a powerful combination as it is possible to combine the two systems in the future. A possible way of integration is to combine the IN logic and database in the CTM network with the IN logic and database in the GSM network, thus deploying a common IN platform for the support of CTM/GSM. In EURESCOM P606, several scenarios have been developed, looking at solutions based on GSM networks without any IN capabilities as well as GSM networks with IN capabilities.

3.4 Virtual Home Environment (VHE)

3.4.1 Background

Within telecommunications in general there is a shift in focus from standardising *services* to standardising *service capabilities*. This means that instead of having standards describing a service in detail, core features are being standardised. These features can be put together to form a wide spectrum of different services, thus allowing for product differentiation between competing service providers.

Service capabilities are also drivers for new services within mobile networks such as CTM and GSM. The question is what to do when a user roams out of his or her home network. To facilitate roaming and still keep your 'home services', the concept of virtual home environment (VHE) has been introduced.

3.4.2 VHE Implementation scenarios

To be able to support VHE the visited network has to, as a minimum, support certain triggers. These triggers will for instance be set off when there is a request for a change to the CTM user's service profile. Then, service control can be carried out in principle in one of two ways:

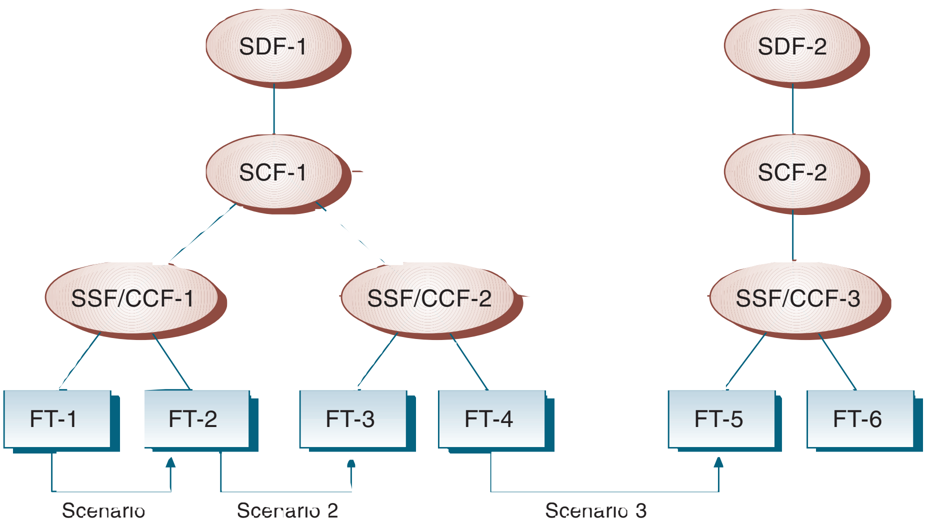
1 Control of the service is released to the home network

2 Relevant data are sent from the home network to the visited network so that control of the service can be carried out from the visited network.

There are also a number of subsets of these two scenarios such that not only control of the service is released to the home network, but also the connection itself is being redirected via the home network.

In general, the choice of implementation comes back to which network should be in control; the home network or the visited network.

Several aspects such as security and signalling load have to be investigated before a solution can be recommended.



CCF = Call Control Function
SDF = Service Data Function
SCF = Service Control Function
SSF = Service Switching Function
FT = Fixed Termination

Figure 6 Handover scenarios. The fixed termination (FT) may represent a single radio base station directly connected to the fixed network, or a cluster of base stations connected via a cluster controller to the fixed network

3.5 Handover

3.5.1 Definitions

Handover is defined as the process of switching a call in progress from one physical channel to another. This process is initiated when there is a degradation of the quality in the original channel caused by for example movement of the portable or some incidental radio disturbance.

Handover may be either internal or external. Internal handover processes are internal to one fixed termination only, whereas in the external case there is a handover between two different fixed terminations. This means that external differs from internal handover in that external handover occurs between two independent systems, where each system has its own lower layers of protocols and an independent set of network layer service access points. For CTM the aim is to support some forms of external handover by providing a common management entity above the respective logical entities in the two fixed terminations.

3.5.2 Why implement external handover?

With the DECT systems on the market today it is possible to provide continuous outdoor radio coverage with internal handover in a geographical area of about 1–2 km^2 in urban environments. However, if either indoor radio coverage or more traffic capacity is to be added, the area with internal handover must be reduced correspondingly. Even though future DECT systems may be capable of providing internal handover in larger geographical areas, external handover obviously still represents a solution which is a lot more desirable from the point of view of the CTM service provider as well as the network planner.

3.5.3 External handover scenarios

Figure 6 shows three different scenarios for external handover, with increasing degree of complexity;

- Scenario 1 – Intra-switch handover
- Scenario 2 – Inter-switch handover with the SSF/CCF in the respective switches connected to the same SCF/SDF
- Scenario 3 – Inter-switch handover with the SSF/CCF in the respective switches connected to the different SCF/SDF.

All of these scenarios have been investigated in the CTM work of EURESCOM.

4 CTM evolution towards UMTS

UMTS (Universal Mobile Telecommunications System) is the name of the ETSI initiative within SMG (Special Mobile Group) to migrate today's second generation mobile system (GSM) towards third generation mobile communication. The existing work program suggests commercial operation in 2002.

There is at the present no exact definition of what UMTS (Universal Mobile Telecommunications System) really is. However, the basic idea behind UMTS is to have a personal telecommunications scenario where the user can access a given set of services anywhere and at any time. The technology behind UMTS should be irrelevant to the user, and could be a mix of, among others, PSTN/ISDN, GSM, CTM, satellite, and so on. Such a scenario requires advanced interworking procedures between different networks, utilising the best from each different technology, to achieve the desired integration of independent services of today.

CTM will have a natural role within the migration toward UMTS, and together with cellular and satellite systems support multimedia applications any time, anywhere with a focus of attention on supporting business and residential users' needs for high bitrate mobile communication. To be able to support such services in an efficient manner, packet data communication will be important. This means that in the evolution towards UMTS, data services in CTM will become increasingly important involving the strong data communications possibilities within the DECT specifications.

Today, the maximum data rate supported by DECT is 552 kbit/s. There is work going on within the DECT community on refining the radio interface so that data rates up to 2 Mbit/s can be supported. This will also be highly interesting in a UMTS environment. Other important aspects of CTM in the migration towards UMTS is higher speech quality and a higher capacity, for example compared to today's GSM networks. This does not imply that CTM and GSM need to be competing technologies. The co-operation between GSM operators and CTM operators represents a powerful combination, because it is possible to deploy a common IN platform for the support of both CTM and GSM. The scenario could be enhanced further by incorporating a new dedicated UMTS radio interface in addition to CTM and GSM, or even by including a satellite component. Co-operation between CTM operators and satellite operators would help the mixed scenario to provide an even better coverage and together with GSM operators, offer communication any time anywhere.

5 Summary and concluding remarks

This article has focused on the technical aspects of CTM, referring to work within ETSI and EURESCOM. It has also been pointed out that CTM enables the fixed network operators to introduce services that are highly attractive, and that CTM might contribute to wiping out the line between mobile and fixed network communications as seen from the end user (Fixed Mobile Convergence – FMC). A CTM-like service has recently been launched in Italy, and the next 1–2 years will bring some highly interesting feedback on the market prospects for CTM.

6 References

1 Lauritzen, G O. Mobility applications integration in Intelligent Networks. *Telektronikk,* 91 (4), 182–184, 1995.

2 ETSI. *IN Architecture and Functionality for the support of CTM.* NA 061302-1 version 7.0.0.

3 ETSI. *IN Architecture and Functionality for the support of CTM; Part 2: CTM IN interworking aspect.* NA 061302-2 version 1.1.10.

4 ETSI. *IN Architecture and Functionality for the support of CTM; Part 3: CTM Interworking between private networks and public Intelligent Networks.* NA 061302-3 version 3.0.

5 ETSI. *Cordless Terminal Mobility (CTM)-Phase 1; Service Description.* DE/NA-010039.

6 ETSI. *Cordless Terminal Mobility (CTM)-Phase 2; Service Description.* DE/NA-010061.

7 *EURESCOM P606 Deliverable 1.*

8 *DRAFT EURESCOM P606 Deliverable 2.*

Ivar Oldervik worked as Research Scientist at Telenor R&D until October 1997 within areas like Intelligent Networks, GSM Phase 2+/CAMEL, as well as Cordless Terminal Mobility. He is now employed by Telenor Mobil working mainly within the area of Intelligent Networks.

e-mail: ivar.oldervik@oslo.mobil.telenor.no

Knut Baltzersen holds an M.Sc in physics and works as Chief Engineer at Telenor R&D, Kjeller. He is currently involved in work on personal communications, with a special focus on value added services.

e-mail: knut.baltzersen@fou.telenor.no

Toril Natvig is Research Scientist at Telenor R&D, Kjeller, working in the field of mobile communication with special interests in network functionality for land mobile systems.

e-mail: toril.natvig@fou.telenor.no

Gisle Pedersen was, at the time of writing this article, working at Telenor R&D, Kjeller, in the field of mobile and personal communications. He is currently employed by Telenor Mobil , where he is in charge of the GSM network development group.

e-mail: gisle.pedersen@oslo.mobil.telenor.no

DECT as a part of personal communication

JOAR LØVSLETTEN

Based on available equipment and standards which are finalized or under preparation, the potential of DECT (Digital Enhanced Cordless Telecommunications) as a part of personal communication / third generation is examined. Non-technical factors influencing the future for DECT are also addressed. This article mainly addresses DECT as an access technology while the necessary support from the network behind is not stressed.

Personal communication scenario

The term personal communication is ambiguous. In an evaluation of how DECT can be a part of personal communication a common understanding of the term is necessary. To gain this common perception a scenario is described and also some consequences of introducing personal communication are included. Two main trends are foreseen as indicated in Figure 1:

- Change from plain old telephony to personal pocket terminals
- Development of pocket terminals ranging from only speech to multimedia capability.

In the Nordic countries with their high penetration of cellulars the use of cellular phones for low bitrate services instead of wired phones is emerging. Also results from a field trial performed by Telenor using DECT to provide local mobility in a small town indicate a willingness to replace the fixed phone with personal terminals. The trial customers stated clearly, however, that small pocket phones had to be personal. If a cordless terminal should replace one fixed phone in a household it gave little added value as the terminal had to be kept stationary. In this trial we also experienced that users of mobile services always want an extended area of coverage.

Both existing and new operators having cellular licences will compete with the fixed network services in a liberalised telecommunication market. Their services will to a large extent be based on the use of pocket terminals and move users from wired phones to personal terminals.

The introduction of pocket phones led to a change in telecommunication behaviour. As long as only fixed phones are available the attitude is that as soon as you get access to a phone, ie. when you get to your desk or when you are back home, you have to remember to make that or those calls. Sometimes you have to hurry to find a phone to make the call at an appropriate time. To deal with incoming calls when you are not at your desk or at home, answering machines, switchboard operators or call forwarding can help.

By carrying a pocket phone you can make calls whenever you want almost wherever you are, but the circumstances have to be taken into account. The same goes for incoming calls. In countries with high penetration of cellulars it can be observed that people use the pocket phone in many different environments. If there is a need to make a call they do it there and then. There is a discussion going about which situations or circumstances use of the phone disturbs or annoys the surroundings and therefore should not be accepted.

For more and more people one personal pocket phone will be the basic telecommunication terminal and use of wired phones will gradually decrease.

Figure 1 Personal terminals replacing plain old telephones

Multimedia pocket terminals replacing personal phones

There is a general trend toward services being offered in the fixed network also being wanted in mobile communications. Also the quality of the service should be comparable.

The emerging pattern of using cellulars whenever and wherever needed can contribute to the creation of a market demand for advanced terminals capable of handling data, video and so on. Making and receiving calls and having access to message services wherever you are will not be sufficient. Users would like to have access to the same services using their pocket terminal as they have using a fixed terminal, for example, being able to access Internet from a pocket terminal will be a feature required by users. A development towards cellular handsets being advanced PDA-like terminals with personal information stored indicates that personal pocket terminals and thereby terminal mobility, will be an essential part of personal communication.

As more and more PC functionality is introduced in terminals the stored information and user profiles will make the terminals an important part of everyday life both at work and for private use. As a consequence of this the same terminals may be used both in job situations and outside working hours. Also the tendency to work from different locations and towards less differentiation between working and non-working hours enhance the need to use only one terminal.

When advanced terminals capable of handling multimedia-like services are available it can be assumed that user patterns seen today among advanced cellular users will be even more apparent. If there is a need to have some kind of information, to do a transaction, order some goods, watch a sports event or remotely control some device, you do it there and then.

Given that personal communication based on pocket terminals becomes the common telecommunication subscription, new markets will emerge. People on the move with their pocket terminals will often be in need of peripheral equipment like printers, large screens and keyboards. From such stand-alone equipment which can be available at convenient points for travellers, access to the home server can be obtained using

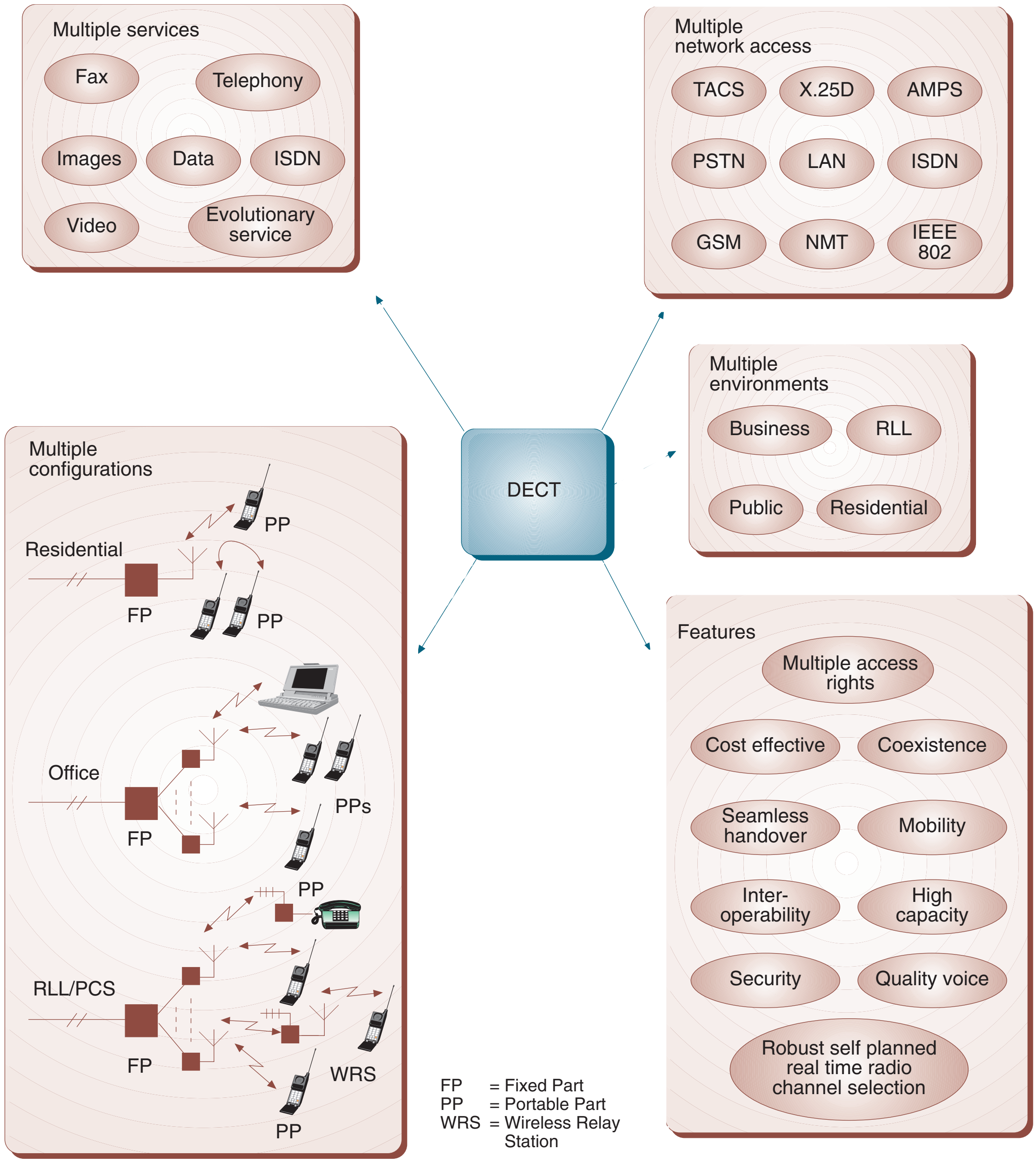


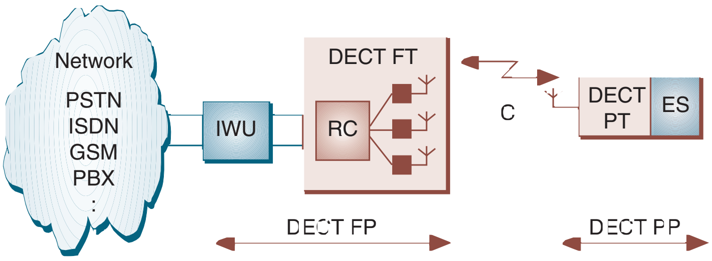
Figure 2 Overview of DECT applications and features (from [1])

pocket terminals. The connection can be obtained by use of a standard interface between the pocket terminal and the equipment in use, eg. some kind of docking station. Paper prints can be taken, documents or other files can be prepared or video watched. The invoice for the use of peripheral equipment and the connection will be related to the pocket terminal, and only one invoice is received.

Demand for capacity

A differentiated demand for capacity dependent on location can be expected. A span in demanded bitrates is anticipated to vary from relatively low bit rates up to 2 Mbit/s.

Users are expected to use their pocket terminals for narrowband services like speech wherever they are. But in some areas more than others, users will be in a position to use their pocket terminals for more advanced services. This will mainly be indoors or where some special activities are taking place outdoors, eg. markets, outdoor restaurants, recreation areas, sports arenas.



CI	:	Common Interface
IWU	:	Interworking Unit
RC	:	Radio Controller or CCFP (Central Control Fixed Part)
PT	:	Portable Termination
FT	:	Fixed Termination
FP	:	Fixed Part
PP	:	Portable Part
ES	:	End System

Figure 3 A general DECT-system

Generally, people moving from one place to another cannot be regarded as users of advanced services, amongst others due to reduced bit rate when on the move.

DECT applications and functionality

DECT is a flexible radio access technology that can support a wide range of services and applications to different kinds of networks as indicated in Figure 2. The DECT standard is based upon TDMA/TDD and includes a mandatory real time Dynamic Channel Allocation which makes frequency planning unnecessary. It also secures that different uncoordinated DECT systems can coexist in the designated DECT frequency band with an efficient use of the frequency spectrum.

A DECT handset can access different networks, eg. public and private, and base stations can be shared between different operators. One part of a public network can host a private user group, or a private system can provide public access for visitors. The same base station can also handle different services, eg. speech and data.

The DECT standard is prepared for the future by the provision of escape codes and a multitude of reserved codes and messages in every layer of the standard so that evolutionary applications can be specified.

A generic diagram of a DECT system is shown in Figure 3. The main standard, DECT CI (Common Interface), in principle covers the air interface between the Fixed radio Termination (FT) and the Portable Part (PP). The CI can be regarded as a tool-box from which selections are made when services and applications shall be specified. DECT is transparent to services provided by the connected network and offers in itself only cordless capability and mobility. The Interworking Unit (IWU) which is network specific and the End System (ES), which may be a microphone, speaker, keyboard or display, are not parts of the DECT CI standard but are subject to specific attachment requirements for the chosen network to secure end-to-end compatibility.

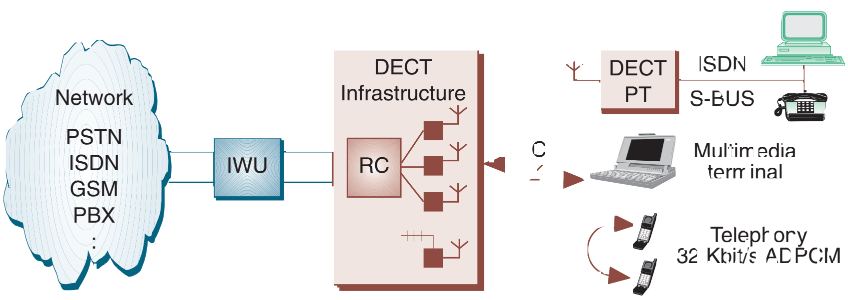


Figure 4 DECT services and applications

A DECT profile is a subset of the DECT CI that gives an unambiguous description of the air interface for specified services and applications. In Figure 4 different services and applications based on DECT profiles are indicated.

The profiles include all requirements for interoperability of equipment from different manufacturers.

The DECT profiles are:

- *Generic Access Profile (GAP)*
  The basic DECT profile for any application supporting 3.1 kHz telephony.
- *Data Profiles*
  A family of profiles exploit the lower layer data services of DECT, specifically oriented towards LAN, multimedia and serial data capability. Each member of the family is optimised for a different kind of user services. Widely varying bandwidths are available, and DECT can support net data throughput of n x 24 kbit/s up to a maximum of 552 kbit/s.
- *DECT/GSM Interworking Profile (GIP)*
  Describes how DECT FP is connected via an IWU to the GSM network which will then see a DECT user as a GSM subscriber.
- *DECT/ISDN Interworking Profiles*
  Two profiles are described:

  The End System (ES) allows a suitable PP to access the ISDN network using DECT signalling. The FP and the PP together appear to the ISDN network as an ISDN terminal.

  The Intermediate System (IS) provides for a wireless link between an ISDN network and ISDN terminals connected to an S interface. The ISDN terminals have transparent access to all network defined services based upon the basic channel structure 2B+D.

- *CTM Access Profile (CAP)*
  The profile will secure interworking with CTM (Cordless Terminal Mobility) networks. CTM will make wide mobility in DECT networks possible.
- *RLL Access Profile (RAP)*
  The profile is divided into two parts:

  Part 1, 'Basic Services', includes telephony services, a 64 kbit/s PCM bearer service and over-the-air OA&M services.

  Part 2, 'Advanced Services', specifies ISDN services (2B+D and 30B+D) and a data port for broadband packet data services up to 552 kbit/s.

Also Wireless Relay Stations (WRS) are standardised for DECT. These can be used to provide extended coverage when use of a base station is inappropriate.

In the ETSI Project DECT there is work going on

- to standardise advanced dual mode terminals DECT/GSM where both modes can be active at the same time
- to standardise implementation of GPRS (General Packet Radio Service) over DECT
- to standardise 2 Mbit/s over DECT.

DECT's potential as a part of personal communication / third generation mobile

Deployment of DECT network

It is assumed that dual mode terminals, DECT/cellular, having one number and being capable of handling advanced services and applications are used. Offering services over such a terminal is a good example of FMC – fixed mobile convergence.

DECT access is provided as a public service in areas with high densities of users wanting to use their pocket terminals for more advanced services than speech. This will be mainly indoors, where people stay for some time. DECT networks may be a mixture of private networks providing access for the public and public networks. A part of the public networks can be virtually private, eg. inside the premises of a company.

As DECT is not designed for velocity coverage, passengers in cars or on public transport may have a problem, even though trials have proved that good speech quality can be obtained in cars exceeding 80 km/h.

In rural areas DECT access can be provided indoors by private DECT systems. There will be a multitude of DECT islands and when moving between these islands the cellular mode of the pocket terminal has to be used. In public DECT islands any DECT terminal can be used. The standard opens for private systems to appear as public systems to visiting users. Thus, the DECT mode of the terminal can in principle be used wherever DECT coverage is provided.

Services

When DECT provides speech 32 kbit/s ADPCM is used and the quality is near the quality of wired phones. A standalone DECT system provides seamless mobility in an area covered by base stations connected to one radio controller. With the implementation of CTM in the fixed network handover between base stations belonging to different radio controllers is also obtained. If a DECT system is connected to a cellular network the mobility management is handled by the cellular network.

One terminal can have access rights to several different kinds of DECT networks and roam between them. The profile standards will ensure interoperability between equipment from different manufacturers. For example, if a terminal is capable of handling voice and a specific data profile used for Internet access, GAP will ensure that the terminal can be used towards any DECT system with voice capability. The data profile secures access to any system with the specific data profile implemented, independent of manufacturers.

ISDN solutions for DECT are on the market. Data profile standards for 552 kbit/s over DECT exist and there is work in progress to standardise 2 Mbit/s. Hence, DECT has a potential to handle multimedia-like services.

However, there will be limitations in the functionality of pocket terminals based on any technology. Terminals should be suited to being carried, eg. in a shirt pocket, hence the size has to be limited. User interfaces like screen and keyboard will suffer. Even if solutions like speech recognition may help to overcome such problems, one can expect the user friendliness of stationary terminals to be better. Especially the need for a larger screen than can be implemented in the pocket terminals will be apparent when graphics or video are presented.

Capacity

A main problem when offering services over radio is the limited bandwidth due to limitation in available frequency resources. If use of a personal terminal capable of handling multimedia-like services is becoming the basic telecommunication utilisation for each individual there will be a bandwidth problem using existing cellular technologies.

DECT may help to solve a capacity problem. Due to short range and the dynamic channel selection feature DECT is capable of handling advanced services in areas with high user densities.

An uncertainty regarding DECT and capacity is that a multitude of DECT systems, both private and public, supporting different kinds of services can coexist, eg. in the same building. This can make the planning of a DECT system regarding traffic capacity somewhat tricky.

Factors influencing DECT's success as part of third generation

DECT products on the market

The DECT CI standard was finalised in 1992, but DECT has still not 'taken off'. One reason may be that it took some time before the DECT profile standards were in place. It seems that DECT residential solutions are popular and DECT business applications has an increasing market share. Products for these applications are made by many manufacturers, and small DECT systems handling ISDN are now available for residential and business applications.

Lately, there has been an increasing interest in DECT as RLL solutions. World-wide there is a huge demand for subscriber lines, and DECT as a standardised access technology is an interesting choice for the access network. Countries with poor telecommunication infrastructure is the main market for DECT RLL, and mainly fixed telephony is provided.

Only a few manufacturers are offering equipment for public DECT networks including mobility. To provide wide mobility in DECT networks a support from the network behind is required. For PSTN/ISDN this means a CTM solution. A standard for CTM is expected to be ready in the near future.

Time gap for DECT

DECT CTM solutions and the GSM/DCS1800 might be competing technologies. DECT has the greatest potential for third generation like services and applications, thus an advantage over GSM/DCS1800. However, in 2002 UMTS products are expected to be available, and one important question is if this time gap is sufficient for building DECT CTM networks to introduce advanced services. Also, one must bear in mind that the GSM technology including DCS1800 will be enhanced to handle more advanced services in the same period, eg. GPRS (General Packet Radio Services). The bitrate of the GPRS technology will be approximately 115 kbit/s which is well below the capability of DECT.

If DECT is going to be part of third generation solutions advanced products have to achieve a certain penetration before competing, advanced GSM and UMTS products are available. The time limit for advanced DECT products to penetrate the market can be discussed. However, there is no doubt that roll-out has to happen soon if DECT shall be a part of third generation.

Also, there is a time gap for substantially lower pricing of services based on DECT compared to GSM, as the cost of using mobile services is steadily falling. However, advanced GSM and UMTS equipment and services can be expected to have a relatively high price in an early phase.

DECT and operators

There has been an interest among European operators for DECT CTM solutions, and Telecom Italia has built DECT networks in 28 cities mainly providing outdoor coverage. The networks are based on proprietary CTM solutions. The service was opened on 1 January 1998, and the experiences made by Telecom Italia can have an important impact on DECT's future as a part of public networks.

Operators that started public DECT trials 3 – 4 years ago were quite enthusiastic, but time has passed without flexible solutions for public DECT becoming available. Hence, the interest has faded among many of them.

It seems that several operators who previously had an interest in public DECT are going for other technologies like DCS1800, but some still go for DECT.

Manufacturers and DECT

Manufacturers making DECT equipment also make GSM/DCS1800, a technology which is a commercial success.

There is still a demand from operators and the corporate market for public and large business systems of handling advanced multimedia-like services, ie. utilizing the full potential of DECT. If this demand is sufficient to convince the manufacturers there is an increased potential for profit when advanced DECT equipment is included in the product portfolio. There is therefore reason to believe that DECT will be part of a third generation solution.

A general impression is that manufacturers have been reluctant to implement functionality beyond speech in their DECT products.

It seems unlikely that manufacturers being big in the GSM industry will promote or be drivers for advanced DECT solutions at their own risk. They find the safest commercial way ahead towards 3rd generation to be enhanced GSM technology and UMTS.

There will be many DECT products for the residential and business market, but the most interesting application seems to be DECT RLL. When operators implementing DECT RLL ask for more advanced services the market can be sufficient for manufacturers to come up with advanced solutions.

Conclusions

Technically, DECT has a potential to become a part of third generation. Especially dual mode DECT/GSM solutions capable of handling multimedia-like services would have qualities suited for being part of a personal communication platform.

However, it seems that other factors than technical ones will influence the future of advanced DECT networks. There is a two-step time gap for advanced DECT, first before enhanced GSM solutions become available and then before UMTS enters the market. DECT needs a certain penetration before this happens. For the time being it seems that the process of implementing advanced functionality in DECT products is rather slow. If DECT is going to be a part of personal communication, the manufacture of advanced DECT products and implementation of these into DECT systems and networks have to happen in the near future.

Reference

1 ETSI. *ETR 178 First edition: Radio Equipment and Systems (RES); Digital European Cordless Telecommunications (DECT); A high level guide to the DECT standardisation.* ETSI, Sophia Antipolis, October 1995.

Joar Løvsletten has been with Telenor since graduating from the Norwegian University of Science and Technology in 1976. He worked in the Network Department, dealing with radio relay systems, until 1990, when he started a group workng on radio access systems. Løvsletten came to Telenor R&D in 1995 and is now Senior Engineer working with personal communications.

e-mail: joar.lovsletten@fou.telenor.no

The DECT radio link

ANDREAS F. MOLISCH, HEINZ NOVAK AND ERNST BONEK

1 Introduction and historical perspective

Mobile radio has given the European industry its most important success in the last decades. Crucial to this success were the standards developed by ETSI (European Telecommunications Standards Institute) that have been adopted not only by European countries, but by many others as well. The best-known example is GSM (Global System for Mobile communications), the most widespread system for cellular communications. It has replaced a multitude of analogue standards throughout Europe, and seems to become the de-facto world-wide standard for cellular communications. The DECT (Digital Enhanced Cordless Telecommunications) standard [1] promises to become a similar success story. It was originally designed in 1991 in order to replace the several analogue standards for cordless telephones that existed in Europe at that time. The first commercial products went on sale in 1994, when about half a million terminals were sold. In 1997, about 8 million terminals were in use, and this number is expected to increase to 30 million by the year 2000.

Since the early 1990s, the market for DECT equipment has increased both geographically and in scope. Geographically, many extra-European countries have accepted the DECT standard (sometimes with slight modifications); this is also reflected in the change of the interpretation of DECT: while the 'E' in DECT was originally translated as *European,* this has been changed to *Enhanced,* to do away with any geographical restrictions. Another important change has occurred in the scope of cordless system. In the analogue days, 'cordless' meant just that the microphone and earpiece were attached to the 'base station' (i.e. the usual phone), not by wire but by a radio link. With the DECT system, many new applications became possible: one base station can serve several mobile terminals, and it is possible to combine several base stations to a 'cellular' system, where the mobile terminal can connect to various base stations during a talk and make handovers. This allows companies to build up a wireless private branch-exchange system (PBX), for example. When we take this feature to its extreme and combine it with capabilities of wired intelligent networks, it also becomes possible to build a low-mobility cellular system based on DECT. Field trials for such systems have been done in Norway [2] and other countries. Finally, DECT is intended to be used in wireless-local-loop (WLL) applications. There, the DECT link replaces the high-installation-cost connection from the subscriber to a central office (CO), or, e.g., to a village centre connected by ordinary cable to the CO.

The new applications of DECT pose new challenges both to the network and to the radio link. The networking aspects have been treated e.g. in [3] and another contribution in this special issue [4]. The present paper will concentrate on the radio link. The quality of the transmission can be limited either by noise, by the time dispersion of the mobile radio channel, or by co- and adjacent channel interference. Various methods have been proposed to mitigate those effects, and we will discuss them below. Section 2 gives the specifications of the air interface as prescribed by the DECT standard. Section 3 describes the performance in noise and co-channel interference. The effects of the time dispersion are analyzed in Section 4. The next two sections discuss diversity and equalizers, two means to overcome the basic limitations on the performance. Section 7 compares DECT to the rival system PHS from Japan. The next two sections deal with practical implementation issues: Section 8 describes a testbed for assessment of the performance of DECT terminals in various environments, and Section 9 compares the popular heterodyne receiver to more advanced direct-conversion receivers. A summary wraps up this paper.

2 The DECT air interface specifications

DECT, like GSM, uses a combination of TDMA and FDMA for multiple access. The 20 MHz band that has been allocated to DECT (1880 – 1900 MHz) is divided into 10 frequency bands; the carrier frequencies are actually separated by 1.728 MHz. On each carrier, 12 users can be accommodated in a TDMA scheme: the time axis is divided into frames which last 10 ms each, and are divided into 24 timeslots. Each user is assigned two timeslots that are separated by 5 ms: the first timeslot is used for the downlink, the second slot for the uplink. This implies that time division duplexing is used for separating uplink and downlink (in contrast to GSM, where frequency division duplexing is applied). We will see later that this has important consequences for diversity, and also for smart-antenna applications. The timeslots are separated from each other by guard times, which are necessary, since the runtimes from the different terminals to the base station may be different. The 52 µs guard time prescribed in DECT can theoretically deal with a 15 km distance between base station and mobile terminal – this is far more than actually needed; even when directional antennas are used, the 'cell radius' is usually less than 3 km. The rest of the guard time is used to give

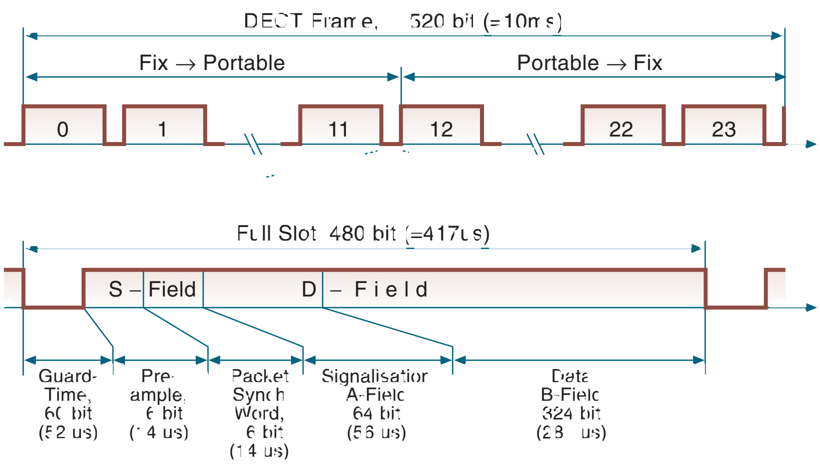


Figure 1 DECT frame and DECT slot structures

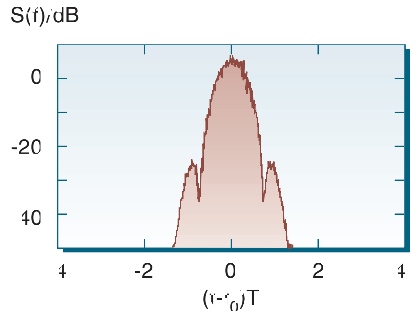


Figure 2 Spectrum of a DECT signal

the transmitter amplifiers time to power up – the faster these amplifiers have to be, the more expensive they become.

Each timeslot begins with a 16 bit preamble and a 16 bit packet synchronization word (the 32 bit S-field). This is followed by the D-field. This D-field contains 64 bits of signalling (A-field) and 324 bits of data. The A-field includes a Cyclic Redundancy Check (CRC) for the detection of packet reception failures; it allows transmission of signalling information with a net rate of 4.8 kbit/s. Finally, the B-field contains 324 bits of data, enabling the transmission of 32 kbit/s ADPCM coded speech in one slot. The overall transmission data rate is therefore 1.152 Mbit/s. This frame/slot structure suggests that DECT terminals can be designed to be far less complex, and thus less expensive, than GSM. DECT does *not* use a complicated speech coder, which explains the fact that it requires 32 kbit/s instead of the 6.5 or 13 kbit/s prescribed by GSM. It also contains no channel coding for the speech data, so that the bit error rate (BER) for tolerable transmission quality has the rather low value of 10^{-3} (as compared to 10^{-1} for GSM). By combining two or several timeslots, DECT can provide much higher bitrates, up to 552 kbit/s, and carry asymmetric traffic, making DECT attractive for data transmission. The modulation format is GFSK (Gaussian Frequency Shift Keying) with nominal modulation index $h_{mod} = 0.5$, which is equivalent to Gaussian Minimum Shift Keying. However, the tolerances in the modulation index are quite large.

The time-bandwidth product of the Gaussian filter B_GT is 0.5, where B_G is the bandwidth of the Gaussian filter, and T is the symbol period. This provides a compromise between emitted spectrum and intersymbol interference (see Figure 2). The resulting RF-channel bandwidth is about 1.56 MHz (for a 99 % power criterion). With the specified channel spacing of 1.728 MHz, adjacent channel emission up to -40 dBc is allowed. Table 1 shows a summary of important DECT parameters.

3 Performance in noise and co-channel interference

The structure of the receiver is not specified in the standard. Three realizations have been at the centre of attention: coherent reception, differential phase detection, and limiter-discriminator detection. Coherent reception is difficult to implement, because the tolerances for the transmitted signal as foreseen in the standard are very large, so that carrier recovery and tracking is difficult. In differential phase detection, we compute the *difference* of the phase between two subsequent samples that are separated by one bit length. It is thus not necessary to recover the carrier phase, because any offset (that remains unchanged during one bit duration) will be eliminated when we take the difference of two phases. Finally, limiter-discriminator detection uses the fact that MSK can be seen as a kind of frequency-shift keying: we just detect the instantaneous frequency with an FM/AM converter and a threshold detector.

Extensive investigations have been done to find out the performance of MSK in a flat Rayleigh-fading channel with noise. Adachi and Parsons [5] found that for differential phase detection and for frequency-discriminator detection, the error probability is about 0.7/SNR (where the signal-to-noise ratio SNR is defined as bit energy over noise power density) when the receiver filter bandwidth is chosen in an optimum way. For differential phase detection, the optimum filter bandwidth is about $B_rT = 0.75$, for limiter-discriminator detection, $B_rT = 0.55$.

Another problem is co-channel interference (CCI). For a system like DECT, cell planning in advance of operation is neither feasible nor desirable. The capacity would be exploited very inefficiently and a complete redesign of the cell structure would be necessary every time someone installs a new base station. So-called dynamic channel allocation avoids these problems. The mobile part listens on the time-slots it does not use whether there is a channel with low CCI available, and moves the user there if there is a need to do so. Also, carrier frequency changes are possible. Of course, such a scheme cannot eliminate all CCI, but only minimize it. It is common to model the CCI (and also adjacent channel interference ACI) as Gaussian (see e.g. [6]), so that it affects the BER in the same way that noise does; the BER is about 0.7/CIR, where CIR is the carrier-to-interference ratio. For CCI, the CIR is not influenced by the receiver filter, while the ACI *is* influenced.

4 Performance in time-dispersive channels

Mobile radio channels are not only time-variant, i.e. fading, but also time-dis-

Table 1 Important parameters of the DECT standard

Transmission format	FDM/TDD/TDMA
Frequency band	1880 – 1900 MHz
Centre frequency for channel N (0..9)	1 897 344 – N * 1728 kHz
Number of frequency channels	10
Number of users per frequency channel	12
Channel spacing	1.728 MHz
Frame length (24 slots)	10 ms
Data rate	1.152 Mbit/s
Modulation method	GFSK, $B_GT = 0.5$
Tolerated bit error rate	< 10^{-3}

persive [7]. When we send a delta pulse from the transmitter, we will receive a series of delta pulses with different delay times, attenuation factors, and phase shifts. Two quantities characterize this time dispersion in a rough approximation: the maximum excess delay, i.e. the difference between the delays of the first and the last echo, and the rms delay spread, i.e. the square root of the second central moment of the square of the absolute value of the impulse response. At the receiver, the different echoes are 'smeared' by filtering and the integration procedures inherent in the receiver. If now the rms delay spread is *very much* smaller than the bit length, then the receiver does not 'notice' the time dispersion of the channel, and we can describe the channel as 'flat fading', the way we did in the previous section. If this condition is not fulfilled, however, then the time dispersion leads to intersymbol interference, and thus to errors that cannot be eliminated by increasing transmitter power. These errors are thus called 'irreducible errors', although this name must not be misinterpreted – they can actually be reduced by equalizers or diversity. The BER associated with these errors is called 'error floor'.

The time dispersion of the channel also makes the synchronization much more difficult. In FDMA systems, the usual way to synchronize a receiver is to perform an eye-pattern analysis, and put the sampling time at the time instant where the eye has its widest opening. Changes in the optimum sampling time are monitored and adjusted by a control loop with a rather low bandwidth (on the order of the Doppler frequency, say, 100 Hz or less). However, this type of synchronization is both difficult and inefficient in TDMA systems with preambles. The design of the S-field in DECT rather suggests a different approach, which works in three steps:

(i) The 'power up' ramp for the relevant slot tells the receiver very roughly (within several bits) that a new timeslot is about to start.

(ii) The sampling time is determined from the preamble (i.e. the first 16 bits of the S-field). The preamble consists of alternating +1/-1 information bits. This means that the received phase difference has the shapes depicted in Figure 3. The sampling time is then put into the middle between two zero-crossings of the received phase difference.

(iii) The word (or packet) synchronization is done with the help of the packet synchronization word, i.e. bits 17–32 of the S-field. Due to the group delay in a time-dispersive channel, the received sequence can be shifted effectively by one bit forward or backward (as compared to a flat-fading channel). By comparing the (known) packet synchronization word with the received sequence, we can detect such a shift, and adjust the receiver timing appropriately.

Actually, the synchronization procedure described above is not optimum. Improvements have been proposed by us [8]; the additional gain in BER is up to a factor of 4. It is important to note that this burst-by-burst synchronization (also called 'adaptive sampling' or 'training-sequence-based determination of the sampling time') differs fundamentally from the pure eye-pattern analysis mentioned above. Eye-pattern analysis is a type of 'blind' synchronization, i.e. can be used when we have no known sequence in our frame structure. It is thus more general, but gives worse results. For DECT, however, the knowledge of the S-field should be used to improve performance.

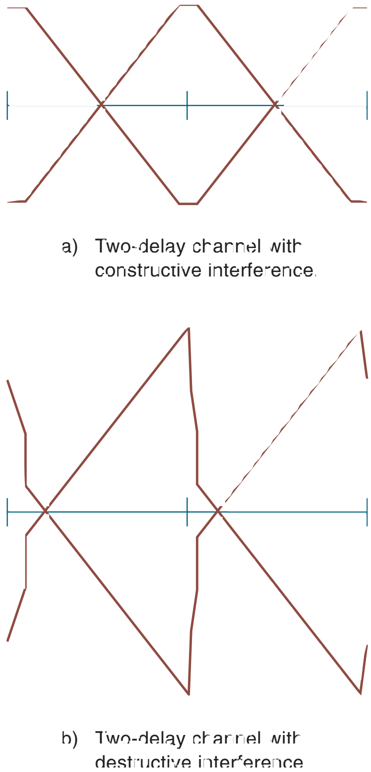


Figure 3 'Eye pattern' (decision variable as function of the sampling time) of the preamble in channels with weak and strong group delay distortions

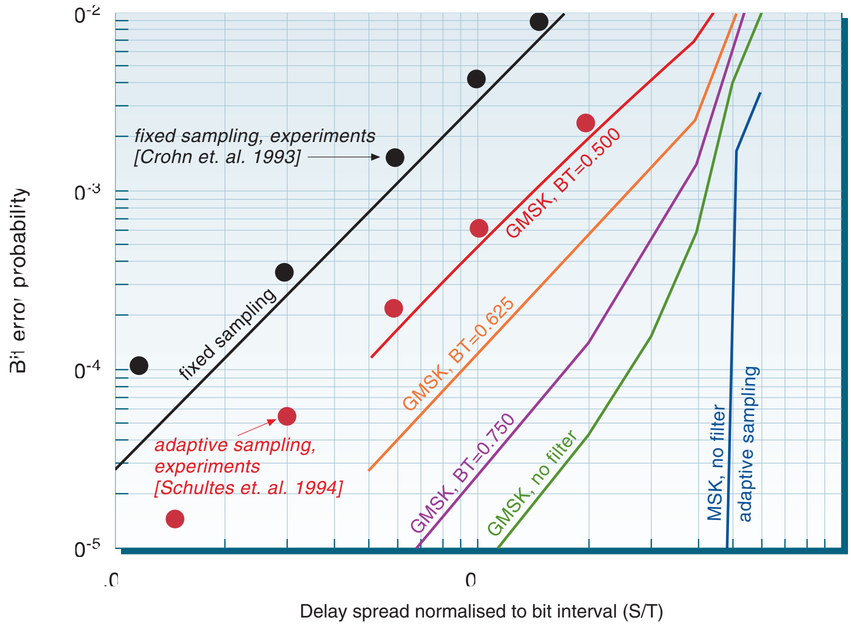


Figure 4 Error floor as a function of the rms delay spread for unfiltered and filtered MSK

For GMSK without any receiver filtering, the error floor obtained with (optimum) adaptive sampling is $K \cdot (S/T)^2$, where K is about 10^{-3}. Any receiver filtering increases K: for $B_rT = 0.625$, K is about 0.05. Figure 4 shows the error floor for various filter widths. This is also an important design parameter for DECT systems: we have seen above that a small receiver bandwidth decreases the effects of noise and adjacent channel interference, while it has little influence in the co-channel interference performance. Now we see that it *increases* the effects of the channel time dispersion. For various environments, different receiver bandwidths will be optimum. This suggests the implementation of various filters in the receiver, with the possibility to (automatically) choose the optimum. Of course, such a strategy is only feasible if the filters are implemented digitally.

In a first approximation, errors due to noise, CCI and time dispersion are independent and their associated BERs can be simply added to give the total BER. This is not strictly true, because the group delay distortions (i.e. what causes ISI-induced errors) are strongest in the fading dips, where also the noise has its strongest effect. However, these are already rather fine points that need not be considered in a first system design.

5 Diversity

DECT demands a BER of 10^{-3} without channel coding in order to give acceptable speech quality. As we have seen, the noise-induced BER is about 1/SNR in a Rayleigh-fading environment, so that the average SNIR must be better than 30 dB in order to fulfil this requirement *on the average*. Furthermore, the indoor environment is often stationary: when a user is sitting in a fading dip (i.e. a point of low received signal strength), he will not move out of it for the duration of the whole conversation. These problems can be greatly reduced by the application of diversity. Usually, DECT systems have two antennas at the base station, and we select the signal of one of them or combine them. Furthermore, we can also have two antennas at the mobile terminal. In the following, we will first consider the uplink, with one antenna at the mobile and two antennas at the base station. Another configuration will be treated later.

The simplest approach to using the signals from antenna branches A and B is the so-called switching diversity. When the transmission starts, we always use branch A. If the transmission quality achieved with that branch is not acceptable (i.e. either the received signal strength is too low, or the CRC check in the A-field tells us that we have a too large BER), then we use antenna B for the reception of the next timeslot. If the reception quality achieved with that antenna is also insufficient, we switch back to branch A, and so on. The advantage of this approach is that we need only one additional antenna and one switch compared to the no-diversity case. This is offset, however, by two serious drawbacks: first, we do not use consistently the ‘better’ signal: if, for example, both antennas give insufficient transmission quality, then we might be switching back and forth between a bad and an even worse signal. This also makes the definition of the threshold (what is ‘sufficient transmission quality’?) difficult. If the threshold is too high, then we are constantly switching back and forth; if it is too low, then we accept signals that sound rather bad to the user. The second drawback of the switching diversity is associated with the inherent time delay. The switching becomes effective only 10 ms after the measurement. In the meantime, the user may have moved so much that the signal from the branch that was formerly bad is now good, and vice versa.

A more involved approach is selection diversity. In this case, we monitor the quality of the signals from both diversity branches, either the received signal strength, or the BER, and choose the branch that gives the better results. This avoids all problems with defining a threshold, and it also becomes possible to do ‘instantaneous switching’. The drawback of this method is that we need more hardware. For RSSI-driven diversity (received-signal-strength-indicator), we additionally need (compared to the switched-diversity case) two sensors for the instantaneous power. For BER-driven diversity, we need additionally a complete receiver (we have to receive and demodulate the signals from both branches). Finally, we could also *combine* the signals from the two branches (not just select from them). However, this approach gives little improvement over the selection diversity.

An additional requirement for the diversity to work properly is that the signals at the two antenna branches are uncorrelated. If the signals are impinging upon the base station from all directions, this is not a big problem: we just require that the antennas are spaced a fraction of a wavelength, e.g. 4 cm – this can easily be done in a base station. If, however, the directions-of-arrival (DOAs) of the incoming waves are all concentrated within a small range, then the necessary distance between the antennas increases tremendously. Such concentrated DOAs usually occur only if the base station is situated above the rooftops – a situation that does not occur in the usual cordless applications, but which might become relevant for WLL links.

Up to now, we have only considered the uplink, but of course we need good speech quality also on the downlink. If we have only one antenna on the mobile terminal, then we must use the base station antennas in such a way that the transmission quality for the downlink is enhanced. This is facilitated by the fact that up- and downlink are on the same frequency. In a stationary environment, we can therefore say that if it is better to use the signal from antenna A in the uplink, it is also better to transmit from antenna A in the downlink (reciprocity theorem). Note that this is not true in GSM, because it uses frequency division duplexing, so that the channels for uplink and downlink are in principle always different. Problems for the branch selection in DECT can only occur if the channel changes strongly within 5 ms, the time between the uplink and downlink transmissions. Such a quick change in the channel can for instance occur in the case of fast-moving scatterers (cars passing by) – the mobile terminal is assumed to be slow-moving anyway. An alternative would be to install two antennas also at the mobile terminal. As mentioned above, the antennas have to be spaced only a few centimetres apart, and even if the distance is smaller, inherent pattern diversity usually leads to good decorrelation of the received signals.

Using diversity considerably improves the performance in noisy or interference-limited environments. The required SNR (for 10^{-3} BER) is decreased by some 10 dB, depending on the exact diversity scheme, if two branches are used. Adding a third branch, however, decreases the required SNR only by an additional 3 dB. This is why only two diversity branches are used in most equipment. Diversity also decreases the effects of time dispersion, see Figure 5 (note that these are theoretical error probabilities under 'ideal' circumstances; measured BERs in actual systems are somewhat higher). However, the tolerable delay spread is still limited; even in the best implementations, it must remain below about 300–400 ns. For environments with larger time dispersion, alternative methods are required, namely equalizers, which will be treated in the next section.

6 Equalizers

The DECT specifications do not foresee an equalizer, because its implementation was deemed to be too expensive at the time the specifications were written. However, for some new applications, delay spreads of 500 ns or larger occur, which cannot be dealt with by 'normal' receivers even when diversity is applied. Thus, an equalizer is required. Since no equalizer was foreseen in the standard, one might think that a 'blind' equalization (adjusting the equalizer coefficients without knowledge of either the channel or the data) is required. Fortunately, however, we can use the simpler and more efficient 'training-sequence-based' equalization by employing the (known) S-field as a training sequence. There is even a subsequence in the S-field (bits 17–27) that has very good autocorrelation properties (the sidelobes have only height 2). This sequence can thus be used for measuring the impulse response of the channel, and for the training of the equalizer.

Essentially, two types of equalizer have been proposed for DECT: Viterbi equalizers [10] [11] [12], and DFEs (decision feedback equalizers) [13]. In a Viterbi equalizer, we first measure the impulse response of the channel. We then determine for each possible input sequence what the output from this channel would be. When the data are received, we analyze what data sequence has been most probably sent. One important parameter for such an equalizer is the so-called 'constraint length', i.e. the anticipated length of the impulse response of the channel. If we anticipate long impulse responses, then the equalizer becomes very complex. On the other hand, if the anticipated length is shorter than the actual length of the impulse response, then the results will become

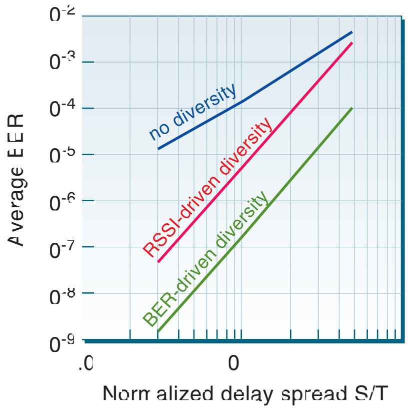


Figure 5 Reduction of error floor by diversity (from Ref. [9])

inaccurate. Up to now, only very short constraint lengths have been proposed (two-state Viterbi equalizers), which were shown by simulations to handle delay spreads up to about 500 ns.

In a DFE, we have essentially a (non-linear) filter, i.e. a tapped delay line with a feedback section. The filter coefficients are adjusted in such a way that during the training phase, the mean squared error between (known) training sequence and the actual received sequence is minimized. Also in this equalizer, we have a trade-off between complexity and performance. One implementation proposed in [13] is a DFE with 3 taps in the feed forward and 2 taps in the feed backward section; its performance (simulated) is shown in Figure 6.

Equalizers have a twofold beneficial effect in time-dispersive channels. The equalizer does away with the error floor, indeed. On the other hand, they also reduce the noise-induced errors in such environments. Some time dispersion in the channel will actually *lower* the BER as compared to the non-dispersive case. This is due to the fact that the time dispersion provides inherent time diversity, which can be used to combat fading. Echoes that arrive at the receiver at different times suffer from different fading:

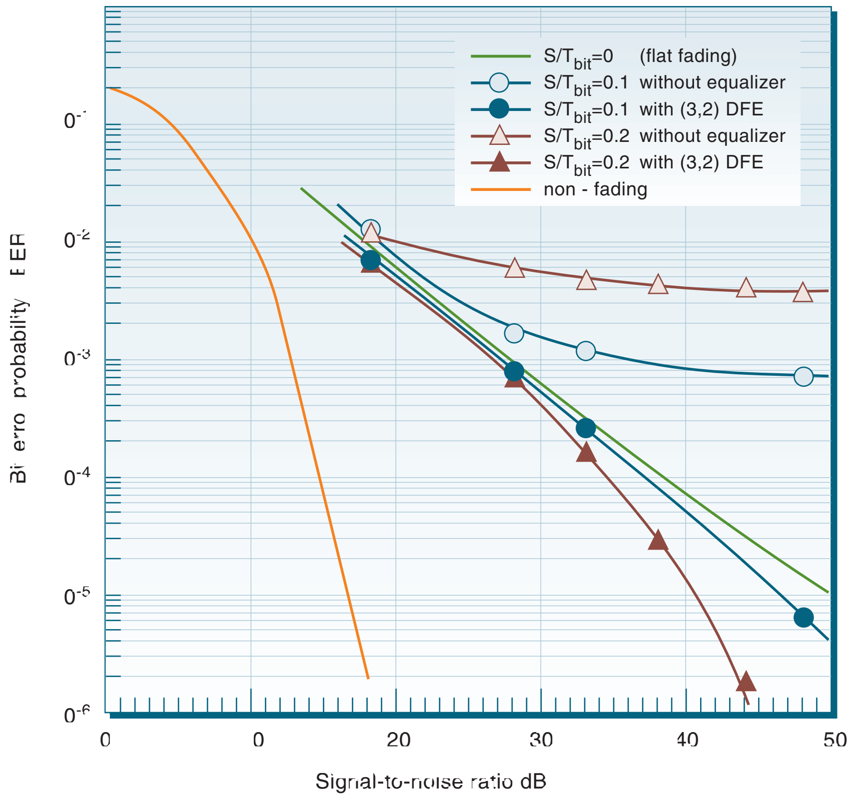


Figure 6 BER as a function of the SNR for a (3,2) DFE for various delay spreads. From [7]

it is very improbable that the receiver is in a fading dip with respect to all the different paths from transmitter to receiver. Thus, the high error probabilities in the fading dips are eliminated. This is also reflected in Figure 6. As the time dispersion increases, the performance with the equalizer becomes more similar to the performance of 'non-fading'.

7 The competition: PHS

While the DECT has found wide acceptance in many countries, Japan has devised its own rival standard, which also has achieved considerable success. This standard is called PHS (Personal Handyphone System), and is mostly identical to DECT in its system design. It is also an FDMA/TDMA/TDD system, and uses 32 kbit/s ADPCM without error correction. However, there are some important differences in the radio link: the modulation format is π/4-DQPSK, so that two bits are sent with each symbol. The pulse shapes used are pulses with a square-root raised-cosine spectrum with $\alpha = 0.5$. There are only eight timeslots (four users) per carrier frequency. The radio channels are only 300 kHz wide, the bit rate is 384 kbit/s.

The change in the modulation format also leads to important changes in the performance of the radio link. The sensitivity of π/4-DQPSK to noise in a Rayleigh-fading channel is about the same as for GMSK for the same E_b/N_o. For co-channel interference, π/4-DQPSK performs worse by about 3 dB. The *bit* error rate due to time dispersion of the mobile radio channel is about 0.5 $(S/T_{symbol})^2$. With the data rates used in PHS, this implies that a delay spread of 230 ns is admissible. This is quite similar to the values that occur for the DECT system; there, we had between 120 ns (measured in a suboptimum system) and 250 ns (computed for an optimum system). At first glance, this similarity is astonishing: PHS uses a symbol duration that is larger by a factor 6 than in DECT, so that we would expect a much smaller sensitivity to time dispersion. However, it can be shown that π/4-DQPSK has an inherently larger susceptibility to time dispersion errors, because adaptive sampling does not reduce ISI-errors as strongly as it does in GMSK [14].

The major advantage of PHS is the large spectral efficiency. While DECT can put 120 users into 20 MHz, PHS has 300 users in 23 MHz. This is caused mainly by the fact that PHS uses a four-level modulation format, which has inherently higher spectral efficiency. On the other hand, the higher sensitivity to co-channel interference requires a larger separation between base stations; this reduces the capacity (compared to DECT) by a factor that lies between 1.4 and 2, so that from a pure 'radio link' point of view, the capacity of a DECT system should be very similar to that of a PHS system. Furthermore, DECT has some capacity advantages due to various *system* aspects, like handovers.

In Japan, the 77 frequency channels have been divided into the 37 lower channels that are used for home and office use (i.e. real cordless and PBX applications) and the 40 upper channels, which are used for a low-mobility PCS system. This system has been very successful in Japan; it already has several million users. On the other hand, this system has also shown the enormous investments that are necessary to build up such a system: the metropolitan area of Tokyo is now covered by several *hundred thousand* base stations. We finally note that in the USA, there is the so-called PWT system that combines the system specifications of DECT (including multiple-access specifications) with the π/4-DQPSK modulation format; this results in a performance that is very similar to DECT.

In any case, the competition between DECT and PHS will probably not be decided in the technical sector, but by regulatory and marketing aspects. Will the Europeans be ready to provide the spectrum from 1900–1920 MHz for a system competing with DECT, and will the Japanese provide the spectrum from 1880–1900 MHz?

Table 2 Important parameters of the PHS standard

Transmission format	FDM/TDD/TDMA
Frequency band	1895 – 1918 MHz
Number of frequency channels	77
Number of users per frequency channel	4
Channel spacing	300 kHz
Frame length (8 slots)	5 ms
Data rate	384 kbit/s
Modulation method	π/4-DQPSK, square-root raised-cosine with $\alpha = 0.5$
Peak mobile transmitter power	80 mW

8 Testing the radio link: the testbed of TU Wien

While most research into the radio link performance of DECT is done by simulations on the computer, there is no substitute for real measurements. The performance of hardware can be best assessed in a testbed, where certain prescribed test procedures can be done under controlled circumstances. At the TU Wien we designed and built such a testbed for the evaluation of the DECT standard's physical and medium access layer specifications and to find the limitations of the DECT link for indoor and outdoor applications especially in dispersive radio channels.

The first generation of the testbed implemented the TDMA and FDMA structure of DECT, and was restricted to a simplex link [15]. The basic arrangement of the testbed is shown in Figure 7. It consists of a single direction DECT RF-link, set up by one desired transmitter (TX1), one interfering transmitter (TX2), the receiver (DRX), and an indoor channel simulator (CAN). The return link is established by a wired low data rate signalization connection from the receiver via a controlling personal computer (DTC) back to the desired transmitter (TX1). Transmitted data, received data and timing as well as synchronization information is transported by a high data rate connection to the error counter (ERC), which calculates the bit error as well as the synchronization error rate. The latter one gives the rate of slots

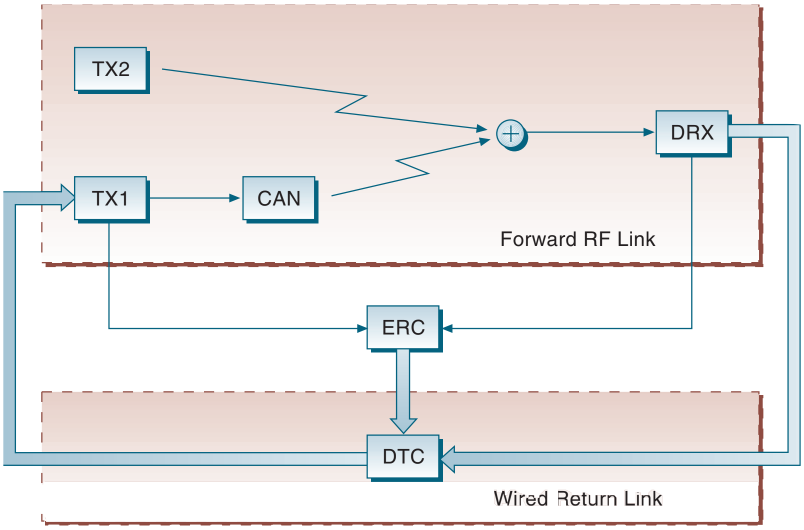


Figure 7 The DECT testbed

which were missed because the receiver could not synchronize. These basic building blocks of the testbed are completed by extensive diagnostics and measurement hardware, which give full hardware and system transparency.

The receiver is an incoherent direct conversion receiver (see section 9). Gain and frequency control may be operated with different algorithms and any combination of internal/external frame, slot, or bit synchronization may be chosen for advanced performance diagnostics. The internal bit synchronization of the receiver uses the algorithm presented in section 4.

The second generation of the testbed currently completed focuses on extensions which may be added to the DECT standard in the future. It also has even more thorough diagnostic possibilities: not only does it record the number of errors which occurred, but the complete received signal in its complex baseband representation (to the computer harddisk). This gives the opportunity to apply off-line postprocessing to the recorded data. To study the effect of variable antenna patterns we implemented a switched beam antenna with a Butler matrix, which allows us to select one of eight antenna beams for transmission and reception at the base station and so to investigate the feasibility and usefulness of switched beam antennas in future DECT systems.

9 Receiver hardware

When looking at cordless communications equipment today, we find mainly two competing transceiver architectures. Heterodyne transceivers have been well established for a long time, but need quite a number of expensive and hardly integrateable RF- and IF-components. Direct conversion architectures, although known in theory for a long time, are relatively new in real implementations but show a high potential for integration and are thus relevant candidates for the implementation of DECT phones [16].

9.1 Heterodyne architecture

The principle of a heterodyne DECT transceiver with a single IF is well known (Figure 8). The receiver RF part consists of a discrete high performance image rejection filter followed by a preamplifier and a simple receiver IF mixer. The receiver IF part operates at an IF of e.g. 110 MHz and consists of a discrete, high performance IF filter followed by an IF amplifier and a discriminator demodulator. The IF filter is one of the most intricate points in performance and costs in the heterodyne transceiver solution. Taking advantage of the constant envelope of GMSK, the IF amplifier is a non-linear limiter amplifier without gain control.

The transmitter IF part consists of an FM modulator on IF level which is controlled by the Gaussian filtered transmit data followed by an IF amplifier and a simple IF low-pass for IF harmonics suppression. The transmitter RF part is built from a simple IF mixer, a discrete, high-performance image rejection filter, a power efficient class C medium power amplifier and a harmonics-low-pass. The synthesizer consists of a voltage controlled oscillator, a dual modulus prescaler and low frequency synthesizer components.

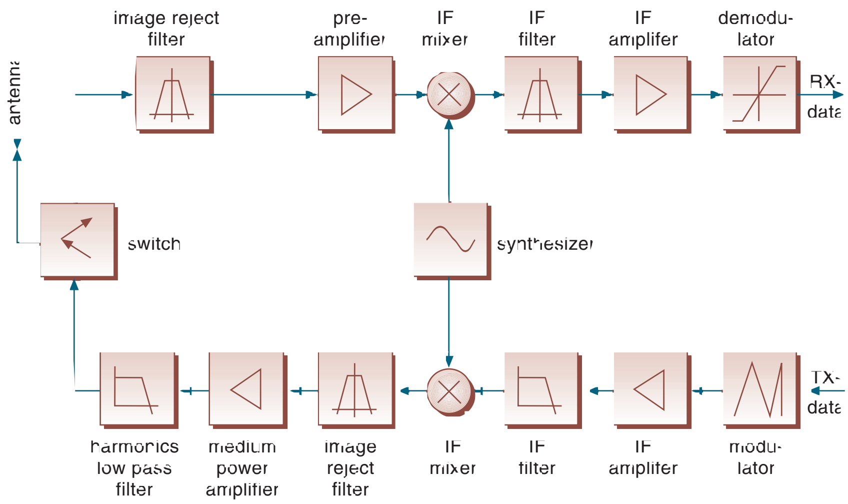


Figure 8 Heterodyne architecture

9.2 Direct conversion architecture

Figure 9 shows a direct conversion receiver, where the signal is transformed from baseband to RF and vice versa in a *single* step without using an intermediate frequency. Here, image frequencies do not appear.

The RF-part of a direct conversion receiver consists of a simple and cheap pre-filter for noise and interference reduction followed by a low-distortion preamplifier with switchable gain to reduce the dynamic range. The down-conversion from RF to baseband has to be performed in a more complex quadrature down-converter. This quadrature converter supplies in-phase and quadrature components of the received signal in baseband.

The receiver baseband part performs channel filtering, level control and A/D conversion by baseband signal processing followed by digital demodulation. To accomplish all these tasks, which were shared between the RF and IF parts in the heterodyne architecture, baseband signal processing can be done in two ways.

The first possibility is to filter the complex baseband signal by two analogue low-pass filters for channel selection, then to amplify the signals in a variable gain amplifier with accurate gain control, so that the dynamic range is highly reduced. Finally, A/D conversion is performed by a low resolution ADC. As a second possibility the signal can be filtered and level adjusted only coarsely by a low-order filter and a simple gain control, which means that both channel selection and gain control have to be done in the digital domain. Thus, a high-resolution ADC must be used for A/D-conversion, and channel filtering is done digitally.

The transmitter baseband-part converts the transmit data stream into complex symbols and performs the shaping of in-phase and quadrature component of the GMSK signal. The transmitter RF-part shifts the complex baseband signals in a quadrature up-converter to RF. Amplification is again performed in a power efficient class C medium power amplifier and a subsequent low-pass removes carrier harmonics. The synthesizer is equal to that in the heterodyne solution.

9.3 Comparison

Both architectures are well suited for DECT transceivers. The differences occur mainly in the receivers, and both impose different challenges for implementation.

Filtering by passive elements with high dynamic range makes the heterodyne receiver uncritical in intermodulation and noise performance. The limiter-amplifier in the IF part and the discriminator-demodulator are simple and low-power solutions and avoid the necessity for an automatic gain control (AGC) by their inherent high dynamic range. The FM modulator in the heterodyne transmitter supports a GFSK signal with a naturally low accuracy of modulation index, but matches quite well to the heterodyne discriminator demodulator. Due to the inaccuracy in modulation index and the non-linear limiting, equalization of delay spread is not possible with the limiter discriminator receiver.

A well known critical point of the RF-part in direct conversion receivers is DC-blocking due to second order non-linearities in the quadrature down-converter at reception of strong signals and self-reception of the local oscillator. This can be removed by AC coupling with carefully selected timing constants. In the receiver baseband part, the direct conversion principle requires linear signal processing from the output of the quadrature down-converter to the demodulator. The accuracy of the automatic gain control can be traded against the resolution of the following A/D converter, both requiring power consuming circuitry.

We have seen that the major drawback of the heterodyne solution is that it uses a number of discrete high-performance filters, which cannot be integrated and make up for a considerable part of the costs of a DECT telephone. The direct conversion DECT transceiver completely avoids costly discrete elements by moving signal processing mainly to the baseband, but also to more complex but integrateable RF circuitry. So far, companies seem to have chosen the low-risk-technology heterodyne handset with its inherently higher production costs. If they choose a direct conversion solution, it will need a higher development effort, but will eventually yield a very low-cost consumer-electronic product.

10 Conclusion

DECT has been designed as a low-cost system which has important consequences for the radio link. Due to the absence of channel coding, the bit error rate in the channel must be quite low, i.e. 10^{-3}. This requires high signal-to-noise ratio, low co-channel interference, and low time dispersion. This last condition is aggravated by the fact that no equalizer is foreseen and that DECT is a high-data-rate system. However, diversity is widely

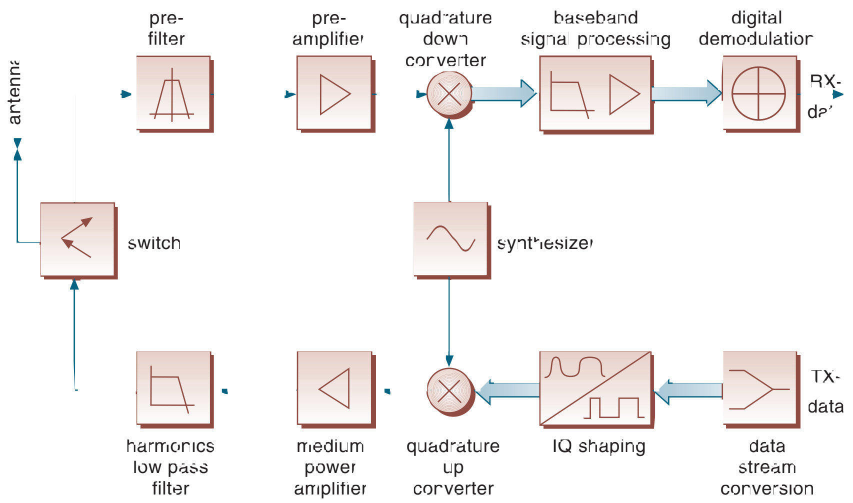


Figure 9 Direct conversion architecture

and successfully applied to improve BER and tolerance to time dispersion. For most PCS and WLL applications, equalization will be necessary to obtain satisfactory performance, but co-existence of hand-helds with or without equalizers is no problem.

On the other hand, the very simplicity of the DECT system allows a straightforward and therefore cheap construction of the hardware. Tolerances in the standard are not very tight, which allows the use of cheap components and even single-chip implementations of complete handsets. These low production costs have been the prerequisite for the phenomenal growth of the DECT market, and allow to anticipate an even brighter future.

References

1 ETSI. *Radio Equipment and Systems Digital European Cordless Telecommunications Common Interface.* DECT Specification, Part 1 – 3, ver. 02.01, 1991.

2 Pettersen, M, Rækken, R H. DECT offering mobility on the local level : experiences from a field trial in a multipath environment. *Proc. 46th IEEE Vehicular Techn. Conf., Atlanta,* 1996, 829–833.

3 Eskedal, B E. The DECT system. *Telektronikk,* 91 (4), 25–29, 1995.

4 Other DECT paper in this issue.

5 Adachi, F, Parsons, J D. Error rate performance of digital FM mobile radio with postdetection diversity. *IEEE Trans. Comm.,* 37, 200–210, 1989.

6. Liu, C L, Feher, K. Bit error rate performance of π/4-DQPSK in a frequency-selective fast Rayleigh fading channel. *IEEE Trans. Vehicular Techn.,* 40, 558–568, 1991.

7 Parsons, J. *The mobile radio channel.* New York, Wiley, 1992.

8 Paier, M, Molisch, A F, Bonek, E. Determination of the optimum sampling time in DECT-like systems. *Proc. EPMCC'97,* 459–465, 1997.

9 Molisch, A F et al. Error floor of unequalized wireless personal communications systems with MSK modulation and training-sequence-based adaptive sampling. *IEEE Trans. Comm.,* 45, 554–562, 1997.

10 Kadel, G. *Performance of the DECT system with a 2-state Viterbi equalizer.* COST 231, TD(93)100.

11 Lopes, L B. *A non-coherent equalizer receiver for DECT-type systems.* COST 231, TD(93)116.

12 Mogensen, P E et al. Evaluation of an advanced receiver concept for DECT. *Proc. 45th IEEE VTC,* 514–519, 1995.

Safavi, S et al. An advanced base station receiver concept for DECT. *Proc. 45th IEEE VTC,* 150–154, 1995.

13 Fuhl, J, Schultes, G, Kozek, W. Adaptive equalization for DECT systems operating in low time-dispersive channels. *Proc. 44th IEEE VTC,* 714–718, 1994.

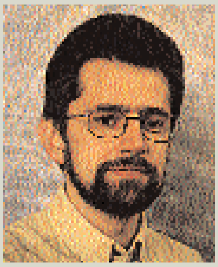
14 Molisch, A F, Bonek, E. Reduction of error floor of differential PSK in mobile radio channels by adaptive sampling. *IEEE Trans. Vehicular Techn.,* in press.

15 Schultes, G et al. A Testbed For DECT Physical and Medium Access-Layer. *Proceedings PIMRC'92,* 349–356, 1992.

16 Schultes, G, Kreuzgruber, P, Scholtz, A L. DECT Transceiver Architectures : Superheterodyne or Direct Conversion? *Proceedings 43rd IEEE VTC,* 953–956, 1993.

Andreas F. Molisch is Assistant Professor at the radio frequency engineering department of the Technische Universität Wien. His current research interests are bit error probabilities of cordless telephones, modelling of the mobile radio channel, and OFDM. He participated in COST 259 'Wireless Flexible Personalised Communications'. Earlier, he worked on surface acoustic wave filters, radiative transfer in atomic vapours, and smart antennas.

e-mail: Andreas.Molisch@tuwien.ac.at



Heinz Novak is Assistant Professor at the radio frequency department of the Technische Universität Wien. He is engaged in the research on transceiver architectures for mobile and cordless communications systems with a strong emphasis on DECT. His present research interests include the implementation of a demonstrator for a third generation communication system, which includes a switched beam adaptive antenna.

e-mail: hnovak@email.tuwien.ac.at

Ernst Bonek is Professor of radio frequency engineering at the Technische Universität Wien. He is interested in radio propagation for cellular mobile radio, 'smart' antennas and mobile radio networks, and serves as a working group chairman for 'Antennas and Propagation' in the COST Action 259 'Wireless Flexible Personalised Communications'. He is also Member of the Board of Post und Telekom Austria.

e-mail: Ernst.Bonek@tuwien.ac.at

Smart antennas

– The answer to the demand for higher spectrum efficiency in personal communications systems

MAGNE PETTERSEN AND PER HJALMAR LEHNE

1 Introduction

Personal and mobile communications systems have experienced an overwhelming increase in number of users in the recent years. With an increasing number of users and also additional services like data services being introduced there is a growing need for capacity.

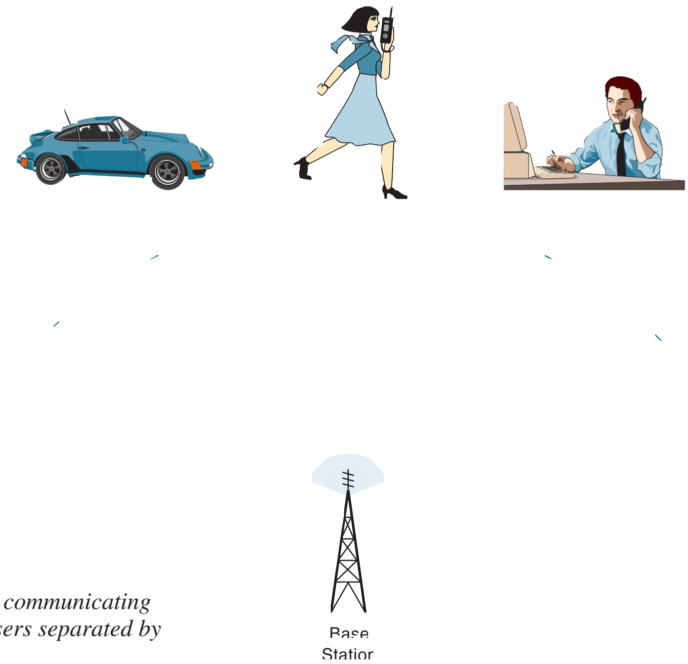


Figure 1 Multiple users communicating with one base station; users separated by frequency, time or code

Capacity can be increased by higher spectrum efficiency or an extended frequency band. Since the frequency band allocated to personal communications systems is limited, a number of techniques is being developed to achieve higher spectrum efficiency. In a cellular system the radio communication is between the user and a *base station* which provides radio coverage within a certain area, which is called a *cell.* In this case the capacity is limited by the cell density, the frequency reuse distance and the number of users which can be served simultaneously by each base station.

The traditional ways of separating users communicating with one base station is by frequency, FDMA – Frequency Division Multiple Access; by time, TDMA – Time Division Multiple Access; or by code, CDMA – Code Division Multiple Access. Traditional base station antennas are omnidirectional or sectored, but this is a 'waste' of power because most of it will be transmitted in other directions than towards the user. In addition, the power radiated in other directions will be experienced as interference by other users. Figure 1 illustrates the concept of a cell and multiple users communicating with one base station.

Spectrum efficiency can be increased by using smaller cells, so-called *microcells,* or by using *frequency hopping*, a technique which disperses interference. All these techniques are used in current digital mobile communications systems (eg. GSM) and will be employed also in future systems. However, in addition to these, the most promising technique is *smart* or *adaptive antennas*. This technique adds a new way of separating users on one base station, namely by space, SDMA – Space Division Multiple Access. An additional capacity gain is achieved by reduced interference levels, making lower frequency reuse distances possible. This will be explained further in chapter 4.

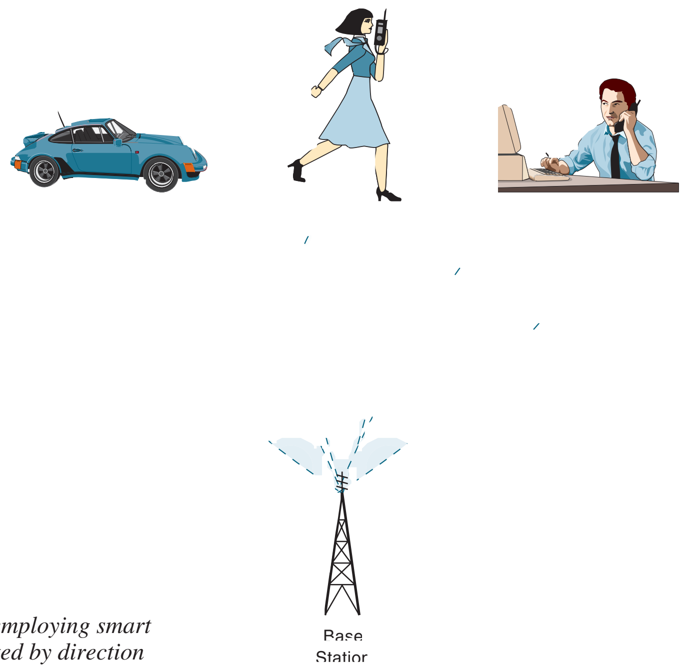


Figure 2 Base station employing smart antennas; users separated by direction

The idea of smart antennas is to use adaptive, directional antennas at the base station. By directing the antenna beam towards the communication partner only, more users can be assigned to the same base station. In this way, a much more efficient usage of the power and the spectrum is achieved, than in the traditional case of omnidirectional or fixed beam antennas. Figure 2 illustrates the concept of SDMA with multiple users communicating with a base station employing

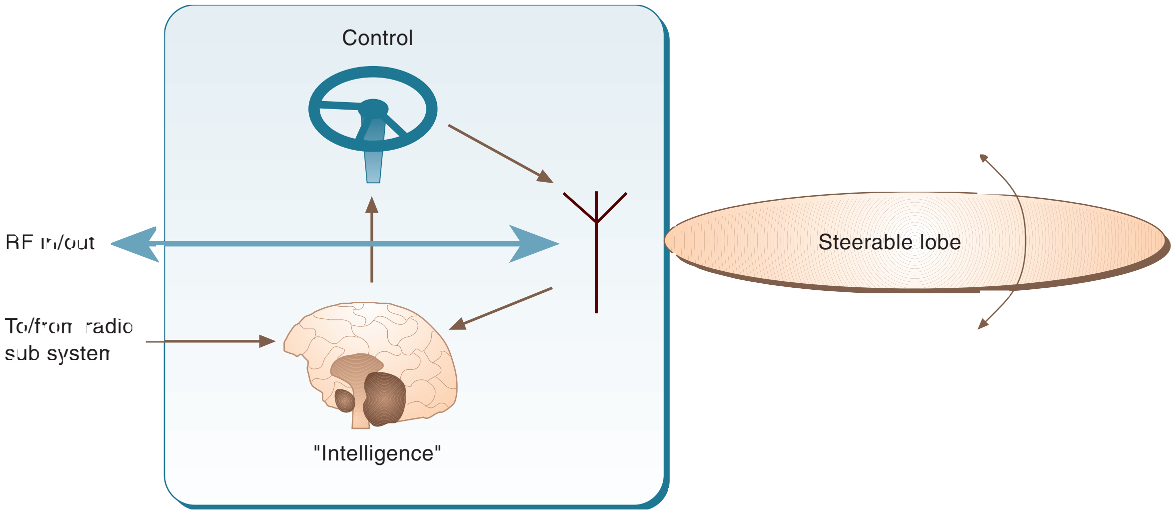


Figure 3 Principle of a smart antenna

smart antennas. In this case the users can use the same physical radio communication channel simultaneously if they are sufficiently separated by direction. As will be clear in chapter 2, SDMA is the final step in a development towards increasingly 'smarter' antennas.

In the last couple of years there has been large activity in the field of smart antennas for mobile and personal communications. So why is the interest appearing now, and not five or ten years ago? The answer probably lies with the fact that until now there has not been much reason to worry about spectrum efficiency. Also, if a base station is to track a large number of users simultaneously, the computational cost will be large, making powerful, expensive signal processors necessary. Only recently have such processors reached a reasonably low cost.

Smart antennas have a number of advantages over traditional omnidirectional or sectored base station antennas in addition to increased capacity, including *increased range, higher level of security* and *possibility for new services*. This will be discussed in more detail in chapter 4.

Although this article focuses on smart antennas for land based mobile communications, one should keep in mind that the technology can and will be exploited also for other types of radio systems.

2 Basic principles

What do we mean by the term 'smart antenna'? The theory behind smart antennas is not new. The technique has for many years been used in electronic warfare (EWF) as a countermeasure to electronic jamming. In military radar systems similar techniques were used as early as World War II. There is in principle a number of ways in which an adaptively adjustable antenna beam can be generated, for instance by mechanically steered antennas. However, the technique almost exclusively suggested for land based mobile and personal communications systems is *arrays,* or group antennas, giving directivity. This technology will be explained in some detail in chapter 3.

The main philosophy is that interferers rarely have the same geographical location as the user. By maximising the antenna gain in the wanted direction and simultaneously placing radiation pattern minima in the directions of the interferers, the quality of the communication link can be significantly improved. In personal and mobile communications, the interferers are other users than the user being addressed.

2.1 The concept of a 'smart' antenna

Several different definitions for smart antennas are used in the literature. One useful and consistent definition can be that the difference between a smart/adaptive antenna and a 'dumb'/fixed antenna is the property of having a steerable and fixed lobe-pattern, respectively. This is the definition to be used in this article. Figure 3 illustrates the concept of a smart antenna. Normally, the term 'antenna' comprises only the mechanical construction transforming free electromagnetic (EM) waves into radio frequency (RF) signals travelling on a shielded cable and vice versa. We may call it the radiating element. In the context of smart antennas, the term 'antenna' has an extended meaning. It consists of the radiating element and a control unit. We can denote the control unit the smart antenna's intelligence and it is illustrated in Figure 3 by the 'brain', normally realised using a digital signal processor (DSP). The processor controls feeder parameters of the antenna, based on several inputs, in order to optimise the communications link. The optimisation criteria can differ as will be explained in the next section. This shows that smart antennas are more than just the 'antenna', but rather a complete transceiver concept.

2.2 Levels of intelligence

Based on the definition above, one can define 'levels of intelligence'. These are described below and illustrated in Figure 4.

- *Switched lobe*
  This is the simplest level of intelligence, and comprises only a basic

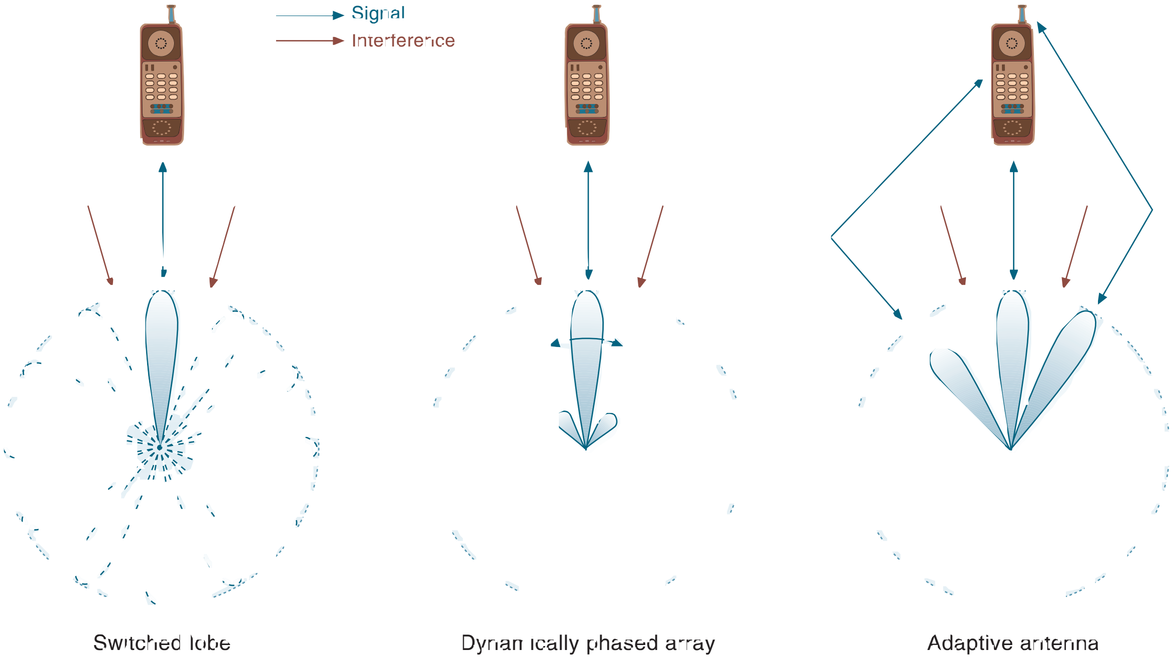


Figure 4 Levels of intelligence

switching function between separate directive antennas. The antenna lobe is chosen which has its angle of direction closest to the direction towards the user. Because of the higher directivity compared to a conventional antenna, some gain is achieved. Such an antenna will be easier to implement in existing cell-structures than the more sophisticated adaptive arrays, but it gives a limited improvement.

- *Dynamically phased array*
  By including a *direction of arrival* (DOA) algorithm for the signal received from the user, a continuous tracking can be achieved. The antenna's main lobe is directed towards the user, and thus the *signal level* is maximised.
- *Adaptive antenna*
  In addition to a DOA algorithm, if algorithms for determining the direction towards interference sources (e.g. other users) are added, the radiation pattern can be adjusted to null out the interferers. In addition, by using special algorithms and space diversity techniques, the radiation pattern can be adapted to receive multipath signals which can be combined. These techniques will maximise the *signal-to-interference-and-noise ratio* (SINR). A complementary gain is achieved in the downlink direction by using the same radiation pattern as found for reception. This requires that the radio channel does not change very rapidly. A more comprehensive description of the technology used is found in chapter 3.

Conventional mobile systems usually employ some sort of antenna diversity (e.g. space or polarisation diversity). Adaptive antennas can be regarded as an extended diversity scheme, using more than two diversity antennas. In this context, phased arrays will have a greater gain potential than switched lobe antennas because all elements can be used for diversity combining [1] [2].

2.3 An evolutionary path for smart antennas

All the described levels of intelligence in the previous section are technologically feasible today. However, in the domain of personal and mobile communications, an evolution can be foreseen in the utilisation of smart antennas towards gradually more advanced solutions. The evolution can be divided into three phases [3]:

1. Smart antennas are used on *uplink* only. (Uplink means that the user is transmitting and the base station is receiving.) By using a smart antenna to increase the gain at the base station, both the sensitivity and range are increased. This concept is called *High Sensitivity Receiver* (HSR) and is in principle not different from the diversity techniques implemented in today's mobile communications systems.
2. In the second phase, directed antenna beams are used in the downlink (base station transmitting and user receiving) direction in addition to HSR. In this way, the antenna gain is increased both on uplink and downlink, which implies a spatial filtering in both directions. Frequencies can be more closely reused, thus the system capacity increases. The method is called *Spatial Filtering for Interference Reduction* (SFIR). It is possible to introduce this in second generation systems, like for example GSM.
3. The last stage in the development will be full *Space Division Multiple Access* (SDMA). This implies that more than one user can be allocated to the same physical channel simultaneously in the same cell, only divided by direction. In a TDMA system, two users will be allocated the same time slot and carrier frequency at the same time and in the

same cell. In phase 2, the capacity is increased due to tighter frequency reuse. In phase 3, an additional increase in capacity is achieved locally per base station. Introducing SDMA in second generation TDMA systems will be difficult, but it will be a natural component in third generation systems, like UMTS.

The 'levels of intelligence' in the previous subsection describes the level of technological development, while the steps described here can be regarded as part of a system evolution.

3 Technology for Smart Antennas

This chapter explains some of the underlying technology employed in order to implement smart antennas. The treatment given here is far from complete, but should give an impression of the basic ideas. The technology is based on array antennas where the radiation pattern is adjusted by altering the amplitude and relative phase on the different array elements. Even though the reception and transmission parts are integrated, and much of the same hardware normally will be used, they will be explained separately. This is because conceptually the uplink and downlink are quite different in terms of smart antennas.

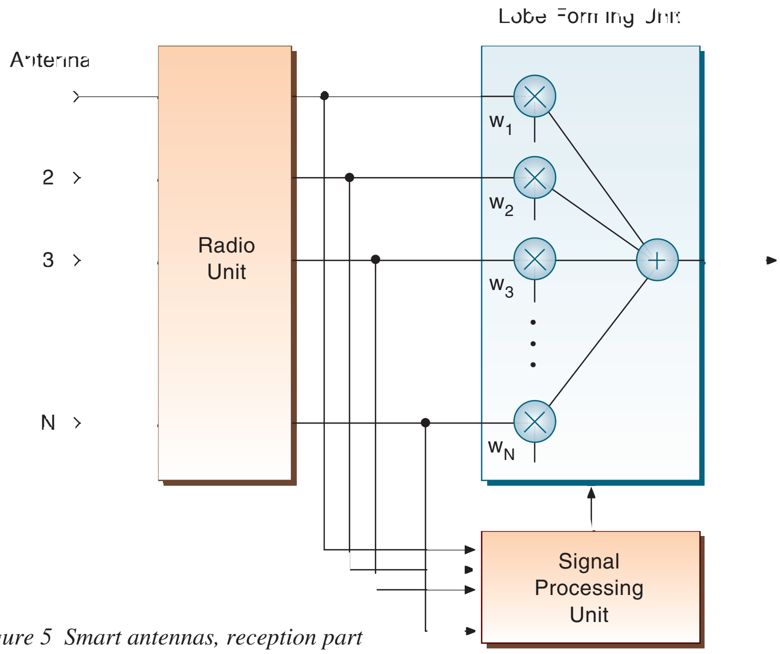


Figure 5 Smart antennas, reception part

3.1 Reception (uplink)

Figure 5 shows schematically the implementation of the reception part of a smart antenna. The antenna array contains *N* elements. The *N* signals are being combined into one signal, which is the input to the rest of the receiver (channel decoding, etc.).

As Figure 5 shows, the smart antenna reception part consists of four units. In addition to the antenna itself it contains a radio unit, a lobe forming unit and a signal processing unit [4].

The array will normally have less than 10 elements in order to avoid too high complexity of the signal processing. The elements could be positioned linearly (linear array) if a sector is to be covered or in a circle (circular array) if symmetrical coverage around the base station is desired. Often, each of the array elements will

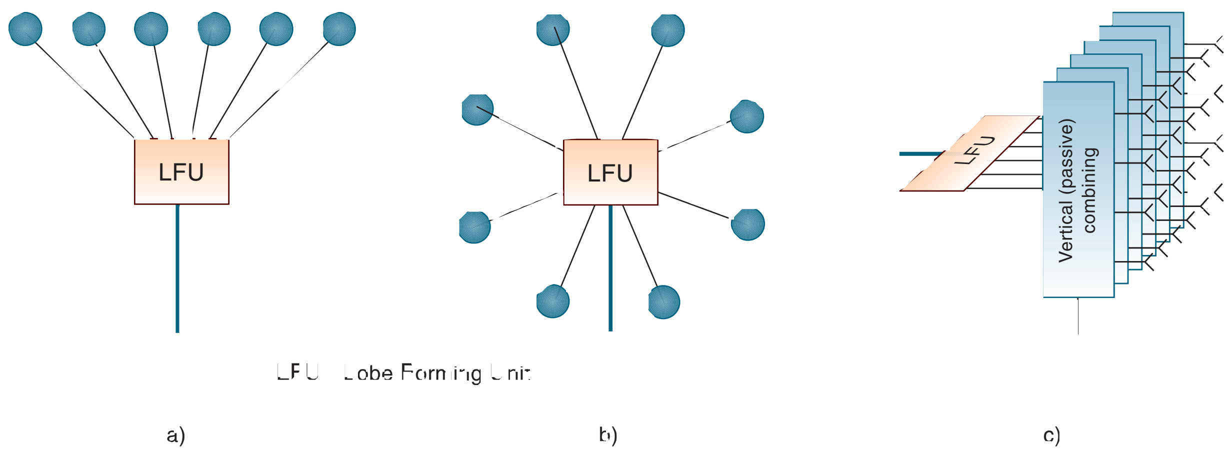


Figure 6 a) Linear array, bird's eye view; b) Circular array, bird's eye view; c) linear array with multiple antenna elements per array element consist of a number of antenna elements separated vertically which are combined passively to achieve a narrower vertical lobe [5]. Each of the array elements will more or less resemble a traditional base station antenna. Figure 6 shows a circular and a linear array, and also a linear array with multiple antenna elements per array element.

The radio unit consists of down conversion chains and (complex) analogue to digital conversion (A/D). There must be *N* down conversion chains, one for each of the array elements.

The signal processing unit will, based on the received signal, calculate the complex weights w_1-w_N with which the received signal from each of the array elements is multiplied. These weights will decide the antenna pattern in the uplink direction. The weights can be optimised from two main types of criteria; maximisation of received signal from the desired user (bullet 2 from sec. 2.2) or maximisation of the SINR by suppressing the signal from interference sources (bullet 3 from sec. 2.2). In theory, with *M* antenna elements one can 'null out' *M*-1 interference sources, but due to multipath propagation this number will normally be lower. Figure 4 (chapter 2) shows a case where the SINR is maximised by placing nulls in the direction of the interference sources.

The methods which maximise SINR do in principle require knowledge of the instantaneous channel response in both time and space from both the desired user and all the interference sources. Common for all the methods which suppress interference is therefore that they require a reference or training sequence, i.e. a known bit sequence must be transmitted periodically in order for the receiver to be able to estimate the channel response. This sequence should be unique for each user.

In the lobe forming unit the actual weighting of the received signal from each of the array elements is performed. In the most advanced case this unit will be an integrated channel equaliser/smart antenna. In this case *N•D* weights will be required, where *D* is the number of symbol periods (depth) in the channel equaliser. This is called a *spatio-temporal filter* [6] [7], because it removes the undesired signal components and keeps the desired ones both in time and space domains.

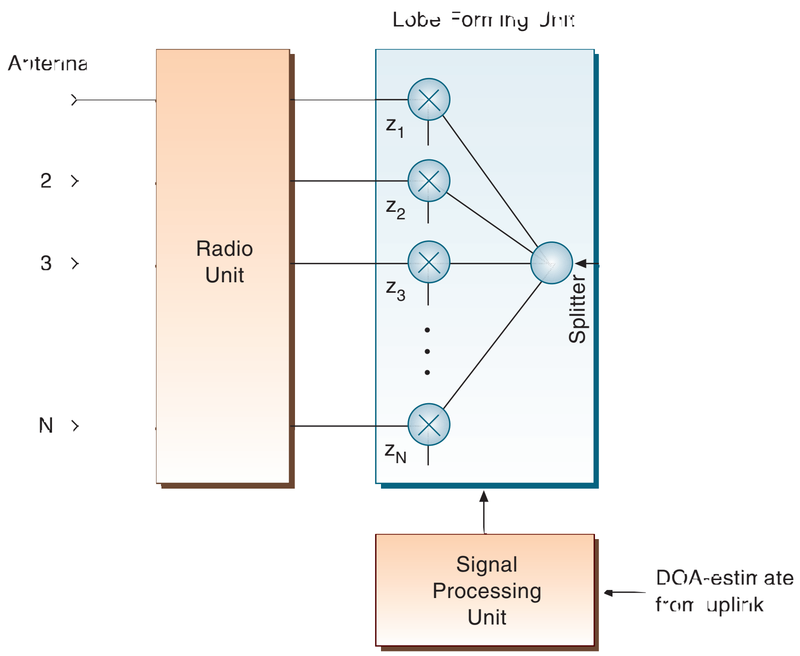


Figure 7 Smart antenna, transmission part

When the lobe forming is performed digitally (after A/D) the lobe forming and signal processing units will normally be integrated in the same unit, namely the DSP. The separation in Figure 5 is done to clarify the functionality. It is also possible to perform the lobe forming in hardware at *radio frequency* (RF) or *intermediate frequency* (IF) [8].

3.2 Transmission (downlink)

The transmission part of the smart antenna will be schematically very similar to the reception part. The transmission part is shown in Figure 7. The transmission signal is split into *N* branches, each being weighted by (complex) weights z_1-z_N in the lobe forming unit. The weights are calculated by the signal processing unit. The weighting decides the radiation pattern in the downlink direction. The radio unit consists of the up conversion chains. In practise, some components, like the antenna itself and the DSP, will be the same as on reception.

The principle difference between uplink and downlink is that since there will be no smart antennas employed to the user terminals (mobile stations), on downlink there is no way of knowing the instantaneous channel response. One could think that the weights calculated on reception (w_1-w_N) would be optimal also on transmission, but this will normally not be the case. If time division duplex is used (TDD, uplink and downlink transmission separated in time), the uplink-weights will be valid on downlink only if the channel does not change during the period from uplink to downlink transmission, i.e. if the user is moving very slowly. This can rarely be assumed in general. If frequency division duplex is used (FDD, uplink and downlink transmission separated in frequency) the weights will generally not be the same, because of the channel response dependency on frequency.

The technique most frequently suggested to solve this problem is the estimation of DOA (chapter 2). On uplink the DOA for the desired user is estimated, i.e. the direction from which the main part of the user signal is received. This direction is used in downlink transmission. This is done by choosing the weights z_1-z_N so that the radiation pattern is a lobe directed towards the desired user, as illustrated for instance in Figure 2.

The DOA can be estimated by a number of well known algorithms, for instance MUSIC, ESPRIT or COSSA [3].

3.3 Technical challenges and critical factors

There are a number of technical challenges which must be overcome in order for smart antennas to achieve their full potential in mobile and personal communications. These include both modem implementation aspects and network aspects. The aspects mentioned in this subsection are only some of the areas where research is done to find proper solutions.

The downlink is one of the major challenges. As described in the previous subsection the lack of knowledge about the instantaneous channel response on downlink can to some extent be compensated for by estimating the DOA and transmit in that direction. But this makes the downlink performance very dependent on the channel response behaviour in space. If there is much multipath propagation, so that the received signal has much angular spread, directing a single lobe in the direction of maximum reception is far from optimum in terms of maximising signal power transferred to the user. Interference suppression, to maximise SINR, is also difficult on downlink for the same reason.

In addition, when the channel is changing rapidly, the optimum direction calculated

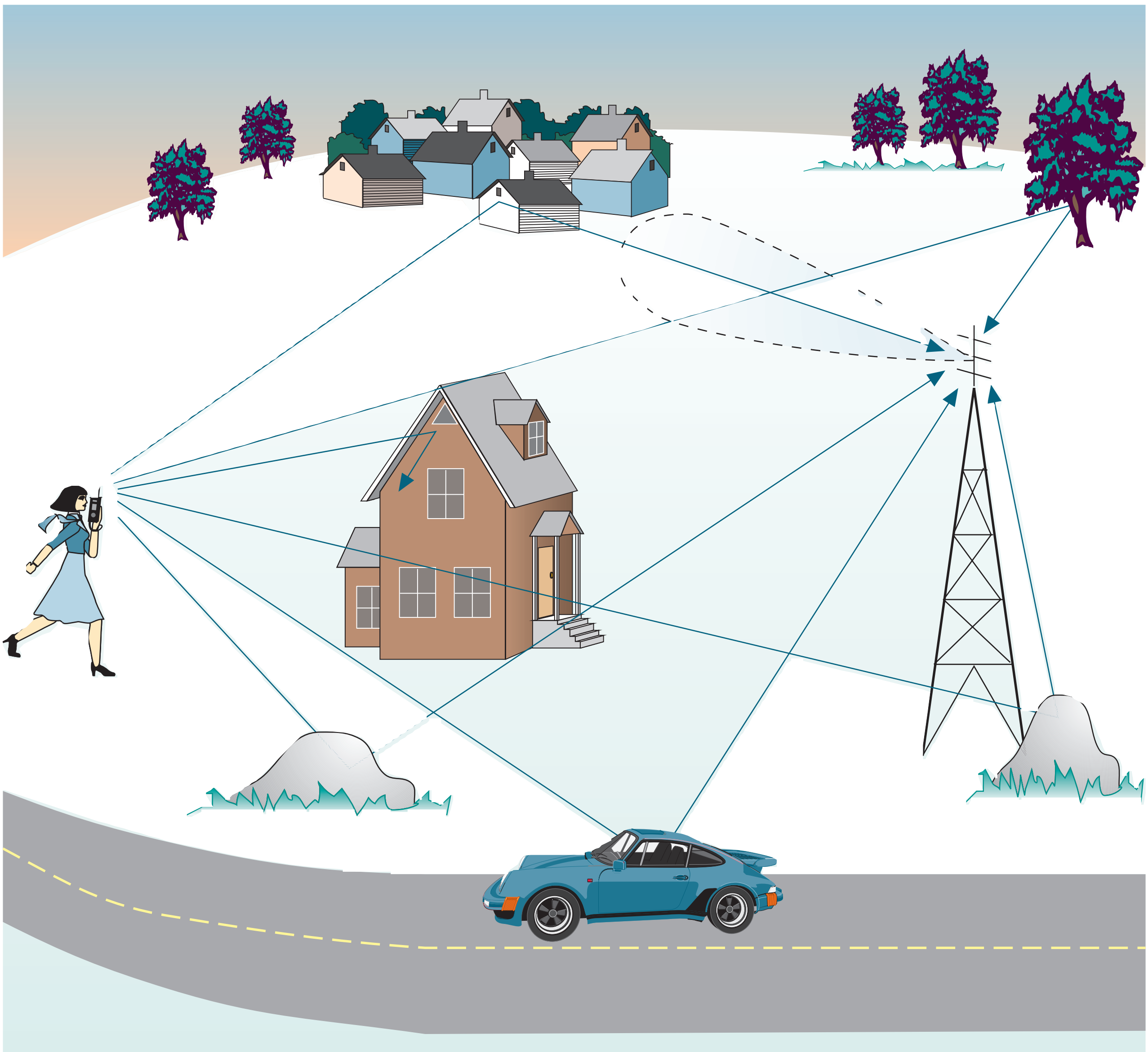


Figure 8 The uplink transmission experiences much angular spread due to multipath propagation. Downlink transmission based on directing a single lobe in the direction of maximum reception (DOA) is not optimum in this case

on uplink may not be valid on downlink transmission. Figure 8 illustrates the case with much multipath propagation and angular spread. Much attention is at the moment given to this problem, and a lot of effort is put into finding possible solutions.

The linearity in reception and transmission chains is another critical factor. If the transfer functions for the up- and down-converter chains are not perfectly known the actual weights will be different from the calculated ones, and the antenna radiation pattern may change significantly because of this. Since the transfer functions of the up- and down-converter chains are temperature dependent and also may change with time, frequent on-line calibration of the radio units of the smart antennas is necessary [8].

One of the network aspects deserving renewed attention when smart antennas are being introduced is connection set-up. When the user initially establishes contact with the base station, no angular information is available and some means for the base station to 'find' the user is necessary. A possible solution is that the base station scans the cell for new users by continuously sweeping through the cell with a 'search' beam. A similarly critical point is handover. When the connection is to be handed over to another base station some means must be found to provide the network with information of which base station to switch to.
A possible solution is to use an external system for positioning, for instance GPS. Information of the location could enable the network to make an 'educated guess' about which cell to hand the connection over to. Another possibility is a sweeping 'handover beam' mapping the user signal quality from different base stations [10].

Upon introduction of full SDMA (chapter 2) there will be new demands on radio resource allocation. For instance in a TDMA system, each user must be allocated not only a carrier frequency and a time slot, but also an angle. When angular collisions occur for two users using the same physical radio channel, one of them must switch to another channel. This must be done in a way so that the connection is not broken. This indicates that a system using SDMA must employ some sort of *dynamic channel allocation*. Figure 9 shows the occurrence of an angular collision.

Figure 9 Angular collision

When using smart antennas a 'spatial isolation' between the desired user and interferers is achieved by directing the antenna lobe in the direction of the desired user [11]. However, if the interference source is much closer to the base station than the desired user, the isolation is lost and communication may break down. This is similar to a 'near-far-problem' and indicates that some sort of *power control* is necessary, so that the transmit power from the user terminal is adjusted according to the distance from the base station.

4 Implications on personal and mobile communications

Introducing smart antenna technology into personal and mobile communications is perhaps the most important new technique to increase capacity in mobile networks since the cellular technology was introduced. It has impact on radio performance and stretches several limits beyond today's feasibility level. In this case, as in most others, there are also drawbacks and cost factors. In addition, it will have impact on radio planning. Also, new demands are put on network functionality. All of these aspects are discussed in this chapter.

4.1 Potential improvements

There are potentially great improvements in using smart antennas in mobile com-

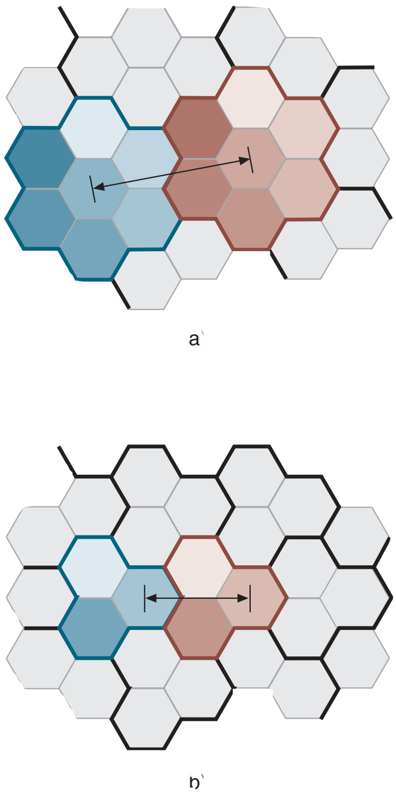


Figure 10 Reduced frequency reuse distance when employing smart antennas can be attained by using smaller cell clusters. a) Traditional 7-cell cluster as used in e.g. GSM; b) Possible 3-cell cluster when using smart antennas

munications. They have a number of advantages over traditional omnidirectional or sectored base station antennas. These include:

- *Increased capacity.* By transmitting electromagnetic power only in the direction of the communication partner, the interference experienced by other users will be reduced, the frequencies can thus be reused more frequently as Figure 10 shows an example of, and capacity is increased.
- *Increased range.* Because the base station antennas will be more directive than traditional ones, the range will increase if the same amount of power is transmitted. This is useful in sparsely populated areas where the base stations can be positioned further apart because capacity is not dictating base station deployment.
- *New services.* The spatial knowledge of the position of the user makes new services possible. These services can for instance include emergency call location or location specific billing.
- *Security.* It is more difficult to tap a connection when smart antennas are used. To successfully tap a connection the intruder must be positioned in the same direction from the base station as the user.
- *Reduced multipath propagation.* When a narrow antenna beam is used, the effect of multipath propagation, due to signal reflections, will be greatly reduced. This could ease the requirements on future modem design.

The most important of the advantages mentioned above is increased capacity. A capacity increase of 3 to 5 times is foreseen for GSM and similar TDMA systems [12]. For CDMA systems the increase is expected to be even higher. As described above, the reduced frequency reuse distance will lead to increased capacity. In future systems, the introduction of SDMA will give an additional capacity gain, because more users can communicate simultaneously via one base station.

Because interference suppression is the most important single factor increasing capacity, let us give an example of the potential. Given a system with N mutually interfering users and an adaptive antenna with $N + K$ antenna elements. For each of the N users, it is then possible to eliminate N - 1 interference sources and simultaneously exploit $K + 1$ multipath signal components [2].

4.2 Drawbacks and cost factors

There is always a price to pay. Increased complexity leads to higher equipment costs. Besides, smart antennas will not be equally suitable in all geographic and demographic scenarios. It is a question whether the higher cost will be compensated for by the gain in all scenarios.

One drawback is the physical size. To obtain a reasonable gain, arrays of several elements are used, having an element spacing of 0.4 – 0.5 wavelengths [9] [13]. (The wavelength for 900 MHz is 34 cm, while it is 17 cm for 1800 MHz.) Thus, these are large antenna systems and for 900 MHz the construction can be impractically large. For example, with 10 elements spaced 0.5 wavelengths apart, the total array width is 1.7 m on 900 MHz. Constraints on the size can also be a deciding factor for the choice of diversity technique (i.e. space vs. polarisation). Smart antenna systems are also computationally intensive. In order to take advantage of the potential benefits, there is a need for powerful and rapid numeric processors and control systems. There is also a need for efficient algorithms for real-time optimising and signal tracking.

4.3 Radio planning

It is necessary to develop a new way of thinking when it comes to choosing good sites for base stations with smart antenna systems. The current rule in placing base stations along a highway, a railway line and so on, is optimised on the basis of a fixed radiation pattern. Angle dependent access methods can, to a small degree, only be exploited in this way. By putting the base stations further away from the road or line, the beam can be swept along the coverage area. In this way, the spatial dimension is much better exploited. This is illustrated in Figure 11. New radio access protocols are necessary to take advantage of the new dimension in e.g. connection set-up and handover.

4.4 Network aspects

Even though smart antennas basically is a radio sub-system technology, it has some implications on the networking functions as discussed in chapter 3. In connection set-up, the base station needs some means to locate the user, and establishing directional information. Also when handover is to be performed, the network needs some positioning information about the user in order to make the right decision. This will require additional signalling information compared to current methods.

4.5 What about GSM?

Smart antennas will be fully exploited in third generation systems. However, parts of the technology can be integrated in current second generation systems like GSM and DCS1800. It is possible to use smart antennas in GSM without making changes to any other components than the Base Transceiver Station[1] (BTS) [14]. It is difficult to implement full SDMA in GSM because it is not possible to allocate more than one user per timeslot and frequency per BTS.

In ongoing trials using DCS1800 (chapter 5), smart antennas will only be used at the BTS. Consequently, the gain is highest on the uplink direction. The SINR is expected to be raised by up to 30 dB! [6] Other sources [16] indicate an improvement of 4 to 8 dB for downlink and 14 dB for uplink. One of the consequences of the improved signal quality can be a capacity increase of 3 to 5 times.

5 Current situation

This chapter contains a brief overview of the current situation in the area of smart antennas with regards to ongoing research and state-of-the-art.

5.1 Availability and state-of-the-art

As mentioned earlier, in addition to becoming an integral part of third generation mobile communications, smart antenna technologies have the potential to significantly improve the capacity in second generation systems, like GSM/DCS1800. Therefore, research is focused

[1] *In GSM, the base station subsystem (BSS) is organised in a controller part (BSC – base station controller) and the radio part (BTS – Base Transceiver Station). The BTS consists of one or more transmitter-receiver units (TRX).*

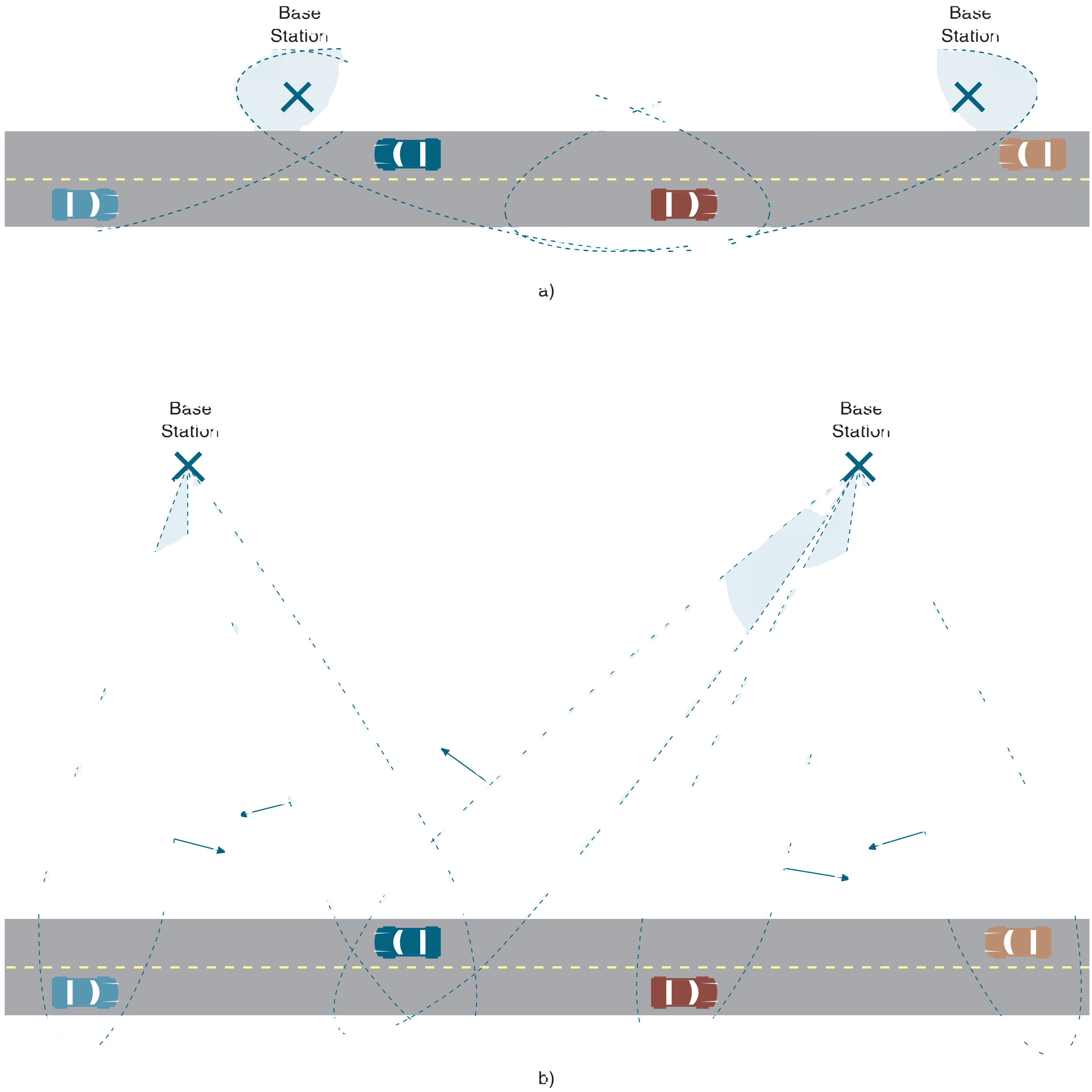


Figure 11 Different radio planning principles must be employed when smart antennas are to be used.
a) Conventional highway coverage technique;
b) Possible method of highway coverage using smart antennas

on current systems as well as on future systems, like UMTS.

Already, simple implementations of smart antennas designed for DCS1800 are available commercially. These are switched lobe solutions (similar to bullet 1, section 2.2), with a small number of lobes and switching based on signal level only [17] [18]. Most likely, many manufacturers are working on more advanced smart antenna solutions for GSM and DCS1800, but for reasons of competition it is difficult to get information about this kind of product development.

Apart from simple switched solutions, the closest smart antennas have come to implementation is testbeds which have been demonstrated on a stand-alone basis, and which are being implemented in a running network at a later stage. The European ACTS-project TSUNAMI [8][11][15] has successfully demonstrated a testbed consisting of one base station and two mobile stations [19]. The testbed is based on DCS1800 technology, and will be implemented into the DCS1800 network of Orange PCS in the UK.

Ericsson have reported impressive results from a similar testbed [14], also based on DCS1800 technology. This testbed is to be implemented into the DCS1800 network of Mannesmann Mobilfunk in Germany.

5.2 Ongoing research

A lot of international research is being done to develop smart antenna solutions. The research ranges from algorithms for DOA-estimation and channel characterisation, investigation of the implications on network aspects like handover and call-set-up, estimations of capacity and range improvements to development of hardware like antennas and receivers. In addition to land based mobile communications there is also research focusing on smart antennas for satellite communications.

A major drive in European research within this area is the aforementioned TSUNAMI-project, which has its main focus on smart antennas for UMTS. Some of the European institutions which have published results are the Universities of Bristol and Bradford in the UK, Aalborg University in Denmark, RWTH Aachen in Germany and also the Royal Institute of Technology in Sweden.

In the United States, the research on smart antennas focuses on benefits for CDMA-systems to a larger extent than is the case for Europe. Some of the institutions reporting results are Lucent Technologies and AT&T Labs Research, Telecom Solutions Inc. (Virginia) in co-operation with universities of Virginia and Texas, MIT Lincoln Lab and Hughes Network Systems and NASA.

In Asia, Kohno Laboratories of the Yokohama National University, Mitsubishi Electric Corporation and the University of Singapore, among others, have reported results. However, the European, American and Asian institutions mentioned above only make up a small percentage of the universities and companies currently performing research within the area of smart antennas.

6 Summary and conclusions

The concept of *smart* or *adaptive antennas* have the potential to significantly improve the capacity in mobile and personal communications systems, almost to the degree where it can be called a revolution. Additional benefits include *increased range, new services, increased level of security* and *protection against multipath propagation*.

The main idea of smart antennas is that the base station antenna radiation pattern is adjusted adaptively in such a way that the antenna beam is directed towards the desired user only, and nulls are directed towards the interference sources. In this way, one base station can separate users on the same physical communication channel by angle. This is achieved by using array antennas on which the radiation pattern can be adjusted by altering the amplitude and relative phase on each of the array elements.

Field trials have shown that smart antennas do work. However, the full potential of the technology remains yet to be demonstrated. There is still a number of technical challenges which must be solved. In addition, there are some drawbacks compared to current systems, like larger antenna constellations and more expensive hardware.

Smart antenna solutions will most likely be an integral part of third generation mobile communications systems, but less powerful solutions can also be implemented in second generation systems, like GSM, to give a significant gain. Smart antennas will influence most aspects of the systems, also network aspects and radio planning.

Few experts seem to doubt that smart antennas will have a large impact on future mobile and personal communications systems, and much research is currently taking place to find proper solutions.

7 References

1 Winters, J H, Gans, M J. The range increase of adaptive versus phased arrays in mobile radio systems. *28th Isilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA,* 31 Oct – 2 Nov 1994, 109–115.

2 Winters, J H, et al. The impact of antenna diversity on the capacity of wireless communications systems. *IEEE Transactions on Communications,* 42 (2/3/4), 1994, 1740–1751.

3 Bull, T, et al. Technology in smart antennas for universal advanced mobile infrastructure (TSUNAMI R2108) : overview. *RACE Mobile Telecommunications Summit 1995, Cascais, Portugal,* 22–24 Nov 1995, 88–97.

4 Barrett, M, Arnott, R. Adaptive antennas for mobile communications. *Electronics and Communications Engineering Journal,* August 1994, 203–214.

5 Johannison, B. Active and adaptive antennas for mobile communications. *Seminar and Demonstration on Adaptive Antennas and Antenna Diversity for Mobile Communications, Aalborg, Denmark,* 19 Jun 1997.

6 Öberg, T et al. Adaptiva antenner för mobil kommunikation. *Nordisk Radioseminarium 1996, Lund, Sverige,* 27–28 Aug 1996, 44–47.

7 Ogawa, Y et al. Spatial-domain path-diversity using an adaptive array for mobile communications. *Proceedings of ICUPC'95 – 4th IEEE International Conference on Universal Personal Communications, Tokyo, Japan,* 6–10 Nov 1995, 600–604.

8 Mogensen, P et al. A Hardware testbed for evaluation of adaptive antennas in GSM/UMTS. *7th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications – PIMRC'96, Taipei, Taiwan,* 15-18 Oct 1996, 540–544.

9 Forssén, U et al. Adaptive antenna arrays for GSM900/DCS1800. *44th IEEE Vehicular Technology Conference – VTC'94, Stockholm, Sweden,* 8–10 Jun 1994, 605–609.

10 Rheinschmitt, R et al. Network aspects of the introduction of adaptive arrays in mobile communications systems. *RACE Mobile Telecommunications Summit 1995, Cascais, Portugal,* 22–24 Nov 1995, 367–371.

11 Clarkson, J W. An advanced antenna system – unplugged. *ACTS Mobile Telecommunications Summit 1996, Granada, Spain,* 27–29 Nov 1996, 302–307.

12 Dam, H. Smart antennas in the GSM system. *TSUNAMI Seminar and Demonstration of Adaptive Antennas, Aalborg, Denmark,* 18 Jan 1996.

13 Wells, M C. Adaptive antennas for frequency reuse in every cell of a GSM network. *IEE Colloquium on 'Mobile Communications towards the Year 2000', London,* 17 Oct 1994, 11/1–11/6.

14 Forssén, U et al. Ericsson GSM field-trial with adaptive antennas. *Nordisk Radioseminarium 1996, Lund, Sweden,* 27–28 Aug 1996, 48–51.

15 Mogensen, P et al. TSUNAMI (II) stand alone testbed. *ACTS Mobile Telecommunications Summit, Granada, Spain,* 27–29 Nov 1996, 517–527.

16 Zetterberg, P. Signal processing algorithms for base station antenna arrays in GSM. *KTH Endagsseminarium över temat 'Avancerade Gruppantenner för kommunikation och radar',* Stockholm, Sweden, 5 Nov 1996, 1–3.

17 Beckman, C. Multibeam antennas. *KTH Endagsseminarium över temat 'Avancerade Gruppantenner för kommunikation och radar', Stockholm, Sweden,* 5 Nov 1996, 1–3.

18 Northern Telecom. *Smart antennas.* Product sheet, Devon, England, 1992.

19 Mogensen, P. Overview of CPK work in TSUNAMI II & system aspects of adaptive antennas for GSM. *Seminar and Demonstration on Adaptive Antennas and Antenna Diversity for Mobile Communications, Aalborg, Denmark,* 19 Jun, 1997.

Abbreviations

A/D	Analogue to digital (conversion)
ACTS	Advanced Communications Technologies and Services, 4th European Union framework programme in telecommunications research in Europe, 1995 – 1998
BTS	Base Transceiver Station (GSM term)
CDMA	Code Division Multiple Access
COSSA	Direction of arrival (DOA) algorithm
DCA	Dynamic Channel Allocation
DCS	Digital Cellular System
DOA	Direction of Arrival
DSP	Digital Signal Processor
EM	Electromagnetic (waves)
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques, Direction of arrival (DOA) algorithm
EWF	Electronic Warfare
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
GSM	Global System for Mobile communications
HSR	High Sensitivity Receiver
IF	Intermediate Frequency
MUSIC	Multiple Signal Classification, DOA algorithm
NASA	National Aeronautics and Space Agency (USA)
PCS	Personal Communications System
RACE	Research in Advanced Communications in Europe, 3rd European Unions framework programme in telecommunications research in Europe. Ended in 1995
RF	Radio Frequency
SDMA	Space Division Multiple Access
SFIR	Spatial Filtering for Interference Reduction
SINR	Signal to Interference and Noise Ratio
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TSUNAMI	Technology in Smart Antennas for Universal Mobile Infrastructure, RACE II project R2108 and ACTS project AC020
UMTS	Universal Mobile Telecommunications System

Magne Pettersen is Research Scientist at Telenor Research & Development, Kjeller. He is working in the field of personal communications, and his special interest is in modelling and characterisation of the mobile radio channel, an area in which he is currently pursuing a PhD at the Norwegian University of Science and Technology.

e-mail: magne.pettersen@fou.telenor.no

Per Hjalmar Lehne is Research Scientist at Telenor Research & Development, Kjeller. He is working in the field of personal communications, with a special interest in antennas and radio wave propagation for land mobile communications.

e-mail: per.lehne@fou.telenor.no

"One can talk about common manners!":

The use of mobile telephones in inappropriate situations

RICH LING

Evil communications corrupt good manners – I Corinthians 15:33.

A man by nothing is so well betrayed, as by his manners – Edmund Spenser.

The use of mobile telephones in various situations has become an element in the definition of socially appropriate/inappropriate behavior. They are causing us to reconsider how we construct our social worlds. Their use demands a reevaluation of the taken-for-granted assumptions of everyday life. This paper examines how people deal with inappropriate mobile telephones use, particularly in restaurants. The data, that is the talk, was collected in focus groups and through participation in electronic discussions on Usenet. The participants from the focus groups included 50 persons – 34 men and 16 women. Of the respondents, 30 reported experience with a mobile telephone while the remaining 20 reported only limited experience. The paper has also drawn on Goffman's notion of drama and staging. From this basis it examines the reasons that restaurants are particularly sensitive to the use of mobile telephones. There is a discussion of, among other things, parallel front stages and coerced eavesdropping. Finally, there is a discussion of the management strategies available when 'threatening' situations arise.

Figure 1 The advent of the mobile telephone has meant that we are experiencing the use of the device in new and unexpected places. As a result we are forced to reexamine our sense of propriety

1 Introduction

Not long ago I attended a church service in my wife's hometown, a small village on the West Coast of Norway. The church is a beautiful old stone building that was built around the year 1000. The inside of the building has white walls with the remains of drawings from the first part of the millennium. In addition, there are beautiful pews and an altar that reflect a long tradition of intricate woodworking. I was sitting near the back of the church not really listening to the singsong of the preaching, only enjoying the sunshine through the windows. I was trying to take in the scene when suddenly, on the other side of the aisle I heard the telltale peeping of a mobile phone. The offender sprang out of the pew in which he was sitting and rushed out of the church. After exchanging glances with several others sitting near me, things settled down again and I retreated back into my inner thoughts.

By now, most of us have been in a bus, an elevator or a restaurant and have heard or seen somebody talking on a mobile telephone.[1] In addition to the locations mentioned above, I have seen, or experienced, people using mobile telephones in places ranging from art galleries to toilet stalls. In a 1937 *Readers Digest* article Smith noted that, "There is no room in the house so private that he (sic.) cannot crash it by telephone" (Smith cited in Fischer 1992, 225). This lament now seems to have been extended to the far corners of the known world. These sightings show just how widespread the technology has become while the awkwardness of these situations points to the ways in which we are having to re-examine, or perhaps re-exert our ideas of propriety.

[1] *In this article I use the term mobile telephone. This is a translation of the Norwegian term* mobiltelefon *and refers to all radio based telephone devices aside from the cordless telephones whose coverage is usually less than 3-400 meters. In Norway the term mobile telephone refers to both devices that are mounted in cars and those carried about. In the US, the similar generic term is cellular telephone.*

In this paper I am interested in examining how it is that people are making sense of these changes and where people place the boundary between the appropriate and the inappropriate use of mobile telephones. I focus particular attention on the use of mobile phones in restaurants since it seems to have become a metaphor for vulgarity. The development of this metaphor is, however, not that surprising from a sociological perspective. After all, it is in restaurants that one sees quite clearly people dealing with the issues of self-presentation, boundary issues, etiquette, and 'stage management'.

This paper is a continuation of my work on the social boundary issues associated with the introduction of video telephony (Ling 1996). The mobile telephone, in many respects, represents the opposite of video telephones. Where videophones require one to be fixed to a specific location, one can roam with a mobile phone. Where one is forced to pay attention to their conversation partner with a video telephone, one is freer to carry out parallel activities with a mobile telephone. Where the video telephone conversation is a well-bounded event, the mobile telephone call is less well defined and can intrude while one is on the bus, in a restaurant ... or in church.

The work analyses data from a series of focus groups held in the Oslo area in addition to data collected in other forums. The data was analyzed using qualitative analysis techniques and the results have been further illuminated through the use of Goffman's dramaturgical approach to small group interaction. The methods of the analysis will be discussed in the next section followed by an examination of Goffman's applicability to this work, an analysis of the social meaning of mobile telephones and finally by the examination of the inappropriate use of mobile phones.

2 Method and theoretical background

2.1 Method

The work here represents an analysis of and reflections upon the information provided by informants in two different fora. The data, that is the talk, was collected in focus groups and through participation in electronic discussions on Usenet. The participants from the focus groups included 50 persons – 34 men and 16 women.[2] Of the respondents, 30 reported experience with a mobile telephone, while the remaining 20 reported only limited experience. It is worth noting that men were far more likely to have reported experience with mobile telephones. The preponderance of men, ca. 75 %, reported having used the devices while only slightly more than 30 % of the women said the same.

The inclusion of information from Usenet is somewhat outside the traditional approach to social science and thus is worthy of special comment. In a way, this form of information gathering is a type of telephone survey in that the interaction is of an anonymous nature. None-the-less, this approach is useful in the collection of natural talk. The data used here was collected by posting a comment in the forum on restaurants asking for the opinions of other participants. The original comment clearly indicated that I was interested in a sociological analysis of the responses. Those who responded, four individuals, were thanked for their input and their comments were added to the database included here. The advantage of this approach is that one can easily take preliminary hunches and concepts, form them into additional questions and gather additional information on these specific questions. This interaction with the data follows in the tradition of qualitative research, though through a new medium (Glazer and Straus 1967; Lofland and Lofland 1984; Spradley 1979).

It is important to be aware of the differences between the data gathered through the two approaches. The first is the degree to which informants felt pressure to respond. In the case of the electronically gathered information, the forum in which the text was produced is open to all with Usenet access. Thus, there is no specific responsibility to respond. By contrast, those who participate in focus groups and in other interview situations are participating in a somewhat unnatural setting where there is perhaps a certain social pressure to contribute. Thus, in the comments from Usenet one might get the loudmouths where in focus groups one may pressure people to formulate ideas that they otherwise would not have considered. Another aspect of the interaction on Usenet is that it is textual, not verbal. While this is a barrier to spontaneous remarks – that are often the most telling – it also means that the remarks are composed. That is, there is an element of reflection in the messages (though some that have read flame wars may be led to wonder about the quality of the reflection). A final difference is that the focus groups were Norwegian where the Usenet input was more global, though dominated by the US. I mention these contrasts to clarify the data, not to assert one source as superior to another.

These two sources of data were examined using standard qualitative analysis techniques. That is, the text was examined, classified, re-examined and further classifications and concepts formed the basis for the material reported below (Spradley 1979; Glaser and Strauss 1967). It will become obvious to the reader that the work of Goffman is deeply integrated in this paper. I must be quick to point out, however, that this is an inductive analysis. The concepts and the taxonomy presented below arose out of the data. This paper is not an attempt to prove Goffman, an impossible task since he left no hypotheses to prove or disprove. Rather, Goffman's insights, along with others from the interactionist school, have shown themselves to be useful for illuminating the concepts and for setting them into relief.

2.2 Theoretical background: the contribution of Goffman

This paper makes use of Erving Goffman's contribution to sociology. It is Goffman, in the tradition of Mead and Blumer, who has focused on the dramaturgical aspects of every-day life. This approach is particularly interesting in that he often gained insight into the structure of society by examining the situation when things went wrong or when there was stress. It is when our ability to maintain façade comes into serious disarray that we see people move to defend issues of importance. In addition, we also see the strategies used to deal with the situation. The results of the focus groups and other inputs have underscored that the use of mobile telephones presents us with a stressing of the normal situation in society.

Before accepting Goffman's work without question, however, one should also consider its shortcomings. At the most obvious level, many have criticized Goffman for what is considered a lack of scientific rigor. In the words of Giddens, Goffman's popularity "derives more from a combination of an acute intelligence and a playful style than from a coordinated approach to social analysis" (1984, 68). Meyrowitz notes that Goffman's work is considered by some to be "a stylistic merger of the scholarly monograph and the novel" (1985, 32).[3] By way of a response, his work seems to have been so evocative that it has been and continues to be a rich source of inspiration for others. Thus, in spite of the fact that his methodology is undocumented and his work is inaccessible, it has inspired others. This speaks to his success in observing and describing social interaction.

Another criticism is that Goffman focused his analysis on only face-to-face interaction. There was little or no analysis of mediated interaction in his work (Giddens 1984, 68-9; Meyrowitz 1985, 32, 345 n. 53). Like the previous point,

---

[2] *The focus groups were conducted in Norwegian and the analysis of the talk was largely undertaken before the material was translated into English for inclusion in this paper.*

[3] *Both of these theorists draw on Goffman in their own work. Thus, this criticism seems to be more of a straw horse upon which one can beat with immunity.*

this limitation has not prevented the re-application of his insights to mediated situations, or as in this case, hybrid situations. His work has served as a point of departure for examining of various forms of electronically mediated interaction.

Looking beyond the issue of style and the failure to examine mediated interaction there are two other criticisms leveled at Goffman. These include the sense Goffman, and the interactionist school to which he belongs, do not adequately describe the reproduction of social structure. In addition, these theorists tend to ignore the idea of motivation by painting a picture of society as populated by cynical manipulators (Turner 1986, 459).

In relation to the first critique there is in Goffman an often implicit – but sometimes explicit – assumption of society's reflexive nature. Reflexivity refers to the mutual assumptions one carries into interaction that are built up and institutionalized over time. This is how interactionists, and to an even greater degree those using the closely related sociology of knowledge, describe the development and maintenance of social structure (Berger and Luckmann 1967, 47–128). For those using these perspectives, institutions, such as the way in which we eat dinner or the way we speak on the telephone, are events that include the reciprocal typification of habitual actions that have built up over time. This may seem like a 'lite' version of analysis for those who see the structure of society in terms of large institutions such as the church, capitalism or kinship. None-theless, the Goffmanian approach does not preclude the inclusion of these types of factors. Rather, it is more concerned with every-day, micro-social issues.

The second point is well taken. To read Goffman is to read about petty schemes. One reads about people who try to present the best face but are continually hampered by masks that do not fit correctly, by irreverent children or naïve idiots who ask the wrong question or by the mobile telephone that rings at the wrong time. According to this reading of Goffman, if we were to take away these contrivances and face the naked ape as it were, there would not be much to take stock of.

There really is more to us than just petty façade management. After all, our very interest in reading Goffman, the empathy that we feel towards the ruses and characters he describes, speaks to a more complete individual than the simple cynical manipulator. In his defense, many disciplines have their more or less complete stick figures (or one might call them crash dummies) upon which they build their theories. There is the economic man of the economists, the Skinnerian response machine, etc. Each is an invention to make a point. Goffman's creation is the same.

While accepting the criticisms outlined here, I find in Goffman a useful and an evocative access point to the analysis of the data that is presented below. At the same time, I have been taken in. It is difficult not to be. Reading his analysis is one of those exciting adventures associated with being social scientists. The insight, the wit, the analysis and the conceptualization all come together in a rich and insightful analysis.

2.3 The dual nature of mobile phones

One last stop before the main analysis will be to consider the social meaning of the mobile telephone. This device has sprung onto the scene over the last decade. In Norway, 40 % of the citizens use a mobile telephone. The combination of access to the devices, the development of an adequate network and various other market related aspects have been important in the growth of this industry (Bakke 1996, 83; see also Haddon 1996b). In addition to the more traditional analyses of market forces, one can also examine the rise of mobile telephony from the perspective of its social meaning. What does the mobile telephone symbolize? In this connection, the work of researchers at the University of Sussex in the UK is of interest. That of Silverstone and Haddon (1996) and Silverstone, Hirsch and Morley (1992) is particularly insightful.

When examining the social meaning of new technology, it is useful to consider how the various devices and services relate to pre-existing social structures. Silverstone et al. (1992) describe the dual nature of communication technologies. On the one hand, they are objects that occupy space and can be displayed as an example of the user's status and position in society. In addition to their status as objects of display, they also have a use value. It is the former in which I am particularly interested. Silverstone et al. note that:

> *All technologies have the potential to be appropriated into an aesthetic environment (and all environments have, in some sense, an aesthetic). And many are purchased as much for their appearance and their compatibility with the dominant aesthetic rationality of the home as for their functional significance* (1992, 23).

The introduction of a new technology into what Silverstone and Haddon call the moral economy of the home also means that they must find their functional role. This role is not necessarily that which the designers had intended. Rather, it is through the combined display and use of the objects that the users come to understand the devices and it is through this process that the users weave them into a part of their own identity (Silverstone and Haddon 1996, 58, see also Marx 1994).

One might, for the moment, consider the mobile telephone a type of jewelry like a belt buckle, a watch, or a broach. That is, the device can have functional features as well as being a type of display. It can be appreciated or scorned. It can be the object of desire and interest or it can be seen as a symbol of vulgarity and bad taste. In order for the display to be successful, however, it must be done at the appropriate time, in the appropriate way with the appropriate élan.[4] The difference between a mobile telephone and a necklace, however, is that the former can be set into life by others who call in at the wrong time.

There are two final points to be made here. The first is that the adoption of new technology is a conservative process in that there is a prerequisite to incorporate the object into users' daily routines. There is in this process, however, a sense in that one can not go back home. Upon the domestication of the object, life changes, the new object and their attached routines nudge their way into the daily life of the user in both obvious and in not so obvious ways (Silverstone and Haddon 1996, 60; Stone 1994).

[4] *These comments are reminiscent of the development of a social norm proscribing against the use of digital watches that 'peep' each hour. When first on the market, many people used the peep function causing a discussion around their use during films and other public gatherings.*

Finally, there is a focus on the domestication of devices such as the TV or the PC within the home in the work outlined above. A mobile telephone is slightly different since the device is, almost by definition, individual and not attached to a physical location. Thus, some of the aspects of the moral economy of the household need to be seen in a slightly broader form. It may be that the device is only used internally within the family, at the cabin or as an internal messaging system. Apart from these alternatives, however, mobile telephones are not necessarily the same type of display or use object as, for example, a TV or a PC. This is expressed in both the use of the object but also in the reorientation of the user to the object itself.

3 Inappropriate use of mobile telephones

Now it is time to turn to the use and misuse of mobile telephones. The data from the focus groups indicated an almost visceral reaction to the inappropriate use of mobile telephones. The respondents were able to offer clearly formulated and well-rehearsed verbal sanctions against those who used mobile telephones inappropriately. At a broad level, the fact that these formulations were readily at hand indicates that the provocation is not at the level of the individual rather it is at the social level. In addition, it is evidence of the debate at a somewhat broader social level over the norms of telephone use. It indicates that we are in the process of sorting out our collective sense of how to deal with this issue (Haddon 1996a).

The respondents in the focus groups were quick to point out various situations where they felt it was inappropriate to use mobile telephones. These included airports, stores, meetings, on trains and busses, at various social functions and in theaters.

> *It irritates me to go to airports and I think it is embarrassing and very pretentious to sit there with that telephone. Sitting there with a cup of coffee with a man beside you with a PC in his lap. It simply irritates me. I think it is unreasonable.*

> *Once [my husband] had [a mobile telephone] in to town and we were in the town and were shopping and right as we came out of a store the telephone rang. I thought it was very disgusting to stand there outside the store and talk and he thought he was going to be so big and because and I saw that. I think that was very disgusting.*

> *Last week I was in a meeting and there were 10 or 12 in the meeting. Four of them had a mobile telephone. There was shuttle traffic out to the hallway. That is unacceptable.*

> *I was at a Christmas party the other evening and there was a lady who flew in and out of the ring [where children sing Christmas carols] 3–4 times. I cannot understand why one will have a mobile telephone at a Christmas party. And this with meetings, there are five men each with their own mobile telephones. They ring just about continually, there should be a rule that one leaves their mobile telephone and pager in the reception area.*

> *I think you have to be considerate when you have a mobile telephone. For example, that you don't have it on when you are at the movies. It is not so nice when they begin to ring in the movies, it is not so pleasant and theater and several other places where you should not call. I think that people need to be considerate and turn them off.*

In their defense, many users of mobile telephones understand that its use is not appropriate in all situations. Some respondents noted that it was embarrassing if the telephone rang in certain situations. Others were sensitive to the need to turn off the device when they were otherwise engaged.

> *If I had it I would have turned it off now. It is not that important. In cinemas and such places as now, I would have turned it off. One can say that there are taboo areas.*

The discussion here illustrates the development of a social norm proscribing the areas in which it is appropriate to use mobile telephones.

3.1 Restaurants as a special social situation

In addition to the situations outlined above, a common theme in among the respondents in the focus groups was the irritation over the use of mobile telephones in restaurants. To understand this irritation it is perhaps useful to examine the social meaning of eating at a restaurant. There are two elements that are of interest here: the territorial nature of restaurants and Goffman's idea of face.

3.1.1 Boundaries in restaurants

Some of the things that make a restaurant special from a social perspective are that

LARS BAHL / 2MAJ / SAMFOTO

Figure 2 The fact that mobile telephones allow for personal communication from and to many locations mean that third parties have audio access to our conversations. The consequent breaching of social boundaries has a resonance for Norwegians who have developed a sense of 'managed unreachability' as a basic cultural trait

one's use is temporary and it has elements of being both a public and a private space (Lipman 1967; Mars and Nicod 1984, 66-69; Silverstone 1996). The tables, booths and the portions of the lunch counter occupied by the patrons become theirs for the period of occupancy.[5] To claim a territory, however, we must go through the rituals of establishing and agreeing upon illusory perimeters or 'symbolic fences' with our fellow patrons (Gullestad 1994).

> *Some [territories] are 'situational'; they are part of the fixed equipment in the setting (whether publicly or privately owned), but are made available to the populace in the form of claimed goods while-in-use. Temporary tenancy is perceived to be involved, measured in seconds, minutes or hours, informally exerted, raising constant questions as to when it terminates. Park benches and restaurant tables are examples* (Goffman 1971, 29).

We become quite accomplished at ignoring others who are in quite close proximity through the use of a fictive curtain between tables that are, in reality, quite close to each other or even the very same table. These barriers allow each dining party to maintain the notion of a type of privacy that is, more or less, open for all to see and hear.

Goffman goes on to note that there are often demarcations, based on architecture and the placement of furniture that stabilize claims to space. These conventions allow for the erection of easily observable boundaries between individuals and parties on an *ad hoc* basis (Goffman 1971, 33–34). In this way, an individual or group can temporarily identify where it is that they belong and also determine their expectations as to where other patrons have the right to intrude. The waiters and waitresses, of course, enjoy a certain freedom of movement within through this complex of fiefdoms.

Within the stability there is also a flexible nature to these settings. Tables can be divided or combined in order to accommodate varying numbers of participants, groups can occupy adjoining tables or counter space and commandeer the area in between, etc. This notion of flexibility varies, however, with the class of the restaurant (Mars and Nicod 1987, 49). In exclusive establishments the tables are only adjusted 'back stage', that is before the arrival of the customers. This minimizes the need for the patrons to readjust their 'borders' with the entrance of other claimants. In other, more proletarian restaurants the customers often take the initiative to divide and re-divide the space as the need arises. While this allows for a more concentrated complex of groups – something in which the proprietor is interested – it also means that there is the need to adjust claims to territory, carry out cross border raids when the need arises for extra chairs or condiments, and in the worst case, it may mean wars of attrition when two parties lay claim to the same territory.[6]

The notion of social boundaries has a particular resonance for Norwegians. Social anthropologists have identified what they call "managed unreachability" as a basic cultural trait. While seen with foreign eyes, behavior may seem cold and aloof. When judged from a Norwegian perspective, however, this is seen in a positive light. The management of contact with others in public situations, and more privately, is seen as a way to protect a sense of self (Gullestad 1994, 167; see also Haugen 1983).

3.1.2 Maintenance of face

The second aspect of interest here is that of façade or 'face'. Goffman suggests that one is maintaining their face when their presentation and management of self is internally consistent. The development and maintenance of face is not a solitary activity, rather it is a social production. The face one presents is only as good as the willingness of the audience to treat it as real. Thus, façades are a crystallization of our willingness to be integrated into society (Goffman 1967, 7). As Goffman notes:

> *A person's performance of face-work, extended by his tacit agreement to help others perform theirs, represents his willingness to abide by the ground rules of social interaction. Here is the hallmark of socialization. If he and others were not socialized in this way, interaction in most societies and most situations would be a much more hazardous thing for feelings and faces* (1967, 31).

Face helps us to integrate ourselves in society, since we base and adjust our presentations on our perceptions of the situation. Without this, behavior would spin off in unpredictable directions and common intersubjective understandings of the interaction would be impossible.

If one considers a restaurant in this context, it can be seen as a dynamic stage upon which one's façade is displayed. It is a special situation where one is asked to combine etiquette and social finesse. It is where there is often a demand to celebrate social unity with one's family, friends or colleagues. At the same time, there are many unexpected turns and twists that a restaurant visit can take. There is a well prescribed set of rules and rituals that must be observed – the correct use of utensils, the way in which one eats, the topics that are available for conversation, etc. There are many possibilities for personal adventure – meeting that special someone in a singles' bar, closing an important deal, savoring a specially cooked meal, celebrating a birthday or other significant transition, etc. Finally, there are many potential hazards – spilling the wine or the soup, making inappropriate remarks, the need to cover over the fact that one received a poorly cooked meal, having to deal with Uncle Fred after he has had too much to drink, meeting others who one had hoped to avoid in that situation, using the wrong

---

[5] *For long term patrons, certain positions may become 'theirs' whenever they are present. To dispute this claim can result in severe sanctioning (Lipman 1967).*

[6] *Singles' bars and restaurants where one eats in order 'to be seen' are in a class for themselves. In the singles' bar, the barriers are there to be opened and closed as the attentions of others are sought or are to be avoided (Collas 1995; Giuffre and Williams 1994; Haavio-Mannila and Snicker 1980; Parker 1988). Common opening ploys – the purchase of a drink by a remote admirer, the use of the line "haven't we met somewhere before", etc. – are both inquiries as to the openness of the local landscape. In the case of restaurants where one appears to be seen, one may even use strategies to underscore their presence, that is they open themselves for observation, i.e. the grand entrance or sending a message to the management to page themselves over the PA system. This latter strategy seems somewhat similar to the characterization of the patron who uses the mobile telephone only as a way to draw attention to themself.*

knife and thus indicating unfamiliarity with 'correct behavior', etc. (Chen 1990–91; Fine 1995; Giuffre and Williams 1994; Parker 1988). All of these demands require that one becomes fluent in the maintenance of face.

One can see that eating at a restaurant is an important social performance by examining the rules of etiquette (Jackson 1952, 325; Duncan 1970, 266–69). In every culture, the rules of eating decorum are extensive and often include quite precise orderings of events and prescriptions of the exact placement of eating equipment, drink and food (Mars and Nicod 1987, 28–29; Rombauer and Becker 1964, 9–25). Manners also indicate status and hierarchy. It is through our use of manners that we indicate the way in which we expect to be treated. Thus, holding the teacup with the little finger extended in the proper way is in the words of Duncan, a "dramatization of the self" and a message to others that we expect a certain type of treatment (1970, 266). Geertz also comments on the way in which manners indicate status. He notes that etiquette is a recripocially built barrier that surrounds the individual. Etiquette protects the inner stability of the occupant and, as one goes up the social ladder, "the thicker the wall of etiquette protecting the emotional life" (1972, 290; see also Gullestad 1992, 165). All of this attention to the process of eating, particularly when it is done in public, speaks to the importance of the event, as well as to the potential for problems. It also speaks to the importance of reaffirming the social order (Cahill 1990, 391).

The final aspect of face in restaurants is that a restaurant visit is an event with a clearly visible price tag. All participants know, with some clarity, the cost of eating out; it is listed for all to see on the menu. Implied in this is the assertion of status and the idea of gifting. Thus, to eat out is to give evidence as to one's position in society. To invite one to a restaurant is, in some respects, to set a price on a social relationship. In a similar way, to accept such an invitation has the implication of indebting one to their host. Thus, untoward things that disturb the experience, such as the ringing of a mobile telephone, not only disrupts the work of maintaining façades, but can also depreciate the exchange between the dining partners. This sentiment was quite common among the informants who noted, for example "I have paid a lot of money for that meal" and "[Hearing a mobile phone in a restaurant] can make you feel like you've wasted money and made a bad choice about what you did that evening."

Thus, the development and maintenance of face in a restaurant is a delicate process. While there are benefits to be enjoyed by its successful achievement, there are also perils. The importance of the event is underscored by the elaborate rituals of etiquette associated which have developed and by the clearly visible economic aspect of the event.

3.2 Use of telephones in restaurants

Against this backdrop it is not surprising that respondents in the focus groups cited the use of a mobile telephone in restaurants as a particularly galling social violation.

> *It is awful to see those who go around, I could throw up! They use it even in restaurants, it looks so dumb.*
>
> *One can talk about common manners! When I was out at Mat & Vinhus [a local restaurant], it costs a lot of money to eat there, and the mobile telephone rang for somebody sitting behind us that had a lot of time to talk and talk. That is not good manners, I mean he could have said, just one second and gone out.*
>
> *I think it is repulsive to sit there and talk with people on a mobile telephone in a restaurant.*

With a point of departure in Goffman's concept of 'face' and that of temporary 'territoriality' one can begin to examine why the use of mobile telephones is so unpalatable. The most basic objection to mobile telephones has to do with the sounds associated with their use. When discussing the 'modalities of violation' of personal space, Goffman brings up 'sound interference'. In this case the violator fills up his or her accorded space and then some. One can violate the territory of others by carrying out an encounter over a longer than proper distance and thus 'obtruding' into the social space of others (1971, 33–34, 51). An aspect of this discussion that Goffman neglects is the type of sound one makes. There are a whole family of sounds that are inappropriate, particularly in restaurants. The unwilling belch or flatulence are only the tip of the iceberg. Any parent or member of a fraternity knows the variety and imaginativeness with which some diners can willingly create 'inappropriate' noises. The ringing of a mobile telephone fits into this family of inappropriate sounds, almost regardless of its volume.[7]

In restaurants where it is the most difficult to make a long term claim to space, i.e. cafeterias, it is the easiest to use a mobile telephone. That is, a certain clash of personal space is taken for granted and people are called upon to manage barriers most actively. In addition, the background noise may be of sufficient volume to cover over the mobile telephone conversation. On the other hand, in those restaurants where the boundaries between parties are the easiest to define, i.e. more formal restaurants where the individual is able to claim a table for a whole evening at the cost of a large bill, there is the assumption that the restaurant, and the other patrons, take responsibility for barrier maintenance.

3.2.1 Ringing

Several respondents noted the abruptness with which the mobile telephones rang. One person said that "They sit on busses and trams and on street corners and in restaurants and the telephones peep and peep". The most abrupt sound produced by the mobile telephone is often their ringing. The peeping or ringing is by nature an intrusive sound not unlike that of an alarm clock. One informant noted

[7] *It needs to be noted that the acceptability of the mobile telephone is somewhat place dependent. In a restaurant where higher levels of noise are part of the setting and where one is not expected to treat their attendance as a particularly special occasion, a telephone is more acceptable because its use is covered by other activities. One can include children's restaurants, fast food restaurants and informal cafes and eating locales in, for example, shopping centers where there is what one might call active background sound, i.e. fountains or musak. On the other hand, in situations where one is expected to attend to the here and now, i.e. the vintage of the wine, the topic of conversation, the quality of the food, the look in her or his eye, etc., then the direction of attention to a telephonic partner would be a greater intrusion.*

"The ring-ring of a phone disturbs the mood. Distracts the diner." While satisfying the demand to alert the attention of the user, the ringing can test one's ability to manage the social situation.

Many other restaurant patrons have developed a set of responses with which they can cover over unexpected sounds. The crashing of glasses or plates is parried with the use of a smile; the interruption of a child is dealt with either through forbearance or a short scolding; a belch is either ignored or laughed at; the abrupt arrival of the waiter is tolerated and the ringing of a traditional telephone in a restaurant is ignored since it is not a signal of importance to the participants. The adroitness with which one can smooth over such social rough spots was called face work by Goffman.

> *Each person, subculture, and society seems to have its own repertoire of face saving practices. It is to this repertoire that people partly refer when they ask what a person or a culture is 'really' like. And yet the particular set of practices stressed by particular persons or groups seem to be drawn from a single logically coherent framework of possible practices. It is as though face, by its vary nature can only be saved in a certain number of ways.* (Goffman 1967, 12–13).

Through the use of poise and *savoir-faire* one minimizes the damage of embarrassing incidents to the social situation. It seems, however, that the ringing of mobile telephones is not adequately routinized such that their ringing causes difficulties in maintaining face. There is, in addition, a complicating factor with mobile telephones. The ringing is often such that it is not just the intended called who must interrupt their conversation to answer the phone. Rather, the growing pervasiveness of mobile telephones means that the ringing occasions the need for all of those with a mobile phone to check if theirs is ringing.

Finally, the ringing of a phone in an inappropriate situation draws attention to that which follows, namely a telephone conversation. The phone conversation will, in effect, take the user out of the social context in which he or she was and place them into another. Thus, the ringing alerts others that the discussion which follows will be different in tone and subject.

3.2.2 Loud talk

The second type of sound that was irritating to the respondents was that of the person talking on the phone. It was common for the informants report that those who used a mobile phone employed 'loud talk'. One noted that: "People talk loud on telephones. Louder than usual, at least. That is as annoying as having a loud party next to your table." This brashness violates territories and makes it difficult to maintain the face:

> *When you have a loud person talking into a [mobile] phone and you (can't help but) hear only half of the conversation, that disrupts the coziness of the restaurant feeling.*

While many attribute the volume of talk to a desire for display or a sign of vulgarity, one can also examine this issue from the perspective of how the technology effects our form and style of interaction. At one level, when speaking on the phone there is the need to replace the visual gesturing of face-to-face communication with a form of verbal gesturing. The telephone conversation is a speech event with distinguishing characteristics. Several of these break with the style and tone of the speech events that are common in face-to-face communication. In face-to-face conversation quite nuanced body language has several functions. Through our use of nods, glances, small sounds and other gestures we indicate attention, the desire to speak, the desire to retain the floor and indicate pauses. We also use these devices to impart meaning and emphasis. All of these gestures are changed in a normal telephone conversation. Visual gestures are replaced by intonation and linguistic structure in 'grounding' the conversation (Rutter 1987, 105, 126; Martin 1991, 95–97; see also Duncan 1972, and Saks, Shegloff and Jefferson, 1974). Instead of relying on body language to control turn taking, pauses, emphasis, etc., these are done with what one might call verbal gestures. We use tones such as 'uh' to replace the lack of eye-contact that controls turn taking, phrases such as 'ah ha' replace nodding and other signals of continued attention on the part of the listener, etc.

In addition to the different style of talk, one is often prompted to speak louder when using the phone. This is because of the need to be heard over background noise on the line and the general socialization that one speaks up on the telephone (Martin 1991, 95–97).[8]

Thus, the need to include verbal gestures and the habit of speaking up on the phone, in addition to the fact that the whole episode is announced by the ringing of the device, makes it easy for others to identify the talk as a telephone conversation. If the listener has preconceived problems with the use of telephones in restaurants the ease of identifying the talk allows them to play out their prejudices.

3.2.3 The management of parallel front stages

Respondents talked about considerations that went beyond the directly audible aspects of mobile telephony. The first of these is that the combination of the private telephone call and the private dinner conversation means that one must juggle two parallel interactions. When one begins to talk on a telephone, they have, at least partially, removed themselves form the ongoing interaction in their physical location (Clark 1995). One informant in the focus groups expressed an annoyance with this potential. This person noted that "It is a little irritating when ... you sit together with someone and they cannot set [the mobile phone] down."

While it is possible to carry on parallel activity while eating at a restaurant or while speaking, these possibilities are limited (Ling 1994; Lohan 1996a). The intrusion of a mobile telephone call threatens the pre-existing situation. At the first level, one must choose which conversation takes precedence. Goffman calls this accreditation.

> *Messages that are not part of the officially accredited flow are modulated so as to not interfere seriously with the accredited messages. Nearby persons who are not participants visibly desist in some way from exploiting their communication position and so modify their own communication, if any, so as to provide interference* (Goffman 1967, 35).

---

[8] *According to Bakke (1996) we expect a certain level of background noise from the phone. If this is not to be heard one feels that the phone is dead and it reduces the sense of co-presence. This has led to the use of 'comfort noise' in digital systems that are, by nature, more free from background noise.*

If the telephone call receives accreditation the mobile phone user has in effect two front stages,[9] that in the physical restaurant and the telephone conversation.[10] Quite often, the behavior appropriate in the one situation is not appropriate in the other. The different relationship that one has with one's dining partners and one's telephone partner means that the negotiation of topics, the depth, passion and emotion with which one can address the one party and the range of common understandings will probably not be the same for the other (Garfinkle 1967, 35–75). The shifting front stages for the phone user makes it possible for others to see the user of the mobile telephone in a type of verbal cubism. While the face-to-face restaurant talk may be, for example, cozy, intimate and integrative, the talk on the mobile phone may be of power relations, fast deals and office politics. Another example of this that has been seen for its comic value is when one's intimate dinner with a lover is interrupted by a call from an irate spouse. The stage management can become quite complex. Like a cubist painting, the speaker on the mobile phone is seen from two perspectives. There is a certain dissonance when judged from the perspective of classical notions of talk, dialogue and narration that present a single perspective to the audience.

In order to manage the stage one can erect various types of partitions. One can, for example, temporarily move from the 'front region' of the restaurant to a reception area or other more removed location in order to allow the harmonious completion of the call. This strategy is, however, only partially successful if the offense to the decorum was done by the ringing of the telephone in the first place. A second strategy discussed by Goffman is that the original audience be granted a temporary back stage status (Goffman 1959, 139). Thus, they will be allowed insight into the staging of another performance. In the short term, this allows the completion of the telephone call: however, once belief in the first performance is suspended, rebuilding its façade may be a difficult proposition, particularly if the original audience is only begrudgingly willing to accept its back stage status.

During the period of the call the dining partners are left in a particularly stressful sort of suspended status in that they are asked to wait. They are not dismissed, rather, they are left hanging. Only after the conclusion of the call can they resume of their earlier status. While waiting they must engage themselves in some type of waiting strategy that is easily discarded when the other summons them back. This is a particularly difficult social juxtaposition.

One possibility is to somehow acknowledge their presence to the caller – i.e. the mobile telephone user indicates the presence of the third person and may even make open side comments to them or, in some cases, the third person may even make open side comments to the caller and the called. Another possibility is that the third person may be called on to indicate that they are not engaged in the conversation at all (Lohan 1996a). An informant described this latter situation for a daily restaurant patron and a frequent mobile telephone user. "He always has a companion with him who is left to peruse the alphabetized beer bottle collection behind the bar for five or ten minutes." Here, the companion is left to maintain his own illusion of indifference.[11] This strategy for dealing with the situation will be discussed immediately below.

3.2.4 Coerced eavesdropping[12]

Another aspect of boundary management is the issue of eavesdropping. The common understanding of this problem is that those inside a conversation fear that those outside the conversation can become privy to secrets. There was a slightly different perspective reported here in that the informants feared that they would hear too much, a sort of coerced eavesdropping. The need to guard against this was seen as a problem.

Several of the respondents noted that the audience to a mobile phone conversation in the restaurant is not only one's immediate dining partners, but also those who are within hearing distance of the conversation. The placeless nature of the call (from the perspective of the remote caller) means that the talk may take up private issues not intended for others to hear. The information exchanged can be, or perhaps more importantly, can appear to be revealing. Thus, the audience is provided involuntary access to portions of the phone user's life, a clear violation of Gullestad's managed unreachability (Gullestad 1994). Many of the respondents talked about unwittingly eavesdropping in on other's telephone conversations. In the words of one respondent:

> *I think it is fine that the telephone is stationary at home. A conversation should be between two persons. I think it is unfortunate that others are there.*
>
> *[Hearing a mobile telephone at a restaurant] breaks up the enclosed social circle that a restaurant offers. Seeing people communicate and be social at different tables is part of the restaurant experience.*

In other cases the topic may be judged to be superfluous as in the case reported by one informant "[The user of a mobile phone] yaks away about (obviously) trivial matters." Regardless, the need to guard against hearing too much was reported to have been disruptive.

---

[9] *A later formulation was developed by Goffman in which he presents the idea of frames (1974). Framing refers to that set of conventions that define 'what is going on here'. While there are many alternative possibilities, social interaction requires that the participants in an interaction agree, at least generally on the same notion of an interaction's content and form. This agreement frames the activity (Giddens 1984, 89–90).*

[10] *This is particularly true if the mobile telephone user has received the call. If he or she has initiated the call they can at least prepare their dining partners for the event, and they also avoid the noise of the ringing telephone. It is also worth noting that these two front stages are not necessarily mutually exclusive. It is, in fact quite possible for others to participate quasi-actively in a telephone conversation. This is particularly true if the caller, the callee and the third person are all familiar with each other and are perhaps planning a common occasion (Lohan 1996a).*

[11] *There are echoes of Simmel's discussion of the stranger in this situation (Simmel 1971, 143–149; Bogard, 1996).*

[12] *The word eavesdrop refers to the ground upon which rainwater drips from the eaves of a house. Presumably if one were to stand inside the eavesdrop they become privy to the communication inside the house.*

3.2.5 The mobile phone as a vulgar metaphor

The fact that one carries on a telephone conversation may have become an inappropriate gesture in its entirety. Again and again the informants went beyond an analysis of the component elements and encapsulated the whole event by describing it as rude or disconcerting.

> *Talking on a cell phone at dinner, in movies, (or at a seminar (yes, I've seen it))* ***is just plain rude****. My family never allowed telephone calls at the dinner hour in my home, because when people are sharing a meal, it is the time to communicate and socialize. (emphasis added)*

> *I think it is a little impolite when people stand there and bellow in a mobile telephone ... or me it is not good manners. It is like standing there and whispering to somebody. That is not polite either.*

There was not a discussion of the mechanics of body language or the non-contextual nature of the conversation, rather, there was a reaction to the event as a whole. Almost as drinking coffee from a saucer, holding a wine glass from the base, eating from one's knife, or in some cultures eating with one's left hand have become a gloss or a characterization of a certain type of person, the use of a mobile telephone in a restaurant has become a metaphor for the vulgar breaching of manners.


LARS BAHL / 2MAJ / SAMFOTO

Figure 3 The use of the mobile telephone in restaurants by employees, but more importantly by patrons, has become a metaphor for the vulgar use of the device

3.3 Management strategies

Once the breach has been made, how do other patrons respond? When face has been threatened, Goffman refers to the need to undertake 'face work'.

> *By* face-work *I mean to designate the actions taken by a person to make whatever he is doing consistent with face. Face work serves to counteract 'incidents' – that is, events whose effective symbolic implications threaten face. Thus poise is one important type of face work, for through poise the person controls his embarrassment and hence the embarrassment that he and others might have over his embarrassment* (Goffman 1967, 12–13).

The mutual management of embarrassment illustrates the integrative function of face work. This is what Berger and Luckmann see as the reciprocal nature of society (1967, 47–91). In this way "one has an interest in a defensive orientation towards saving his own face and a protective orientation saving the face of others" (Goffman 1967, 14). Thus, we are likely to ignore minor infractions in the spirit of the occasion.

There are other situations, however, that go beyond the simple management of embarrassment. Setting this type of a situation aright is an important issue since its continuation can threaten the situation and can insult our sense of identity. Duncan, in fact, asserts that the flagrant ignoring of manners can be seen as a serious threat to order.

> *Anger over ill manners of others arises out of the belief that not following our manners is a way of telling us that we are not really important in the eyes of the transgressor. We excuse a* faux pas *made out of ignorance (and soon corrected) because we still feel the importance of our manners as a social bond. We laugh at comic depictions of vulgarity so long as the majesty of what we hold important is not threatened. But we do not laugh at savage ridicule or continued vulgarity, because they endanger the social principle upon which our manners are based* (Duncan 1970, 267).

When a breach has been made and there is the need for face work, the audience has several responses available to them. At the simplest level Goffman suggests that one employs studied non-observance. Beyond this one might challenge the offender or vilify the offender *in abstentia*.

3.3.1 Civil inattention

One of the most common strategies for dealing with a relatively minor offense, such as talking on a mobile telephone, is to engage in what Goffman calls 'civil inattention' (Goffman 1963, 85–86). This is described as the willingness to turn a blind eye towards behavior that represents a potential threat to face. It is an attempt to gloss over the disturbance and to carry on.

Via socialization in manners and courtesy, we understand the intrusiveness paying attention to the conversations of others. Thus, the public use of a mobile telephone tests the courtesy of patrons in a restaurant. As discussed in the section on eavesdropping, those outside the conversation may fear that they will hear too much. The reaction is to become embarrassed for the potential embarrassment of the telephone user. Thus, we must parry our own interest and replace that with a set of staged stances that are socially defined, e.g. making a display of reading the menu, looking at the decor, suddenly becoming engrossed in the papers one is carrying in their hand or, as reported above, looking at beer bottle collections. This is studied non-observance. As the mobile telephone becomes normalized we will likely need to develop a repertoire of suitable inattentive postures which we can assume.

The goal of this is to not disturb the scene.

> *The combined effect of the rule of self-respect and the rule of considerateness is that the person tends to conduct himself during an encounter so as to maintain both his own face and the face of the other participants ... A state where everybody temporarily accepts everyone else's line is established. This kind of mutual acceptance seems to be a basic structural feature of interaction, especially the interaction of face-to-face talk* (Goffman 1967, 11).

The fact that we generally give no outward sign of reacting to the use of mobile telephones indicates that the offense is not as significant as others.

3.3.2 Challenges and responses

Beyond civil inattention one has the ability to sanction others using various types of challenges. In the case of "Extreme impropriety ... is likely to result in his being stared at or studiously not seen" (Goffman 1963, 87).

In these cases one may make a transition from civil inattention to the slightly stronger studied non-observance (1967, 17–18). The differences between the two seem to be the degree to which the observer acknowledges the existence of the observed. Civil inattention includes the idea of a token recognition followed by focusing on other matters. By contrast, studied non-observance seems to be making a great display of refusing to acknowledge that the observed exists. The former seems to be more of a strategy of interaction for those who are more or less in agreement on the agenda while the latter may indicate more of a disagreement as the nature of the setting. Civil inattention is a courtesy and studied non-observance is a negative sanction.

Beyond these relatively discrete sanctions one can also employ strategies such as glaring, sighs, poorly disguised comments etc. Further steps include direct confrontations, ask the waiter or other authority to intercede, and the like. These strategies may or may not succeed. If the offender refuses to curtail their behavior in the face of the sanctions, then the ball moves back to the person who issued the sanction. It may be that their face has been put into question. This means that either they are left to bluster, they can retaliate or they "can withdraw from the undertaking in a visible huff-righteously indignant, outraged, but confident of ultimate vindication" (Goffman 1967, 22–23).

There is obviously a balancing point in all of this. If one begins to make more of a disturbance in their sanctioning than the situation allows, they may become the object of sanctions themselves.

> *Improper conduct, however, does not automatically release others from the obligation of extending civil inattention to the offender, although it often weakens it. In any case, civil inattention may be extended in the face of offensiveness simply as an act of tactfulness, to keep an orderly appearance in the situation in spite of what is happening* (Goffman 1963, 87).

3.3.3 Attribution

Judging from the responses of informants, a final strategy for dealing with the use of mobile telephones in restaurants and other public places seems to be *ex post facto* stereotyping of the offender. While not addressing the situation on the spot, these stereotypes help to form a general attitude toward mobile telephone use. These are the building blocks in the rise of new social norms. They are the formulation and rehearsal of preventative sanctions. It is as if to say "If you do this, people like me will have a poor opinion of you". This follows in the tradition of Gluckmann (1963) who notes that informal discussion and gossip can function to hold a group together.

The stereotypes or attributions used by informants included the notion of over-important status display, the pointlessness of the talk and finally, illusions as to the character of the offender. In terms of the first, it was common to hear that those who use mobile phones inappropriately were showing off or that they were some type of a status seeking yuppie (Haddon 1996a, 3).[13]

> *I know a patron at my lunch time spot who apparently hires a school kid to call him at set intervals during lunch every time he is there, lest we be (blissfully!) unaware of his being the proud owner of a mobile phone.*

The second type of attribution had to do with the non-instrumental nature of the discussions. This type of an assertion is well-grounded in commentaries from an earlier era of the telephone. When system capacity was limited and calls were expensive social talk was discouraged (Ling 1994, 13–14). Those who used the telephone for inappropriate calls were likely to be seen as gossips or wags. Similar glosses were applied to those who made open use of mobile phones. In the words of one informant.

> *One sees people standing around on the street and there is a lot of baloney that is said there. Unnecessary discussions so I don't see the point of it.*

Finally, some respondents went beyond rather neutral descriptions to the assertion that the mobile telephone was a device used to facilitate illicit activity. This is seen in the following sequence of comments.

> Person 1: *It is a fact that more jobless have a mobile telephone now than those who have a job. And a large portion of these are as a matter of fact immigrants.*
>
> Person 2: *But how is it that you say that the jobless teens and youth have mobile telephones?*
>
> Person 1: *Jobless on paper. I would not say they are.*
>
> Person 3: *No.*
>
> Person 1: *I would suggest that there are many pushers that have mobile telephones ... Criminals, real criminals.*
>
> Person 3: *Oh, that doesn't sound too good.*
>
> Person 1: *No, but it is true. I come from Oppsal and there is big time crime and there are all the worst types and they go around with mobile telephones. They call all the time and there is no let up.*

---

[13] *In a brochure published by Telenor describing the use of mobile telephones in restaurants, there is the title "It is only dumb to be a Yuppie seven years too late."*

Here the use of a mobile telephone is not just a small social *faux pas*, rather it has become a symptom of, in the speaker's mind, dangerous social trends, i.e. immigration and drug use.

Regardless of the stereotype applied to the offender, the application of a gloss is an attempt to make sense of the way in which technology is being used. The necessity for this is being pressed upon us by the growing use of these technologies. In many respects, one can not remain neutral. They must take up a position on one or the other side of the barricade. It is here that the social bond is in danger of becoming unraveled, and it is here that the most intense strategies of social maintenance need to be deployed.

4 Conclusion

In this paper, based on the examination of focus group data, I have examined how people make sense of the inappropriate use of mobile telephones, particularly in restaurants. I have looked at the social meaning of mobile phones and restaurants. In addition, I have looked at how the use of the former in the latter presents both the user and other patrons with a difficult social situation. The analysis has drawn on the work of Goffman's notion of drama and staging in order to place the analysis in a larger framework. Goffman's concepts of social boundaries, face, front stage / back stage, civil inattention, studied non-observance and sanctioning via attribution have been used.

This paper is a further examination of the ways in which technology has shifted social boundaries. Devices such as the video telephone and the Internet have meant that we need to reconsider how it is that we construct our social worlds. They have made demands on the taken-for-granted assumptions of everyday life. These developments mean that we can communicate in new ways at new times in new places. While there are unimagined possibilities there are also unimagined complications. In sorting these out we will start to see technologies' effects on power relations, gender and age differences.

5 References

Bakke, J W. 1996. Competition in mobile telephony. *Telektronikk,* 92 (1), 83–88.

Bakke, J W. Forthcoming. Technologies and interpretations : the case of the telephone. Research in science and technology studies : material culture. S. Gorenstein (ed.). *Research in science and technology studies,* vol. 10. Greenwich, CN, JAI Press.

Berg, A J. 1996. *A mobile phone is a mobile phone, is a mobile phone, is a mobile phone?* Presented at the COST 248 meeting 9.7.96, Brussels, Belgium.

Berger, P, Luckmann, T. 1967. *The social construction of reality : a treatise in the sociology of knowledge.* New York, Anchor.

Bogard, W C. 1996. *Simmel in cyberspace : Distance and strangeness in postmodern communications.* Paper presented at the Amercian Sociological Association meeting 19 Aug. 1996.

Cahill, S E. 1990. Childhood and public life : reaffirming biographical divisions. *Social Problems,* 37 (3), 390–402.

Chen, V. 1990–91. Mien Tze at the Chinese Dinner Table : A study of the interactional accomplishment of face. *Research on Language and Social Interaction,* 24, 109–140.

Clark, J. 1995. Personal discussion. Boston, MA.

Collas, S F. 1995. *Some of my best friends are ... : constructing borders and boundaries between self and other, us and them.* Presented at the 1995 Meeting of the American Sociological Association.

Duncan, H D. 1970. *Communication and the social order.* London, Oxford.

Duncan, S. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology,* 23 (2), 283–292.

Fine, G A. 1995. Wittgenstein's kitchen : Sharing meaning in restaurant work. *Theory and Society,* 24, 245–269.

Fischer, C. 1992. *America calling : a social history of the telephone to 1940.* Berkeley, CA., University of California.

Garfinkle, H. 1967. *Studies in ethnomethodology.* Oxford, Polity.

Geertz, C. 1972. Linguistic etiquette. In: *Readings in the sociology of language.* Fishman, J A (ed.). The Hague, Mouton, 282-295.

Geertz, C. 1973. *The interpretation of culture.* New York, Basic.

Giddens, A. 1984. *The constitution of society : Outline of the theory of structuration.* Berkeley, Univ. of California.

Giuffre, P A, Williams, C L. 1994. Boundary lines : labeling sexual harassment in restaurants. *Gender and society,* 8 (3), 378–401.

Glaser, B, Strauss, A. 1967. *The discovery of grounded theory.* Chicago, Aldine.

Gluckman, M. 1963 Gossip and scandal. *Current Anthropology,* 4, 307–316.

Goffman, E. 1959. *The presentation of self in everyday life.* New York, Doubleday.

Goffman, E. 1963. *Behavior in public places : note on the social organization of gatherings.* New York, Free Press.

Goffman, E. 1967. *Ritual Interaction : Essays on face-to-face behavior.* New York, Pantheon.

Goffman, E. 1974. *Frame analysis.* New York, Harper.

Gullestad, M. 1992. *The art of social relations : essays on culture, social action and everyday life in modern Norway.* Oslo, Universitetetsforlaget.

Haavio-Mannila, E, Snicker, R. 1980. Traditional sex norms and the innovative function of afternoon dancing. *Research in the Interweave of Social Roles,* 1, 221–245.

Haddon, L. 1996a. *Mobile telephony issues : discussion paper for COST 248, Mobile sub-group.* Presented at COST 248 meeting 11.04.96, University of Sussex, Brighton, UK.

Haddon, L. 1996b. *The development of the UK mobile telephone market.* Presented at the COST 248 meeting 09.07.96, Brussels, Belgium.

Haugen, I. 1983. *Tilgjengelighet og sosiale selv.* Arbeidsnotat, Institutt for samfunnsforskning, Oslo.

Jackson, T. 1952. Some variables in role conflict analysis. *Social forces,* 30, 323–327.

Ling, R. 1994. *A taxonomy of telephone conversations based on the relationship of the callers : preliminary results of a qualitative examination.* Kjeller, Telenor R&D. (TF.R 33/94.)

Ling, R. 1996. The technological definition of social boundaries : Video telephony and the constitution of group membership. *Telektronikk,* 92, (1), 61–73.

Lippman, A. 1967. Chairs as territory. *New Society,* 20, 564–6.

Lofland, J, Lofland, L H. 1984. *Analyzing social settings.* Belmont, Wadsworth.

Lohan, E M. 1996a. *The domestic phone space : the telephone space in the home and the domestic space in telephony : a participant observation case study of the telephone in three Dublin households.* COST 248.

Lohan, E M. 1996b. *Key points for COST 248 project on home and work : the incorporation of new technologies into the private sphere.* Presented at the COST 248 meeting 09.07.96, Brussels, Belgium.

Mars, G, Nicod, M. 1984. *The world of waiters.* London, George Allen & Unwin.

Martin, M. 1991. *Hello, Central? Gender, technology and culture in the formation of telephone systems.* Montreal, McGill-Queens.

Marx, G. 1994. New telecommunications technologies require new manners. *Telecommunications policy,* 18 (7), 538–551.

Meyrowitz, J. 1985. *No sense of place : The impact of electronic media on social behavior.* Oxford, New York.

Miller, H. *The presentation of self in electronic life : Goffman in the Internet.* http://www.ntu.ac.uk/ soc/psych/miller/ goffman.htm

Parker, S. 1988. Rituals of gender : a study of etiquette, public symbols and cognition. *American anthropologist,* 90 (2), 372–384.

Rombauer, I, Becker, M. 1964. *The joy of cooking.* New York, Signet.

Rutter, D. 1987. *Communicating by telephone.* Oxford, Pergamon.

Saks, H, Schegloff, E A, Jefferson, G. 1974. The simplest systematics for the organization of turn-taking for conversations. *Language,* 50 (4), 696–735.

Silverstone, R. 1996. Personal conversation. 15.07.1996, Brussels, Belgium.

Silverstone, R, Haddon, L. 1996. Design and domestication of information and communication technologies : technical change and everyday life. *Communication by design : the politics of information and communication technologies.* Mansell, R, Silverstone, R (eds.). London, Oxford.

Silverstone, R, Hirsch, E, Morley, D. 1992. Information and communication technologies and the moral economy of the household. *Consuming technologies : media and communication in domestic spaces.* Silverstone, R, Hersch, E (eds.). London, Routledge.

Simmel, G. 1971. *Georg Simmel : on individuality and social forms.* Levine, D N (ed.). Chicago, Chicago.

Spradley, J P. 1979. *The ethnographic interview.* New York, Holt, Rinehart and Winston.

Stone, A. 1994. *The Eyes of the Vampire : Desire, Bodies, and Legibility in Cyberspace.* At:Virtual Reality Oslo 1994. Oslo, Norway, 18 – 20 Aug.

Turner, J H. 1986. *The structure of social theory.* Chicago, Dorsey.

Rich Ling is a sociologist working at Telenor R&D, Kjeller. He received his Ph.D. in sociology from the University of Colorado, Boulder. Since coming to Norway he has worked at The Resource Study Group founded by Jergan Randers and has been a partner in the consulting firm Ressurskonsult. For the past five years he has worked with Telenor and has been active in researching issues associated with new information technology and society.

e-mail: richard.ling@fou.telenor.no

Status

International Research and Standardization Activities in Telecommunications

Editor: Per Hjalmar Lehne

Introduction

PER HJALMAR LEHNE

This issue of the Status section contains four papers, ranging from standardisation and research on broadband access networks, quality of service (QoS) and ETSI's 10th anniversary, 1988 – 1998.

EURESCOM stands for 'European Institute for Research and Strategic Studies in Telecommunications'. It is organised as a German GmbH, and has its headquarter in Heidelberg, Germany. Telenor is currently participating in 15 projects which are shown in the table on the right. EURESCOM originally organised only the Public Network Operators (PNOs) in Europe, but in the last year membership has been opened for all fixed and mobile network operators as well as service providers. It is interesting, though, that no new members have applied so far.

First, Mr. Umberto Ferrero presents project P614 – 'Implementation strategies for advanced access networks'. Mr. Ferrero is project leader of P614 which started in October 1996 and will end in November 1998. In his paper he explains the goals of the project and different broadband access technologies and architectures. The project aims at providing guidelines for the Public Network Operators (PNOs) for the implemenation of broadband access technologies.

The second paper addresses P806-GI – 'A common framework for QoS/Network performance in a multi-provider environment', and is written by Mr. Ola Espvik, Mr. Terje Jensen and Mr. Bjarne Helvik. Providing a defined quality of service (QoS) to the end-user has become more complicated after the liberalisation of the telecommunications market because several network operators and service providers may be involved in providing the final service. The P806-GI project tries to investigate the challenges and define a common framework for QoS upon which the PNOs can agree.

The third paper also adresses QoS, but from a user point of view. Mr. Magnus Krampell discusses the efforts done in the ETSI project TIPHON – 'Telecommunications and Internet Protocol Harmonization Over Networks'. EP TIPHON has defined so-called Quality Classes, which are used as examples of a generic quality model. Mr. Krampell's paper introduces a general methodology for a user-oriented quality model and proposes an extended set of Quality Classes.

The *European Telecommunications Standards Institute – ETSI* – was founded in 1988 from the reorganisation of CEPT (Conférence des Administration Européennes de Postes et Télécommunications). In conjunction with the General Assembly, ETSI arranged a one day seminar to celebrate its 10th anniversary under the title 'The State of the Art in European Telecommunications Standardisation'. The seminar was held in Nice, France on 25 March 1998. *Telektronikk* was present and a report from the seminar is the last paper in this issue. A strong Norwegian delegation, counting 10 people, was present.

Table 1 EURESCOM projects in which Telenor is currently participating (April 1998)

Project no	Duration	Name
P 605	Feb 96 – Feb 98	Jupiter – Joint Usability, Performability and Interoperability Tests in Europe
P 608	Aug 96 – Apr 98	TINA concepts for 3rd generation mobile systems
P 614	Oct 96 – Nov 98	Implementation strategies for advanced access networks
P 616	Nov 96 – Mar 98	Enhanced ATM Implementation Issues
P 617	Mar 96 – May 98	Evaluation of High Performance Databases for Interactive Services
P 619	Feb 96 – Mar 98	PNO – Suppliers Technical Interfaces
P 710	Feb 97 – Dec 98	Security for the TMN X-interface
P 714	Feb 97 – Apr 98	CSCW Tools for Telecommuting
P 801	Mar 98 – Jun 99	Local Information Infrastructure to Support Community Communication Networks (CCN)
P 802	Jan 98 – Jun 98	Strategic study into sustainability with respect to social & economic impacts of telecom services
P 806	Mar 98 – Sep 99	A common framework for QoS/Network performance in a multi-provider environment
P 807	Mar 98 – Mar 99	Jupiter II – Usability, performability & interoperability trials in Europe
P 813	Apr 98 – Jun 99	Technical Development and Support for European ATM Service Introduction
P 816	Mar 98 – Sep 99	Implementation frameworks for integrated wireless-optical networks

Per Hjalmar Lehne is Research Scientist at Telenor Research & Development, Kjeller. He is working in the field of personal communications, with a special interest in antennas and radio wave propagation for land mobile communications.

e-mail:
per.lehne@fou.telenor.no

EURESCOM Project P614

– Implementation strategies for advanced access networks

UMBERTO FERRERO

The access network is the portion of the telecommunication infrastructure linking every single customer to the closest local exchange: traditionally dimensioned to deliver telephony, it now calls for a thorough upgrade to accommodate new midband and broadband services, exploiting the extensive range of available technologies. EURESCOM P614 represents the third generation of access network related EURESCOM Projects, after P306 'Access network evolution and preparation for implementation' [1,2] and P413 'Optical networking' [3].

1 Introduction: what is EURESCOM P614?

The Project started in October 1996 and involves over 50 experts from 12 European Operators. P614 addresses the access network dilemma with a more holistic view, and reflects three main ideas developed during the previous Projects and lessons learned. First, the future access network will feature a number of alternative implementations and several architectures, and will extend its domain towards longer reach allowing for node consolidation: European Operators need to understand and cope with a number of architectures and systems with a single, comprehensive overall network perspective to assure the effective exploitation of the heavy investment involved. Second, the access network evolution is quickly moving towards real deployment, and co-operative projects have to focus on practical issues, such as specifications, outside plant technologies and interoperability of new and old technologies. Third, the monitoring and the contribution to standardisation bodies, together with the techno-economic appraisal of technical implementation need to be carried out on an ongoing basis by EURESCOM members, to quickly and effectively react to the changing access network scenario.

2 Broadband access networks requirement and specifications

New standardization bodies and groups of interest are sprouting every so often, and it is increasingly difficult to monitor all of them, particularly for small operators. P614 provides members with the accurate coverage of the work going on within all the relevant study groups. Moreover, sensitive topics can be jointly addressed and supported, elaborating common papers to speed up subsequent processing of established bodies. P614 members are identifying areas where standards are urgently required, with the purpose of contributing as appropriate in a co-ordinated way (Deliverable 9).

Significant cost reduction is achieved by the mass market of one (or few) solutions. EURESCOM access network related Projects identified ATM-based passive optical networks (ATM-PON) complemented by short-range high-speed copper based transmission as the preferred option. This choice, which dates back to mid 1994, facilitated the quick convergence of FSAN (Full Services Access Network) members to this solution, enabling the successful development of the specification requirements and its downstreaming to established standardisation bodies. Specific targets are selected as areas requiring further and quick study or areas showing inconsistencies. So far, the first area for common effort is the Service Node Interface requirements identification, analysing advantages and disadvantages of the different architectures incorporating such interfaces (Deliverable 1). Another critical study area is the identification of requirements for the broadband home networks (Deliverable 11).

3 Evolution towards broadband: the end of a single architecture

The dramatic technological evolution and the regulatory change-out are leading to the installation of a variety of broadband access systems. On the one hand, Operators look at target solutions: a wider time frame scenario enables the investigation of fibre rich implementations, taking advantage of potential benefits in terms of global network rationalisation and optimisation. The overall access network optimisation potentially enables significant cost savings, improved service quality and eventual integration of services for business customers, provided that specifications and standard solutions are being developed.

The target solutions encompass both ATM-PON and ATM point-to-point links; one or several evolution paths from a set of existing access networks architectures to the identified target architecture is being addressed (Deliverable 2). On the other hand, the compelling need to enter new markets and provide new services requires the full exploitation of existing infrastructures, re-using twisted pairs, coaxial cables and, under certain circumstances, even powerlines. The broad diversity of technical implementation is able to match the diversity of service offering and acceptance, existing infrastructures and other local constraints, enabling the investment optimisation (Deliverable 6). The results offer Operators the possibility to identify the relevant parameters to be considered when planning an upgrade of their own access network, understanding the capability and limits of the existing initial architectures and giving full knowledge of the established target solution (including performance and reliability figures).

4 Going wireless: broadband radio access

Broadband radio technologies are becoming more and more close to real field deployment, and are offering very promising applications for both incumbent operators and new competitors. But the big, potential, cost savings stimulate a big hype of interest, boosted by an aggressive marketing campaign. P614 reviews the radio access network architectures and technologies giving B-ISDN/ATM access to residential and business customers; a Request For Information (RFI) was issued. P614 is also reviewing ongoing research and standardisation activities and estimate the type of technologies available in the long term.

P614 highlights the truth on time for available systems and technical capabilities. Many systems are in their infancy and the wide diversity of technical implementation and regulatory constraints are making the successful exploitation of such technologies more difficult. The need for common air interface and radio-to-fibre interfaces have been identified as a key enabler for a real and extensive use of broadband radio solutions; the visions elaborated on are now being downstreamed in the appropriate bodies (Deliverable 4). A promising development is the possibility to exploit broadband radio technologies on top of fixed, fibre-rich access networks as an ideal complement to by-pass costly final drop technologies and home network related problems.

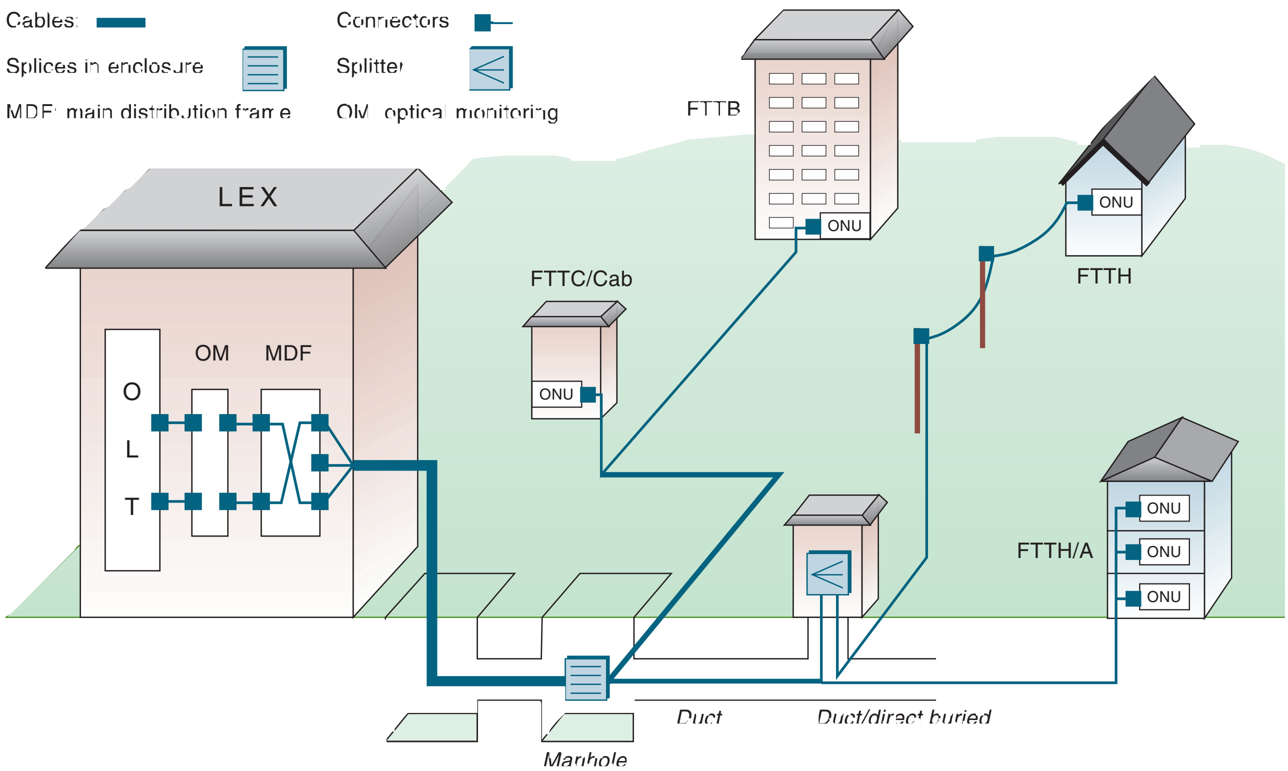


Figure 1 FTTX infrastructures

5 Basic technologies and practical implementation aspects

It is widely known that 70 % of the access network costs is due to infrastructure – how can this burden be reduced? Re-use of copper wires is well known, but other solutions include adopting innovative and relatively little known installation techniques that make it possible to re-use ducts, simplify installation, make the infrastructure more flexible, or simply adopting ultimate solutions whenever possible.

The deployment of new broadband technologies in the access network implies a number of practical implementation issues such as powering, testing, cabling and outside plant technology. P614 promotes the understanding of these technologies with the organisation of the Workshop on Access Network Cabling and Installation Techniques (ANCIT, 30-31 March 1998). P614 focuses on feasibility evaluation and cost estimates, supporting Operators choices with the necessary technological understanding (Deliverable 5).

6 Optical technologies in the access today and tomorrow

New optical low cost technologies are emerging, such as cheap point-to-point optical transmission sub-systems, allowing for several applications as drop, home cabling or alternative optical distribution network. Moreover, the increasing availability of optical fibres in the access enables the fruitful exploitation of technologies that have already proved their effectiveness in other network segments, such as wavelength multiplexing and sophisticated measurements and maintenance techniques (Deliverable 7).

FTTH has been discussed since the early 1980s. Now, with the real deployment of hybrid systems and major technological breakthrough, the exploitation of the FTTH concept becomes really viable, at least for some applications and in some environments. P614 members are willing to stimulate the cost reduction of FTTH solutions in such a way that it can be used as early as possible for avoiding intermediate evolutionary steps at least in green-field areas (Deliverable 8). The idea is to identify function limitations and Operator requirements that could make FTTH configuration cheaper, and to justify from an economic point of view a service offer corresponding to the FTTH

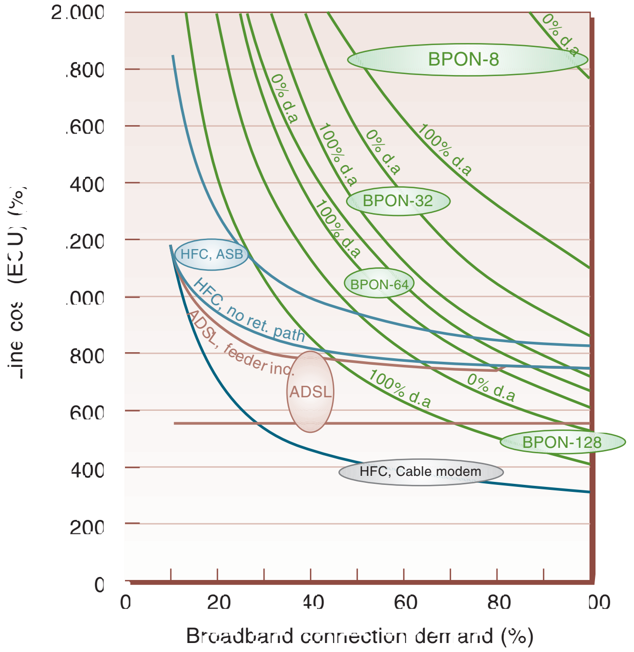


Figure 2 Line cost: 2 Mbit/s asymmetric connection [5]

configuration: a cheap broadband FTTH access might be a suitable complement to the narrowband mobile service offering (Deliverables 12, 13 and 14).

7 The price of broadband evolution

EURESCOM Projects P306 'Access network evolution and preparation for implementation' and P413 'Optical networking' elaborated on extensive techno-economic evaluations which proved very useful as commonly agreed guidelines for the definition of network evolution strategies. So, one of the P614 objectives is to identify economically viable implementation strategies for an effective broadband access network introduction, maintaining a comprehensive and updated cost evaluation of system and architectural alternatives. P614 periodically updates results evaluating new service assumptions and including new technological developments (Deliverable 3).

Marketing people tend to propose extremely low cost figures when the technology is being developed, and much higher prices when the contract is approaching ... So, from reading reports and magazines the latest technology always seems to be the cheapest, and drawbacks seem to be very minor. P614 reports bring back to earth the (easy) enthusiasm, recommending under which circumstances solutions can really be cost-effective. This activity represents an ongoing concern for EURESCOM, supporting Shareholders with independent and critically assessed estimations of the economic implications of the broadband access network deployment. Studies are made in such a way as to make the application of the results by the Operators easy, taking into account country and area specific parameters, such as social profile, service offering, civil work costs, duct availability, labour costs etc.

8 Relationship outside the EURESCOM community

P614 is becoming increasingly popular as a reference source of independent broadband access network recommendations, as witnessed by the number of hits collected by the web-site and the request for reprints and documents from operators and manufacturers all over the world: access network experts from US, Brazil, Korea, Taiwan, Indonesia and New Zealand, as well as some European suppliers and consultants, have already requested P614 reports. P614 has established links with several access network related Projects. A bilateral agreement is established with ACTS OPTIMUM (Optimized Network Architectures for Multimedia Services) Project to use the OPTIMUM tool and exchange cost estimates.

9 Dissemination of results

EURESCOM Project P614 is committed to open dissemination of results, reflected in the choice to make all the deliverables public (i.e. accessible by people outside EURESCOM Shareholders). The effort to raise consensus on the Project results is confirmed by the plan to favour co-operation with other bodies (international standardisation organisations and manufacturers) and with the preparation and running of periodical seminars and workshops. P614 presents its results at several international conferences and maintains a public web-site where a full range of information on the Project is made available; papers and Deliverables can be easily downloaded and P614 members can be contacted for further information.

10 Conclusions and future work

The Operators tend to swing between the idea of doing nothing or everything as far as broadband access network is concerned. P614 results recommend what to do (something), when to do it (with specific phasing, following the maturity of different technologies), and where (in areas of well defined technical and service characteristics).

P614 tries to spot, and tackle, some areas that have been out of focus so far, increasing the awareness of their potential and giving answers to 'hot', much discussed questions, or to counteract some 'myths' on technical capabilities and trends.

P614 is becoming a recognised independent and reliable source of information on broadband access network evolution. The next challenge can only be the actual field deployment of some of the available systems.

References

1 Warzanskyj, W, Ferrero, U. Access network evolution in Europe : a view from EURESCOM. In: *Proc. ECOC 1994, Florence, Italy,* 25–29 September 1994.

Table 1 EURESCOM P614 Deliverables

	Title	Planned issue date
D1	An analysis of the relative benefits of proposed SNI standards	April 1997
D2	Target B-ISDN access network architectures	December 1997
D3	Techno-economic analysis of major factors of B-ISDN/ATM upgrades	September 1997
D4	Opportunities for broadband radio technologies in the access network	January 1998
D5	Optical technologies for advanced access networks: early results	December 1997
D6	Evolution paths towards target B-ISDN access networks	July 1998
D7	Optical technologies for advanced access networks: final results	September 1998
D8	Elaboration of common FTTH guidelines	July 1998
D9	Contributing to standardisation bodies and other fora	April 1998
D10	Techno-economic evaluation of B-ISDN access networks architectures, scenarios and business cases	September 1998
D11	Evaluation of Broadband Home Networks for residential and small business	September 1998
D12	FTTH: Definition of the suitable powering architectures	* February 1998
D13	FTTH: Definition of access network quality and cost	* February 1998
D14	FTTH: Definition of related service offer strategies	* February 1998

* Deliverables to be produced following approval by the Board of Governors.

2 Ims, L Aa et al. Multiservice access network upgrading in Europe: a techno-economic analysis. *IEEE Communications Magazine,* December 1996.

3 Ferrero, U. Broadband optical access network : cooperative work among European PNOs. In: *Proc. of ECOC 1996, Oslo, Norway,* 15–19 September 1996.

4 Ims, L Aa et al. Key factors influencing investment strategies of broadband access network upgrades. In: *Proc. ISSLS 98, Venezia, Italy,* 22–27 March 1998

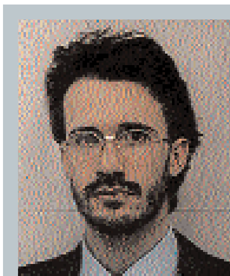
All the papers published by access network related EURESCOM Projects, together with all the P614 Deliverables (as listed in Table 1) and information on the workshops, are available from the Internet at:

http://www.cselt.it/Cselt/euresc/P614/welcome.html

Acknowledgement

This document is based on results achieved in a EURESCOM project; this does not imply that it reflects the common technical position of all the EURESCOM Shareholders/Parties. The author gratefully acknowledges the support of EURESCOM for carrying out this work. The author wishes to express a special thanks to all P614 participants, since this paper reflects the results of their work.

EURESCOM P614 members are: BT, CNET-France Telecom, CSELT, DTAG, Hungarian Telecom Company, OTE, Portugal Telecom, Swisscom, Telecom Ireland, Sonera, Telefónica I+D, Telenor.



Umberto Ferrero graduated in Telecommunication Engineering from the Politecnico di Torino in February 1992. He joined CSELT in 1991, where he now works as Senior Engineer. He is EURESCOM P614 Project Leader and he is involved in access network architecture design and economic evaluation. Ha has published more than 40 papers on access networks related topics.

e-mail: umberto.ferrero@cselt.it

EURESCOM project P806-GI:

– A Common Framework for QoS / Network Performance in a multi-Provider Environment

OLA ESPVIK, TERJE JENSEN AND BJARNE HELVIK

The multitude of operators/providers together with the wide range of services emerging in telecommunication networks, request for general descriptions dealing with quality of service issues. Such descriptions should be established between the end-user and the provider as well as between the providers. The latter is more pronounced as a number of providers could co-operate in order to fulfil a service request.

1 Introduction

It is essential to observe that the quality of a service is basically different from the attributes of the service. The quality (defined by the value of a set of quality parameters) of a service tells the customer whether an ordered service is delivered within specifications of an agreement. The point of having parameters describing Quality of Service (QoS) is that they – when given numerical values – enable the customer to have a well defined opinion of what he is paying for, and allow him to make comparisons in the market. This requires a common understanding of QoS and how it is measured.

Typically, for an Internet type of service, the customer (and service provider) has to rely on services from different network operators, to fulfil his mission. For the years to come it is foreseen that more customers will require more explicit guarantees of QoS for the Internet services. QoS is therefore becoming increasingly more important to competing public network operators (PNOs) / service providers and customers.

A majority of PNOs apply the ITU QoS framework (ITU-T Rec. E-800) [1] in their quality work. However, there are various other international organisations like ISO, ETSI etc that address the issue of QoS, each having to some extent their own collection/system of parameters. Having many collections of QoS parameters in use – some of them even being a mixture of the others together with some 'droplets' of attributes and 'personal inventions' – makes life difficult both for the customer/user and the service provider – as well as creating inefficiency in research programs addressing aspects of QoS.

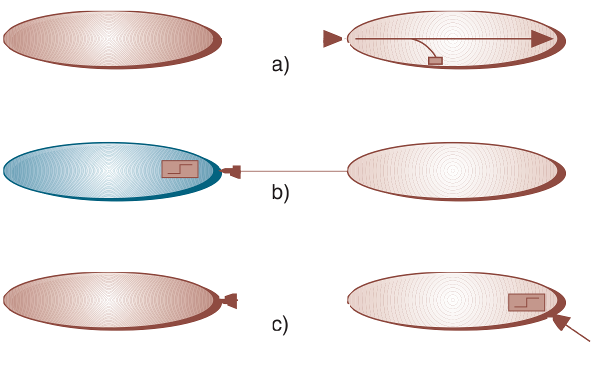


Figure 1 Interconnect agreements must consider aspects of: a) guarantee for originating part; b) protection for destinating party; c) protection for originating party

Imperative to the future is to avoid solutions to frameworks and parameters dependent upon the ever increasing number of services and actors in the telecommunication market. The EURESCOM project P806-GI: *A Common Framework for QoS/Network Performance in a multi-Provider Environment* is going to investigate the challenges of QoS in a multi-provider service provision with the aim of reaching an agreement among the EURESCOM shareholders.

The challenges are further elaborated in Section 2. Section 3 then deals with existing frameworks and approaches, prior to outlining the steps undertaken in Section 4.

2 The multi-provider service agreement interface challenges

Operating telecommunication networks have long traditions for interworking. That is, it should be possible to transfer information between users in networks managed by different organisations. Interconnect agreements must be described correspondingly. Traditionally, this is done among a few actors of the same kind (the national PNOs) and where 'the QoS impairment budget' is subdivided between the actors according to ITU standardised reference connections.

Considering the increasing number of actors as well as the different types of actors, (eg. platform operators, access providers, transport provider, content hosts and users/customers) and the corresponding identification of network domains, multiple combinations of competition and co-operation are expected. Further emphasis on regulations concerning interconnect, in particular looking at dominating operators, is also anticipated.

It is foreseen that QoS/performance agreements should be established related to most of the interfaces between actors. In particular, any requirements for unbundling the network elements and functional groups could lead to a considerable number of non-traditional interfaces.

Ensuring availability and successful calls/information transfers originated in one network domain and destined for other domains is also an issue. In particular, the users may require explanations and possible compensations by the operator/provider dealing with the 'originating' side (of which he is a customer). Therefore, interconnect agreements have to incorporate means that the relevant operator/provider can apply in order to avoid deteriorating situations. That is, on the transfer/call level both outgoing and incoming situations must be considered and the views of both parties of an interface must be taken on. This implies that the contents of such agreements should adequately cover issues arising for the purpose of protection and guarantee between actors, see Figure 1. Guarantee issues include statements from relevant actors promising a sufficient high level of QoS such that the end-users are believed to be satisfied. Protection issues include precautions against receiving traffic flows severely reducing the QoS within the domain considered. The protection aspects may, again, be seen from more parties; either these could be stated in order to limit the load from neighbouring domains (as seen from the left domain in Figure 1 b)) or these could be stated in order to avoid degradation of QoS due to load generated from connected subscribers (as seen from the right domain in Figure 1 c)).

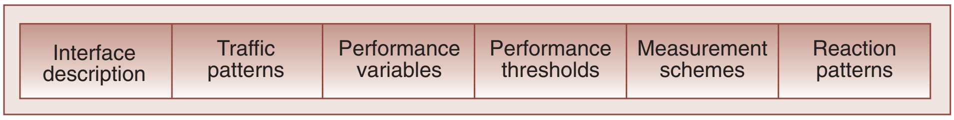


Figure 2 Possible structure of an interconnect agreement

An agreement should include aspects like the conditions for operation, how to assess these conditions and actions to be undertaken in case agreed levels on any of the conditions are exceeded. Naturally, the particularities on interfaces have to be considered when describing these aspects. For instance, delays may be more essential on packet switched relationships while blocking probabilities are used on circuit switched connections. In the general case, variables describing delays as well as variables describing blocking are relevant.

The approach to be followed may begin by describing the potential roles present in the provisioning of services. This may also look at the different phases of the service life cycle. Contributions from regulatory bodies should also be taken into account as a motivation for identifying interfaces between actors.

Such interfaces may be located between horizontally separated domains (domains with similar functionality) and between vertically separated strata (strata where the lower one provides certain services for the one above). That is, horizontal means interconnecting network domains at the same functional level, eg., between two exchanges. Vertical means connecting different levels in a functional view or network architectural view, the interface between a content host and a platform operator, or between a platform operator and a transport provider. An example of the latter, related to Intelligent Networks, is connections between Service Control Points and Service Data Points. Functional relationships would exist between the different network elements, like when service logic/data are executed on a platform provided by another actor. That is, an interface may be more diffuse than a physical link between network elements. Naturally, when interfaces are incorporated within an element, assessing the conditions may become more involved unless external equipment can be used.

In addition, an actor would also like to know the characteristics of the loads at the ingress points. For instance, load with higher variability or having severe correlation may influence the network such that the performance is significantly degraded compared to situations where such effects are not present. Therefore, descriptions of traffic characteristics should also be given. Alternatively, counteracting measures could be taken if the offered traffic load deviates from the characteristics. It might be tempting to either avoid such situations or make the relevant traffic flow confirm with the better behaving characteristics (eg., by applying shaping). Activating management mechanisms, like call gapping, in order to smoothen the impact on the network conditions would lower the QoS for the service concerned. Therefore, the way management procedures affect QoS must be taken into account. Naturally, if the situation is such that no significant reduction of performance is foreseen, the traffic flow could be treated as it is. This may, in one way, seem similar to a 'best effort' manner for treating the traffic load and could be stated as such in a contract. In that way the thresholds in the agreement may be regarded as minimum values which are commonly met and where better values can be found when the network state allows for it. However, multiple thresholds could be given where different reaction patterns are described for each interval.

Based on the description of roles and interfaces, a framework for interconnect agreements related to QoS is defined. Other aspects, like legal, economical, technical protocol specific etc, should also be considered; however, these are not treated in P806-GI. In principle, the contents of an agreement looked at may include, ref. Figure 2:

- Interface description
- Traffic patterns expected (with relevant parameters in order to characterise the traffic flows)
- QoS parameters and thresholds
- Evaluation schemes including procedures for carrying out measurements
- Reaction patterns.

Varying aspects of these topics may be pronounced for the different interfaces. For this project we point out the need of a general framework to be applied. As part of the evaluation procedures, exchange of data between interconnected actors will be referenced.

Naturally, the network operator will usually be the only instance having a complete view of the network state. In order to limit any disturbances following a network condition, it is a sound principle to choke the relevant traffic flows at the edges of the network. Considering interconnect, this means that suitable mechanisms should be present in the network elements associated with the ingress points.

Often, having described the conditions to appear on the interface, belonging measurement schemes are identified. Values for variables of the arrival processes, service mixture and the resulting performance are to be captured. As for every measurement, decisions of when, where and what to measure must be made. That is, topics like time and duration, interface/location and events have to be specified. These are measurements that may be carried out by both parties of an interface. It must also be decided whether or not continuous measurements are to be performed. As measuring could be regarded as sampling, it is also to be agreed upon if terms in a contract can be questioned based on a single measurement period or if a number of measurement periods have to be done of which several indicate that the terms can be questioned before the contact is renegotiated or other means are applied. This is also seen as a trade-off between the time for reactions (responsiveness of a scheme) and the effort needed for preparing and carrying out the reactions.

In addition, measures treating sudden changes in the arrival processes must be present. Load control is an example of such quick response measures. Which measures to apply for an interface should also be stated in the interconnect contract. A number of measures, operating on a range of time scales, may be thought of.

Load control implying rejecting or delaying traffic is considered as a feature utilised during operation. One of the purposes may be to avoid that a single group of services/call types seizes too large a fraction of the capacity leading to degradation of the quality for other services/call types. This may be particularly relevant when mass calling services are introduced.

Another example where load control between operators may be requested is when one operator chooses to reroute traffic through network domains[1] that are not prepared for this. This may happen when the more direct connection is not available, eg. due to link failure. Although accounting rules may treat this situation by introducing financial compensations, using measures for not degrading the QoS for the remaining services could be more fruitful in order to keep one's reputation.

3 Existing QoS frameworks and approaches

Various international bodies have made efforts towards making general QoS frameworks, eg. see [4]. Some of these are:

- ITU-T E-800 where part of this standard is adopted by IEC as terminology standard IEV 191
- *ETSI framework,* [2]. This framework is based on the work of the FITCE Study Commission [3]
- The *service failure concept* of EURESCOM P307. The attributes of a service failure is regarded independent of the service it affects and its cause in the network [7]
- *QoS in layered and distributed architectures*
  - The ISO/OSI QoS framework [6]
  - The Telecommunication Information Networking Architecture Consortium (TINA-C) QoS framework [14]
- *ETNO Working Group 07/95 on QoS.* This group is working toward a consistently defined set of common European QoS parameters (QoS indicators). The aim is harmonised European QoS definitions and possibly performance targets for pan-European services, in order to facilitate comparison of the results of the measurements. The work is based on the approach of the FITCE Study Commission and ETSI. Up to now, the work has concentrated on voice telephony.

In addition, EURESCOM P603 [8] has studied the bearer service part of some actual QoS frameworks. It was concluded, [4], that the frameworks were partially overlapping but essentially supplementary, since they focus on different aspects. Although concepts may be different, and there are considerable differences in terminology, a unification seems feasible. None of these are primarily intended as a basis for measurements, nor are they intended for a quantitative comparison of QoS aspects across different types of bearer networks.

Aagesen [5][12][13] made a thorough discussion on QoS in open distributed processing (ODP) systems. One conclusion is that there are no fundamental contradictions between the OSI, ITU-ISDN, ODP, and TINA QoS frameworks. Common features are: i) objects offering services and performing QoS handling functions, ii) QoS parameters/attributes, and iii) contracts between objects. However, the focus and the use of concepts differ.

A brief description of a collection of EURESCOM projects having addressed or QoS/NP aspects is given in [11].

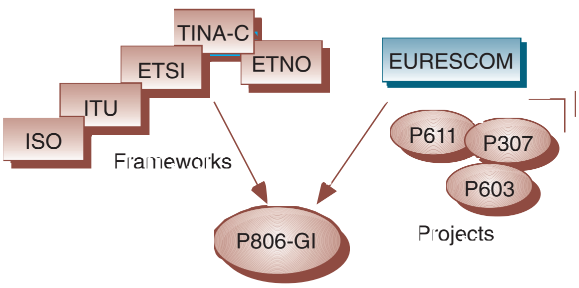


Figure 3 Input and basis for elaborating generic framework for multi-provider environments

4 How to proceed

From the studies so far is seems feasible to reach a common understanding as regards a first common framework applicable in a multi-provider environment, where the providers may operate in different

- domains and/or
- strata.

The most important issue is to reach a core structure and minimum set of parameters that the various actors can agree upon. Based upon the experiences gained in P307 [15] and P619 [9] together with the recommendations of P611 [10], see Figure 3, EURESCOM – in its project P806-GI during 1998–1999 – addresses the following topics:

- A preliminary common end-user – provider QoS interface is to be established. This includes a well defined mapping of the end-user's terminology/parameters into 'top level' parameters of a common harmonised QoS/network performance framework to be chosen/constructed. A QoS agreement between end-user and provider is to be structured and defined. The interface, framework and agreement should be independent of specific services and network/provision technologies. The usefulness of the framework in connection with service degradation analysis is also observed.
- After having investigated the above mentioned scheme in an IN and Internet scenario, a generic service structure, customer – provider QoS interface, QoS parameter structure, and structure for service level agreement applicable at the interface between provider(s) and end-user will be proposed. A Memorandum of Understanding (MoU) related to the structure and definitions of the previous objective and items is also to be prepared.
- Including the experience gathered in EURESCOM projects P307 and P619 as regards data dissemination and common targets, adequate procedures for collecting and exchanging QoS data in a multi-provider environment will be investigated and structured.

P806-GI is organised as three main tasks. In the first of these the common framework is to be elaborated. As previously described, this is done in two phases. After the first phase, in the

[1] *Under the control of one or several other operators.*

second task, the framework is to be applied for IN and Internet configurations. For these applications the structure of interconnect agreements as depicted in Figure 2 is to be examined together with the framework. A third task deals with management of measurement and data handling. During the second phase of the first task a (possible) refined framework is described together with a draft MoU.

The project started in March 1998 and is scheduled to end in September 1999. Currently, there are nine PNOs involved, sharing the 100 man-months allocated. Two workshops/seminars are planned, the first one around November 1998 and the second one around October 1999. Three deliverables are to be issued, edited by the three main tasks.

5 Conclusion

The public network operators have for a number of years used ITU-T E.800 as their conceptual framework with respect to QoS and network performance. In addition, the ETSI and ETNO frameworks are in use. Although important aspects of existing frameworks are still valid, the challenge is to agree upon a common harmonised QoS framework sufficient in an ODP/multi-provision handling of services. The EURESCOM project P806-GI suggests to take an initiative towards a more widely applicable QoS framework aiming at a consensus among all major organisations and bodies dealing with QoS issues. In the work outlined above, the project P806-GI team will also communicate with a wide range of actors, both from the provider side and from the customer side.

6 References

1 ITU-T. *Terms and definitions related to Quality of Service and Network Performance including Dependability.* ITU-T Recommendation E.800, August 1994.

2 ETSI TC-NA. *Network Aspects : general aspects of Quality of Service and Network Performance.* ETSI Technical Report ETR 003 (ref. RTR/NA-042102), October 1994.

3 FITCE Study Commission. *The study of Network Performance considering Customer Requirements, Final Report,* 1993.

4 Helvik, B. Elements of a Measurement Framework for a Bearer Services. *Telektronikk,* 93, (1), 7–25, 1997.

5 Aagesen, F A. QoS frameworks for open distributed processing systems. *Telektronikk,* 93, (1), 26–41, 1997.

6 ISO/IEC JTC1-SC21. *Quality of Service Framework.* Working Draft No. 2, SC21-WG1 meeting, Yokohama, 1993.

7 EURESCOM. Project P307 Reliability Engineering; Deliverable D3. *An approach to dependability modelling of telecommunication networks and services,* November 1995.

8 EURESCOM. Project P603 Quality of Service : Measurement Method Selection; Deliverable D1. *QoS overview, services and measurement methods-selected services and methods,* October 1997.

9 EURESCOM. Project P619 PNO-Supplier Technical Interfaces; Deliverable D6. *Recommendations to Improve the Technical Interfaces between PNO and Suppliers,* 1998.

10 EURESCOM. Project P611 Overall Strategic Studies; Task 3.11 Report. *Quality of Service and Network Performance,* 1997.

11 Groen, H et al. EURESCOM and QoS/NP related Projects. *Telektronikk,* 93, (1), 184–195, 1997.

12 Aagesen, F A. *Quality of Service in layered architecture.* Trondheim, SINTEF, 1994. (SINTEF report no. STF40 F94088.)

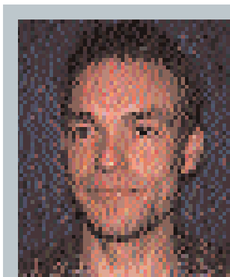
13 Aagesen, F A. Architectures for the modelling of QoS functionality, *Telektronikk,* 91, (2/3), 56–68, 1995.

14 TINA-C (Telecommunication Information Networking Architecture Consortium). *Quality of Service framework.* Doc. no. TR_MRK.001_1.0_94, TINA consortium, 1994.

15 EURESCOM. Project P307 Reliability Engineering; Deliverable D5. *European Quality Measurements,* August 1995.

Ola Espvik is Senior Research Scientist and editor of Telektronikk. He has been with Telenor R&D since 1970 doing research in traffic engineering, simulation, reliability and measurements. His present research focus is on Quality of Service and Quality Assurance mainly related to various EURESOMC projects.

e-mail: ola.espvik@fou.telenor.no



Terje Jensen is Senior Research Scientist at Telenor R&D, Kjeller. He is mainly engaged in activities related to performance analysis, dimensioning and service quality for a number of networks and systems. He is Task Leader of P806-GI.

e-mail: terje.jensen@fou.telenor.no

Bjarne Helvik, Dr. Techn., is Professor at the Norwegian University of Science and Technology (NTNU), Department of Telematics. His fields of interest are QoS, Dependability (modelling, analysis, fault-tolerance, fault-management), Teletraffic (modelling, analysis, simulation, measurements) and Telematics system architecture.

e-mail: Bjarne.E.Helvik@item.ntnu.no

The ETSI project TIPHON as a way towards a Harmonised User-Oriented Quality Model?

MAGNUS KRAMPELL

Quality of Service (QoS) is an ambiguous term. Often it is regarded as equivalent to reserved/guaranteed bandwidth. This reflects a technology oriented view of telecommunications and data networks. The end users often knows nothing about bandwidth, jitter, packet loss and other terms which they are confronted with.

What the average user would like to discuss is *Quality,* using terms which they are familiar with. A generic Quality Model is thus very much needed. Especially when considering a future scenario when users daily need to assess and control the Quality of the abundance of services available.

The Quality Classes defined in the ETSI project TIPHON are used as a good example. This paper introduces a generalised methodology to define a user-oriented Quality Model and also proposes an extended set of Quality Classes.

1 Introduction

1.1 Definitions of QoS and Service Quality

QoS is a term that is often misused. The ITU-T Rec. E.800 [1] defines QoS as: *"The collective effect of service performance which determines the degree of satisfaction of a user of the service"*.

This definition can be understood as covering *all* aspects of service provisioning that can affect the users' perception of Quality, eg. provision, repair, billing, help desk, etc.

However, the term QoS is sometimes being used for describing the ATM transfer capabilities (eg. CBR, VBR, UBR, ABR) [2]. QoS is also sometimes used for describing the Resource Reservation Protocol for IP, RSVP [3].

Apparently, there is a need to structure the discussion on QoS and make a separation between what the user perceives as Quality, what the user's application implements (and the requirements it puts to the network) and what the network implements. (The latter is often referred to as Network Performance.)

The ETSI document ETR 003, 'General aspects of QoS and Network performance' [4] gives generic and particular definitions of QoS, Network Performance and their interrelationships. The document describes a method for the management of network performance related to QoS and it describes a generic framework for capturing the user's QoS requirements, mapping them to QoS service performance parameters, and consequently to Network Performance parameters. The process of managing the feed-back which results in the QoS perceived is also described.

This paper discusses a simpler methodology with a similar objective but focusing on the Information transfer phase (as defined in [10]) of a service only. The method can also be used to define Quality Classes for other phases.

1.1.1 Components of QoS

QoS depends on:

- the Support, ie. the secondary services such as information, provisioning, billing, account management, etc.
- the Security, ie. the security aspects related to both network/services and customers' expectations
- the Operability, ie. the ease of use, etc.
- the Serviceability, ie. the access and retention of the service and transmission aspects, etc.

The user of a service generally identifies two 'bodies' when using a service:

- the Administration, ie. the 'human related aspects'
- the Network, ie. the 'technical related aspects'.

The 'human related' aspects are concerned with 'Support' and 'Security', while 'technical related' aspects are mostly concerned with 'Operability' and 'Serviceability' (including Accessibility, Retainability and Integrity (transmission)) (from [1] and [5]). We will in this paper concentrate on Operability and Serviceability.

1.1.2 User-oriented concepts of Service Quality

The concepts presently provided by the standards bodies are not suitable for end-users. For example, the ETSI ETR 138 [6] defines 18 parameters (for PSTN and ISDN only!). The number of possible combinations of these is far beyond what a user can handle (even if the individual parameters themselves were understandable to the average user!).

When users use (or are about to use) a service, it would be beneficial if the 'Quality' (user perceived Quality related to network performance) were selected from a limited set of Quality Classes, all well known to the user (this implies the need for a generic Quality Model where the same Quality Classes apply to all services). What is needed is:

- A small set of Quality Classes, (7–9 is seen as a target number. This would allow users to keep a good understanding of all Classes).
- Quality Classes that relate to what the users perceive when they use a service.
- Quality Classes defined in terms of parameters that have a value domain understandable by the user.

The parameters should relate to human 'limitations' (which change slowly, if at all ...) rather than to technological development (which changes rapidly). Some examples are *Delay* (ETSI ETR 250 [9] contains a division of delays into sub-domains based on what humans can distinguish), *Lip-synch* (how well video and sound must be synchronised), *video resolution*, and *video frame rate*.

Of course, there may also be other parameters (eg. *bit-error rate*) to cover other aspects (eg. correctness).

1.1.3 Provider-oriented concepts of Service Quality

The provider is in a different situation. For the provider it is vital that the Quality provided can be measured, so that it can be proven that the Quality agreed with the customer was provided. Parameters defined by standard bodies do fit the providers well.

What is needed is thus, apart from the user-oriented Quality Classes, a mapping to provider (or network) oriented parameters.

1.2 Problems of having many models

A likely future scenario is that users will have a multitude of services and would prefer *one* Quality Model that applies to all services rather than being faced with a different model for each service (or provider).

Different models for different services (and providers) would mean that users would have to remember the context in which a certain Quality Class is used. In the worst case, one Quality Class may mean different things, depending on where it is used!

Obviously, users would be the losing party in such a situation.

1.3 Why is a harmonised user-oriented model needed?

In PSTN the goal of the network planning used to be to provide the highest possible Quality within the cost limits set up. This was helped by the fact that most Public Network Operators enjoyed a monopoly situation.

The present situation with the opening up of the market for telecommunications warrants a different strategy. Customers will be provided with what they pay for, no more and no less. This will shift the responsibility to the customer – what are they really paying for?

Advanced users (mostly big companies) will always be able to define their needs and negotiate the relevant Quality required. In fact, most operators already have SLAs (Service Level Agreements) with their major customers.

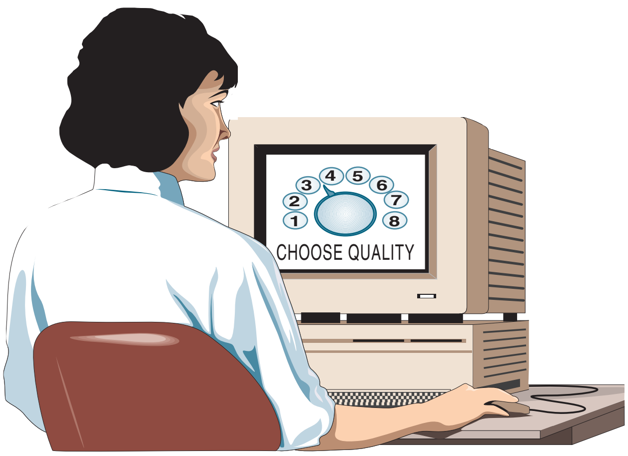


Figure 1 Would it not be convenient if this question was easily answered?

However, the average residential user and SME (Small and Medium Enterprise) users cannot afford to have the competence and equipment needed to define and negotiate their needs.

If a commonly known Quality Model existed, the users could select a Quality Class, confident that they know what to expect. The provider would also know what is required, since measurement methods and best practices would soon be created for the (relatively few) Quality Classes concerned. For each new service that is introduced, the provider would have to define how the different Quality Classes would be supported. Compare this to the ad-hoc situation (largely existing today) where each provider defines the Quality independently for each service!

Table 1 Proposed model from the ETSI TIPHON Project
(NB. All delay parameters represent an upper bound for 90 % of the connections over the TIPHON System)
The exact look of the table is still under discussion as this is written

		4 (Best)	**3 (High)**	**2 (Medium)**	**1 (Best Effort)**
Speech Quality (one way, non-interactive measurement)		Equivalent or better than G.711 for all types of signals	Equivalent or better than G.726 for all types of signals	Equivalent or better than GSM-FR for all types of signals	
End-to-End Delay (mouth to ear)		< 150 ms	< 250 ms	< 450 ms	
Call Set-up time	Direct IP	< 1.5 s	< 4 s	< 7 s	
	E.164 Number	< 2 s	< 5 s	< 10 s	
	E.164 Number via clearing house or roaming	< 3 s	< 8 s	< 15 s	
	Email alias lookup	< 4 s	< 13 s	< 25 s	

So, the successful introduction of a Quality Model with a limited number of Quality Classes would lead to:

- Increased customer satisfaction (since users would know what they agree to and may have a way to verify that they get what was promised)
- Saved cost for the providers (since a Quality assurance model with measurements etc. that is reusable can be developed. Also, there would be thousands of customers per Quality Class rather than individual settings – this would make planning easier).

2 Existing user-oriented Quality Models

Some attempts at creating user-oriented Quality Models exist. Two such models are presented here.

2.1 ETSI Project TIPHON as an example

The ETSI project TIPHON [7] has proposed four (4) Quality Classes for voice services (over IP), see Table 1.

It can be noted from the table that TIPHON has chosen only a few 'technical' parameters as affecting the Quality perceived by the user. Other parameters may be important, but do not affect Quality directly.

2.2 Other examples

The Multimedia Communications Forum (MMCF) has proposed three (3) Quality Classes for multimedia applications [8], see Table 2.

The MMCF has chosen to define their Quality Classes using more parameters. The disadvantage of this approach is that the Quality Classes are less 'service independent'.

3 How can a harmonised user-oriented Quality Model be reached?

There are similarities between the models described above. However, few Generic Quality Models exist that can be applied to *any* service. [11] contains one such model; it does not, however, contain any concrete Quality Classes.

3.1 A methodology

This paper proposes a methodology for developing a user-oriented Quality Model that can be applied to any service. The methodology presented has been applied (partly) in the ETSI TIPHON project [7] presented above.

The methodology is based on the following steps, that are iterated until a reasonably stable set of Quality Classes has been reached (see also Figure 2):

1 Create a table.

2 Find relevant Network Parameters. Make these rows in the table.

3 Find the value domain of each Network Parameter.

4 Divide each Network Parameter value domain into a number of discrete sub-domains. The division into sub-domains should be based on human aspects, eg. reaction times.

Table 2 The model proposed by MMCF

QoS Parameter	QoS Class 3	QoS Class 2	QoS Class 1
Audio transfer delay with echo control	< 150 ms	< 400 ms	< 400 ms
Audio frequency range	> 0.05 – 6.8 kHz	> 0.3 – 3.4 kHz	> 0.3 – 3.4 kHz
Audio level (typical)	– 20 dBm0	– 20 dBm0	– 20 dBm0
Audio error free interval	> 30 min.	> 15 min.	> 5 min.
Video transfer delay	< 250 ms	< 600 ms	< 10 s (still image only)
Video/audio differential delay	> 150 and < 100 ms	> 400 and < 200 ms	N/A
Video frame rate	> 25 frames/s	> 5 frames/s	N/A
Video resolution	> 352 x 288	> 176 x 144	N/A
Video error free interval	> 30 min.	> 15 min.	N/A
Audio differential delay	< 100 ms	< 200 ms	< 1s
Error free interval	> 30 min.	> 15 min.	> 5 min.
Data rate	> 500 kb/s	> 50 kb/s	> 5 kb/s

5 Define 7–9 reasonably disjoint user-oriented Quality Classes. Make these columns in the table. These are really ‘profiles’, and mnemonic names are recommended to make them more tangible.

6 Define a mapping between each Quality Class and the different Network Parameter value sub-domains by grouping value sets from different Network Parameters together. The mapping of each Quality Class to Network Parameters is seen when looking down the corresponding column. (Several sub-domains may be placed in one Quality Class if needed, see Figure 2.)

3.2 Applying the methodology

In order for the results to be generic, the methodology should be applied in a service-independent way. The assumption here is that all application specific data is at some point transported as bits. The differentiating factors can thus be expressed in terms of bandwidth, bit error rate, delay, jitter, etc. This is the kind of Network Parameters foreseen.

The value domains of such Network Parameters are often well defined. Concerning the division of the Network Parameter value domains into sub-domains there is much material in the Psychology and Sociology disciplines that should be helpful. A reasonable set of Quality Classes should not be very difficult to define and agree on. (This step should, however, include discussions with users and user groups.) The difficult part is deciding which Network Parameter sub-domains to group together in the different Quality Classes. Several iterations and trial implementations may be needed.

Table 3 Segmentation of Quality classes

	Low Importance	High Importance
High Urgency	Quick but not necessarily Safe	Quick and Safe
Low Urgency	Best Effort type	Safe but not necessarily Quick

Table 4 A combined table with 7 classes

	Low Importance		High Importance	
High Urgency	Best	3	Best	7
High Urgency	High		High	
High Urgency	Medium	2	Medium	6
High Urgency	Best Effort		Best Effort	
Low Urgency	Best	1	Best	5
Low Urgency	High		High	
Low Urgency	Medium		Medium	4
Low Urgency	Best Effort		Best Effort	

4 Possible extensions to the TIPHON model

The four Quality Classes proposed by TIPHON (Best, High, Medium and Best Effort) may be generic enough to be applicable to many services. However, sometimes you need to indicate how Quality should be changed. Table 3 shows a possible way to segment the requirements into Importance and Urgency.

For some applications the choice of segment could be automatic (eg. Voice Over IP favours Urgency before Importance; File transfer using FTP would favour Importance before Urgency). In other cases the user will have to give some indication on how Important/Urgent the data being transferred is (eg. when browsing the web, it is sometimes desirable to get the web page quickly just to see if it contains anything important; in other cases the web page contains critical information that must not be lost).

The application of the four segments to the four Quality classes (from TIPHON) would lead to 16 Classes, see Table 4.

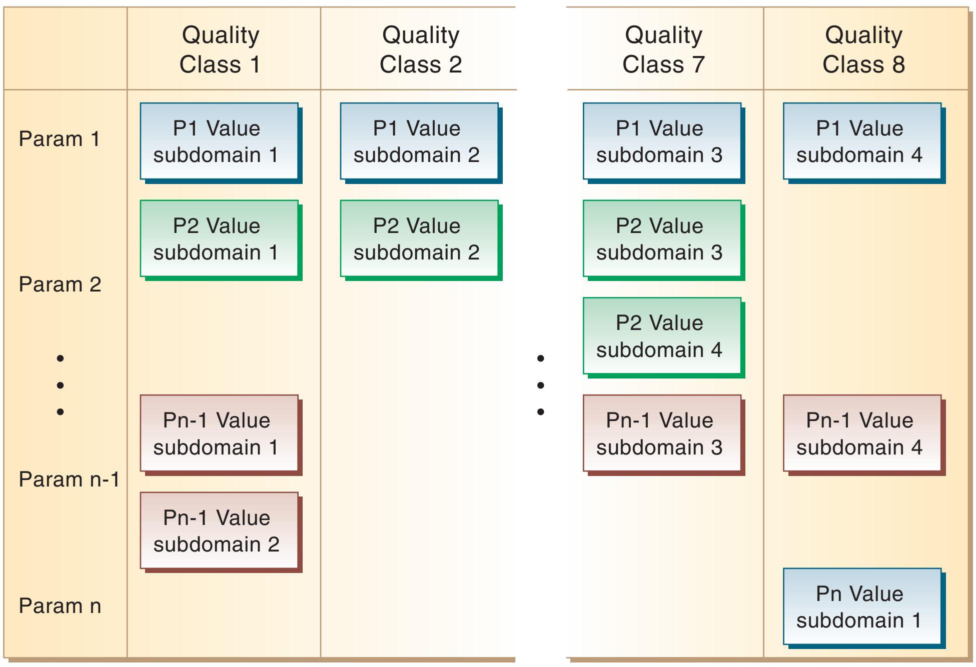


Figure 2 The principles of the methodology presented

If some of the resulting 16 classes are combined, we could end up with 7 Classes (see also Table 4):

1 Best-Effort	'Best Effort'
2 Quick-but-not-so-Safe	'xxx'
3 Quickest-but-not-so-Safe	'yyy'
4 Safe-but-not-so-Quick	'zzz'
5 Safest-but-not-so-Quick	'mmm'
6 Quick-and-Safe	'nnn'
7 Quickest-and-Safest	'Best'

It remains to put better mnemonics to the Classes (the xxx, yyy, etc.) in order to make them easily understandable and easy to remember. It also remains to be tested whether the serialisation used here is suitable (another possible sequence would be 1-4-5-2-3-6-7). Finally, it remains to try the Quality Classes out on a set of services and see whether relevant mappings to network parameters can be made.

5 Conclusions

This paper points out the irregular use of the term Quality of Service (QoS) and identifies the need for a harmonised Quality Model which is understandable to users and at the same time mappable to Network Parameters that providers use to ensure the Quality delivered. (Using Quality Classes does in no way restrict the provider from offering a fine granularity of Quality levels if it so chooses.)

The ETSI Project TIPHON is used as an example of an attempt to create such a model. The proposed Quality Classes are generic and easily understood. However, they are perhaps too generic to be applicable to all kinds of services.

The TIPHON Quality Model is also used as a basis for some initial ideas on how the model can be extended.

6 References

1 ITU. *Terms and Definitions related to Quality of Service and Network Performance including Dependability.* ITU-T Recommendation E.800, Geneva 1994.

2 EURESCOM. *European switched VC-based ATM Network Studies – stage 2, Deliverable 1 : Specifications, Volume 1 of 11 : Outline of the Specification.* EURESCOM Project P515, www.eurescom.de.

3 IETF draft-ietf-rsvp-spec-11, *Resource Reservation Protocol, RSVP – Version 1, Functional Specification,* February, 1996.

4 ETSI. *Network Aspects (NA) : General Aspects of Quality of Service (QoS) and Network Performance (NP).* ETSI Technical Report ETR 003, 1994.

5 EURESCOM. *Quality of Service : measurement method selection, Deliverable D1 : QoS overview, services and measurement methods – selected services and methods.* EURESCOM Project P603, www.eurescom.de, 1997.

6 ETSI. *Network Aspects (NA) : Quality of Service indicators for Open Network Provisioning (ONP) of voice telephony and Integrated Services Digital Network (ISDN).* ETSI Technical Report ETR 138, 1994.

7 ETSI Project TIPHON, Working Group 5, DTR/TIPHON-05001, version 1.2.2, www.etsi.fr, 1998.

8 Multimedia Communications Forum (MMCF). *Multimedia Communications Quality of Service, Part I : Framework; Part II : Multimedia Desktop Collaboration Requirements.* 1995 (www.mmcf.org).

9 ETSI. *Transmission and Multiplexing (TM); Speech communication quality from mouth to ear for 3.1 kHz handset telephony across networks.* ETSI Technical Report ETR 250, 1996.

10 ITU-T. *Integrated Services Digital Network (ISDN). Overall network aspects and functions. General aspects of Quality of Service and Network Performance in digital networks, including ISDNs.* ITU-T recommendation I.350, Geneva, March 1993.

11 Oodan, A, Ward, K, Mullee, T. *Quality of service in telecommunications.* IEE Publishing (UK), 1998. ISBN 0-85296-919-8.

Magnus Krampell is Project Supervisor at EURESCOM GmbH , supervising projects in the areas of IP/Internet and QoS. He is presently on leave from Telia AB, where he has been employed since 1990. Experiences include: Development of TMN systems, TINA work at Bellcore in USA and research and development of broadband systems and services. Mr. Krampell received his B.Sc in Mathematics in 1984 and his Lic.Sc in Computer Science in 1988.
e-mail: krampell@eurescom.de

European Telecommunications Standardisation

– ETSI’s 10th Anniversary 1998

PER HJALMAR LEHNE

The *European Telecommunications Standards Institute – ETSI* was established in 1988. The first General Assembly (GA) was held in March the same year. On the occasion of the 10th Anniversary, a special one-day conference was held in Nice, France on 25 March 1998 in conjunction with the General Assembly. The conference subtitle was raising expectations as to the contents: “The State of the Art in European Telecommunications Standardization”.

The conference was opened with a speech on ETSI’s history and role by Dr. Karl Heinz Rosenbrock, Director of ETSI since 1990. Then there were seven speeches on specific topics:

- Universal Mobile Telecommunications System (UMTS)
- Broadband Radio Access Networks (BRAN)
- Terrestrial Trunked Radio Access (TETRA)
- Voice Telephony on the Internet (TIPHON)
- Multimedia Applications & Communications
- European ATM Services Interoperability (EASI)
- Telecommunications Security.

The seminar was closed by the chairman of the General Assembly, Antonio Castillo, who summed up the status in one sentence: *“What is state of the art?”* The answer was: *“Everything is changing.”*

The seminar was clearly divided into two parts, one covering radio based communications and the other fixed network services, systems and applications. Thus the ‘hottest’ subjects were covered: Mobile communications, Internet and Multimedia.

Table 1 Some facts about ETSI as per April 1998

- 485 full membership organisations
- Members from 34 European countries
- 81 observing organisations
- 48 associated member organisations
- 1471 Standards (ETS) produced since 1988
- 432 Technical Reports (ETR)

Introduction

Dr. *Karl Heinz Rosenbrock* has been Director of ETSI since 1990. He made the opening speech by summarizing the story of European telecommunications standardisation since the founding of CEPT in 1959. The title of the speech reflected his views on the importance of ETSI: “ETSI’s role in standardisation – the past, the present & the future”. He summed up the facts in numbers, which are shown in Table 1. ETSI’s technical organisation as per October 1997 is shown in Figure 1.

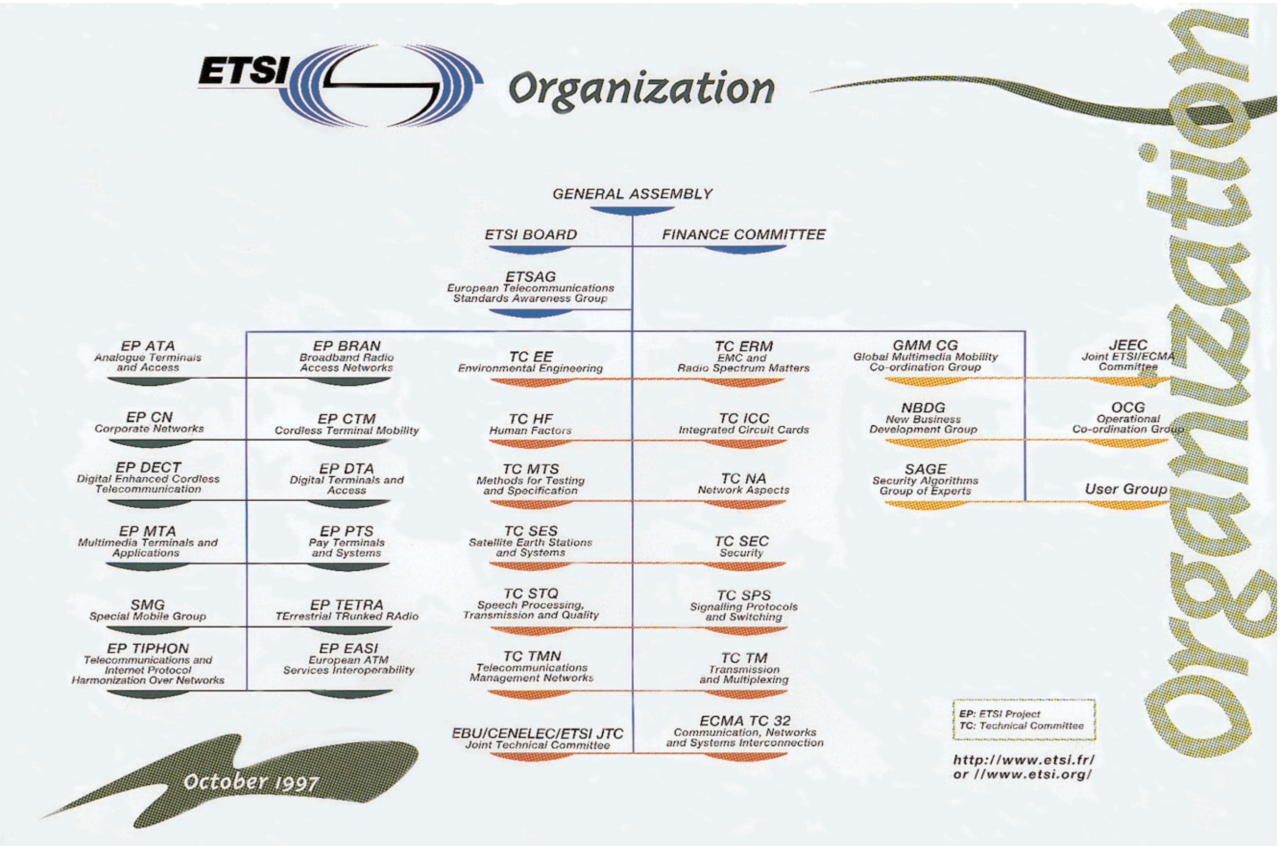


Figure 1 ETSI technical organisation as per October 1997

Table 2 *Change of focus for UMTS [1]*

	"Old" UMTS	"New" UMTS
Core ideas	Integration of all existing and new services into one universal network	1) Focus on innovative new services 2) Support of GSM services
Partner-networks	Broadband ISDN	Intranets and Internet
Introduction	Migration from existing networks	Evolution from GSM and ISDN networks
Roaming	New development, INAP based	Evolution of GSM roaming, MAP based
Standardization	FPLMTS as one monolithic standard in ITU	IMT 2000 family in ITU and ANSI, ETSI, ARIB/TTC

Rosenbrock's 'message' was about the connection between ETSI's main goals and the inner market of Europe. This is clearly stated in one of ETSI's original objectives:

> *"To produce the technical standards which are necessary to achieve a large & unified European telecommunications market."*

Later, this point has been emphasised in June 1995 to "Making international standards happen first in Europe".

The 'efficiency' of ETSI has improved during the years, at least measured in relation to the number of published reports per year or the time needed to bring forward a new standard. Rosenbrock spent some time emphasising this point with graphs and figures. For example, the number of published deliverables per year has increased from approximately 30 in 1990 to over 600 in 1997, while the cost per deliverable has decreased from over 500 kECU in 1990 to 30 kECU in 1997. Some people would probably argue that such figures are not the only relevant measure of success.

Table 3 EP BRAN types of broadband radio networks

HIPERLAN/2	A complement to HIPERLAN/1 providing communication between portable computing devices and broadband core networks. User mobility is supported within the local service area, 50 to a few hundred metres. Up to 25 Mb/s
HIPERACCESS	An outdoor, high speed radio access network, providing fixed radio connections to customer premises. Up to 25 Mb/s
HIPERLINK	A very high speed radio network for infrastructure-like applications; a typical use is the interconnection of HIPERACCESS networks and/or HIPERLAN Access points into a fully wireless network. Up to 155 Mb/s.

The past and present are easy to fill with facts and figures, while the future is definitely more 'in the mist'. The issue is the role of ETSI in the future. Will ETSI be the main driving force in future European telecom standardisation? This part of the speech was as misty as the future might seem. Anyone who can create a critical mass around a technology is in the position of producing a standard, and the key to success is inventiveness and creativity. The word 'convergence' was much used. While Rosenbrock's postulate was that convergence of technologies will come by itself, what is needed now is the 'convergence of people', whatever is meant by that.

Radio based communications

The fact that there were three speeches on radio communications standards shows what a high priority this has.

Friedhelm Hillebrand, chairman of Technical Committee *Special Mobile Group (TC SMG)* went through the status for the work towards next generation mobile communication, in ETSI terms *UMTS – Universal Mobile Telecommunications System*. From the 'global' GSM to UMTS 'covering the Universe', was his opening words. The framework for UMTS was set up in July 1996 with the conclusions from the work with *Global Multimedia Mobility – GMM*. The focus was shifted for UMTS, and reference is often made to 'old' and 'new' UMTS. The changes are summed up in Table 2.

Most important to notice is the focus on Internet and IP, as well as UMTS now being an evolution from GSM instead of a 'new' network. The first phase of UMTS is the so-called 'Release 99' having elements from GSM Phase 2+.

Fixed radio systems are also having a great momentum. *Jan Kruys* is chairman of the ETSI project *Broadband Radio Access Networks (EP BRAN)*, which started up in April 1997. Fixed radio access technology is directly competing with for example high-speed photonic networks, so why spend so much resources on this? Radio technology has its force in a high degree of flexibility, fast and cheap installation and low cost in general. The focus of the work is very much on deregulation and multimedia services. User service demands vary greatly over a very short period of time leading to large variations in bandwidth demand. Additionally, users are very sensitive towards the response times. EP BRAN has taken over the responsibility for HIPERLAN/1, ETSI's high speed wireless LAN, from previous TC RES (Radio Equipment and Systems) and this is now part of a series of specifications for broadband radio access standards. Table 3 shows the three types of broadband radio access networks that EP BRAN is working towards. The first specifications are expected to be ready in June 1999, while the whole project is scheduled to end in June 2002. It is all about radio technology with high performance and an open-ended structure.

The work on *TETRA – TErrestrial TRunked Radio* – started as early as 1989 in RES6. TETRA will eventually become a family of standards, and the work is performed in four phases, where Phases 1 and 2 are finished. Phase 3 is expected to finish in the third quarter of 1998, while the work on Phase 4 has just started. *Brian Oliver* is chairman of EP TETRA and among other things he focused on the question of success for TETRA. The Memorandum of Understanding (MoU) has been signed by 66 organisations from 19 countries. One of the results from EP TETRA is the specification of the Digital Advanced Wireless Service (DAWS). DAWS is based upon TETRA's advanced PDO (Packet Data Optimised) protocol. It is possible to obtain full ATM speed of 155 Mb/s with full mobility. EP TETRA is co-operating closely with EP BRAN and EP TIPHON (see next section), as well as TC SMG.

Fixed network standards, systems and services

The second main session of the seminar focused on future high speed networks and services.

Internet Telephony, or Voice over IP, is a hot topic. ETSI has recognised this fact and started *EP TIPHON – Telecommunications and Internet Protocol Harmonization Over Networks*. *Helmut Schink* is chairman for TIPHON which started in April 1997. The focus of the project is on voice- and voice-band communications over IP and the interoperability with PSTN/ISDN/GSM. Voice over IP is based on ITU recommendation H.323 from ITU-T SG16 [2], and there are strong relations to VoIP-Forum/IMTC (International Multimedia Teleconferencing Consortium), IETF (The Internet Engineering Task Force) and ITU-T. Several possible scenarios were presented, eg. 'IP to PSTN/ISDN/GSM' and 'PSTN/ISDN/GSM to PSTN/ISDN/GSM over IP', as shown in Figure 2. One of the problems currently being focused is the addressing of TIPHON IPs. Traditional IP addresses are both complex and dynamic, and not very useful from a standard telephone keyboard. The solution is to use ITU's E.164 numbering scheme. There is a close co-operation with TC NA (Network Aspects) on this.

Rolf Rüggeberg, chairman of ETSI Project Multimedia Terminals and Applications (EP MTA) opened his speech by setting up a definition of Multimedia:

> *"Integration of text, data, graphic, audio, voice and video in a uniform user platform by simultaneous usage of more than one of these elements."*

He went on to show a list of different bodies working with multimedia standardisation: 11 global, 9 European and 3 US regional, and the list was probably not exhaustive. The point in showing this was probably to focus on the interest for the topic, but also show how difficult it can be to work on standards for multimedia. The main focus of this speech was to show some possible future visions (or horror-scenes?), as for example:

The Networked Home:

11 March 2009:

You are approaching your home, the Home Area Network (HAM) reports:

All domestic appliances o.k.
No unusual occurrences.
Call up security report.
You have 5 items of mail waiting, including a registered letter.
What would you like for dinner?
Please enter your mealtime.

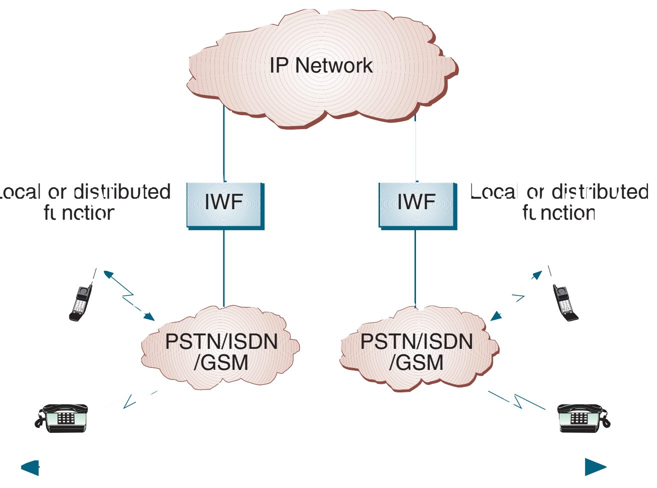


Figure 2 Scenario 'PSTN/ISDN/GSM to PSTN/ISDN/GSM over IP'

It was pointed out that this is not necessarily the future we want, but it could be possible.

The Interoperability of ATM networks and services are dealt with by the ETSI Project *European ATM Services Interoperability (EP EASI)*. The chairman, *Mike Bexon*, focused in his speech very much on ATM MoU. EP EASI started as late as October 1997. The ATM MoU Working Group (WG) was set up by ETNO (European Public Telecommunications Networks Operators' Association) in January 1997, and in co-operation with EURESCOM, ETSI and the ACTS project JAMES (Joint ATM Experiment on European Services), the MoU was ready in December 1997. The goal of EP EASI is to work out two European Norms (ENs) and two so-called guideline documents. The ENs are meant to be 'Supe-ICSs' (Implementation Conformance Statements).

Security in future telecommunications

Technical Committee Security (TC SEC) works on security matters in modern, digital telecommunications. TC SEC is chaired by Michael Walker, and his speech addressed the different aspects and the importance of security. Security is important to avoid eavesdropping and fraud. Previously, this has been solved by using a special hardware module and crypto-chips to 'secure' protocols and firewall technologies. The big change came with the migration from analogue to digital mobile telephony where security also encompasses authentication and encryption, and not least, the Subscriber Identity Module (SIM). The encryption also functions to authorise the use of the radio channel. In electronic commerce, security is very important, and there are great challenges on secure electronic transactions from mobile and/or personal terminals. Today the payment terminals are fixed installations in shops, at petrol stations and so on. A lot of the standards work is carried out outside ETSI, as for example SET – Secure Electronic Transaction, and a lot of the basics on encryption techniques have been put into the X.400 and X.500 recommendations from ITU.

Final words

Antonio Castello, chairman of ETSI General Assembly (GS) closed the seminar by, among other things, postulating that the real 'State of the Art' is that everything is changing. What holds today, does not necessarily hold tomorrow, and ETSI as an organisation must relate itself to that fact.

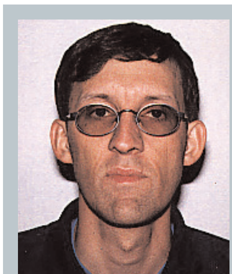
References

1 Hillebrand, F. Universal Mobile Telephone System. In: *ETSI 10th Anniversary Conference : The State of the Art in European Telecommunications Standardization.* Nice, France, 25 March 1998.

2 Ulseth, T. ITU-T multimedia standardisation activities. *Telektronikk,* 93, (2), 87–89, 1997.

Abbreviations

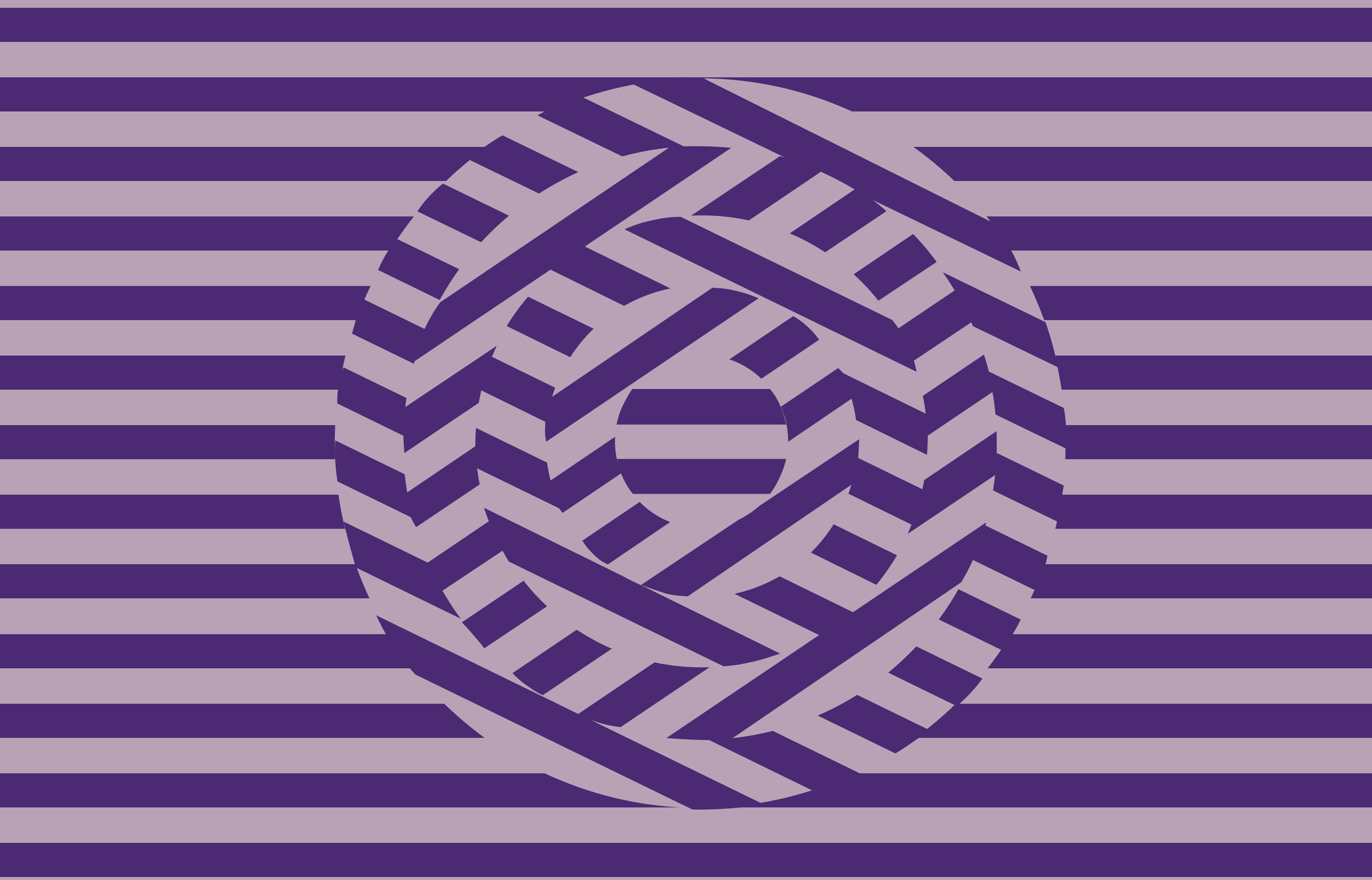
ANSI	American National Standards Institute
ARIB	Association of Radio Industries and Businesses
ATM	Asynchronous Transfer Mode
BRAN	Broadband Radio Access Network
DAWS	Digital Advanced Wireless Service
EASI	European ATM Services Interoperability
EN	European Norm
EP	ETSI Project
ETNO	European Public Telecommunications Networks Operators' Association
ETR	ETSI Technical Report
ETS	ETSI Technical Specification
FPLMTS	Future Public Land Mobile Telecommunications System
GMM	Global Multimedia Mobility
GSM	Global System for Mobile Communications
HIPERLAN	High Performance Radio Local Area Network
IETF	Internet Engineering Task Force
IMT-2000	Integrated Mobile Telecommunications 2000
IMTC	International Multimedia Teleconferencing Consortium
INAP	Intelligent Networks Application Protocol
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ITU	International Telecommunications Union
MAP	Mobile Application Protocol
MTA	Multimedia Terminals and Applications
PSTN	Public Switched Telephone Network
RES	Radio Equipment and Systems
SIM	Subscriber Identity Module
SMG	Special Mobile Group
TC	Technical Committee
TETRA	TErrestrial TRunked Radio Access
TIPHON	Telecommunications and Internet Protocol Harmonization Over Networks
UMTS	Universal Mobile Telecommunications System



Per Hjalmar Lehne is Research Scientist at Telenor Research & Development, Kjeller. He is working in the field of personal communications, with a special interest in antennas and radio wave propagation for land mobile communications.

e-mail:
per.lehne@fou.telenor.no

Kaleidoscope



On the calculation of switches in an automatic telephone system

An investigation regarding some points in the basis for the application of probability theory on the determination of the amount of automatic exchange equipment

B Y T. E N G S E T, K R I S T I A N I A 1 9 1 5

T R A N S L A T E D B Y A R N E M Y S K J A

Contents

Some comments on the translation are found on page 44.

I Introduction

1. We will try to apply probability calculation to determine the necessary number of switches in a switch group and within a section of co-operating switches in a group.

The mathematical concept of 'probability' is commonly defined in the following way:

The probability of an event E (Ereignis, événement) is the ratio between the number of cases that are favourable for the occurrence of this event and the total number of possible events, given that all cases are equally likely and mutually independent.

Given, for example, an urn with N balls, of which there are m white and n black, $m + n = N$, and if those balls by a suitable action are mixed so that they are mutually non-orderly located, then the *probability* that a single ball picked is a *white* ball $= m/N$. Likewise, the *probability* of picking a *black* ball $= n/N$.

In the following we shall apply the two main theorems of probability theory.

The *first* theorem states that among different events $E_1, E_2, E_3, \ldots, E_n$, that depend on the probabilities $p_1, p_2, p_3, \ldots, p_n$, the *probability* $P_{(1+2+3+\ldots+i)}$ that *either* E_1 *or* E_2 *or* E_3 ... *or* E_i will occur $= p_1 + p_2 + p_3 + \ldots + p_i$, i.e. equal to the *sum* of the probabilities of the single events.

If this is applied to an urn with balls, e.g. n_1 white, n_2 red, n_3 green, n_4 black, altogether $n_1 + n_2 + n_3 + n_4 = N$ balls, and if we set $n_1/N = p_1$, $n_2/N = p_2$, $n_3/N = p_3$, $n_4/N = p_4$, then the *probability* of picking *either one white or one green ball* (irrespective of which) is: $P_{(1+3)} = p_1 + p_3$.

The *second* main theorem states that if different events $E_1, E_2, E_3, \ldots, E_n$, depend on, respectively, $p_1, p_2, p_3, \ldots, p_n$, then the *probability* $P_{1.2.3\ldots i}$ that *both* E_1 *and* E_2 *and* E_3 ... *and* E_i will occur – i.e. a combination of those events will occur – $= p_1 p_2 p_3 \ldots p_i$, being the *product* of the single event probabilities.

At trials with the last mentioned urn this means that the *probability* of picking a red ball and, after replacing it, picking a green ball (those two pickings being considered a combination or commonality) will be: $P_{2.3} = p_2 p_3$.

We will also have to consider *complementary* probabilities, i.e. probabilities that exclude one another, i.e. the probability p that an event *will* occur and the probability $(1 - p)$ that it will *not* occur.

In the first mentioned example of white and black balls m/N and n/N are complementary probabilities Their sum is $(m+n)/N = 1$. In the second example among others $(p_1 + p_3)$ and $(p_2 + p_4)$ are complementary probabilities. Similarly, at the combined pickings, $p_2 p_3$ and $(1 - p_2 p_3)$ are complementary probabilities.

II Determination of the connection probability of single subscribers

2. When one intends to apply probability theory to the local telephone correspondence, one may proceed in various ways. One may, for instance, put the main weight on a presentation taking into account that calls from different subscribers to the exchange in general occur at different time instants, and one may try to calculate the probability that a certain number of calls arrive during a certain interval, corresponding to the mean conversation time within some specified period, for instance the busiest hour of an ordinary workday. This approach, however, will incur substantial difficulties, unless arbitrary simplifications are made, the influence of which hardly can be estimated in an obvious way.

As the intention is to obtain an insight into the process of the *aggregated* telephone traffic at any instant within working hours – be it for the whole subscriber network or for certain larger or smaller groups of subscribers – we have preferred to direct our attention to what happens or can be assumed to happen at an arbitrary epoch within a specified work period. In other words, we direct the focus of our investigation at *simultaneous* events at each moment of the period under consideration.

We name this period the *selected* time T_c. Within T_c, which may be the busiest hour or the busiest quarter-hour or some other specified work time, traffic is assumed to flow fairly evenly, and for a longer time space T_t show, for a given exchange, a certain mean number of calls r per T_c with a certain mean duration t per call.

T_t is measured in days, T_c and t in seconds. The product rt thus indicates the mean aggregate call (or conversation) seconds within the period T_c, and $rt / T_c = y$ indicates the *mean number of simultaneous* local telephone connections at the exchange during the period T_c. It is assumed that r and t are found by a very large number of observations within the telephone network concerned under current, satisfactory manual telephone operation during a lengthy period.

3. We will assume that n subscribers are connected to the exchange. If throughout T_c there were exactly y connections, then according to the given probability definition above, the general mathematical probability p_g of finding a subscriber busy with an own initiated call at a random epoch within T_c would be exactly y / n. Since y is only an average number, as the real number of simultaneous connections in T_c can vary between 0 and substantially more than y (eg. $2y$ or even more), one cannot give any exact, real probability p_r for finding a calling subscriber by a single sample. One must be satisfied by an approximate indication by setting $p_r \approx y / n$.

Thus $p_r \approx y / n$ in a way indicates the average likelihood during time T_c to find any particular subscriber in the course of a conversation initiated by himself. We can, however, not accept right away the quantity y / n as a measure of our subscribers' proneness to make calls. We know there are great differences between different subscribers' use of the telephone. We must therefore make a judgement as to whether the estimate of y / n (or some differing value, in any case a small proper fraction less than one) implies any *harmful* error, ie. an error leading to *underestimation* of traffic requirements by calculating a too small number of switches.

By our considerations we will remember the great regularity that is apparent in the values of r and t at one particular telephone network through a lengthy period, and the value r's dependence on the number of subscribers and the time of day (the concentration). These average values emerge as a sum or as an extract of the total calling activity of the whole set of subscribers. The single subscribers act in the calling sense quite uncoordinated. In the long run, however, each subscriber must satisfy a certain requirement that is particular for him.

A particular subscriber will be denoted A_i. The total time spent by this subscriber during the selected period T_c of each day through a lengthy time space of $T_t = D$ days shall be denoted T_i, measured in seconds. $T_i / (D \cdot T_c) = p_i$ denotes the ratio between the time spent on calling and the total time available for this subscriber.

One may also view the case in such a way as to assume the total time under observation, $D \cdot T_c$, to be subdivided in a very large number *(N)* of very short time intervals Δt. Thus $D \cdot T_c = N \cdot \Delta t$. For subscriber A_i's telephone calling there will, out of the total time $N \cdot \Delta t$, be m_i favourable intervals and $(N - m_i) = n_i$ unfavourable intervals Δt. Here we set $m_i \cdot \Delta t = T_i$ and consequently $n_i \cdot \Delta t = (DT_c - T_i)$.

From the above follows

$$p_i = \frac{T_i}{D \cdot T_c} = \frac{m_i \cdot \Delta t}{N \cdot \Delta t} = \frac{m_i}{N}, \tag{1a}$$

$$\left(1 - p_i\right) = 1 - \frac{m_i}{N} = \frac{N - m_i}{N} = \frac{n_i}{N}. \tag{1b}$$

4. We will now assume that we take momentary test samples at the exchange whether A_i's line is occupied by an outgoing call, as we suppose such a sample technically realised in a suitable way. We will let the instant of any sample be completely random, with no pre-knowledge whether the line is occupied or not. According to the nature of the sampling one can assume that all cases are equally likely.

Furthermore, as we can assume that any start and close of a connection are independent relative to the sampling points, we must be allowed to consider the sample results as independent, in spite of the fact that a very large number of 'connection cases' or intervals with connection (corresponding to a single call) will follow immediately after one another.

5. In order to make sure that we do not expose ourselves to any self-delusion by these considerations, we shall make a comparison with the previously treated example of white and black balls. Instead of placing these balls in an urn where they by some suitable movement are mixed, we will assume the balls fastened along the rim of a circular disk that can be turned around a vertical axis in the centre of the disk. The disk may be set in an even rotation by some mechanism. The balls, all of the same size, are densely positioned along a circle line with centre in the axis of the disk. They are hidden behind a screen with a small slot in front of the ball ring, through which a single ball may be observed as it passes the slot during rotation. The slot is normally closed by a slide. By a quick opening of the slide at

any moment a white or a black ball may be observed through the slot. At a simultaneous view of the adjacent parts of two balls it is an established rule that the ball leaving the slot is not considered. The slide is only kept open a short moment, so that only a part of a ball can pass the slot during opening.

When now, as in the first mentioned example, the number of white balls is m, the number of black balls is n and the joint number $m + n = N$, then the probability of observing a white ball in an arbitrary moment is $p = m / N$ and the probability of observing a black ball is $(1 - p) = n / N$, exactly like the probability of picking a white or a black ball from the before mentioned urn. By repeated samples with the rotating ball ring the choice of observation moments must be made without knowledge of the speed of rotation, so that the conditions of the probability definition, concerning even likelihood of occurrence for all cases and their mutual independence, are fulfilled. Under this condition we may consider the ball urn and the hidden rotating ball ring as fully equivalent mechanisms for illustration of the concept of mathematical probability.

6. The analogy between the ball ring with m white and n black balls, altogether N balls, and the case of total time $N \cdot \Delta t$ with m_i connection intervals and n_i non-connection intervals, altogether N intervals Δt is obvious. There is, however, the difference that a traversed time space $N \cdot \Delta t$ with its telephone calls never repeats itself, whereas the ball ring for each round presents the same order, unless it is changed by some particular arrangement. Thus telephone traffic presents a picture of more disorder and randomness, implying that there is less need to fear that the conditions of the probability definition are or will be unfulfilled, when the events of connection are observed. On the other hand one should be aware that there is no constant probability related to the single subscriber line such that that of the urn or ball ring. More about this later.

7. It appears from the discussion that the quantity of $p_i = m_i / N$ for subscriber A_i's call activity, in analogy with the quantity $p = m / N$ for the ball trial, has the character of a probability, namely the probability that subscriber A_i at an arbitrary instant is busy with a self-initiated call, and that the quantity $(1 - p_i) = n_i / N$ for the same line – corresponding to the ball trial's $(1 - p) = n / N$ – should be considered as the complementary probability, ie. the probability of finding the same line *not* busy with a self-initiated call. (Any incoming call from other parties to A_i are neglected).

8. According to the preceding arguments we feel justified to present the following definition: A subscriber A_i's connection probability p_i is the ratio between his assumed time T_i of outgoing call occupancy and the total time $D \cdot T_c$.

In order to avoid misunderstanding we have here used the term 'assumed' rather than 'favourable'[1)], that is usually applied in the general definition of probability.

[1)] *The Norwegian text applies the terms 'paaregnelig' and 'gunstig'.*

III General calculation of efficiency and degree of hindrance

9. We will now consider a group of subscribers $A_1, A_2, A_3, \ldots, A_i, \ldots, A_n$, whose connection probabilities in the selected period are $p_1, p_2, p_3, \ldots, p_i, \ldots, p_n$, respectively, and which have jointly available for connection through the exchange a certain number x of adjacent circuits with the necessary equipment.

Assuming that $p_1, p_2, p_3, \ldots, p_i, \ldots, p_n$ are known and exactly given constant quantities, we shall calculate the probabilities $P_0, P_1, P_2, \ldots, P_r, \ldots, P_x$ of finding, respectively, 0, 1, 2, 3, ..., r, ..., x out of the available circuits busy with calls at an arbitrary moment within the selected interval T_c on a similarly arbitrary day within T_t. Furthermore we shall calculate the probabilities $P_{x+1}, P_{x+2}, \ldots, P_{x+(n-x)}$ of finding $x + 1, x + 2, \ldots, n$ circuits busy in case one had $x + 1$ or $x + 2$ or ..., or n circuits available at the exchange instead of just x.

In order to facilitate the process of this calculation we shall apply a couple of easily understandable product symbols. We shall establish a product of r out of the n given quantities $p_1, \ldots, p_n$

and denote this by $\prod\limits_{(n)}^{(r)} p_i$, where any quantity p_i appears

only once in any product. According to combination theory

there will be altogether $\dfrac{n(n-1)\ldots(n-r+1)}{1 \cdot 2 \cdot 3 \ldots r} = \dbinom{n}{r}$ such

products. Each of these products can be identified by a number

within the series $1, 2, \ldots, j, \ldots \dbinom{n}{r}$ by putting the number

behind a parenthesis in front of the product expression:

$(j\prod\limits_{(n)}^{(r)} p_i$. The notation (n) under $\prod$ may be omitted when

no ambiguity can occur. When we have formed a product of r probabilities p, eg. $p_1 p_2 p_3 \ldots p_r$, we will form another product $(1 - p_{r+1})(1 - p_{r+2}) \ldots (1 - p_n)$ of the complementary probabilities of the remaining $(n - r)$ probabilities p. In general we

shall denote such a product by $\prod\limits^{(n-r)} \left(1 - p_k\right)_{k \neq i}$. The number

such counterproducts will likewise be $\dbinom{n}{r}$ since for each

product of r probabilities there will be a product of complementary probabilities corresponding to the remaining $(n - r)$ probabilities. We will furnish each of the such given complementary products with the same number as the primary product.

Thus to the primary product $(j\prod\limits_{(n)}^{(r)} p_i$ corresponds the

counterproduct $(j\prod\limits_{(n)}^{(n-r)} \left(1 - p_k\right)_{k \neq i}$. We will, however,

omit the latter number indication as superfluous when we present the product of both those products, thus writing:

$$(j\prod_{(n)}^{(r)} p_i \cdot \prod_{(n)}^{(n-r)} \left(1-p_k\right)_{k\neq i}$$

A sum of all products formed this way for a given r will be denoted in the usual way:

$$\sum_{j=1}^{j=\binom{n}{r}} (j\prod_{(n)}^{(r)} p_i \cdot \prod_{(n)}^{(n-r)} \left(1-p_k\right)_{k\neq i}$$

10. We will now calculate P_0, the probability that all x circuits are found to be free by a sample test, ie. the probability of *no* connection.

As subscriber A_i's connection probability is p_i, then his non-connection or idleness probability is $(1 - p_i)$. Likewise A_j's idleness probability is $(1 - p_j)$. As the matter in question here is the probability that all of A_1 and A_2 and A_3 and ... and A_i and A_j and ... and A_n are without a telephone connection, the second probability theorem implies that the requested probability is given by the product of all the idleness probabilities:

$$P_0 = \left(1-p_1\right)\left(1-p_2\right)\left(1-p_3\right)\mathrm{K}\ \left(1-p_k\right)\mathrm{K}\ \left(1-p_n\right) = \overset{(n)}{\Pi}\left(1-p_k\right) \qquad (2,_0)$$

In order to find the probability of observing only *one* connection a more compound operation is needed. First we seek the probability of finding A_1 busy and all the other subscribers idle. This probability will be denoted $p_{(1)}$. According to the same theorem we have:

$$p_{(1)} = p_1(1 - p_2)\ (1 - p_3)\ ...\ (1 - p_n)$$

In the same way are found:

$$p_{(2)} = (1 - p_1)p_2\ (1 - p_3)\ ...\ (1 - p_n)$$

...

$$p_{(k)} = (1 - p_1)\ (1 - p_2)\ ...\ p_k\ ...\ (1 - p_n)$$

...

$$p_{(n)} = (1 - p_1)\ (1 - p_2)\ ...\ (1 - p_k)\ ...\ (1 - p_{n-1})p_n$$

As the question is whether the first or the second or the third or ... or the n^{th} case will be observed, one will according to the first theorem have:

$$P_1 = p_{(1)} + p_{(2)} + \mathrm{K}\ + p_{(j)}\mathrm{K}\ + p_{(n)}$$
$$= \sum_{j=1}^{j=n} p_{(j)} = \sum_{j=1}^{j=n} (j\overset{(1)}{\Pi} p_i \overset{(n-1)}{\Pi}\left(1-p_k\right)_{k\neq i} \qquad (2,_1)$$

In order to find the probability of observation of *two* connections we will first seek the probability $p_{(1,2)}$ of finding A_1 and A_2 in connection and simultaneously all the others without. Then we get:

$$p_{(1,2)} = p_1 p_2\left(1-p_3\right)\left(1-p_4\right)\mathrm{K}\ \left(1-p_n\right)$$
$$= (1\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)_{k\neq i}$$

In the same way we obtain:

$$p_{(1,3)} = p_1\left(1-p_2\right)p_3\left(1-p_4\right)\left(1-p_5\right)\mathrm{K}\ \left(1-p_n\right)$$
$$= (2\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)$$

$$p_{(1,4)} = p_1\left(1-p_2\right)\left(1-p_3\right)p_4\left(1-p_5\right)\left(1-p_6\right)\mathrm{K}\ \left(1-p_n\right)$$
$$= (3\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)$$

... ...

$$p_{(1,n)} = p_1\left(1-p_2\right)\left(1-p_3\right)\left(1-p_4\right)\mathrm{K}\ \left(1-p_{n-1}\right)p_n$$
$$= ((n-1)\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)$$

Furthermore:

$$p_{(2,3)} = \left(1-p_1\right)p_2 p_3\left(1-p_4\right)\left(1-p_5\right)\mathrm{K}\ \left(1-p_n\right)$$
$$= (n\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)$$

... ...

$$p_{(2,n)} = \left(1-p_1\right)p_2\left(1-p_3\right)\left(1-p_4\right)\mathrm{K}\ \left(1-p_{n-1}\right)p_n$$
$$= ((2n-2)\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)$$

... ...

$$p_{(j,n)} = \left(1-p_1\right)\left(1-p_2\right)\mathrm{K}\ \left(1-p_{j-1}\right)p_j\left(1-p_{j+1}\right)\mathrm{K}\ p_n$$
$$= (q\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)$$

... ...

$$p_{(n-1,n)} = \left(1-p_1\right)\left(1-p_2\right)\mathrm{K}\ \left(1-p_{n-2}\right)p_{n-1}p_n$$
$$= (\frac{n(n-1)}{1\cdot 2}\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)$$

Hereby we obtain:

$$P_2 = p_{(1,2)} + p_{(1,3)} + \mathrm{K}\ + p_{(1,n)} + p_{(2,3)} + \mathrm{K}\ + p_{(2,n)} + p_{(j,n)}$$
$$+\mathrm{K}\ + p_{(n-1,n)} = \sum_{\alpha=1}^{\alpha=n-1}\sum_{\beta=2}^{\beta=n} p_{(\alpha,\beta)_{\alpha<\beta}}$$
$$= \sum_{j=1}^{j=\binom{n}{2}} (j\overset{(2)}{\Pi} p_i \overset{(n-2)}{\Pi}\left(1-p_k\right)_{k\neq i} \qquad (2,_2)$$

... ...

In a similar way is obtained:

$$P_3 = \sum_{j=1}^{j=\binom{n}{3}} (j\overset{(3)}{\Pi} p_i \overset{(n-3)}{\Pi}\left(1-p_k\right)_{k\neq i} \qquad (2,_3)$$

$$P_4 = \sum_{j=1}^{j=\binom{n}{4}} \left(j \prod^{(4)} p_i \prod^{(n-4)} \left(1-p_k\right)\right)_{k\neq i} \qquad (2,_4)$$

... ...

$$P_x = \sum_{j=1}^{j=\binom{n}{x}} \left(j \prod^{(x)} p_i \prod^{(n-x)} \left(1-p_k\right)\right)_{k\neq i} \qquad (2,_x)$$

$$P_{x+1} = \sum_{j=1}^{j=\binom{n}{x+1}} \left(j \prod^{(x+1)} p_i \prod^{(n-(x+1))} \left(1-p_k\right)\right)_{k\neq i} \qquad (2,_{x+1})$$

... ...

$$P_n = \sum_{j=1}^{j=\binom{n}{n}} \left(j \prod^{(n)} p_i \prod^{(0)} \left(1-p_k\right)\right)_{k\neq i}$$

$$= \prod_{(n)}^{(n)} p_i = p_1 p_2 p_3 \ldots p_i \ldots p_n \qquad (2,_n)$$

The last equation expresses the probability P_n of finding at an arbitrary moment all n subscribers occupied by self-initiated calls on a set n of exchange circuits, irrespective of possible call completion due to the state of called subscribers. This probability is equal to the product of the connection probabilities of all subscribers, a result that also follows directly from the second theorem.

11. With the probabilities $P_0, P_1, P_2, ..., P_x, ..., P_n$ all possibilities with n internal connections are exhausted, since any test will find either 0 or 1 or 2 or 3 or ... or n circuits occupied, in case there are altogether n circuits. Thus the sum of all those probabilities must express certainty:

$$P_0 + P_1 + P_2 + ... + P_x + ... + P_n = 1.$$

If there are only x circuits available, then the alternative probability of finding either 0 or 1 or 2 or 3 or ... or x circuits occupied in an arbitrarily chosen moment is $P_0 + P_1 + P_2 + ... + P_x$. The probability of finding all x circuits busy and simultaneously 1 or 2 or ... or $(n - x)$ 'standing' calls without possibility of progress is $P_{x+1} + P_{x+2} + ... + P_{x+(n-x)} = 1 - (P_0 + P_1 + P_2 + ... + P_x)$, where $P_0, P_1, P_2, ..., P_x, ..., P_n$ are determined by the expressions above, given as functions of $p_1, p_2, p_3, ..., p_n$, which quantities are assumed known, and indicate the individual probabilities for each subscriber of being connected through the exchange of an ideal telephone system with no restrictions. In the determination of $p_1, p_2, p_3, ..., p_n$ enter all calls and all telephone use, even calls that encounter busy or no-answer subscribers.

As can be seen, by the applied method under the condition mentioned, we have been able to calculate an expression of system efficiency, given by the probability $P_{(0x)} = P_0 + P_1 + P_2 + ... + P_x$, with full mathematical exactness without any hypothesis concerning the single call or connection duration or concerning the average duration of calls or connections.

12. We shall now proceed to apply the obtained exact and general results in an appropriate manner. First of all we notice that the quantities $P_0, P_1, P_2, ..., P_x, ..., P_n$ indicate the probabilities of occurrence of certain states. We shall, therefore, call them state probabilities. P_0, for instance, gives a measure of how often we can expect to find all x (or n) circuits free. If we take a very large number of samples, N, and if the number of 'favourable' cases for total non-occupancy is q_0, then according to our calculations and definitions $P_0 = \prod^{(n)} \left(1-p_k\right) = q_0 / N$. In other words: among the N samples we expect to find q_0 cases without any connection. In reality the result will differ to a larger or smaller extent from that, however, if we gradually increase the number of samples, then the fraction q_0/N will gradually approach $\prod_{(n)}^{(n)} \left(1-p_k\right)$. In a similar way sample tests of other cases will behave:

q_1 / N will approach the calculated value of P_1

q_2 / N will approach the calculated value of P_2

...

q_x / N will approach the calculated value of P_x

...

q_n / N will approach the calculated value of P_n

We have not, however, by this obtained any direct expression for the relative amount of telephone calls that pass unimpeded with the application of x circuits for a set of n subscribers. But this can now theoretically be carried out by means of the found expressions.

13. We will again consider the total time $D \cdot T_c = N \cdot \Delta t$, where N is a very large or extremely large number and Δt is a very small or extremely small time interval. We assume that samples are taken momentarily. If a test falls exactly between two intervals, we will assume, like in the ball ring test, that it falls in the second interval. The intervals are assumed to be so short that within one interval there will not be more than one state change. If a test falls exactly between two different states, (it occurs simultaneously with a state change), we will consider it belonging to the second state. We choose N so large, and hence the interval Δt so small that two adjacent tests will not fall within the same interval, ie. we let Δt converge towards the negligible time that is necessary to record a certain state. With these conditions a non-ambiguous marking of the state of any interval will be ascertained.

As will be remembered, for the total time $N \cdot \Delta t$ there will for each subscriber A_i be a favourable or presumed connection time of $m_i \cdot \Delta t$ seconds, as $m_i \cdot \Delta t = p_i \cdot N \cdot \Delta t$. The sum of presumed connection times of all subscribers during the total time is $T_s = (p_1 + p_2 + p_3 + ... + p_i + ... + p_n) \cdot N \cdot \Delta t$ seconds.

$$T_S / n = \sum_1^n p_i \cdot N \cdot \Delta t / n$$

is the average connection time per subscriber and

$$\frac{T_S}{N \cdot \Delta t} = \sum_1^n p_i$$

gives the average number of simultaneous connections per interval Δt.

On the other hand we will now study how the various connection states are distributed over the total time $N \cdot \Delta t$. As indicated, the number of cases or intervals that are favourable or presumed for the number of (i) simultaneous connections is q_i, where $q_i/N = P_i$, ie. $q_i = N \cdot P_i$. Consequently, the total duration of corresponding states is $t_i = q_i \cdot \Delta t = P_i \cdot N \cdot \Delta t$. Hereby the total time $N \cdot \Delta t$ will be subdivided in time sections $t_0, t_1, t_2, ..., t_i, ..., t_n$, corresponding to the $n + 1$ different connection or call states $P_0, P_1, P_2, ..., P_i, ..., P_n$. In other words, in an ideal system one can presume that out of $N \cdot \Delta t$ seconds there will be:

$t_0 = P_0 \cdot N \cdot \Delta t$ seconds with no connection,

$t_1 = P_1 \cdot N \cdot \Delta t$ seconds with one connection,

$t_2 = P_2 \cdot N \cdot \Delta t$ seconds with 2 connections, and so on up to

$t_n = P_n \cdot N \cdot \Delta t$ seconds with n connections.

From this follows the presumption that there in the time

	t_0 must be carried	$0 \cdot P_0 \cdot N \cdot \Delta t$ connection seconds
	t_1 must be carried	$1 \cdot P_1 \cdot N \cdot \Delta t$ connection seconds
	t_2 must be carried	$2 \cdot P_2 \cdot N \cdot \Delta t$ connection seconds
	t_3 must be carried	$3 \cdot P_3 \cdot N \cdot \Delta t$ connection seconds
	...	
	t_i must be carried	$i \cdot P_i \cdot N \cdot \Delta t$ connection seconds
	...	
	t_x must be carried	$x \cdot P_x \cdot N \cdot \Delta t$ connection seconds
	...	
	t_n must be carried	$n \cdot P_n \cdot N \cdot \Delta t$ connection seconds
Sum:	$N \cdot \Delta t$	$T_s = (p_1 + p_2 + p_3 + ... + p_i + ... + p_n) \cdot N \cdot \Delta t$ seconds

Hereby we have obtained to get the whole presumed 'traffic' subdivided on $n + 1$ sections of the total time. Each time section t_i is of course in reality composed of a number of disjoint time intervals, which should not prevent a compound view of all intervals characterised by the same connection state.

14. With the results thus obtained we are now able to indicate the amount of presumed traffic that we are likely to carry in an ideal system with x internal circuits, and consequently the likely amount that will be prevented because of the lack of circuits or switches.

The sum of the required traffic, measured in connection seconds, that is carried during the time $t_0 + t_1 + t_2 + ... + t_x$ amounts to $(1 \cdot P_1 + 2 \cdot P_2 + 3 \cdot P_3 + ... + x \cdot P_x) \cdot N \cdot \Delta t$ connection seconds. During this time there are no hindrances, as the number of simultaneous connections do not surpass x. In the remaining time $t_{x+1} + t_{x+2} + ... + t_{n-1} + t_n$ a traffic of $(x \cdot P_{x+1} + x \cdot P_{x+2} + x \cdot P_{x+3} + ... + x \cdot P_{n-1} + x \cdot P_n) \cdot N \cdot \Delta t$ will be carried, since all x circuits are busy with traffic during this time, whereas $(1 \cdot P_{x+1} + 2 \cdot P_{x+2} + 3 \cdot P_{x+3} + ... + (n - x) P_n) \cdot N \cdot \Delta t$ connection seconds during this time will be prevented from being carried.

The ratio between the part of traffic that can be carried by x coordinate exchange circuits and the entire presumed traffic from n subscribers will be designated the *efficiency* of x internal circuits for n subscribers and denoted $\underline{\varphi_{nx}}$. The ratio between the part of the traffic that is *not* carried by x such circuits and the entire presumed traffic from n subscribers will be designated the *degree of hindrance* of x internal circuits for n subscribers and denoted $\underline{\psi_{nx}}$.

From the presentation above now follows:

$$\begin{aligned}\underline{\varphi_{nx}} &= (1 \cdot P_1 + 2 \cdot P_2 + 3 \cdot P_3 + \ldots + (x-1)P_{x-1} \\ &\quad + x(P_x + P_{x+1} + P_{x+2} + \ldots + P_n))\frac{N\Delta t}{T_s} \\ &= (1 \cdot P_1 + 2 \cdot P_2 + 3 \cdot P_3 + \ldots + (x-1)P_{x-1} \\ &\quad + x(P_x + P_{x+1} + P_{x+2} + \ldots + P_n))\frac{1}{(p_1 + p_2 + \ldots\ p_n)},\end{aligned} \tag{3}$$

where $P_1, P_2, ..., P_x, ..., P_n$ are given the values calculated above as functions of $p_1, p_2, ..., p_n$.

Likewise is obtained:

$$\begin{aligned}\underline{\psi_{nx}} &= (1 \cdot P_{x+1} + 2 \cdot P_{x+2} + 3 \cdot P_{x+3} + \ldots + (n-x)P_n) \\ &\quad \cdot \frac{1}{(p_1 + p_2 + \ldots + p_n)},\end{aligned} \tag{4}$$

where $P_{x+1}, P_{x+2}, ..., P_n$ are given the values calculated above as functions of $p_1, p_2, ..., p_n$.

15. Hence the problem that we have presented is theoretically solved under the conditions given, according to which the probabilities $p_1, p_2, p_3, ..., p_n$ are known, constant quantities and thus during the traffic process are assumed to remain non-influenced by hindrances because of lack of circuits. The last statement, however, is virtually the case only when the number of such hindrances is very small compared with the number of 'busy' messages.

As it appears from the above, the quantities $\underline{\varphi_{nx}}$ and $\underline{\psi_{nx}}$ are proper fractions such that $\underline{\varphi_{nx}} + \underline{\psi_{nx}} = 1$. If one wants efficiency and hindrance expressed in percent or per thousand, one can use the notations

$\underset{nx}{\varphi_{oo}}\underset{nx}{\varphi_{ooo}}$, and $\underset{nx}{\psi_{oo}}$ or $\underset{nx}{\psi_{ooo}}$, where we have

$$\underset{nx}{\varphi_{oo}} = 100 \cdot \underline{\varphi_{nx}},\quad \underset{nx}{\varphi_{ooo}} = 1000 \cdot \underline{\varphi_{nx}}$$

$$\underset{nx}{\psi_{oo}} = 100 \cdot \underline{\psi_{nx}},\quad \underset{nx}{\psi_{ooo}} = 1000 \cdot \underline{\psi_{nx}}.$$

IV The first simplification of the calculations

16. When we next will try to bring the found general expressions to application on actual problems of a practical nature, we will, as is easily seen, encounter a multitude of difficulties. These can only be bypassed by sacrificing a fully exact calculation.

If in a network there is a large number of subscribers, it will be practically impossible to evaluate each single subscriber's connection probability. Even if we were able to, by an enormous amount of counting and measurement, to determine the quantities p_i, ($i = 1, 2, 3, ..., n$), then by the calculation of the state probabilities P_i ($i = 0, 1, 2, 3, ..., x, ..., n$) even for a limited subscriber number such as $n = 100$, the number of multiplication and summation operations would be so huge as to make the calculation virtually impossible.

Since the single subscribers' connection probabilities, for reasons unnecessary to discuss, will not stay constant, whereas the traffic intensity of the collective set of subscribers by experience is seen to stay nearly constant through years, save periodic variations over day and season, we will direct our attention at the *average traffic intensity* per subscriber during the selected time (eg. the busiest hour). This is a quantity that can relatively easily be determined with rather great exactness.

17. Let us assume that we have found the average connection time of T seconds per subscriber during the selected time of T_c seconds in a system with no restrictions. The average connection time per subscriber then is

$$p = T / T_c.$$

Thus p is exactly determined.

Instead of the real set of n subscribers with in general different connection probabilities we consider an imagined set of n subscribers with each the same connection probability p. We will now calculate $\underline{\varphi}_{nx}$ and $\underline{\psi}_{nx}$ for this subscriber set. That is done by setting $p_1 = p_2 = p_3 = ... = p_n = p$. In order to indicate the altered character of the calculated quantities we will put a bar over the letters that indicate these quantities. Hereby we get:

$$\overline{P}_0 = (1-p)^n$$

$$\overline{P}_1 = \frac{n}{1}(1-p)^{n-1} \cdot p$$

$$\overline{P}_2 = \frac{n(n-1)}{1 \cdot 2}(1-p)^{n-2} \cdot p^2$$

$$\overline{P}_3 = \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3}(1-p)^{n-3} \cdot p^3$$

$$\ldots$$

$$\overline{P}_x = \frac{n(n-1)\ldots(n-x+1)}{1 \cdot 2 \ldots x}(1-p)^{n-x} \cdot p^x$$

$$\ldots$$

$$\overline{P}_n = p^n. \tag{5,0-n}$$

Consequently we obtain:

$$\overline{\varphi}_{nx} = (1 \cdot \frac{n}{1}(1-p)^{n-1} \cdot p + 2 \cdot \frac{n(n-1)}{1 \cdot 2}(1-p)^{n-2} \cdot p^2$$
$$+3 \cdot \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3}(1-p)^{n-3} \cdot p^3 + \ldots$$
$$+(x-1)\frac{n(n-1)\ldots(n-(x-1)+1)}{1 \cdot 2 \cdot 3 \ldots (x-1)}(1-p)^{n-(x-1)} p^{(x-1)}$$
$$+x(\frac{n(n-1)\ldots(n-x+1)}{1 \cdot 2 \cdot 3 \ldots x}(1-p)^{n-x} \cdot p^x + \ldots + p^n))\frac{1}{n \cdot p} \tag{6}$$

$$\overline{\psi}_{nx} = (1 \cdot \frac{n(n-1)\ldots(n-(x+1)+1)}{1 \cdot 2 \cdot 3 \ldots (x+1)}(1-p)^{n-(x+1)} \cdot p^{(x+1)}$$
$$+2 \cdot \frac{n(n-1)\ldots(n-(x+2)+1)}{1 \cdot 2 \cdot 3 \ldots (x+2)}(1-p)^{n-(x+2)} \cdot p^{(x+2)}$$
$$+3 \cdot \frac{n(n-1)\ldots(n-(x+3)+1)}{1 \cdot 2 \cdot 3 \ldots (x+3)}(1-p)^{n-(x+3)} \cdot p^{(x+3)}$$
$$+ \ldots + (n-x)p^n)\frac{1}{n \cdot p}. \tag{7}$$

It should be remembered here that p denotes the fraction of the selected time (the busiest hour) that on average occupies each subscriber during outgoing calls. np thus denotes the average number of existing connections over that time.

$\overline{\psi}_{nx}$ denotes the fraction of telephone calling that during this time 'gets lost' by being conveyed on x circuits, while the different subscribers are assumed to have the same average use of the telephone, but otherwise use the telephone randomly.

18. At present the quantity $\overline{\psi}_{nx}$ is our primary interest, and we must try to get an idea whether it by a suitable degree of accuracy can replace the presumed exact quantity of $\underline{\psi}_{nx}$, and also whether the quite essential simplification done will lead to the target of obtaining practically useful formulas.

It is thus a matter of clarifying how the quantities P_0, P_1, P_2, ..., P_x, ..., P_n, considered as functions of $p_1, p_2, ..., p_n$, change by equalising $p_1, p_2, ..., p_n$, given that the sum $p_1 + p_2 + ... + p_n = np$ is constant. By the application of the theory for maxima and minima on the $n + 1$ equations (2, 0-n) that determine P_0, P_1, P_2, ..., P_x, ..., P_n as functions of $p_1, p_2, ..., p_n$, that are considered variable under the condition $p_1 + p_2 + ... + p_n = np$ = constant, one finds that each of the quantities $P_0, P_1, P_2, ..., P_x, ..., P_n$ have *either* maximum *or* minimum when $p_1 = p_2 = p_3 = ... = p_n = p$ (apart from certain exceptions to be discussed below). To the extent that the quantities $P_{x+1}, P_{x+2}, ..., P_n$ have their maxima for $p_1 = p_2 = p_3 = ... = p_n = p$, then in any case $\overline{\psi}_{nx} > \underline{\psi}_{nx}$, and the equalization of the connection probabilities leads to no harmful error for the system operation, since the calculated hindrance is greater than the one obtained by exact calculation. As we need to be certain not to miscalculate, we must try to clarify when the quantities P_0, P_1, P_2, ..., P_x, P_{x+1}, ..., P_n reach their maximum and minimum values.

A more detailed analysis indicates that the necessary and sufficient condition to make P_r a *maximum* when $p_1 = p_2 = p_3 = \ldots = p_n = p$ is:

$$-\binom{n-2}{r-2}(1-p)^2 + 2\binom{n-2}{r-1}(1-p)p - \binom{n-2}{r}p^2 < 0. \tag{8}$$

Here r denotes any number from 0 up to n.

$\binom{n-2}{r-2}$, $\binom{n-2}{r-1}$ and $\binom{n-2}{r}$ are the known expressions for certain binomial coefficients.

The condition for a *minimum* value of P_r under the given assumptions is:

$$-\binom{n-2}{r-2}(1-p)^2 + 2\binom{n-2}{r-1}(1-p)p - \binom{n-2}{r}p^2 > 0. \tag{9}$$

The conditions for maximum or minimum, respectively can also be expressed by

$$-\left(\binom{n-2}{r-2} + 2\binom{n-2}{r-1} + \binom{n-2}{r}\right)p^2 \tag{8'}$$

$$+2\left(\binom{n-2}{r-2} + \binom{n-2}{r-1}\right)p - \binom{n-2}{r-2} \lessgtr 0, \tag{9'}$$

where the upper inequality sign indicates maximum and the lower one indicates minimum. By setting equality sign the above mentioned exception case ('undetermined') is obtained.

By solving the resulting quadratic equation we obtain:

$$\frac{-2\left(\binom{n-2}{r-2}+\binom{n-2}{r-1}\right) \pm \sqrt{4\left(\binom{n-2}{r-2}+\binom{n-2}{r-1}\right)^2 - 4\left(\binom{n-2}{r-2}+2\binom{n-2}{r-1}+\binom{n-2}{r}\right)\binom{n-2}{r-2}}}{-2\left(\binom{n-2}{r-2}+2\binom{n-2}{r-1}+\binom{n-2}{r}\right)}$$

$$= \frac{r}{n}\left(1 \mp \sqrt{1 - \frac{(r-1)}{r} \cdot \frac{n}{(n-1)}}\right) = \frac{r}{n}\left(1 \mp \sqrt{\frac{n-r}{r(n-1)}}\right) \begin{matrix} > p \text{ for maximum} \\ < p \text{ for minimum} \end{matrix} \tag{10}$$

We remark that the lower sign in front of the square root corresponds to an interchange of complementary probabilities (where p is replaced by $(1-p)$ and $(1-p)$ by p).

Furthermore, by a closer look one recognises that the mentioned exceptions, that appear as undetermined, in reality correspond to $P_r = \overline{P}_r$, ie. the equalization neither increases nor decreases the probability in case.

If, as a rule, we apply the upper sign, we will be on the safe side as long as we make sure to have

$$r\left(1 - \sqrt{1 - \frac{(r-1)}{r} \cdot \frac{n}{(n-1)}}\right) = r\left(1 - \sqrt{\frac{n-r}{r(n-1)}}\right) \geq np \tag{11}$$

The equality sign is only applied for the lowest r that appears in the calculation. If this condition is fulfilled for $r = x + 1$, then it is fulfilled for all higher values of r, and thus the required safety is obtained for all terms in $\overline{\psi}_{nx}$. For x circuits available for n subscribers with an average connection probability of p during the selected time (ie. on average *np simultaneous* 'speech connections' in the period) the condition for the simplified calculation method is:

$$(x+1)\left(1 - \sqrt{1 - \frac{x}{(x+1)} \cdot \frac{n}{(n-1)}}\right) = (x+1)\left(1 - \sqrt{\frac{n-x-1}{(x+1)(n-1)}}\right) \geq np \tag{12}$$

At this point we will assume this condition to be fulfilled. Should there be a question of choosing x below the given limit, one would have to check by closer scrutiny whether that would lead to any harmful error in the whole calculation. We have *before* the development of this general formula by an examination of all conceivable combinations for sets of 4, 5, 6, 7, 8 subscribers calculated the values that np for above-given reasons must not exceed for $r = 2, 3, \ldots, n-2$ ($n = 4, 5, 6, 7, 8$) and have found the results shown in the table below left, where C denotes the critical value of np.

As is easily seen, there is full agreement with the above formula, in which the value under the square root sign is never negative.

19. It is now possible for certain small subscriber sets to calculate the degree of hindrance for given values of $p = y/n$, where y as repeatedly mentioned is the average number of simultaneously existing telephone calls, and for given values of $x = r - 1$, where x is always a positive number. The results can for a given number n (n = number of subscribers) be presented in a table where the left column contains the values of x and the top row gives the values of y. If a value of $\overline{\psi}$, say $\overline{\psi} = 1/1000$ is given, and we want to find the values of y corresponding to certain values of x, then this can be done by interpolation, using

		r = 2	r = 3	r = 4	r = 5	r = 6
n = 4	C =	$2-2\sqrt{\frac{1}{3}}$				
n = 5	C =	$2-\sqrt{\frac{3}{2}}$	$3-\sqrt{\frac{3}{2}}$			
n = 6	C =	$2-2\sqrt{\frac{2}{5}}$	$3-3\sqrt{\frac{1}{5}}$	$4-2\sqrt{\frac{2}{5}}$		
n = 7	C =	$2-\sqrt{\frac{5}{3}}$	$3-\sqrt{2}$	$4-\sqrt{2}$	$5-\sqrt{\frac{5}{3}}$	
n = 8	C =	$2-2\sqrt{\frac{3}{7}}$	$3-\sqrt{\frac{15}{7}}$	$4-\sqrt{\frac{16}{7}}$	$5-\sqrt{\frac{15}{7}}$	$6-2\sqrt{\frac{3}{7}}$

the values of $\overline{\psi}$ in the table closest to the given value of $\overline{\psi}$ in the table closest to the given value of $\overline{\psi}$ for each x.

V The second simplification. Calculation of the degree of hindrance

20. It appears obvious from the formula for $\overline{\psi}$ that the calculation of a such table will be an increasingly long-lasting and tedious work with increasing n, and one will not reach very high numbers before the task becomes insuperable. Despite the simplified formula that only gives approximate results, the practical demands will necessitate further simplifications of the expression for the degree of hindrance. Here we will apply the procedure of Bortkewitsch.

21. Let us consider the first terms of $\overline{\psi}_{nx}$, namely

$$\overline{P}_{nr} = \binom{n}{r} p^r (1-p)^{n-r}$$

Here we will set $np = y$, ie. $p = y / n$. Then we will imagine that y is given a constant value equal to the given average number of simultaneous connections in busy time, however so that n exceeds any limit while p approaches 0. In other words, we imagine that the given set of n subscribers with a certain average requirement p of telephone use is replaced by a gradually increasing set of subscribers with a gradually decreasing requirement each, until we approach an infinite number of subscribers with an infinitely small requirement of telephone use each, however so that the average collective traffic is exactly the same as that of the real set of subscribers. Then we get:

$$\underset{n\to\infty}{\overline{P}_{nr}} = \frac{n(n-1)\ldots(n-r+1)}{1\cdot 2\ldots r}\left(\frac{y}{n}\right)^r\left(1-\frac{y}{n}\right)^{(n-r)}$$

$$= \underset{n\to\infty}{\frac{1}{r!}}\cdot\frac{n}{n}\cdot\frac{(n-1)}{n}\cdot\frac{(n-2)}{n}\ldots$$

$$\frac{(n-r+1)}{n}\cdot y^r\cdot\left(1-\frac{y}{n}\right)^{\frac{(n-r)}{(-y)}(-y)}$$

$$= \frac{1}{r!}\cdot y^r\cdot\left(1-\frac{y}{n}\right)^{\left(-\frac{n}{y}\right)(-y)} = \frac{1}{r!}y^r e^{-y}$$

where e is the base of the natural logarithm. The present expression:

$$\underset{n\to\infty}{\overline{p}} = \frac{1}{r!}y^r e^{-y} = p_{\infty r} \tag{13}$$

is Bortkewitsch' formula.

By introducing in the formula for $\overline{\psi}_{nx}$ the thus obtained values of $P_{\infty(x+1)}, P_{\infty(x+2)}, \ldots, P_{\infty(n-x)}$ we get for $n = \infty$:

$$\overline{\psi}_{\infty x} = \frac{1}{(x+1)!}y^{(x+1)}e^{-y}\Big(1+\frac{2y}{(x+2)}+\frac{3\cdot y^2}{(x+2)(x+3)} + \frac{4\cdot y^3}{(x+2)(x+3)(x+4)}+\ldots \frac{(n_0-x)}{(x+2)(x+3)\ldots(n_0)}y^{(n_0-x-1)}\Big)\frac{1}{y} \tag{14}$$

where n_0 denotes the real number of subscribers in the considered set. By this we have obtained a further and very essential simplification of the calculated degree of hindrance.

22. However, we must even here try to assess whether this change of calculation has any harmful effect or not, ie. whether $\overline{\psi}_{\infty x}$ is greater or not greater than $\overline{\psi}_{n_0 x}$. For this purpose we will once more study the expression

$$\overline{P}_{nr} = \binom{n}{r}p^r(1-p)^{(n-r)}$$

$$= \frac{1}{r!}\frac{n(n-1)(n-2)\ldots(n-r+1)}{n\cdot n\cdot n\ldots n}y^r\left(1-\frac{y}{n}\right)^{n\left(1-\frac{r}{n}\right)}$$

$$= \frac{y^r}{r!}\cdot\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{r-1}{n}\right)\left(1-\frac{y}{n}\right)^{n\left(1-\frac{r}{n}\right)}$$

From this is obtained:

$$\log\overline{P}_{nr} = \log\frac{y^r}{r!}+\log\left(1-\frac{1}{n}\right)+\log\left(1-\frac{2}{n}\right)\log\left(1-\frac{3}{n}\right) +\ldots+\log\left(1-\frac{r-1}{n}\right)+n\left(1-\frac{r}{n}\right)\log\left(1-\frac{y}{n}\right) \tag{15}$$

Furthermore ($\log P$ denotes here log nat P):

$$
\begin{aligned}
&\frac{d}{dn}\log \overline{P}_{nr}\\
&=\frac{1}{n^2}\left(\frac{1}{\left(1-\frac{1}{n}\right)}+\frac{2}{\left(1-\frac{2}{n}\right)}+\frac{3}{\left(1-\frac{3}{n}\right)}+\ldots+\frac{(r-1)}{\left(1-\frac{r-1}{n}\right)}\right)\\
&+\frac{y}{n}\frac{\left(1-\frac{r}{n}\right)}{1-\frac{y}{n}}+\log\left(1-\frac{y}{n}\right)\\
&=\frac{1}{n^2}(1+\frac{1}{n}+\frac{1}{n^2}+\frac{1}{n^3}+\ldots\\
&+2\left(1+\frac{2}{n}+\frac{2^2}{n^2}+\frac{2^3}{n^3}+\ldots\right)\\
&+3\left(1+\frac{3}{n}+\frac{3^2}{n^2}+\frac{3^3}{n^3}+\ldots\right)\ldots\\
&+(r-1)\left(1+\frac{r-1}{n}+\frac{(r-1)^2}{n^2}+\frac{(r-1)^3}{n^3}+\ldots\right))\\
&+\frac{y}{n}\left(1+\frac{y}{n}+\frac{y^2}{n^2}+\frac{y^3}{n^3}+\ldots-\frac{r}{n}\left(1+\frac{y}{n}+\frac{y^2}{n^2}+\frac{y^3}{n^3}+\ldots\right)\right)\\
&-\left(\frac{y}{n}+\frac{y^2}{2n^2}+\frac{y^3}{3n^3}+\ldots\right)\\
&=\frac{1}{n^2}\left(1+2+3+\ldots+(r-1)+y^2-ry-\frac{y^2}{2}\right)\\
&+\frac{1}{n^3}\left(1+2^2+3^2+\ldots+(r-1)^2+y^3-ry^2-\frac{y^3}{3}\right)\\
&+\frac{1}{n^4}\left(1+2^3+3^3+\ldots+(r-1)^3+y^4-ry^3-\frac{y^4}{4}\right)\\
&+\ldots\\
&+\frac{1}{n^{s+1}}\left(1+2^s+3^s+\ldots+(r-1)^s+y^{s+1}-ry^s-\frac{y^{s+1}}{s+1}\right)+\ldots\\
&=\frac{1}{n^2}\left(\frac{(r-1)r}{1\cdot 2}+\frac{1}{2}y^2-ry\right)\\
&+\frac{1}{n^3}\left(\frac{(r-1)r(2r-1)}{1\cdot 2\cdot 3}+\frac{2}{3}y^3-ry^2\right)\\
&+\frac{1}{n^4}\left(\left(\frac{(r-1)r}{1\cdot 2}\right)^2+\frac{3}{4}y^4-ry^3\right)\\
&+\ldots\\
&+\frac{1}{n^{s+1}}\left(symb.\frac{((r-1)+B)^{s+1}-B^{s+1}}{s+1}+\frac{s}{s+1}y^{s+1}-ry^s\right)\\
&+\ldots,
\end{aligned}
\tag{15'}
$$

where in the expression after 'symb.' the notation B^s, B^{s-1}, B^{s-2}, ..., B^1 should be replaced by B_s, B_{s-1}, B_{s-2}, ..., B_1, which is the notation for Bernoulli numbers. If n in the expression for $\frac{d}{dn}\log \overline{P}_{nr}$ is made sufficiently large, then the value of the expression will be essentially determined by the first term

$$\frac{1}{n^2}\left(\frac{(r-1)r}{1\cdot 2}+\frac{1}{2}y^2-ry\right).$$

This term is positive when

$$y^2 - 2ry + (r-1)r > 0.$$

By resolving the equation

$$y^2 - 2ry = -r^2 + r$$

is obtained

$$y = r \pm \sqrt{r}.$$

According to the nature of the problem we have $y < r$, and hence we must choose the solution

$$y = r - \sqrt{r}, \text{ or}$$

$$r = y + \frac{1}{2} + \sqrt{y + \frac{1}{4}} \tag{16}$$

Thus for these values of y and r the term mentioned is equal to 0.

If $y < r - \sqrt{r}$ or $r > y + \frac{1}{2} + \sqrt{y + \frac{1}{4}}$, then the term is positive; if $y > r - \sqrt{r}$ or $r < y + \frac{1}{2} + \sqrt{y + \frac{1}{4}}$, then the term is negative.

If in the second term $y < r - \sqrt{r}$, then this term is positive when $r > (4/3)^2$. In a similar way one can find corresponding conditions for the following terms to be positive. There is, however, no need for a closer study of this, since the character of the task implies that $n >> r$, and besides, the following terms can be given the form $(r/n)^4 \cdot \Theta_4$, $(r/n)^5 \cdot \Theta_5$, ..., $(r/n)^{s+1} \cdot \Theta_{s+1}$, ..., where Θ_4, Θ_5, ..., Θ_{s+1}, ... are proper fractions, so that the sum of these terms can be neglected in comparison with the first two terms. Thus we content ourselves by establishing as a sufficient condition for

$$\frac{d}{dn}\log \overline{P}_{nr} > 0, \text{ that}$$

$$\left(\begin{array}{l} y < r - \sqrt{r}, \text{ or} \\ r > y + \frac{1}{2} + \sqrt{y + \frac{1}{4}}, \text{ and } r << n \end{array}\right. \tag{17}$$

However, if $\frac{d}{dn}\log \overline{P}_{nr} > 0$, then $\overline{P}_{nr}$ increases with increasing n, and under the given conditions we have

$$P_{\infty r} > \overline{P}_{nr}.$$

From this follows immediately $\overline{\psi}_{\infty x} > \overline{\psi}_{n_0 x}$, and when the stated conditions are fulfilled we can be sure that the simplified calculation does not lead to any harmful error.

23. It now remains to add the series that is contained in $\overline{\psi}_{\infty x}$. As can be seen it concerns a finite series. Since $n_0 - x$ normally is a number of more than one digit, the series may contain a substantial number of terms. Also, in some likely practical cases the initial part of the series may consist of increasing terms from left to right. Thus, in general the calculation must include a considerable number of terms in order to obtain sufficient exactness. We must, therefore, find a way of approximate calculation of this series in order to determine the degree of hindrance for commonly occurring (greater) values of n_0, y and x.

To this end we will study the infinite series:

$$
\begin{aligned}
&1+\frac{2y}{(x+2)}+\frac{3y^2}{(x+2)(x+3)}+\frac{4y^3}{(x+2)(x+3)(x+4)} \\
&+\frac{(n_0-x)}{(x+2)(x+3)\ldots(n_0)}\cdot y^{n_0-x-1}+\ldots \\
&\frac{n_0}{(x+2)(x+3)\ldots(x+n_0)}y^{n_0-1}+\ldots\infty \\
&=\frac{\overline{\psi}_{\underline{\infty}x}(x+1)!e^y}{y^x} \\
&=1+\frac{2y}{(x+k)}+\frac{3y^2}{(x+k)^2}+\frac{4y^3}{(x+k)^3}+\ldots \\
&+\frac{n_0 y^{n_0-1}}{(x+k)^{n_0-1}}+\frac{(n_0+1)y^{n_0}}{(x+k)^{n_0}}+\ldots\infty \\
&=1+\frac{y}{(x+k)}+\frac{y^2}{(x+k)^2}+\frac{y^3}{(x+k)^3}+\frac{y^4}{(x+k)^4}+\ldots\infty \\
&\quad+\frac{y}{(x+k)}+\frac{y^2}{(x+k)^2}+\frac{y^3}{(x+k)^3}+\frac{y^4}{(x+k)^4}+\ldots\infty \\
&\qquad+\frac{y^2}{(x+k)^2}+\frac{y^3}{(x+k)^3}+\frac{y^4}{(x+k)^4}+\ldots\infty \\
&\qquad\quad+\frac{y^3}{(x+k)^3}+\frac{y^4}{(x+k)^4}+\ldots\infty \\
&\qquad\qquad+\frac{y^4}{(x+k)^4}+\ldots\infty \\
&\qquad\qquad\qquad+\ldots\infty \\
&=\frac{1}{1-\frac{y}{(x+k)}}+\frac{y}{(x+k)}\cdot\frac{1}{1-\frac{y}{(x+k)}}+\frac{y^2}{(x+k)^2}\cdot\frac{1}{1-\frac{y}{(x+k)}} \\
&+\frac{y^3}{(x+k)^3}\cdot\frac{1}{\left(1-\frac{y}{(x+k)}\right)}+\ldots\infty \\
&=\frac{1}{\left(1-\frac{y}{(x+k)}\right)^2},\ \text{where}
\end{aligned}
\tag{18}
$$

$$2<k<\frac{n_0-x}{4}+\frac{3}{2}, \tag{19}$$

as $\frac{n_0-x}{4}+\frac{3}{2}$ is the arithmetic mean of the n_0-x-1 values $\left(\frac{2}{1}\right), \left(\frac{2+3}{2}\right), \left(\frac{2+3+4}{3}\right), \ldots, \left(\frac{2+3+4+\ldots+(n_0-x)}{n_0-x-1}\right)$.

Thus we have:

$$\overline{\psi}_{\underline{\infty}x}=\frac{1}{(x+1)!}\cdot\frac{y^x}{e^y}\cdot\frac{1}{\left(1-\frac{y}{(x+k)}\right)^2} \tag{20}$$

By choosing k such that equation (18) is satisfied we get:

$$
\begin{aligned}
\overline{\psi}_{\infty x}&=\frac{1}{(x+1)}\cdot\frac{y^x}{e^y}\left(1+\frac{2y}{(x+2)}+\frac{3y^2}{(x+2)(x+3)}+\ldots+\frac{(n_0-x)y^{n_0-x-1}}{(x+2)\ldots(n_0)}\right) \\
&=\overline{\psi}_{\underline{\infty}x}\left(1-\left(\frac{n_1 y^{n_1-1}}{(x+k)^{n_1-1}}+\frac{(n_1+1)}{(x+k)^{n_1}}y^{n_1}+\frac{(n_1+2)}{(x+k)^{n_1+1}}y^{n_1+1}+\ldots\infty\right)\right) \\
&=\overline{\psi}_{\underline{\infty}x}\left(1-\frac{y^{n_1-1}}{(x+k)^{n_1-1}}\cdot\frac{n_1}{\left(1-\frac{y}{x+k}\right)^2}\left(1-\frac{y}{x+k}\left(1-\frac{1}{n_1}\right)\right)\right).
\end{aligned}
$$

where $n_1 = n_0 - x - 1$.

Thus we have:

$$
\begin{aligned}
\overline{\psi}_{\infty x}&=\frac{1}{(x+1)!}\cdot\frac{y^x}{e^y}\cdot\frac{1}{\left(1-\frac{y}{x+k}\right)^2} \\
&\left(1-\frac{y^{n_1-1}}{(x+k)^{n_1-1}}\cdot\frac{n_1}{\left(1-\frac{y}{x+k}\right)^2}\left(1-\frac{y}{x+k}\left(1-\frac{1}{n_1}\right)\right)\right)
\end{aligned}
\tag{21}
$$

We can always imagine a number h determined so that:

$$\overline{\psi}_{\infty x}=\frac{1}{(x+1)!}\cdot\frac{y^x}{e^y}\cdot\frac{1}{\left(1-\frac{y}{x+h}\right)^2} \tag{22}$$

It is evident that the number h must be greater than k, and consequently from the previous $h > 2$. The number 2 may thus be considered as a 'lower limit' for h. It is of course also possible to find an 'upper limit' for h by means of the previously given limitation of k. We will, however, be content with a limitation that can be immediately put forward by studying the expression for $\overline{\psi}_{\infty x}$ as a function of k, as at least for considerable values of n_1 and x, ($y < x << n_1$), h will be less than $2k$, so

$$h<\frac{n_0-x}{2}+3$$

Hence, we have got useful means available for calculation of $\overline{\psi}_{\infty x}$, as we have:

$$\overline{\psi}_{\infty x} < \frac{1}{(x+1)!}\cdot\frac{y^x}{e^y}\cdot\frac{1}{\left(1-\frac{y}{x+2}\right)^2} \quad \text{and} \tag{23}$$

$$\overline{\psi}_{\infty x} > \frac{1}{(x+1)!}\cdot\frac{y^x}{e^y}\cdot\frac{1}{\left(1-\frac{y}{x+\frac{n_0-x}{2}+3}\right)^2} \tag{24}$$

The first of these two formulas is seen to be independent of n_0. It can, therefore, be used for any number of subscribers. The expression to the right of the inequality sign gives us, by insertion of actual values of x and y, always a number that is larger than the real degree of hindrance. Thus the application will never lead to a harmful error in the calculation. The second formula is of less importance, apart from the possibility of finding the number of decimals that can be correctly calculated by means of the first expression. In order to clarify this the number of subscribers must be known. Besides, it must be pointed out that the upper limit of h is set very high. For both reasons the second formula will probably be of little practical importance. For the present we will only abide by the first formula. As we now omit as superfluous the indication of a fictitious infinite number of subscribers, each with an infinitely little connection probability, and also consider that we have a function of two variables x and y (x always an integer, $0 < y < x$), we will write:

$$\overline{\psi}_{xy} = \frac{1}{(x+1)!}\cdot\frac{y^x}{e^y}\cdot\frac{1}{\left(1-\frac{y}{x+2}\right)^2} \tag{25}$$

where we then have $\overline{\psi}_{xy} > \overline{\psi}_{\infty x}$.

y = x =	0,5	1	1,5	2	2,5	3	...
1	$\overline{\psi}_{1(0,5)}$	$\overline{\psi}_{1,1}$	...	...	...	$\overline{\psi}_{1,3}$	...
2	$\overline{\psi}_{2(0,5)}$	$\overline{\psi}_{2,1}$	...	...	...	$\overline{\psi}_{2,3}$	...
3	$\overline{\psi}_{3(0,5)}$	$\overline{\psi}_{3,1}$	...	...	...	...	...
4	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

24. We have now obtained an expression that is well applicable for an approximate calculation of the degree of hindrance for various values of x and y. The results can be presented in a table like the one below left.

Since in practice only small values of $\overline{\psi}_{xy}$ will be prescribed, for instance between 1 % and 0.1 ‰ (0.1 per mill), the calculations need only comprise certain corresponding values of x and y selected in such a way as to fulfil the previously considered conditions.

When a suitable part of this table is filled in, it will be possible for all integer x to determine by interpolation the values of y that correspond to a degree of hindrance of 1 ‰, 2 ‰, 3 ‰ etc. By this means one can construct a 1 ‰ curve, a 2 ‰ curve, a 3‰ curve etc. If for a certain set n_0 of subscribers the value of y ($y = n_0 \cdot p$) is known, one can easily from the given curve read the number of exchange circuits or switches that corresponds to the degree of hindrance that is considered suitable.

25. In order to simplify the calculation when it is not done by successive use of the values of 2!, 3!, 4!, ..., $(x + 1)!$, one can for higher values of x use Stirling's formula:

$$(x+1)! = \sqrt{2\pi(x+1)}\,(x+1)^{x+1}\,e^{-(x+1)+\frac{\Theta}{12(x+1)}},$$

where Θ is a proper fraction. Hence we get:

$$\begin{aligned}\overline{\psi}_{xy} &= \frac{1}{\sqrt{2\pi(x+1)}\,(x+1)^{x+1}}\,\frac{e^{x+1-\frac{\Theta}{12(x+1)}}}{e^y}\cdot\frac{y^x}{\left(1-\frac{y}{x+2}\right)^2}\\ &= \frac{1}{\sqrt{2\pi(x+1)}}\cdot\left(\frac{e}{x+1}\right)^{x+1}\cdot\frac{y^x}{e^{y+\frac{\Theta}{12(x+1)}}}\cdot\frac{1}{\left(1-\frac{y}{x+2}\right)^2}\end{aligned} \tag{26}$$

Thus we have:

$$\overline{\psi}_{xy} < \frac{1}{\sqrt{2\pi(x+1)}}\left(\frac{e}{x+1}\right)^{x+1}\cdot\frac{y^x}{e^y}\cdot\frac{1}{\left(1-\frac{y}{x+2}\right)^2} \tag{26'}$$

and

$$\overline{\psi}_{xy} > \frac{1}{\sqrt{2\pi(x+1)}}\left(\frac{e}{x+1}\right)^{x+1}\cdot\frac{y^x}{e^{y+\frac{1}{12(x+1)}}}\cdot\frac{1}{\left(1-\frac{y}{x+2}\right)^2} \tag{26''}$$

We set:

$$\psi_{xy} = \frac{1}{\sqrt{2\pi(x+1)}}\,\frac{y^x}{(x+1)^{x+1}}\,\frac{e^{x+1-y}}{\left(1-\frac{y}{x+2}\right)^2}, \tag{27}$$

where then $\psi_{xy} > \overline{\psi}_{xy}$.

Neither with the application of this in some cases more convenient expression do we commit any harmful error. In the calculation of $\overline{\psi}_{xy}$ as well as ψ_{xy} we make use of logarithms, as we have:

$$\log \overline{\psi}_{xy} = -\log 2 - \log 3 - \mathrm{K} - \log(x+1) + x \log y$$
$$-y \log e - 2\log(x+2-y) + 2\log(x+2), \qquad (25')$$

$$\log \psi_{xy} = -\frac{1}{2}\log \pi - \frac{1}{2}\log 2 - \frac{1}{2}\log(x+1)$$
$$-(x+1)\log(x+1) + x\log y + (x+1-y)\log e$$
$$-2\log(x+2-y) + 2\log(x+2) \qquad (27')$$

If one wants to know the number of correct decimals in the calculated degree of hindrance, one can apply the given expression where Θ is set equal to 1. This control may with sufficient usefulness for practical applications be limited to values for the degree of hindrance in the vicinity of the prescribed magnitude.

By means of formula (14) in the present exposition and Bortkewitsch' tables[2] we have calculated a table (Annex I) over the degree of hindrance for all values of x from 1 to 19 with application of the values of y = 1/2, 1, 1.5, ..., 10 that imply a degree of hindrance close to that prescribed. Since Bortkewitsch only allows for 4 decimals, and because of that only a few terms of the series in question can be expressed by means of his tables, and since, as explained, the series has only a weak convergence, our table can not present the degree of hindrance with great exactness. In order to further elucidate the matter we have, therefore, calculated $\overline{\psi}_{xy}$ with 7 decimals by means of formula (25) for x = 4, 7, 10, 13, 16, 19, 22 and some of the mentioned values of y.

As can be seen from the thus produced table (Annex II) the deviations between corresponding values in the two tables are not significant. From this we can conclude that we do not cause any great error when producing a 1 ‰ curve by applying formula (25), that is deduced from the previously given more exact expressions by setting $h = k = 2$.

VI Other procedures for elucidation of the likely efficiency. Comparison of the different expressions for the allowable hindrance

26. We have in the foregoing presented a method, by which one can express the ability or lack of ability of an automatic telephone exchange to carry a certain amount of traffic. There exist also other methods that can be applied for this purpose.

Instead of considering the part of traffic that can be expected to be carried or hindered when offered to a certain number of parallel central exchange circuits or switches, one can for instance attempt to find the probability of all those circuits being occupied, ie. the *probability of full utilisation*[3], or one can, if preferred, seek the probability that under full utilisation there are requirements that can not be satisfied, ie. the *probability of overfill*. For both of those probabilities apply, although not to the same extent, that the smaller they are, the greater is the carrying capacity of the system.

The first mentioned full utilisation probability with x circuits for traffic y in busy hour will be denoted $P_{(xy)}$, and the last mentioned overflow probability with x circuits for traffic y in busy hour P_{xy}.

(28). Let us regard the first of these probabilities. As we keep to what is stated in the sections of the first and the second simplifications we will have:

$$P_{(xy)} =$$
$$\frac{1}{x!}\frac{y^x}{e^y}\left(1 + \frac{y}{(x+1)} + \frac{y^2}{(x+1)(x+2)} + \frac{y^3}{(x+1)\mathrm{K}\,(x+3)} + \mathrm{K}\right) \qquad (28)$$

This series is strongly converging, and it will therefore be of practically no significance whether the addition within the parenthesis is done up to and inclusive the term

$\frac{y^{(n_0-x)}}{(x+1)(x+2)\mathrm{K}\; n_0}$, where n_0 denotes the real subscriber number, or one regards the series as infinite and seeks the corresponding sum. One can even in many cases obtain sufficient accuracy by only adding a limited number of terms.

As stated, $P_{(xy)}$ gives us the probability that all x circuits or switches are occupied. Such a state can be caused by either x or $(x + 1)$ or $(x + 2)$ or ... or n_0 simultaneous conversations or calls. $P_{(xy)}$ does not directly give a measure of the ratio of the telephone activity that is hindered. This appears from the fact that the first and largest term of the expression, $\frac{1}{x!}\cdot\frac{y^x}{e^y}$, gives the probability of x calls, of which none are hindered, since there are precisely the sufficient number of x circuits, while the terms following the second do not indicate the number of calls that are hindered under the probability states given by those terms. The first term indicates the probability, ie. the relative frequency, of a state where the utilisation limit is reached, which may not be exceeded without traffic hindrances, and the following terms indicate the probabilities of the states that are associated with such hindrances. In combination these probabilities represent a measure of a certain admissible shortcoming of the system, and the expression for $P_{(xy)}$ may thus very well be applied for evaluation of the usefulness of the system, when one fully realises what it expresses. $P_{(xy)} = 0.001$ thus does not indicate an average likely loss of 1 call out of 1000, but the likelihood that all x circuits or switches are fully occupied an average of 1 second out of 1000. In a similar way as that previously dealt with one can construct 1 ‰, 2 ‰, 3 ‰, ..., r ‰ curves for the probability or risk of full load $P_{(xy)}$ for determination of the number of switches x for a certain given traffic y. This is already known to have been carried out by Master Erlang and engineer Christensen in Copenhagen[4].

As an annex to the present account a table is given, by which a 1 ‰ curve is drawn for ψ_{xy} and a 1 ‰ and a 4 ‰ curve are

[2] *See 'Das Gesetz der kleinen Zahlen'.*

[3] *In XII Summary called 'blocking'.*

[4] *See 'Elektrotechnische Zeitschrift', H. 46, 1913.*

drawn for $P_{(xy)}$. It can be seen that the curves for ψ_{xy} and $P_{(xy)}$ have the same character, however to a given x and a given ‰ value corresponds a significantly larger y for ψ_{xy} than for $P_{(xy)}$.

It appears from the foregoing that there is no mutual disagreement between the expressions for ψ_{xy} and $P_{(xy)}$. They have both their justification, their theoretical and practical significance. The difference rests in the fact that they are measures of different things. It is a matter of opinion what measure to choose and what degree of hindrance or full load risk one will allow for the establishment of a telephone system.

28. We will now consider the overload probability:

$$P_{xy}=\frac{1}{(x+1)!}\frac{y^{x+1}}{e^{y}}\left(1+\frac{y}{(x+2)}+\frac{y^{2}}{(x+2)(x+3)}+\frac{y^{3}}{(x+2)\ldots(x+4)}+\ldots\right)$$

$$=P_{(xy)}-\frac{1}{x!}\frac{y^{x}}{e^{y}} \tag{29}$$

As already mentioned, $\frac{1}{x!}\cdot\frac{y^{x}}{e^{y}}$ denotes the probability of exactly x simultaneous telephone calls (conversations). We now assume that in the course of these x conversations a new call arrives. Then one of two cases will occur: Either will this call, when it gradually reaches the x circuits end tests if they are busy, happen to replace a conversation just being concluded, in which case no hindrance will occur. Or there will not occur such a conclusion, as all x conversations continue and thus prevents the new call from obtaining a connection. In the latter case we say that the call is lost. In order to succeed a *new* call must be initiated, provided there is no remedy, by which new tests can be made without a new call. More about this below. However, in the case of x conversations in progress and one call unsuccessfully searching for a free circuit or switch, then the matter is no longer the state, whose probability of existence is

$$\frac{1}{x!}\cdot\frac{y^{x}}{e^{y}}$$

but a state whose probability of existence is

$$\frac{1}{(x+1)!}\cdot\frac{y^{x+1}}{e^{y}}$$

This is the state that implies the first case of hindrance, The second and somewhat less frequent case is implied by the probability

$$\frac{1}{(x+2)!}\cdot\frac{y^{x+2}}{e^{y}}$$

The third case:

$$\frac{1}{(x+3)!}\cdot\frac{y^{x+3}}{e^{y}} \quad \text{Etc.}$$

The sum of these probabilities gives the alternative or total probability of overflow or for the occurrence of hindrances:

$$P_{xy}=\frac{1}{(x+1)!}\frac{y^{x+1}}{e^{y}}\left(1+\frac{y}{(x+2)}+\frac{y^{2}}{(x+2)(x+3)}+\ldots\right) \tag{29}$$

This quantity thus gives a measure of the likely frequency of real hindrances. As a concept it is therefore to be preferred for $P_{(xy)}$. The two expressions are also different because they indicate different things. The difference between them is the first term of $P_{(xy)}$.

29. It is interesting to observe that

$$P_{xy}=\frac{1}{(x+1)!}\frac{y^{x+1}}{e^{y}}\left(1+\frac{y}{(x+2)}+\frac{y^{2}}{(x+2)(x+3}+\ldots\right)$$

$$=1-e^{-y}\left(1+\frac{y}{1}+\frac{y^{2}}{1\cdot 2}+\frac{y^{3}}{1\cdot 2\cdot 3}+\ldots+\frac{y^{x}}{1\cdot 2\cdot 3\ldots x}\right)$$

[5]

$$=\int_{0}^{y}\frac{1}{x!}e^{-y}y^{x}\,dy$$

The state probability of x simultaneous telephone conversations for a traffic intensity y:

$$p_{xy}=\frac{1}{x!}e^{-y}y^{x},\quad \left(p_{xy}=p\infty x\right)$$

is thus generative with respect to the whole overflow probability. Assume a rectangular co-ordinate system with axis x and ordinate p_{xy} and construct the curve corresponding to $x!\cdot p_{xy}=f_x(y)=e^{-y}y^x$, ($x$ denotes a *constant,* positive integer). Then according to the given integral $P_{xy}=1/x!$ times the area that is limited by the curve, the abscissa and the ordinate p_{xy} for the given value of y. This is assumed not to exceed the earlier given limits. Having constructed a set of such curves on graph paper, corresponding to relevant x values, one can easily, by counting the squares related to the given values of y obtain a fairly exact indication of the concerned values of P_{xy}.

We have derived P_{xy} independently of this integration formula. It is, however, a question, that need not to be answered here, whether the same clarity of the relevant phenomena can be obtained by this integration, with traffic intensity varying between 0 and y, as with the previously presented operations based on general considerations of various state probabilities under a given average traffic intensity y.

As P_{xy} can be formed by integration from 0 to y of Bortkewitsch' formula, so also integration of $P_{(xy)}$ dy between the same limits leads to $y\cdot\overline{\psi}_{\underline{\infty}xy}$, where $\overline{\psi}_{\underline{\infty}xy}$ has the previously indicated meaning, and $y\cdot\overline{\psi}_{\underline{\infty}xy}$ thus indicates the total number of telephone conversations that on average are hindered or lost in any moment of the busy hour. This can easily be verified by differentiation:

[5] *Remark: The author of the present report has been made aware of this relation by a representative of Western Electric Co., after having derived P_{xy} by general probability considerations.*

$$\frac{d}{dy}\left(y\cdot\overline{\psi}_{\underline{\overline{\infty}}xy}\right)$$

$$=\frac{1}{(x+1)!}\frac{d}{dy}\left(\left(y^{x+1}+\frac{2}{(x+2)}y^{x+2}+\frac{3}{(x+2)(x+3)}y^{x+3}+\ldots\right)e^{-y}\right)$$

$$=\frac{1}{(x+1)!}e^{-y}\Big(\big((x+1)y^{x}+\frac{2(x+2)}{(x+2)}y^{x+1}+\frac{3(x+3)}{(x+2)(x+3)}y^{x+2}$$

$$+\frac{4(x+4)}{(x+2)\ldots(x+4)}y^{x+3}+\ldots-y^{x+1}-\frac{2}{(x+2)}y^{x+2}$$

$$-\frac{3}{(x+2)(x+3)}y^{x+3}-\ldots\Big)$$

$$=\frac{1}{x!}e^{-y}\left(y^{x}+\frac{y^{x+1}}{(x+1)}+\frac{y^{x+2}}{(x+1)(x+2)}+\frac{y^{x+3}}{(x+1)\ldots(x+3)}+\ldots\right)$$

$$=P_{(xy)}$$

Thus

$$\int_0^y P_{(xy)}dy=y\cdot\overline{\psi}_{\underline{\overline{\infty}}xy}$$

The quantity $y\cdot\overline{\psi}_{\underline{\overline{\infty}}xy}$ is thus generated by definite integration of $P_{(xy)}\,dy = P_{x-1,y}$ dy between 0 and y, which in a mathematical sense can be said to illustrate the significance of $P_{(xy)}$.

According to the preceding we have:

$$\int_0^y \frac{y^{(x-1)}}{(x-1)!}e^{-y}dy=P_{x-1,y}$$

Thus we have the following remarkable relation:

$$\frac{1}{y}\int_0^y dy\int_0^y \frac{y^{(x-1)}}{(x-1)!}e^{-y}dy=\overline{\psi}_{\underline{\overline{\infty}}xy};$$

$$\frac{d^2}{dt^2}\left(y\cdot\overline{\psi}_{\underline{\overline{\infty}}xy}\right)=\frac{y^{x-1}}{(x-1)!}e^{-y}.$$

This integral can be given a certain geometric interpretation such that the degree of hindrance $\overline{\psi}_{\underline{\overline{\infty}}xy}$ is depicted as $\frac{1}{y}$ times the volume of a spatial region, the form and dimensions of which are determined by the expression $\frac{y^{x-1}}{(x-1)!}e^{-y}$. (The region is for instance limited by the three planes of a three-dimensional rectangular co-ordinate system and a curved surface). The state probability given by this expression, namely the probability that $(x-1)$ circuits are occupied by the traffic y, thus has a remarkable mathematical significance in relation to the phenomena dealt with here. We will, however, not go any further into this matter. By means of tools for mechanical integration this integral formula might be of practical use for derivation of tables and curves that we need. Thus one could also avoid the errors in the calculation due to the fact that one cannot in advance set an exact value for the quantities h and k discussed earlier. These have been set equal to 2, which results in a calculated degree of hindrance somewhat larger than the probable value.

VII Objections to the applicability of the calculations. Calculations on a different basis

30. Until now we have during our discussion of the topic at hand mainly had in view the demands that the total traffic from a set of subscribers can be expected to make at any moment of the selected time or the busy hour. We have with the degree of exactness that we have been able to obtain tried to find expressions for determination of the part of the traffic that due to the lack of circuits can not be expected to be carried. We have not set forth any conditions with respect to the technical arrangement that should make access possible, or which in case of insufficient equipment for the total traffic should turn away the hindered traffic. Furthermore we have not, as already stated, made any hypothesis about the duration of each telephone connection, nor have we distinguished between the single time intervals within each connection. There might be a question whether we during our efforts of treating the topic in a most general way have defined the task so that the derived solution also corresponds to what the technical equipment can attain in reality. It cannot be denied that some objections may be made. We shall explain this in more detail.

We have calculated the probability, approximately, that all available circuits are occupied by traffic from the set of subscribers. It is clear that as long as this state prevails all excess traffic is hindered. If, however, the system is arranged in a way that a subscriber who is not given access at once is allowed to *wait* for access without making a new call, then the hindrance will cease as soon as a circuit becomes free, and the hindrance will not lead to a ‘lost’ call. Considering the calculated degree of hindrance as a measure of what is lost and thus must be compensated for by new calls, then our calculation will not, if interpreted as an anticipated loss, correspond to the performance of a system that permits such waiting. A system of this type can even be arranged so that within certain traffic limits practically no calls are lost. This does not mean that there are no restrictions in such a system. On the contrary, the restrictions may in certain cases be serious enough for subscribers that have to wait for access. In that sense the calculated degree of hindrance may be applied to illustrate the efficiency of these systems.

We have on the other hand also calculated the probability that not all common circuits are occupied by traffic, and for the time that such a state prevails there is not assumed any hindrance. However, does this in general correspond to what happens under the actual traffic in busy hour? – We shall consider a simple case.

A subscriber sends a call request at a moment when all common circuits are occupied. If the system is not arranged to permit waiting, then the call and the connection that in an ideal manual system would follow immediately would get lost. Let us now imagine the case when a circuit becomes available immediately after the rejection of the call. Then there is in our calculation no longer any hindrance for the following portion of the call, which nevertheless in the actual system gets lost as a whole. Thus it seems like our calculation involves waiting, and even for a duration similar to that of an ordinary communication.

It should, however, be noted that with regard to this disagreement with the assumptions for the present system a certain compensation takes place. Certainly, the refused call is lost, and the planned conversation cannot take place unless a new call is made. But the real hindrance that occurred will in general benefit some other call to arrive shortly after, that will now succeed. If the calculated situation were realised, then the mentioned succeeding call would have been hindered. Yet so far there is doubt whether this compensation is complete. However, it should be taken into account that the degree of hindrance as estimated under the conditions comprehensively described earlier, which always will be fulfilled for small values of this degree, contain a significant security coefficient.

31. In order to clarify whether the mentioned compensation really can be assumed to be included in our expression for the degree of hindrance, we will now briefly carry out a calculation on the second basis that is available according to §2, regarding the sequence of calls and the ensuing occupation of the available circuits. The average duration of the connections plays a significant role in the calculations.

Let us again consider the total time DT_c. We assume that subscriber A_i wants $\hat{m}_i$ conversations during this time, so in DT_c he will initiate $\hat{m}_i$ calls. We subdivide DT_c in N_i equally long intervals such that the length of each interval is equal to the average duration of his connections, and this duration is denoted A_i's conversation (or connection) period t_i. In order that subscriber A_i may have his conversation demand satisfied, ie. that each call is followed by a conversation, there will for him in DT_c be only N_i possible calls, of which there will be for his communication $\hat{m}_i$ favourable or expected calls.

We denote the ratio $\frac{\hat{m}_i}{N_i}$ the *call probability* $\hat{p}_i$ of subscriber A_i. Thus we have

$$\hat{p}_i = \frac{\hat{m}_i}{N_i}$$

The *no-call probability* is:

$$1 - \hat{p}_i = \frac{N_i - \hat{m}_i}{N_i}$$

In the same way the call and no-call probabilities of any subscriber are found.

As we have earlier calculated the probability of the various connection states in randomly chosen *moments* in the total time DT_c, so we can now calculate the probability of the various connection states in randomly chosen *intervals* in DT_c of a duration equal to the average connection period $\bar{t}$ of the n subscribers.

The new calculations can be carried out in exactly the same way as the previous ones, and the state expressions $\hat{P}_0, \hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n$, that can be found as functions of $\hat{p}_1, \hat{p}_2, \hat{p}_3, \ldots, \hat{p}_n$ have exactly the same form as the expressions $P_0, P_1, P_2, \ldots, P_n$.

To the equalisation of $p_1, p_2, p_3, \ldots, p_n$, that is done for calculation purposes, corresponds the equalisation of $\hat{p}_1, \hat{p}_2, \hat{p}_3, \ldots, \hat{p}_n$, expressed by $\hat{p}_1 + \hat{p}_2 + \hat{p}_3 + \ldots + \hat{p}_n = n \cdot \hat{p} = \frac{\hat{m}_1}{N_1} + \frac{\hat{m}_2}{N_2} + \frac{\hat{m}_3}{N_3} + \ldots + \frac{\hat{m}_n}{N_n} = \left(\hat{m}_1 t_1 + \hat{m}_2 t_2 + \hat{m}_3 t_3 + \ldots + \hat{m}_n t_n\right) \frac{1}{DT_c} = y,$

since the total traffic divided by the total time gives the average number of simultaneous connections during this time. So we have

$$n\hat{p} = y = np,$$

meaning: y not only denotes the average number of *simultaneously* existing connections in the considered time, but also the average number of calls in the course of the average period $\bar{t}$ of this time.

32. We can now apply the same simplifications as those in the previous section for our calculations concerning the call events, as we have $\hat{p} = \frac{y}{n} = p$. We can let $\hat{p}$ converge towards zero while we let n increase towards infinity, but so that y is kept constant. Thus again we arrive at Bortkewitsch' formula, and by means of the results already obtained we can express the probability of 0, 1, 2, 3, ..., x, $x + 1$, ..., n_0 calls during the period $\bar{t}$, as well as the number of calls that under any of the characterised states would result in unhindered calls in an ideal manual system. The conditions for the applicability of these expressions for the purpose are also the same as before.

33. The question is now: How many of these calls will be refused in a telephone system with x common circuits without any waiting arrangement?

It is obvious that what blocks a subscriber's call (apart from possible circuit or device faults) is the occupation state of the x common circuits when the call is made.

As frequently mentioned we have calculated the probability of finding, at a random moment, more than x occupations in case there are more than x circuits, and the expression of this probability is derived without making any conditions about the duration of the individual connections. We have only assumed that the average traffic intensity is y, ie. the total number of traffic seconds divided by the total time remains unchanged. The state probability thus calculated is therefore independent of the duration of the *individual* connections, but dependent on the collective duration. Nothing prevents us, therefore, from applying our expressions for connection states on a hypothetical traffic, where all connections last equally long, while the average duration is the same as for the real traffic. The mathematical state probabilities will also be the same for the two types of traffic.

What is said here with respect to the duration of connections also applies to the state probabilities relating to the various call states. Even these probabilities remain the same if we allocate the same duration $\bar{t}$ to all calls. This is also immediately evident when we remember the definition of a call probability and the corresponding determination of a call state.

34. We will now consider our hypothetical traffic with respect to call states as well as connection states. If $x + 1$ calls have arrived in a certain period $\bar{t}$, then between the x^{th} and $(x + 1)^{\text{th}}$ call there must necessarily be a connection state in which all x common circuits are occupied. If there are no more than x such circuits, then the $(x + 1)^{\text{th}}$ call must necessarily be refused, given that waiting is not admissible. The same will happen to all following calls in this period.

There is of course the possibility that one or more of the first x calls in the considered period will be refused because of connections initiated in the previous period. Here, however, it must be remembered that the starting point of our subdivision in periods is arbitrarily chosen, and nothing prevents us from shifting the whole subdivision. This can be done in infinitely many ways. In all cases when such a shift in relation to unhindered calls makes it possible to assemble more than x calls within a single period, refusal will occur for all calls arriving after the x^{th}. But the number of such cases is determined by the expressions of the probability of the relevant call states, as previously dealt with. On this basis the number of calls one must expect to be refused in the total – or if preferred, the selected – time can be calculated in fully the same way as that used for the previous calculations of the amount of the hindered traffic. If this number is divided by the total number of calls in the relevant time, then we have the degree of hindrance for calls. That is exactly equal to the degree of hindrance for the whole traffic.

Herewith it is fully clarified that our calculations apply to systems without waiting arrangement. The new calculation method also has the advantage that it makes obvious when there are hindrances and when not.

The question that was raised during the criticism of the first calculation method can now be answered: The compensation to be considered because of a certain divergence between the mathematical and the technical conditions is complete, even without considering the security factor that the applied method offers.

35. The task that we presented ourselves was: To derive on the basis of most general considerations a useful expression for the degree of hindrance of a telephone plant with a limited number of common exchange circuits or connection possibilities for a certain number of subscribers with a given communication request. Furthermore to prove the validity of the derived result on a probability theory basis, which of course presumes that the main theorems of probability are applicable to the considered phenomena. Despite the fact that we have solved the task mentioned, we will for the sake of completeness briefly consider a third method that can be applied and has been applied by others for the calculation of the probability of hindrances to telephone traffic due to lack of connection circuits.[6)]

One assumes the time $\bar{t}$, which is equal to the average duration of the telephone connections of the given set of subscribers, subdivided in such a large number of small intervals that only one call can arrive in any interval. It is assumed that no two or more calls arrive at exactly the same moment. (The opposite is still not unthinkable.) Insofar as a call arrives exactly on the limit between two intervals, one can for non-ambiguity assume that it belongs to the latter interval. The number of intervals in $\bar{t}$ is denoted n. The average number of calls from the subscribers in question in the busy time T_c can be denoted N. The average number of calls in the time section $\bar{t}$ will be $m = N \cdot \bar{t} / T_c$. There will thus in $\bar{t}$ be m intervals that are favourable for calls from the given number of subscribers and $n - m$ intervals that are non-favourable for calls. The probability of finding by random choice an interval affected by a call will be m/n, and the probability of finding an interval without a call will be $1 - m/n$.

Let us now assume that we some way or other can mark any interval affected by a call, so that at the end of $\bar{t}$ we can observe how the calls are distributed over the intervals. Then we can at the end of a randomly chosen period of length $\bar{t}$ calculate the probability of finding x marked intervals and accordingly $n - x$ non-marked intervals. Those x intervals may be distributed over $\bar{t}$ in many different ways.

Let us first find the probability for a fixed distribution, eg. the first x intervals marked and the following $n - x$ unmarked. The probability of this is according to the 2nd theorem equal to $\left(\frac{m}{n}\right)^x \left(1 - \frac{m}{n}\right)^{n-x}$. If we assume another distinct distribution of the x marked intervals, then the probability of the corresponding marking can also be expressed by exactly the same product, so it will also be equal to $\left(\frac{m}{n}\right)^x \left(1 - \frac{m}{n}\right)^{n-x}$. Actually, the factors for illustration of the phenomenon ought to be set in the same order as that of the distribution. However, this is of no significance for the result.

As will be known, there are $\frac{n(n-1)(n-2)\ldots(n-x+1)}{1 \cdot 2 \cdot 3 \ldots x}$ different ways of selecting x out of n items. The number of combinations of markings in the given example will thus be equal to

$$\frac{n(n-1)(n-2)\ldots(n-x+1)}{1 \cdot 2 \cdot 3 \ldots x} = \binom{n}{x}.$$

The total probability that either the first or the second or the third etc. or the $\binom{n}{x}^{\text{th}}$ marking combination will be observed, according to the first probability theorem will be:

$$\tilde{P}_x = \frac{n(n-1)(n-2)\ldots(n-x+1)}{1 \cdot 2 \cdot 3 \ldots x} \left(\frac{m}{n}\right)^x \left(1 - \frac{m}{n}\right)^{n-x}$$

The number n, being a definite integer, is chosen so that certain conditions for the intervals formed are fulfilled. There is nothing to prevent the choice of a higher number than the first

[6)] *Remark: Ie. by Grinsted. See also: Dr. F. Spiecker: Die Abhängigkeit des erfolgreichen Fernsprechanrufes von der Anzahl der Verbindungsorgane.*

$$\tilde{P}_x = \frac{n(n-1)(n-2)\ldots(n-x+1)}{1\cdot 2\cdot 3\ldots x}\left(\frac{m}{n}\right)^x\left(1-\frac{m}{n}\right)^{n-x}$$

The number n, being a definite integer, is chosen so that certain conditions for the intervals formed are fulfilled. There is nothing to prevent the choice of a higher number than the first adopted. However, in general $\tilde{P}_x$ does not remain unaffected by this. On the contrary, under certain conditions (see the section 'The second simplification') $\tilde{P}_x$ will increase with increasing n and approach a definite limit for $n = \infty$.

We will keep m unchanged and let n approach ∞. This leads to:

$$\tilde{P}_x = \frac{1}{x!}\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\ldots\frac{n-x+1}{n}$$
$$\cdot m^x\left(1-\frac{m}{n}\right)^{\left(-\frac{n}{m}+\frac{x}{m}\right)(-m)}$$
$$=\frac{1}{x!}m^x e^{-m},\quad (\lim n=\infty),$$

where e is the base of the natural logarithm.

This is again Bortkewitsch' formula, that we now have encountered a third time after having dealt with the subject from widely different aspects.

As indicated above $m = N\cdot\bar{t}/T_c$ and thus equal to the average traffic intensity, ie. we have $m = y$. Consequently, we obtain:

$$\tilde{P}_x = \frac{1}{x!}y^x e^{-y}$$
$$n\to\infty$$

36. We have thus again arrived at the expression $\frac{1}{x!}y^x e^{-y}$ for the probability of finding x calls during the period $\bar{t}$. 7)

This value of the state probability in question is derived without any consideration of the number of subscribers or the individual demands of each subscriber. We have only presumed the knowledge of or the consideration of the average number of calls in the selected time or the busy hour as well as the average duration of connections. Those two averages determine the average traffic intensity.

The probability of receiving more than x calls in the period $\bar{t}$ is also on this basis given by (nothing is presumed about the number of subscribers, it may be imagined to be a very great number):

$$\tilde{P}_{xy} = \frac{y^{x+1}}{(x+1)!}e^{-y}\left(1+\frac{y}{(x+2)}+\frac{y^2}{(x+2)(x+3)}+\ldots\right)$$

In order to determine the degree of hindrance only on the basis of this development all connections must be considered to be of equal duration.

Hence the degree of hindrance is obtained:

$$\tilde{\psi}_{xy} = \frac{y^{x+1}}{(x+1)!}e^{-y}\left(1+\frac{2y}{(x+2)}+\frac{3y^2}{(x+2)(x+3)}+\ldots\right)\frac{1}{y},$$

by dividing the hindered calls by the total number of calls.

We have put the mark ~ over the letters denoting the given probabilities and the degree of hindrance in order to emphasise the difference between the present and the previous developments. As one can observe, the final results are identical for the three procedures in question.

VIII Calculation of the degree of hindrance for the daily traffic

37. So far we have been engaged in the obstacles for the traffic process that occur within the selected time. We will now try to apply the obtained results for a calculation of the traffic volume that on average can be expected to be hindered over twenty-four hours, and the corresponding degree of hindrance. We will denote this the *general* or *daily* degree of hindrance, contrary to that previously discussed, which suitably may be denoted the *particular* or *momentary* (*short term*), or, as far as it is referred to the busiest time, the *maximum* degree of hindrance.

As can be remembered, when traffic intensity is y and the number of circuits or switches is x, we have found a usable expression of the traffic volume that can be expected to be hindered in the time unit (one second) within the selected time:

---

7) *Remark:* Dr. Spiecker *assumes a* constant *number* n *of calls per hour and thus arrives at a result for the state probability (formula (9)):*

$$W_a = \binom{n}{a}\left(\frac{1}{m}\right)^a\left(1-\frac{1}{m}\right)^{n-a},$$

which with the notation used by us can be expressed as follows (by Spiecker n/m *corresponds to our* y *and* a *to* x*):*

$$W_a = \binom{n}{a}\left(\frac{1}{m}\right)^a\left(1-\frac{1}{m}\right)^{n-a} =$$
$$\underline{P}_x = \binom{n}{x}\left(\frac{y}{n}\right)^x\left(1-\frac{y}{n}\right)^{n-x} =$$
$$\frac{1}{x}\frac{n(n-1)\ldots(n-x+1)}{(n-y)^x}y^x\frac{1}{\left(\left(1-\frac{y}{n}\right)-\frac{n}{y}\right)^y}$$

The value of this expression increases with increasing n *and is only transformed to* $\frac{1}{x!}y^x\frac{1}{e^y} = \tilde{P}_x$ *for* $n = \infty$.

Thus for any actually appearing traffic intensity his calculated probability $W_a = \underline{P}_x$ *is* smaller *than* $\tilde{P}_x$, *that is derived by assuming a number of calls only given as an* average *value in the busy hour.*

$$H_c = y\psi_{xy} = e^{-y}\left(\frac{y^{x+1}}{(x+1)!} + 2\frac{y^{x+2}}{(x+2)!} + 3\frac{y^{x+3}}{(x+3)} + \ldots\right)$$

$$= e^{-y}\frac{y^{x+1}}{(x+1)}\frac{1}{\left(1-\frac{y}{x+2}\right)^2} \text{ (approximately)}$$

Here, according to our first representation, H_c denotes the number of hindered traffic seconds ('conversation seconds', 'phone seconds') that can be expected in one second. The number can equally well denote the number of hindered traffic hours ('conversation hours', 'phone hours') that can be expected per hour. According to the representation in section VII, H_c can even denote the number of hindered calls in the period $\bar{t}$.

As the period T_c is arbitrarily chosen, the expression above may be considered valid for any period of the day as long as that period is short enough to grant only very small changes in the traffic relations.

As the hindered traffic in one second thus is H_c connection seconds, then in an infinitesimal time interval dt the hindered connection seconds are $H_c dt$, and in the period from $t = 0$ to $t = h$ within one day:

$$H_h = \int_0^h H_c dt \text{ connection seconds} \tag{31}$$

This gives an average per second:

$$\overline{H}_h = \frac{1}{h}\int_0^h H_c dt \text{ connection seconds} \tag{32}$$

Dividing this by the average second-traffic in the period h, which we will denote by y_h connection seconds, then the general degree of hindrance for the period h will be:

$$\psi_{xy_h} = \frac{\overline{H}_h}{y_h} = \frac{1}{hy_h}\int_0^h H_c dt = \frac{1}{hy_h}\int_0^h e^{-y}\left(\frac{y^{x+1}}{(x+1)!} + 2\frac{y^{x+2}}{(x+2)!} + \ldots\right)dt \tag{33}$$

If y could be given as a known time function of such form that the integration could be carried out, then the last equation would determine the general degree of hindrance ψ_{xy_h}.

In case y could be considered a linear function of t, this integration would be a simple matter. However, as well known the traffic rises and falls in a way that only approximately can be expressed by a function of a more complicated nature. By means of a large number of observations in a telephone exchange one might possibly propose a mathematical expression for y's dependence on t on average over the day, or one might, without any consideration of such a possible approximate mathematical relation construct a twenty-four hour traffic curve. The integration could then be carried out either by series expansion or by subdivision of the curve into a certain number of approximately linear parts. Both procedures will, apart from the carrying out of the observations, demand quite a substantial work. And the result of such work would hardly suffice for the estimation of the circumstances at other telephone plants.

38. Since we in the present report have set ourselves the task of treating the subject at issue in a most general way, we will even here try to find a solution that presumably will satisfy the demand of a wide applicability.

The problem may be expressed as follows, as we still consider the traffic of one day (twenty-four hours), regarded as the average of the traffic of many days:

When the traffic intensity y at time $t = 0$ is 0 and at time $t = h$ is $\bar{y} = y_{\max}$, how large is the traffic intensity $y = f(t)$ at an arbitrary time in between?

We assume for better overview that t is measured in hours.

It is sensible to search for the answer by means of Gauss' so-called error curve $y = ke^{-p^2x^2}$, and thus set $y = \bar{y}e^{-p^2t^2}$. Naturally, we must set the time so that we have $y = \bar{y}$ when $t = 0$. However, by closer study it turns out that if p is chosen to give a value close to zero during the night, then the concentration to the busiest hour of the day (ie. the ratio between the traffic of this hour and the total twenty-four hour traffic) will be too large.

Furthermore, one can proceed tentatively with Bortkewitsch' function by setting $y = \bar{y}qe^{-t}t^h\frac{1}{h!}$. Utilizing Stirling's formula one can write $y = \bar{y}e^{-t}t^h\left(\frac{e}{h}\right)^h$. Here $y = 0$ for $t = 0$ and $y = \bar{y}$ for $t = h$, as indicated in our question. The curve that according to this expression describes y as a function of t in a rectangular co-ordinate system has its peak at $t = h$, its zero points at $t = 0$ and $t = +\infty$, and its two inflexion points at $t = h \pm \sqrt{h}$. Thus it ought to have characteristics to make it useful for the planned calculations. However, when carried out it appears that it has similar drawbacks as Gauss' error curve. If h is determined so as to obtain a concentration according to normal traffic conditions, then the curve shows too high traffic at the end of day. Setting h so that traffic at the end of day is practically zero leads to an unreasonably high concentration.

These discrepancies are of course connected with the fact that the functions in question are derived or can be derived from pure probability considerations pertaining to a collection of items or events around a maximum point, without introducing in the formulas any coefficient taking into account the distribution of phenomena effected by certain unknown causes. We will in the modified Bortkewitsch' formula introduce such a coefficient, writing:

$$y = \bar{y}e^{-ct}t^{ch}\left(\frac{e}{h}\right)^{ch} \tag{34a}$$

To simplify we set:

$c = k/h$, leading to

$$y = \bar{y}\left(\frac{e}{h}\right)^k e^{-\frac{k}{h}t} t^k \tag{34b}$$

We will denote k the concentration coefficient.

39. It can be seen from the given expression for y that $y = 0$ for

for $t = 0$, $y = \bar{y}$ for $t = h$, $y = y_{\max}$ for $t = h$, while by

setting $\frac{d^2y}{dt^2} = 0$ the inflexion points of the curve are found

at $t = h\left(1 \pm \frac{1}{\sqrt{k}}\right)$.

It is just by this displacement of the inflexion points effected by the variation of k that concentration can be increased or

decreased (without changing the other parameters $\bar{y}$ and h)

so that the curve can be adapted to the existing or assumed conditions, at least for the time interval $t = 0$ to $t = h$. Thus one can set $k = 1, 2, 3, 4$, etc. and get the inflexion points for $t = h \pm h$,

$$t = h\left(1 \pm \frac{1}{\sqrt{2}}\right),\quad t = h\left(1 \pm \frac{1}{\sqrt{3}}\right),\quad t = h \pm \frac{h}{2},\ \text{etc.}$$

We will now imagine the working day subdivided in two parts, the first part from $t = 0$ to $t = h$ with increasing traffic, and the second part from $t = h$ to $t = 3h$ with decreasing traffic.

Let us set $k = 3$. The traffic volume for the day according to the

curve $y = \bar{y}\left(\frac{e}{h}\right)^3 e^{-\frac{3}{h}t} t^3$ will be:

$$M = \int_0^{3h} \bar{y}\left(\frac{e}{h}\right)^3 e^{-\frac{3}{h}t} t^3 dt = \bar{y}h\left(\frac{2e^3}{27} - e^{-6}\left(9 + 3 + \frac{2}{3} + \frac{2}{3^3}\right)\right)$$
$$= \bar{y}h \cdot 1.45623\text{K} \tag{35}$$

Assuming $h = 5$, corresponding to a workday of 15 hours, we

obtain $M = \bar{y} \cdot 7.282\text{K}$, so $\bar{y} = \frac{M}{7.28\text{K}} = M \cdot 0.137\text{K}$, giving

a concentration of 13.7 % which is a sensible ratio. If we assume $h = 4^2/_3$, corresponding to a workday of 14 hours, the concentration will be 14.7 %, which is higher than the usual level. If we assume a workday of 15 hours, then at the end of this time we will have a traffic intensity:

$$y = \bar{y}\left(\frac{e}{5}\right)^3 e^{-9} 3^3 5^3 = \bar{y}e^{-6} \cdot 3^3 = \bar{y} \cdot 0.00247875\text{K} \cdot 27$$
$$= \bar{y} \cdot 0.0669\text{K}$$

The curve thus shows that it has decreased to approx. 6.7 % of the maximum intensity. This last number must, however, be considered too high under normal circumstances. The curve does not fall fast enough for the evening traffic.

If, on the contrary, a concentration coefficient of 4, then we will get a lower final traffic, namely:

$$y = \bar{y}\left(\frac{e}{5}\right)^4 e^{-12} 3^4 5^4 = \bar{y}e^{-8} \cdot 3^4 = \bar{y} \cdot 0.00033546\text{K} \cdot 81$$
$$= \text{approx. } 2.7\% \text{ of } \bar{y}.$$

On the other hand will the concentration be far too high. Thus the curve will not give a fully satisfactory illustration of the fading of the traffic in the evening and at night.

Because of this we will only utilise the first part of the curve from $t = 0$ to $t = h$. For the afternoon traffic we will reverse the

curve (by turning it around the ordinate $\bar{y}$), letting the

time $t = 0$ indicate the end of the day traffic and considering time as positive *backward* towards $t = h$. Furthermore we will consider h twice as big with the latter curve as the former.

We now set $M = M_{h_1} + M_{h_2}$, where M_{h_1} and M_{h_2} denote

morning and afternoon traffic volumes respectively. We will assume k to be equal to 4.

We then get:

$$M_{h_1} = \int_0^{h_1} \bar{y}\left(\frac{e}{h_1}\right)^4 e^{-\frac{4}{h_1}t} t^4 dt$$
$$= \bar{y} \cdot h_1\left(\frac{6e^4}{256} - \left(\frac{1}{4} + \frac{1}{4} + \frac{3}{16} + \frac{6}{64} + \frac{6}{256}\right)\right) = \bar{y} \cdot h_1 \cdot 0.475$$ [8)]

$$M_{h_2} = \int_0^{h_2} \bar{y}\left(\frac{e}{h_2}\right)^4 e^{-\frac{4}{h_2}t} t^4 dt = \bar{y}h_2 \cdot 0.475$$
$$M = \bar{y}(h_1 + h_2) \cdot 0.475\text{K} = \bar{y} \cdot 15 \cdot 0.475 = \bar{y} \cdot 7.125 = \bar{y}\frac{100}{14.2}$$

Thus we have here a concentration of approx. 14.2 %. Outside the 15 working hours, which may be assumed to be from 7.30 a.m. until 10.30 p.m., the traffic is considered to be equal to 0.

From the expression for M we see that it may be set equal to

$$M_{3h} = \int_0^{3h} \bar{y}\left(\frac{e}{3h}\right)^4 e^{-\frac{4}{3h}t} t^4 dt = \bar{y} \cdot 3h \cdot 0.475$$

Thus we can apply one and the same curve

$$y = \bar{y}\left(\frac{e}{ph}\right)^k e^{-\frac{k}{ph}t} t^k$$

for an approximate calculation of the average day traffic during certain fractions of the working time as well as the whole time, when we know the average maximum intensity (the number of telephone hours per busy hour), the duration ph of the ordinary daily working hours and the average concentration coefficient.

[8)] *The number is rounded upwards.*

40. Hereby means are made available for calculating the aggregate traffic hindrance or the daily degree of hindrance. It would, however, lead too far to carry out this calculation with minute exactness, since the integral to be solved

$$\int_0^{ph} e^{-y}\left(\frac{y^{x+1}}{(x+1)!}+2\frac{y^{x+2}}{(x+2)!}+3\frac{y^{x+3}}{(x+3)!}+\mathrm{K}\right)dt \qquad (36)$$

where the relation between y and t is given by

$y=\bar{y}\cdot\left(\frac{e}{ph}\right)^{k} e^{-\frac{k}{ph}t}\, t^{k}$, only can be found by laborious series expansion. We will therefore make a detour and let this be as short and straightforward as possible, while we make sure to be on the safe side.

In the rectangular co-ordinate system where we assume y in a practical way given as a function of t we will through the zero point draw a straight line in the plane of the traffic curve. Let the equation of this line be:

$$y=\frac{\bar{y}}{a-r}t \qquad (37)$$

where $a = ph$ and r denotes a constant of such property that the straight line does not intersect the curve outside the zero point, but either is a tangent to the curve between its first inflexion point and the end point given by $t = a$, or that it remains outside or on the upper side of the curve.

In a well known way we find that r must be such that:

$$\frac{1}{a-r}\geq\frac{e}{a}\left(\frac{k-1}{k}\right)^{k-1} \qquad (38)$$

If we stick to $a = 15$ and $k = 4$, then we have to choose r so that

$$\frac{1}{15-r}\geq\frac{2.71828}{15}\cdot\frac{27}{64},\ \text{or}$$

$$15-r\leq\frac{320}{2.71828\cdot 9},\quad r\geq 1.919\mathrm{K}$$

Hence we set $r = 2$.

Through the endpoint $(a,\bar{y})$ of the curve we draw a parallel to the t-axis until it intersects the line $y=\frac{\bar{y}}{a-2}t$. The distance between the intersection point and the end point mentioned is equal to 2. Instead of the found curve we will now employ the thus constructed line $(0,0)\to(a-2,\bar{y})\to(a,\bar{y})$ to obtain a simpler calculation when carrying out the necessary integration. Since the broken line entirely lies outside the curve between its end points, it is obvious that the traffic volume and the corresponding amount of hindrances that we obtain are not too small. On the other hand, it is evident that since the greatest deviation between the line and the curve occurs around the lower part of the curve, while around the upper part the deviation is relatively quite small, the result will not be essentially too large.

41. Let us now determine the hindered communication $H_{(a-r)}$ during the time $a-r$, when $y=\frac{\bar{y}}{a-r}t$. We have:

$$H_{(a-r)}=\int_0^{a-r} e^{-y}\left(\frac{y^{x+1}}{(x+1)!}+2\frac{y^{x+2}}{(x+2)!}+3\frac{y^{x+3}}{(x+3)!}+\mathrm{K}\right)dt,\ \text{and}$$

$$dt=\frac{a-r}{\bar{y}}dy.$$

At time $t=0$, $y=0$ and at $t=a-r$, $y=\bar{y}$.

Hence we get:

$$H_{(a-r)}=\int_0^{\bar{y}} e^{-y}\left(\frac{y^{x+1}}{(x+1)!}+2\frac{y^{x+2}}{(x+2)!}+3\frac{y^{x+3}}{(x+3)!}+\mathrm{K}\right)\frac{(a-r)}{\bar{y}}dy=$$

$$\begin{aligned}\frac{(a-r)}{y}[e^{-y}(&\frac{y^{x+2}}{(x+2)!}+\frac{y^{x+3}}{(x+3)!}+\frac{y^{x+4}}{(x+4)!}+\frac{y^{x+5}}{(x+5)!}+\frac{y^{x+6}}{(x+6)!}+\mathrm{K}\\ &+2\frac{y^{x+3}}{(x+3)!}+2\frac{y^{x+4}}{(x+4)!}+2\frac{y^{x+5}}{(x+5)!}+2\frac{y^{x+6}}{(x+6)!}+\mathrm{K}\\ &+3\frac{y^{x+4}}{(x+4)!}+3\frac{y^{x+5}}{(x+5)!}+3\frac{y^{x+6}}{(x+6)!}+\mathrm{K}\\ &+4\frac{y^{x+5}}{(x+5)!}+4\frac{y^{x+6}}{(x+6)!}+\mathrm{K}\\ &+5\frac{y^{x+6}}{(x+6)!}+\mathrm{K}\\ &+\mathrm{K}\ )]_0^{\bar{y}}\end{aligned}$$

Thus we obtain:

$$\begin{aligned}H_{(a-r)}<\frac{a-r}{\bar{y}}[e^{-y}(&\frac{y^{x+2}}{(x+2)!}\frac{1}{\left(1-\frac{y}{x+3}\right)}+2\frac{y^{x+3}}{(x+3)!}\frac{1}{\left(1-\frac{y}{x+4}\right)}\\ &+3\frac{y^{x+4}}{(x+4)!}\frac{1}{\left(1-\frac{y}{x+5}\right)}+4\frac{y^{x+5}}{(x+5)!}\frac{1}{\left(1-\frac{y}{x+6}\right)}\\ &+5\frac{y^{x+6}}{(x+6)!}\frac{1}{\left(1-\frac{y}{x+7}\right)}+\mathrm{K}\ )]_0^{\bar{y}}\end{aligned}$$

$$<[\frac{(a-r)e^{-y}}{\bar{y}\left(1-\frac{y}{(x+3)}\right)}\left(\frac{y^{x+2}}{(x+2)!}+2\frac{y^{x+3}}{(x+3)!}+3\frac{y^{x+4}}{(x+4)!}+4\frac{y^{x+5}}{(x+5)!}+\mathrm{K}\right)]_0^{\bar{y}}$$

$$<[\frac{(a-r)e^{-y}}{y\left(1-\frac{y}{x+3}\right)^3}\cdot\frac{y^{x+2}}{(x+2)!}]_0^{\bar{y}}=\frac{(a-r)}{\left(1-\frac{\bar{y}}{x+3}\right)^3}\cdot\frac{\bar{y}^{x+1}}{(x+2)!}e^{-\bar{y}}$$

The traffic volume that is hindered during working hours $a – r$, ie. *outside* the r busiest hours, thus on average per hour amounts to:

$$\frac{H_{(a-r)}}{(a-r)}<\frac{\bar{y}^{x+1}}{(x+2)!}\cdot e^{-\bar{y}}\cdot\frac{1}{\left(1-\frac{\bar{y}}{x+3}\right)^3}\ \text{telephone hours.} \qquad (39)$$

As the traffic volume that the calculation is based on amounts

We will apply the expression on the right side of the inequality sign as a practical measure of the degree of hindrance in question, which we will denote $\psi_{xy_{(a-r)}}$, so that we have:

$$\psi_{xy_{(a-r)}} = \frac{\bar{y}^{x}}{(x+2)!} \cdot e^{-\bar{y}} \cdot \frac{2}{\left(1-\frac{\bar{y}}{x+3}\right)^3} \tag{40b}$$

Thus we have the following relation between this degree of hindrance and the earlier calculated ψ_{xy}, that we will denote $\psi_{xy_{(r)}}$ in order to better indicate the time that is referred to:

$$\psi_{xy_{(a-r)}} = \frac{\bar{y}^{x}}{(x+1)!} \cdot e^{-\bar{y}} \cdot \frac{1}{\left(1-\frac{\bar{y}}{x+2}\right)^2} \cdot \frac{2}{(x+2)} \cdot \frac{\left(1-\frac{\bar{y}}{x+2}\right)^2}{\left(1-\frac{\bar{y}}{x+3}\right)^3}$$

$$= \psi_{xy_{(r)}} \cdot \frac{2}{(x+2)} \cdot \frac{(x+3)^3}{\left(x+3-\bar{y}\right)^3} \cdot \frac{\left(x+2-\bar{y}\right)^2}{(x+2)^2} \tag{40c}$$

As $\frac{(x+3)}{\left(x+3-\bar{y}\right)} < \frac{(x+2)}{\left(x+2-\bar{y}\right)}$, we have also:

$$\psi_{xy_{(a-r)}} < \psi_{xy_{(r)}} \cdot \frac{2}{\left(x+2-\bar{y}\right)} \tag{40d}$$

So if in the busiest time the intensity is $\bar{y}=9$ over a circuit bundle of $x = 17$, which gives an average degree of hindrance of approx. 1 per mill in the busiest hour, then the degree of hindrance *outside the two busiest hours* is less than 0.2 ‰.

If we calculate the value of the expression

$$\frac{2}{(x+2)^3} \cdot \frac{(x+3)^3}{\left(x+3-\bar{y}\right)^3} \cdot \left(x+2-\bar{y}\right)^2$$

for $x=17$ and $\bar{y}=9$, then we get 0.175‰ instead of 0.2‰ as obtained by the last mentioned simplification. This is then in the present case tantamount to increasing the fourth decimal of the fraction 0.000175 and to delete the succeeding decimals, to make the fraction equal to 0.0002.

If we choose another example:

$\bar{y}=4$ and $x=10$, we get $\frac{2}{x+2-\bar{y}} = \frac{1}{4}$, corresponding to a degree of hindrance *outside the two busiest hours* of 0.25 ‰.

On the other hand is $\frac{2}{(x+2)^3} \cdot \frac{(x+3)^3}{(x+3-\bar{y})^3} \cdot (x+2-\bar{y})^2 = 0.231$, corresponding to a little more than 0.23 ‰. This can also be considered a rounding off, namely upwards to the nearest decimal divisible by 5. We can therefore even for rather small traffic values consider the expression

$$\psi_{xy_{(a-r)}} = \psi_{xy_{(r)}} \cdot \frac{2}{\left(x+2-\bar{y}\right)} \tag{40e}$$

a usable formula in practice.

42. The objection can be made against the present calculation that although it is granted that the calculated amount of traffic is not to small and neither is the degree of hindrance too small for the amount of traffic that is the basis for calculation, it is still not given that this degree of hindrance is sufficiently large for the traffic that is described by our curve, and which is assumed to correspond very closely to the real traffic with its given maximum intensity and its given concentration. The situation is that in order to be able to carry out the integration in a convenient manner, we have assumed the traffic increased in such a way that the area in the *ty*-plane that gives a measure of the traffic can be limited by means of straight lines. We have certainly not extended the base line a of the area, on the contrary we have increased the area all the way to its upper part, which after all influence the hindrance, and in that respect we should have reason to assume that we have found an approximately correct calculation. We will, however, as the matter also is to find the average degree of hindrance for the entire day traffic, carry out a calculation on the basis of the real aggregate day traffic according to the curve without increase or decrease.

43. Through the inflexion point of our traffic curve and the point $t = r_0$ on the abscissa axis we draw a straight line that intersects the previously mentioned parallel to the abscissa axis in a point at a distance r_1 to the left of the curve endpoint, ie. in the point $\left(a-r_1, \bar{y}\right)$. We now choose r_0 in a way that:

$$\left(a-r_0+r_1\right)\bar{y}/2 = \int_0^a \bar{y}\left(\frac{e}{a}\right)^4 e^{-\frac{4t}{a}} t^4 \, dt = M \tag{41}$$

Thus we depict the traffic volume of the day by means of a trapeze of height $\bar{y}$ and the parallel sides $a-r_0$ and r_1.

If we denote the ordinate of the inflexion point by y_0, then we have:

$$y_0 = \bar{y}\left(\frac{e}{a}\right)^4 e^{-\frac{4}{a}\cdot\frac{a}{2}} \cdot \frac{a^4}{2^4} = \bar{y} \cdot 0.461816\ldots,$$

as the abscissa for this point is $a/2$.

If we denote the number 0.461816 as u and $1 - u$ as v, then we have the following ratios:

$$\frac{\frac{a}{2}-r_0}{\frac{a}{2}-r_1} = \frac{u}{v}$$

Hence we get:

$$r_0 = \frac{r_1 u + \frac{a}{2}(v-u)}{v}; \quad r_1 = \frac{r_0 v + \frac{a}{2}(u-v)}{u}.$$

By means of one of these equations and the equations

$$\frac{\frac{a}{2}-r_0}{\frac{a}{2}-r_1}=\frac{u}{v}$$

Hence we get:

$$r_0=\frac{r_1 u+\frac{a}{2}(v-u)}{v}; \quad r_1=\frac{r_0 v+\frac{a}{2}(u-v)}{u}.$$

By means of one of these equations and the equations

$a=15,\ \left(a-r_0+r_1\right)=2\frac{M}{\bar{y}}=2m=14.248\mathrm{K}$, we get:

$$r_0=\frac{2mu-\frac{a}{2}(3u-v)}{v-u}=2.94\mathrm{K}$$

$$r_1=\frac{2mv-\frac{a}{2}}{v-u}=2.19\mathrm{K}$$

The curve, the t-axis and the ordinate y_0 enclose an area:

$$\int_0^{\frac{a}{2}} \bar{y}\left(\frac{e}{\bar{y}}\right)^4 e^{-\frac{4t}{a}} t^4\, dt=\bar{y}\cdot 1.01\mathrm{K}$$

The area that is enclosed by the same axis and ordinate plus the slanting side of the trapeze is:

$$\frac{1}{2}\left(\frac{1}{2}a-r_0\right)y_0=\bar{y}\cdot 1.05\mathrm{K}$$

The difference between these two areas, with the here given length of a is approx. $\bar{y}\cdot 0.04$. Exactly that size is the difference between the segment of the curve area above the inflexion point that is cut off by the slanting side of the trapeze, and the triangular area that is enclosed by the previously mentioned parallel with the t-axis, the same trapeze side and the upper part of the curve. The segment is thus approx. $\frac{4\bar{y}}{100}$ greater than the latter area. Nevertheless is the influence of this area on the degree of hindrance greater than that of the segment, namely because of its position and because of its size that is substantial compared to the mentioned difference. The difference between the two lower areas also contribute, if only to a minor degree, to increasing the efficiency, and here the effect of the difference is in the same direction as the position.

What is said here can of course be proved mathematically by means of the equations and a more detailed discussion for the given curve, essentially determined by the concentration coefficient $k = 4$. However, we consider it unnecessary to go further into the matter, as anybody by construction and closer examination of the curve and the trapeze, and by consideration of the expression for the degree of hindrance, may convince himself of the correctness.

Consequently, we are on the safe side when we for calculation of the degree of hindrance apply the straight line through the inflexion point given by:

$$y=\frac{\bar{y}}{(a-r_0-r_1)}t-\frac{r_0 y_0}{\left(\frac{a}{2}-r_0\right)}$$

$$=\frac{y_0}{\left(\frac{a}{2}-r_0\right)}t-\frac{r_0 y_0}{\left(\frac{a}{2}-r_0\right)} \tag{42}$$

as well as the line $y=\bar{y}$.

By means of the equation of the former line we get:

$$dt=\frac{(a-r_0-r_1)dy}{\bar{y}}=\frac{\left(\frac{a}{2}-r_0\right)dy}{y_0}$$

We will now determine the traffic volume $\overline{H}_{(a-r_0-r_1)}$ that is hindered under the conveyance of the traffic that is remodelled to increase linearly from $y = 0$ at $t = r_0$ to $y=\bar{y}$ at $t=a-r_1$.

In a similar way as before we obtain:

$$\overline{H}_{(a-r_0-r_1)}=$$

$$\int_0^{\bar{y}} e^{-y}\left(\frac{y^{x+1}}{(x+1)!}+2\frac{y^{x+2}}{(x+2)!}+3\frac{y^{x+3}}{(x+3)!}+\mathrm{K}\right)\frac{(a-r_0-r_1)}{\bar{y}}dy$$

$$<\frac{(a-r_0-r_1)}{\left(1-\frac{\bar{y}}{x+3}\right)^3}\frac{\bar{y}^{x+1}}{(x+2)!}e^{-\bar{y}}$$

Dividing by the traffic volume concerned, $\frac{\bar{y}}{2}\left(a-r_0-r_1\right)$, we get the degree of hindrance

$$\overline{\psi}_{xy_{(a-r_0-r_1)}}<\frac{\bar{y}^x}{(x+2)!}\cdot e^{-\bar{y}}\cdot\frac{2}{\left(1-\frac{\bar{y}}{x+3}\right)^3}$$

Furthermore:

$$\overline{\psi}_{xy_{(a-r_0-r_1)}}<\psi_{xy_{(r)}}\cdot\frac{2}{(x+2-\bar{y})}$$

Hence we get quite the same values as before, which must be expected, as we in both cases have dealt with a linearly increasing traffic from 0 to $\bar{y}$.

It is not, however, in order to get this self-evident confirmation of the calculation of the partial degree of hindrance that we have carried out this new calculation, but rather to obtain a satisfactory determination of the hindered traffic volume, so as to indicate with certainty an upper limit for the average *daily*

$$H=\overline{H}_{(a-r_0-r_1)}+H_{r_1}$$

$$<\frac{(a-r_0-r_1)}{\left(1-\frac{\bar{y}}{x+3}\right)^3}\cdot\frac{\bar{y}^{x+1}}{(x+2)!}\cdot e^{-\bar{y}}$$

$$+\frac{r_1}{\left(1-\frac{\bar{y}}{x+2}\right)^2}\frac{\bar{y}^{x+1}}{(x+1)!}e^{-\bar{y}}$$

$$=e^{-\bar{y}}\frac{\bar{y}^{x+1}}{(x+1)!}\left(\frac{a-r_0-r_1}{\left(1-\frac{\bar{y}}{x+3}\right)^3(x+2)}+\frac{r_1}{\left(1-\frac{\bar{y}}{x+2}\right)^2}\right)$$

$$=e^{-\bar{y}}\frac{\bar{y}^{x+1}}{(x+1)!}\frac{1}{\left(1-\frac{\bar{y}}{x+2}\right)^2}\left(\frac{(a-r_0-r_1)(x+2-\bar{y})^2}{(x+3-\bar{y})^3}\frac{(x+3)^3}{(x+2)^3}+r_1\right),$$

Now dividing H by $M=\frac{1}{2}(a-r_0+r_1)\bar{y}$ we get:

$$\frac{H}{M}<2e^{-\bar{y}}\frac{\bar{y}^{x}}{(x+1)!}\frac{1}{\left(1-\frac{\bar{y}}{x+2}\right)^2}$$

$$\cdot\left(\frac{(a-r_0-r_1)}{(a-r_0+r_1)}\frac{(x+2-\bar{y})^2}{(x+3-\bar{y})^3}\cdot\frac{(x+3)^3}{(x+2)^3}+\frac{r_1}{(a-r_0+r_1)}\right) \tag{44a}$$

By inserting the calculated values of r_0 and r_1, and setting as assumed $a = 15$, then we get:

$$\frac{H}{M}<\psi_{xy_{(r)}}\left(1.385\frac{(x+2-\bar{y})^2(x+3)^3}{(x+3-\bar{y})^3(x+2)^3}+0.307\right) \tag{44b}$$

We will denote the *daily* degree of hindrance with concentration coefficient k by $\psi_{xy_{(d,k)}}$, and so we get:

$$\psi_{xy_{(d,4)}}<\psi_{xy_{(r)}}\left(1.385\frac{(x+2-\bar{y})^2}{(x+3-y)^3}\frac{(x+3)^3}{(x+2)^3}+0.307\right), \tag{44c}$$

where $\psi_{xy_{(r)}}$ as earlier mentioned denote the degree of hindrance in the busiest hour.

Applying this on the two previously given examples 1) $x = 17$ and $y = 9$, and 2) $x = 10$ and $y = 4$, then we get for 1):

$\psi_{xy_{(d,4)}}<\frac{43}{100}$‰ and for 2): $\psi_{xy_{(d,4)}}<\frac{46}{100}$‰.

The quantity that has a dominating influence in the expression for $\frac{H}{M}$ is r_1. We have had to choose this equal to 2.19 hours in order to be on the safe side at the integration. This relatively large number is, with the simple procedure used, due to the particular form of the curve. Had this form implied that r_1 could be set close to 0, then we would have obtained an expression for the daily degree of hindrance corresponding to that of a linearly increasing traffic. The substantial accumulation of traffic around the two busiest hours thus has the effect that the low degree of hindrance while the traffic is low or is strongly increasing or decreasing is of little influence.

One can get an idea of the influence of the form of the traffic curve on the daily degree of hindrance (compared to that of the busiest hour) by considering two extreme cases, that of k being negligibly small and that of k being immensely large. In the former case the traffic curve ascends steeply from $t = 0$ and reaches in infinitesimal time close to the straight line $y=\bar{y}$, which it follows until the point $t = a$ where the traffic curve is considered finished. This corresponds to a uniform traffic during the whole working day and gives a daily degree of hindrance that only differs from the maximum degree of hindrance by a negligibly small fraction. In the latter case the traffic curve stays very close to the abscissa axis until close to the point $t = a$, from where it ascends steeply towards the line $y=\bar{y}$ and after a short bend reaches the end point, where the curve ordinate is $\bar{y}$. In this case practically all traffic is accumulated in the busiest hour, and the daily degree of hindrance will also here be very close to that of the busiest hour. Between these two cases one can imagine an innumerable number of intermediate cases, corresponding to values of k between those given above. With $k = 4$ (concentration 14.2 %) we have an intermediate stage corresponding to normal traffic, and for this case we have found that the daily degree of hindrance for practical values of x lie substantially below half of the day maximum. How much it will be below the calculated values of $\frac{43}{100}$ and $\frac{46}{100}$ in the two examples dealt with above can only be determined by a more elaborate calculation. We will not here carry out such a calculation – which might be done by subdividing the curve in approximately linear parts – but only do a comparison with the daily degree of hindrance that results from the previously discussed linear increase from $t = 0$ to $t = a - r = a - 2$ and the subsequent constant traffic course until time $t = a$.

Here we have:

$$H=H_{(a-r)}+H_{(r)}<\left(\frac{(a-2)}{\left(1-\frac{\bar{y}}{x+3}\right)^3}\frac{\bar{y}^{x+1}}{(x+2)!}+\frac{2}{\left(1-\frac{\bar{y}}{x+2}\right)^2}\frac{\bar{y}^{x+1}}{(x+1)!}\right)e^{-\bar{y}}$$

$$=e^{-\bar{y}}\frac{\bar{y}^{x+1}}{(x+1)!}\frac{1}{\left(1-\frac{\bar{y}}{x+2}\right)^2}\left(\frac{(a-2)(x+2-\bar{y})^2(x+3)^3}{(x+3-\bar{y})^3(x+2)^3}+2\right)$$

and

$$M=\frac{1}{2}(a+2)\bar{y}$$

Consequently:

$$M=\frac{1}{2}(a+2)\bar{y}$$

Consequently:

$$\frac{H}{M}<e^{-\bar{y}}\frac{\bar{y}^{x}}{(x+1)!}\frac{1}{\left(1-\frac{\bar{y}}{x+2}\right)^{2}}\left(\frac{2(a-2)\left(x+2-\bar{y}\right)^{2}(x+3)^{3}}{(a+2)(x+3-\bar{y})^{3}(x+2)^{3}}+\frac{4}{(a+2)}\right)$$

Setting now 1) $x = 17$, $y = 9$ and 2) $x = 10$ and $y = 4$, and inserting the stipulated value of a, we get the ratios:

$$\frac{2(a-2)(x+2-\bar{y})^{2}(x+3)^{3}}{(a+2)(x+3-\bar{y}^{3}(x+2)^{3}}+\frac{4}{a+2}=$$

1) $$\frac{2\cdot 13\cdot 10^{2}\cdot 20^{3}}{17\cdot 11^{3}\cdot 19^{3}}+\frac{4}{17}=0.369K = \text{approx. } \frac{37}{100}$$

2) $$\frac{2\cdot 13\cdot 8^{2}\cdot 13^{3}}{17\cdot 9^{3}\cdot 12^{3}}+\frac{4}{17}=0.405K = \text{approx. } \frac{41}{100}$$

As earlier suggested we cannot assume any certainty that this calculation is on the safe side. On the other hand we know that the previously found ratios under 1) and 2), $\frac{43}{100}$ and $\frac{46}{100}$ respectively, are greater than the likely values. If, on the other hand, we set $r = 1$, which for the applied curve must be expected to give too small ratios, then we get the values respectively:

1) $$\frac{2\cdot 14\cdot 10^{2}\cdot 20^{3}}{16\cdot 11^{3}\cdot 19^{3}}+\frac{2}{16}=0.278K = \text{approx. } \frac{28}{100}$$

2) $$\frac{2\cdot 14\cdot 8^{2}\cdot 13^{3}}{16\cdot 9^{3}\cdot 12^{3}}+\frac{2}{16}=0.320K = \text{approx. } \frac{32}{100}$$

Altogether we have after the previous reason to consider the ratio $\left(\frac{2(a-2)}{(a+2)}\cdot\frac{(x+2-\bar{y})^{2}}{(x+3-\bar{y})^{3}}\cdot\frac{(x+3)^{3}}{(x+2)^{3}}+\frac{4}{(a+2)}\right)$ to be in very good agreement with the real or likely expression for useful values of x and y. We can bring the value closer to that calculated by means of the quantities r_0 and r_1 by increasing the factor $\frac{(x+2-\bar{y})^{2}(x+3)^{3}}{(x+3-\bar{y})^{3}(x+2)^{3}}$, which can be replaced by $\frac{1}{(x+3-\bar{y})}$ or by $\frac{1}{(x+2-\bar{y})}$.

In the latter case we get the ratio

$$\left(\frac{2(a-2)}{(a+2)(x+2-\bar{y})}+\frac{4}{(a+2)}\right)=\frac{1}{17}\left(\frac{26}{(x+2-\bar{y})}+4\right).$$

This gives for

1) $$\frac{1}{17}\left(\frac{26}{(x+2-\bar{y})}+4\right)=\frac{1}{17}\left(\frac{26}{(17+2-9)}+4\right)$$

$$=0.388K = \text{approx. } \frac{39}{100}$$

2) $$\frac{1}{17}\left(\frac{26}{(x+2-\bar{y})}+4\right)=\frac{1}{17}\left(\frac{26}{(10+2-4)}+4\right)$$

$$=0.426K = \text{approx. } \frac{43}{100}$$

These ratios are, respectively, 4/100 and 3/100 less than the ratios that give full certainty that we do not calculate too small values. The resulting differences with a degree of hindrance of 1 ‰ in the busiest hour thus amount to 4 hundredths and 3 hundredths of 1 ‰, ie. 4/100000 and 3/100000 of the total day traffic, which must be considered sufficient exactness.

For a normal course of traffic (with $k = 4$ or a concentration of 14.2 %) we can in accordance with the above approximately calculate the daily degree of hindrance by means of the already calculated degree of hindrance for the busiest hour by setting:

$$\psi_{xy_{(d,4)}}<\psi_{xy_{(r)}}\cdot\frac{1}{17}\left(\frac{26}{(x+2-\bar{y})}+4\right) \tag{45}$$

If $\psi_{xy_{(r)}}=1$ ‰, then straightforward we get:

$$\psi_{xy_{(d,4)}}=\frac{1}{17}\left(\frac{26}{(x+2-\bar{y})}+4\right) \text{‰}$$

By increasing the corresponding values of x and $\bar{y}$ in a way that $x+2-\bar{y}$ upwards approaches 26, then the daily degree of hindrance decreases towards $\frac{5}{17}$ per mill.

44. If we want to construct a curve for a useful daily degree of hindrance, say for 4/10 ‰, then according to the present development one can apply the formula:

$$\psi_{xy_{(d,4)}}=e^{-y}\frac{y^{x}}{(x+1)!}\cdot\frac{1}{\left(1-\frac{y}{x+2}\right)^{2}}\cdot\frac{1}{17}\left(\frac{26}{(x+2-y)}+4\right) \tag{46}$$

For $\psi_{xy_{(d,4)}}=4/10$ ‰ this curve will approach the curve for $\psi_{xy_{(r)}}=\psi_{xy}$ not far from the point $x=10$, $y=4$.

For values of x greater than 10 the ordinates of the former curve will be higher than that of the latter, and the differences will increase the more x is increased. That implies: For a given concentration the curve of a given daily degree of hindrance, say 4/10 ‰, will for a certain higher value of x = number of circuits or switches, say $x > 10$, allow a greater maximum traffic intensity $y=\bar{y}$ than the curve for the corresponding degree of hindrance for the busiest hour, say 1 ‰, will allow for the same value of x. By the expression 'the corresponding degree of hindrance for the busiest hour' is understood the degree of hindrance in the busiest hour that for a number of circuits $x = 10$ results from the same traffic that would be carried by 10 circuits with the same daily degree of hindrance.

There is thus for the two curves a difference for higher values of x that should not be left out of consideration when evaluating the traffic requirements.

IX Calculation of the degree of hindrance in systems with skewing of circuits (slip multiple)

45. So far we have only considered the cases where the x circuits or switches that are arranged for common application by a certain number of subscribers or a certain amount of traffic are all equally accessible for different calls demanding service.

There exist, however, arrangements where two or more bundles of x circuits work together in order to reduce the hindrances, partly by a peculiar skewing of the circuit arrangements, partly by placement of a certain number of common spare circuits, or by a combination of such skewing and spare arrangements.

46. We will first consider the skewing system, and particularly with regard to co-operating groups of preselectors.

Without skewing one will usually have that each group of x^2 forward searching primary preselectors will be able to search for access by x fully co-ordinate secondary preselectors (or primary group selectors), so that each group has its own complete system of common circuits. *With* skewing, on the other hand, certain circuits will be accessible for two or more groups of x^2 preselectors. We will initially stick to the simplest and most surveyable type of skewing, characterised by one step forward for each circuit. By the combination of two groups of x^2 primary preselectors in this way, numbering these from 1 to $2x^2$ and numbering the circuits from 1 to $2x$, we will get:

Preselector number				has unlimited access to ckt. no.
on and after	1	up to and including	x^2	1
on and after	$x + 1$	up to and including	$x^2 + x$	2
on and after	$2x + 1$	up to and including	$x^2 + 2x$	3
on and after	$3x + 1$	up to and including	$x^2 + 3x$	4
	...			
on and after	$(x - 1)x + 1$	up to and including	$x^2 + (x - 1)x$	x
on and after	$x^2 + 1$	up to and including	$2x^2$	$x + 1$
on and after	$x^2 + x + 1$	up to and including	$2x^2$ and 1 to x	$x + 2$
on and after	$x^2 + 2x + 1$	up to and including	$2x^2$ and 1 to $2x$	$x + 3$
on and after	$x^2 + 3x + 1$	up to and including	$2x^2$ and 1 to $3x$	$x + 4$
	...			
on and after	$x^2 + (x - 1)x + 1$	up to and including	$2x^2$ and 1 to $(x - 1)x$	$2x$

The system, consisting of two mutually skewed groups, by the arrangement of the multiple connections in this manner forms a closed ring.

In a similar way one can combine 3 or 4 or 5 ... or r groups.

47. We will now try to show how one can calculate approximately hindrances and degree of hindrance for such an arrangement, consisting of r groups, each of x^2 primary preselectors with rx skewed circuits or secondary preselectors, as well as with an average traffic intensity y for each group in the busy time.

Let us initially set $r = 2$.

We will denote by R_s the risk that because of the systematic arrangement a hunting selector *bypasses* all of the *s free* out of the existing $2x$ circuits. This risk is equal to the ratio of the number of fully occupied circuits to the total number of circuits that can be hunted for. R_s is thus a proper fraction and may be denoted the *systematic probability of hindrance*. If $s \geq x + 1$, then this probability is equal to 0. Thus we have $R_{2x} = R_{2x-1} = R_{2x-2} = ... = R_{x+1} = 0$. If $s < x + 1$, then there is always a possibility that the hunting occurs on a fully occupied set of circuits and thus that the call 'gets lost'. By a detailed examination of the combination problem at hand we find the following values of R_s for a skewed system as described above for a combination of *two* groups:

$$R_x = 1 \cdot \frac{1}{\binom{2x}{x}}$$

$$R_{x-1} = \left(2 + \binom{x-2}{1}1\right)\frac{1}{\binom{2x}{x-1}}$$

$$R_{x-2} = \left(3 + \binom{x-3}{1}2 + \binom{x-2}{2}1\right)\frac{1}{\binom{2x}{x-2}}$$

$$R_{x-3} = \left(4 + \binom{x-4}{1}3 + \binom{x-3}{2}2 + \binom{x-2}{3}1\right)\frac{1}{\binom{2x}{x-3}}$$

$$\ldots$$

$$\begin{aligned} R_{x-i} = {} & ((i+1) + \binom{x-(i+1)}{1}i + \binom{x-i}{2}(i-1) \\ & + \binom{x-(i-1)}{3}(i-2) + \ldots + \binom{x-(i-p+2)}{p}(i-p+1) \\ & + \ldots + \binom{x-2}{i} \cdot 1)\frac{1}{\binom{2x}{x-i}} \end{aligned}$$

$$\ldots$$

$$\begin{aligned} R_3 = {} & ((x-2) + 2(x-3) + 3(x-4) + \ldots \\ & + p(x-p+1) + \ldots + (x-3)2 + (x-2) \cdot 1)\frac{1}{\binom{2x}{3}} \end{aligned}$$

$$R_2 = ((x-1) + (x-2) + (x-3) + \ldots + 2 + 1)\frac{1}{\binom{2x}{2}}$$

$$R_1 = \frac{1}{2} \qquad \binom{47a}{1 \ldots x}$$

These expressions, where the numerator is greater than 1, are compressed to give:

$$R_x = 1 \cdot \frac{1}{\binom{2x}{x}}$$

$$R_{x-1} = \binom{x+1}{1}\frac{1}{\binom{2x}{x}}$$

$$R_{x-2} = \binom{x+2}{2}\frac{1}{\binom{2x}{x}}$$

$$R_{x-3} = \binom{x+3}{x}\frac{1}{\binom{2x}{x}}$$

$$\ldots$$

$$R_{x-i} = \binom{x+i}{i}\frac{1}{\binom{2x}{x}}$$

$$\ldots$$

$$R_3 = \binom{2x-3}{x-3}\frac{1}{\binom{2x}{x}} = \frac{(x-2)(x-1)}{2(2x-2)(2x-1)}$$

$$R_2 = \binom{2x-2}{x-2}\frac{1}{\binom{2x}{x}} = \frac{x-1}{2(2x-1)}$$

$$R_1 = \binom{2x-1}{x-1}\frac{1}{\binom{2x}{x}} = \frac{1}{2} \qquad \binom{47b}{1 \ldots x}$$

We now set:

$$\begin{aligned} & S_1 = R_x \\ & S_2 = R_x + R_{x-1} \\ & S_3 = R_x + R_{x-1} + R_{x-2} \\ & \ldots \\ & S_{i+1} = R_x + R_{x-1} + R_{x-2} + \ldots + R_{x-i} \\ & S_{x-2} = R_x + R_{x-1} + R_{x-2} + \ldots + R_3 \\ & S_{x-1} = R_x + R_{x-1} + R_{x-2} + \ldots + R_3 + R_2 \\ & S_x = R_x + R_{x-1} + R_{x-2} + \ldots + R_3 + R_2 + R_1 \end{aligned} \qquad \binom{48a}{1 \ldots x}$$

These sums can be expressed in the following way:

$$S_1 = 1 \cdot \frac{1}{\binom{2x}{x}}$$

$$S_2 = \binom{x+2}{1}\frac{1}{\binom{2x}{x}}$$

$$S_3 = \binom{x+3}{2}\frac{1}{\binom{2x}{x}}$$

$$S_4 = \binom{x+4}{3}\frac{1}{\binom{2x}{x}}$$

$$S_5 = \binom{x+5}{4}\frac{1}{\binom{2x}{x}}$$

K

$$S_{i+1} = \binom{x+i+1}{i}\frac{1}{\binom{2x}{x}}$$

K

$$S_{x-2} = \binom{2x-2}{x-3}\frac{1}{\binom{2x}{x}}$$

$$S_{x-1} = \binom{2x-1}{x-2}\frac{1}{\binom{2x}{x}}$$

$$S_x = \binom{2x}{x-1}\frac{1}{\binom{2x}{x}} = \frac{x}{x+1} \qquad \binom{48\text{b}}{1\mathrm{K}\;x}$$

The calculated quantities R and S apply as mentioned to a system of 2 groups with each x^2 primary preselectors and together $2x$ secondary preselectors.

For a triple system of corresponding type we obtain:

$$R_{2x} = 1\frac{1}{\binom{3x}{x}}$$

$$R_{2x-1} = \binom{x+1}{1}\frac{1}{\binom{3x}{x}}$$

$$R_{2x-2} = \binom{x+2}{2}\frac{1}{\binom{3x}{x}}$$

$$R_{2x-3} = \binom{x+3}{x}\frac{1}{\binom{3x}{x}}$$

K

$$R_{2x-i} = \binom{x+i}{i}\frac{1}{\binom{3x}{x}}$$

K

$$R_3 = \binom{3x-3}{2x-3}\frac{1}{\binom{3x}{x}}$$

$$R_2 = \binom{3x-2}{2x-2}\frac{1}{\binom{3x}{x}}$$

$$R_1 = \binom{3x-1}{2x-1}\frac{1}{\binom{3x}{x}} = \frac{2}{3} \qquad \binom{49}{1\mathrm{K}\;2x}$$

The corresponding sums S take the following form, as we indicate the triple system by means of a separate index:

$$S_{\substack{3\\1}} = 1 \cdot \frac{1}{\binom{3x}{x}}$$

$$S_{\substack{3\\2}} = \binom{x+2}{1}\frac{1}{\binom{3x}{x}}$$

$$S_{\substack{3\\3}} = \binom{x+3}{2}\frac{1}{\binom{3x}{x}}$$

K

$$S_{\substack{3\\2x-2}} = \binom{3x-2}{2x-3}\frac{1}{\binom{3x}{x}}$$

$$S_{\substack{3\\2x-1}} = \binom{3x-1}{2x-2}\frac{1}{\binom{3x}{x}}$$

$$S_{\substack{3\\2x}} = \binom{3x}{2x-1}\frac{1}{\binom{3x}{x}} = \frac{2x}{x+1} \qquad \binom{50}{1\mathrm{K}\;2x}$$

For an r-tuple system we get:

$$R_{(r-1)x} = 1\frac{1}{\binom{rx}{x}}$$

$$R_{(r-1)x-1} = \binom{x+1}{1}\frac{1}{\binom{rx}{x}}$$

$$R_{(r-1)x-2} = \binom{x+2}{2}\frac{1}{\binom{rx}{x}}$$

K

$$R_3 = \binom{rx-3}{(r-1)x-3}\frac{1}{\binom{rx}{x}}$$

$$R_2 = \binom{rx-2}{(r-1)x-2}\frac{1}{\binom{rx}{x}}$$

$$R_1 = \binom{rx-1}{(r-1)x-1}\frac{1}{\binom{rx}{x}} = \frac{(r-1)}{r} \qquad \binom{51}{1\text{K }(r-1)x}$$

$$S_{\underset{1}{r}} = 1 \cdot \frac{1}{\binom{rx}{x}}$$

$$S_{\underset{2}{r}} = \binom{x+2}{1}\frac{1}{\binom{rx}{x}}$$

$$S_{\underset{3}{r}} = \binom{x+3}{2}\frac{1}{\binom{rx}{x}}$$

K

$$S_{\underset{(r-1)x-2}{r}} = \binom{rx-2}{(r-1)x-3}\frac{1}{\binom{rx}{x}}$$

$$S_{\underset{(r-1)x-1}{r}} = \binom{rx-1}{(r-1)x-2}\frac{1}{\binom{rx}{x}}$$

$$S_{\underset{(r-1)x}{r}} = \binom{rx}{(r-1)x-1}\frac{1}{\binom{rx}{x}} = \frac{(r-1)x}{x+1} \qquad \binom{52}{1\text{K }(r-1)x}$$

48. Hereby we have got means available for calculation of hindrances and degree of hindrance in an ordinary skewing system consisting of any number of equally large groups of preselectors under the condition that the arrangement constitutes a 'ring', and that the skewing for each group is also fully cyclic.

We shall now keep to the representation of traffic based on consideration of a sequence of calls. The traffic intensity is still denoted y, which here indicates the average number of calls from x^2 primary preselectors in the course of the average connection duration $\bar{t}$ in the busy hour. y is assumed to be on average approximately equally large in each group.

In order to be able to carry out the calculations in a fully general way we will assume that the system in question is so equipped in relation to the traffic that the hindrances in any case are very small. We will, therefore, in general not let the probability of one particular call being hindered be of any advantage to succeeding calls. It is easily seen that this simplification of the calculation does not cause any harmful error.

We have already calculated the probability of the various call states as well as the number of calls during $\bar{t}$ corresponding to each state. The probability of x, $x + 1$, $x + 2$, ..., rx, ... calls, respectively, in $\bar{t}$ under an aggregate traffic intensity ry is:

$$\overline{P}_x = \frac{e^{-ry}(ry)^x}{x!}$$

$$\overline{P}_{x+1} = \frac{e^{-ry}(ry)^{x+1}}{(x+1)!}$$

$$\overline{P}_{x+2} = \frac{e^{-ry}(ry)^{x+2}}{(x+2)!}$$

K

$$\overline{P}_{rx} = \frac{e^{ry}(ry)^{rx}}{(rx)!} \qquad \binom{53}{1\text{K }rx\text{K}}$$

K

Under the state probability $\overline{P}_x$ there are no traffic hindrances. They only begin under $\overline{P}_{x+1} = \frac{e^{-ry}(ry)^{x+1}}{(x+1)!}$, when $x + 1$ calls arrive, of which the x first occupy one circuit each in the system. There is then 1 call with a risk of passing the free circuits and thus get lost. This risk is: $R_{(r-1)x} = \frac{1}{\binom{rx}{x}}$.

The loss that one can estimate will be caused by this state and risk thus amounts to:

$$q_1 = \overline{P}_{x+1} \cdot R_{(r-1)x} = \overline{P}_{x+1} \cdot S_{\underset{1}{r}} = \frac{e^{-ry}(ry)^{x+1}}{(x+1)!} \cdot \frac{1}{\binom{rx}{x}} \qquad (54,1)$$

Under the state with probability $\overline{P}_{x+2}$, when $x+2$ calls arrive, the risk for the first call after the x^{th} is $R_{(r-1)x}$ and for the following $R_{(r-1)x-1}$. The likely loss then becomes:

$$q_2 = \overline{P}_{x+2}\left(R_{(r-1)x} + R_{(r-1)x-1}\right) = \overline{P}_{x+2} \cdot S_{\underset{2}{r}}$$

$$= \frac{e^{-ry}(ry)^{x+2}}{(x+2)!} \cdot \frac{\binom{x+2}{1}}{\binom{rx}{x}} \qquad (54,2)$$

In the same way we get, respectively, for 3, 4, ..., $(r - 1)x$ calls in excess of x in $\bar{t}$ the following likely losses:

$$q3=\overline{P}_{x+3}\left(R_{(r-1)x}+R_{(r-1)x-1}+R_{(r-1)x-2}\right)=\overline{P}_{x+3}S_{\substack{r\\3}}$$

$$=\frac{e^{-ry}(ry)^{x+3}}{(x+3)!}\frac{\binom{x+3}{2}}{\binom{rx}{x}}$$

$$q_4=\overline{P}_{x+4}\cdot S_{\substack{r\\4}}=\frac{e^{-ry}(ry)^{x+4}}{(x+4)!}\frac{\binom{x+4}{3}}{\binom{rx}{x}}$$

$$\mathrm{K}$$

$$q_{(r-1)x}=\overline{P}_{rx}\cdot S_{\substack{r\\(r-1)x}}=\frac{e^{-ry}(ry)^{rx}}{(rx)!}\frac{(r-1)x}{x+1} \qquad \binom{54}{3\mathrm{K}\ (r-1)}$$

Under the state corresponding to this last equation there will arrive exactly as many calls as there are circuits. Despite the fact that some of these calls will be blocked, so that one or more circuits will be left free, we will still, in consistency with what has been said above, not take advantage of this circumstance when deciding the loss for higher number of calls than *rx* in $\bar{t}$. So we assume the risk of hindrance to be equal to 1 for each call exceeding *rx*. Thus we get:

$$q_{(r-1)x+1}=\overline{P}_{rx+1}\cdot\left(S_{\substack{r\\(r-1)x}}+1\right)=\frac{e^{-ry}(ry)^{rx+1}}{(rx+1)!}\left(\frac{(r-1)x}{x+1}+1\right)$$

$$q_{(r-1)x+2}=\overline{P}_{rx+2}\left(S_{\substack{r\\(r-1)x}}+2\right)=\frac{e^{-ry}(ry)^{rx+2}}{(rx+2)!}\left(\frac{(r-1)x}{x+1}+2\right)$$

$$q_{(r-1)x+3}=\overline{P}_{rx+3}\left(S_{\substack{r\\(r-1)x}}+3\right)=\frac{e^{-ry}(ry)^{rx+3}}{(rx+3)!}\left(\frac{(r-1)x}{x+1}+3\right)$$

$$\mathrm{K} \qquad \binom{55}{1,2,3\mathrm{K}}$$

The aggregate loss in $\bar{t}$ thus amounts to:

$$Q_{\substack{r\\xy}}=q_1+q_2+q_3+q_4\mathrm{K}+q_{(r-1)x}+q_{(r-1)x+1}$$

$$+q_{(r-1)x+2}+q_{(r-1)x+3}+\mathrm{K}$$

$$=\frac{e^{-ry}}{\binom{rx}{x}}\frac{(ry)^{x+1}}{(x+1)!}\cdot(1+\frac{ry}{1}+\frac{(ry)^2}{2!}+\frac{(ry)^3}{3!}+\mathrm{K}$$

$$+\frac{(ry)^{(r-1)x-1}}{\left((r-1)x-1\right)!})$$

$$+\frac{e^{-ry}(ry)^{rx+1}}{(rx+1)!}\frac{(r-1)x}{(x+1)}$$

$$\cdot(1+\frac{ry}{(rx+2)}+\frac{(ry)^2}{(rx+2)(rx+3)}$$

$$+\frac{(ry)^3}{(rx+2)(rx+3)(rx+4)}+\mathrm{K}\ )+\frac{e^{-ry}(ry)^{rx+1}}{(rx+1)!}$$

$$\cdot(1+2\frac{ry}{(rx+2)}+3\frac{(ry)^2}{(rx+2)(rx+3)}$$

$$+4\frac{(ry)^3}{(rx+2)(rx+3)(rx+4)}+\mathrm{K}\ )$$

$$=\frac{e^{-ry}}{\binom{rx}{x}}\cdot\frac{(ry)^{x+1}}{(x+1)!}\cdot(e^{ry}-\frac{(ry)^{(r-1)x}}{\left((r-1)x\right)!}(1+\frac{ry}{\left((r-1)x+1\right)}$$

$$+\frac{(ry)^2}{\left((r-1)x+1\right)\left((r-1)x+2\right)}+\mathrm{K}$$

$$+\frac{(ry)^{2x-1}}{\left((r-1)x+1\right)\left((r-1)x+2\right)\mathrm{K}\left((r-1)x+2x-1\right)}+\mathrm{K}\ ))$$

$$+\frac{e^{-ry}(ry)^{rx+1}}{(rx+1)!}\cdot\frac{(r-1)x}{(x+1)}\cdot\frac{1}{\left(1-\frac{ry}{(rx+2+k_1)}\right)}$$

$$+\frac{e^{-ry}(ry)^{rx+1}}{(rx+1)!}\cdot\frac{1}{\left(1-\frac{ry}{(rx+2+k_2)}\right)^2}$$

$$<\frac{e^{-ry}}{\binom{rx}{x}}\frac{(ry)^{x+1}}{(r+1)!}\cdot(e^{ry}-\frac{(ry)^{(r-1)x}}{\left((r-1)x\right)!}(1+\frac{ry}{\left((r-1)x+1\right)}$$

$$+\frac{(ry)^2}{\left((r-1)x+\frac{3}{2}\right)^2}+\frac{(ry)^3}{\left((r-1)x+\frac{4}{2}\right)^3}+\frac{(ry)^4}{\left((r-1)x+\frac{5}{2}\right)^4}+\mathrm{K}$$

$$+\frac{(ry)^{2x-1}}{\left((r-1)x+\frac{2x}{2}\right)^{2x-1}}))$$

$$+\frac{e^{-ry}(ry)^{rx+1}}{(rx+1)!}\left(\frac{(r-1)x}{(x+1)}\cdot\frac{1}{\left(1-\frac{ry}{(rx+2)}\right)}+\frac{1}{\left(1-\frac{ry}{(rx+2)}\right)^2}\right)$$

$$<\frac{e^{-ry}}{\binom{rx}{x}}\cdot\frac{(ry)^{x+1}}{(x+1)!}(e^{ry}-\frac{(ry)^{(r-1)x}}{\left((r-1)x\right)!}(1+\frac{ry}{\left(rx-\frac{x-1}{2}\right)}$$

$$+\frac{(ry)^2}{\left(rx-\frac{x-1}{2}\right)^2}+\frac{(ry)^3}{\left(rx-\frac{x-1}{2}\right)^3}+\mathrm{K}+\frac{(ry)^{2x-1}}{\left(rx-\frac{x-1}{2}\right)^{2x-1}}))$$

$$+\frac{e^{-ry}(ry)^{rx+1}}{(rx+1)!}\left(\frac{(r-1)x}{(x+1)}\cdot\frac{1}{\left(1-\frac{ry}{(rx+2)}\right)}+\frac{1}{\left(1-\frac{ry}{(rx+2)}\right)^2}\right)$$

$$=\frac{e^{-ry}}{\binom{rx}{x}}\cdot\frac{(ry)^{x+1}}{(x+1)!}\left(e^{ry}-\frac{(ry)^{(r-1)x}}{\left((r-1)x\right)!}\cdot\frac{1-\frac{(ry)^{2x}}{\left(rx-\frac{x-1}{2}\right)^{2x}}}{1-\frac{ry}{\left(rx-\frac{x-1}{2}\right)}}\right)$$

$$\psi_{\substack{4\\10;5}} = \frac{e^{-20}}{\binom{40}{10}} \cdot \frac{(20)^{10}}{11!} \left(e^{20} - \frac{20^{30}}{30!} \cdot \frac{\left(\left(\frac{71}{2}\right)^{20} - 20^{20} \right)}{\left(\frac{71}{2}\right)^{19} \frac{31}{2}} \right)$$

$$+ \frac{e^{-20} 20^{40}}{41!} \left(\frac{30}{11} \cdot \frac{42}{22} + \frac{42^2}{22^2} \right)$$

$$= \frac{30!}{40!} \cdot \frac{20^{10}}{11} - \frac{e^{-20}}{40!} \cdot \frac{20^{40}}{11} \cdot \frac{71^{20} - 40^{20}}{71^{19} \cdot 31}$$

$$+ \frac{e^{-20} 20^{40}}{41!} \cdot \frac{42 \cdot 3}{11^2 \cdot 2} (10+7)$$

$$= \frac{30!}{40!} - \frac{20}{11} \cdot \frac{e^{-20}}{40!} \cdot \frac{20^{40}}{11} \cdot \frac{\left(71 \left(1 - \left(\frac{40}{71}\right)^{20} \right) \right)}{31}$$

$$+ \frac{e^{-20} 20^{40}}{41!} \cdot \frac{63}{11^2} 17$$

$$= \frac{e^{-30} 30^{30} \sqrt{2\pi 30}}{e^{-40} 40^{40} \sqrt{2\pi 40}} \cdot \frac{20^{10}}{11} - \frac{e^{-20} 20^{40}}{e^{-40} 40^{40} \sqrt{2\pi 40}} \cdot \frac{71}{11} \cdot \frac{99999}{31 \cdot 10^5}$$

$$+ \frac{e^{-20} 20^{40} \cdot 63 \cdot 17}{e^{-40} 40^{40} \sqrt{2\pi 40} \cdot 41 \cdot 11^2}$$

$$= \frac{e^{10} 3^{30} \sqrt{30} \cdot 2^{10}}{4^{40} \sqrt{40} \cdot 11} - \frac{e^{20} 2^{40} 71 \cdot 99999}{4^{40} \sqrt{2\pi 40} \cdot 11 \cdot 31 \cdot 10^5}$$

$$+ \frac{e^{20} \cdot 2^{40} \cdot 61 \cdot 17}{4^{40} \sqrt{2\pi 40} \cdot 41 \cdot 11^2}$$

$$= \frac{e^{10} 3^{30} \sqrt{3}}{2^{71} 11} - \frac{e^{20} 71 \cdot 99999}{2^{41} \sqrt{2\pi 10} \cdot 11 \cdot 31 \cdot 10^5}$$

$$+ \frac{e^{20} \cdot 63 \cdot 17}{2^{41} \sqrt{2\pi 10} \cdot 41 \cdot 11^2} = A - B + C$$

We have:

$\log A = 10 \log e + 30 \log 3 + 1/2 \log 3 - 71 \log 2 - \log 11$

$$\begin{array}{rlrr}
 & 4.3429450 & & \\
+ & 14.3136390 & - & 21.3731300 \\
+ & \underline{0.2385607} & - & \underline{1.0413927} \\
+ & 18.8951447 & - & 22.4145227 \\
 & & + & \underline{4 \qquad\qquad -4} \\
 & & - & 18.4145227 - 4
\end{array}$$

$= 0.4806220 - 4.$

Num. $\log A = 0.000302428$

$\log B = 20 \log e + \log 71 + \log 99999$
$\quad - 41 \log 2 - 1/2 \log (2\pi 10) - \log 11$
$\quad - \log 31 - 5 \log 10 =$

$$\begin{array}{rlrr}
 & 8.6858900 & - & 12.3422300 \\
+ & 1.8512583 & - & 0.8990900 \\
+ & \underline{4.9999957} & - & 1.0413927 \\
 & 15.5371440 & - & 1.4913617 \\
 & & - & \underline{5.} \\
 & & - & 20.7740744 \\
 & & + & \underline{6 \qquad\qquad -6} \\
 & & - & 14.7740744 - 6
\end{array}$$

$= 0.7630696 - 6$

Num. $\log B = 0.0000057952.$

$\log C = 20 \log e + \log 63 + \log 17$
$\quad - 41 \log 2 - 1/2 \log (2\pi 10) - \log 41 - 2 \log 11$

$$\begin{array}{rlrr}
= & 8.6858900 & - & 12.3422300 \\
+ & 1.7993405 & - & 0.8990900 \\
+ & \underline{1.2304489} & - & 1.6127839 \\
+ & 11.7156794 & - & \underline{2.0827854} \\
 & & - & 16.9368893 \\
 & & + & \underline{6 \qquad\qquad -6} \\
 & & - & 10.9368893 - 6
\end{array}$$

$= 0.7787901 - 6$

Num. $\log C = 0.0000060089.$

Hence we have:

$$\begin{array}{lcl}
A + C - B & = & 0.000302428 \\
 & + & \underline{0.000006009} \\
 & & (0.000308437) \\
 & - & \underline{0.000005795} \\
 & = & 0.000302642
\end{array}$$

The degree of hindrance is thus:

$\psi_{\substack{4\\10;5}} = 0.000302642$, ie. approx. 0.3‰.

We will also carry out the calculation for $y = 5.5 = 11/2$, $x = 10$, $r = 4$.

Then we get:

$$\psi_{\substack{4\\10;\frac{11}{2}}} = \frac{30!\ 10! \left(\frac{44}{2}\right)^{10}}{40!\ 11} - \frac{\left(\frac{44}{2}\right)^{10} \left(\frac{44}{2}\right)^{30} 10!}{40!\ 11! e^{\frac{44}{2}}} \frac{\left(\left(40 - \frac{9}{2}\right)^{20} - \left(\frac{44}{2}\right)^{20} \right)}{\left(40 - \frac{9}{2}\right)^{19} \left(40 - \frac{44}{2} - \frac{9}{2}\right)}$$

$$+ \frac{\left(\frac{44}{2}\right)^{40}}{40!\ 41\ e^{\frac{44}{2}}} \left(\frac{30 \cdot (40+2)}{11 \left(40 - \frac{44}{2} + 2\right)} + \frac{(40+2)^2}{\left(40 - \frac{44}{2} + 2\right)^2} \right)$$

$$= \frac{30!\ 22^{10}}{40!\ 11} - \frac{22^{40} \left(71^{20} - 44^{20}\right)}{40!\ 11 \cdot e^{22} 71^{19} \cdot 27} + \frac{22^{40} \cdot 3 \cdot 42}{40!\ 41 e^{22} \cdot 20} \left(\frac{10}{11} + \frac{7}{10} \right)$$

$$= \frac{30!\ 22^{10}}{40!\ 11} - \frac{22^4\ 71 \left(1 - \left(\frac{44}{71}\right)^{20} \right)}{40!\ 11 e^{22} 27} + \frac{22^{40} \cdot 3 \cdot 42 \cdot 177}{40!\ 41 e^{22} \cdot 20 \cdot 110}$$

$$= \frac{30^{30} \sqrt{2\pi 30} e^{-30} \cdot 22^{10}}{40^{40} \sqrt{2\pi 40}\ e^{-40} 11} - \frac{22^{40} 71 \left(1 - \left(\frac{44}{71}\right)^{20} \right)}{40^{40} \sqrt{2\pi 40}\ e^{-40} e^{22} 11 \cdot 27}$$

$$+\frac{22^{40}3\cdot 2\cdot 21\cdot 177}{40^{40}\sqrt{2\pi 40}e^{-40}e^{22}\cdot 41\cdot 20\cdot 110}$$

$$=\frac{e^{10}3^{30}\sqrt{3}11^{9}}{2^{71}10^{10}}-\frac{e^{18}11^{39}\cdot 71\cdot 99993}{2^{41}10^{40}\sqrt{2\pi 10}\cdot 27\cdot 10^{5}}$$

$$+\frac{e^{18}11^{39}\cdot 63\cdot 177}{2^{38}10^{42}\sqrt{2\pi 10}\cdot 41}=A-B+C$$

Here we have:

$$\begin{array}{lllll}
\log A & = & 10\log e & = & 4.3429450 \\
 & + & 30\log 3 & = & 14.3136390 \\
 & + & 1/2\log 3 & = & 0.2385607 \\
 & + & 9\log 11 & = & \underline{9.3725343} \\
 & & & & 28.2676790 \\
 & - & 71\log 2 & = & -21.3731300 \\
 & - & 10\log 10 & = & -10.0000000 \\
 & & & & \underline{+4 \qquad\qquad -4} \\
 & & & = & +0.8945490-4
\end{array}$$

Num. $\log A = 0.00078442$

$$\begin{array}{llll}
\log B = & & & \\
18\log e & = & 7.8173010 & \\
+39\log 11 & = & 40.6143153 & \\
+\log 71 & = & 1.8512583 & \\
+\log 99993 & = & \underline{4.9999696} & +55.2828442 \\
-41\log 2 & = & -12.3422300 & \\
-40\log 10 & = & -40. & \\
-1/2\log(2\pi 10) & = & -0.8990900 & \\
-\log 27 & = & -1.4313638 & \\
-5\log 10 & = & \underline{-5.} & -59.6726838 \\
 & & & \underline{+\ 5. \qquad -5} \\
 & & & =\ 0.6101604-5
\end{array}$$

Num. $\log B = 0.000040753$

$$\begin{array}{llll}
\log C = & & & \\
18\log e & = & 7.8173010 & \\
+39\log 11 & = & 40.6143153 & \\
+\log 63 & = & 1.7993405 & \\
+\log 177 & = & \underline{2.2479733} & +52.4789301 \\
-38\log 2 & = & -11.4391400 & \\
-42\log 10 & = & -42. & \\
-1/2\log(2\pi 10) & = & -0.8990900 & \\
-\log 41 & = & \underline{-1.6127839} & -55.9510139 \\
 & & & \underline{+\ 4. \qquad -4} \\
 & & & =\ 0.5279162-4
\end{array}$$

Num. $\log C = 0.00033722$.

Thus we get:

$$\begin{array}{lll}
A+C-B & = & +0.00078442 \\
 & & \underline{+0.00033722} \quad 0.00112164 \\
 & & \qquad\qquad\qquad \underline{-0.00004075} \\
 & & \psi_{10;\frac{11}{2}}^{\ 4} = 0.00108089
\end{array}$$

The degree of hindrance accordingly is very close to 1 ‰.

As we by earlier calculations have found that the degree of hindrance for an ordinary system without skewing with $x = 10$ and $y = 4$ is also approximately equal to 1 ‰, and as we with this skewed system of $x = 10$ circuits have been allowed to increase the traffic to $y = 5.5$ without any increase of the degree of hindrance, then we have obtained a traffic gain of (5.5 – 4) / 4 = 3/8 = 0.375, ie. an increase of the efficiency of the secondary selector arrangement of 37.5 percent.

With a given traffic and a suitable section of an exchange the average case will be that an efficiency increase of $\underline{a}$ percent permits a saving of $\underline{b}$ percent in the number of selectors. Then we have $b=\frac{a}{1+\frac{a}{100}}$, as the number of selectors for a given traffic can be assumed to be inversely proportional to their efficiency. For $a=37.5\%$ we thus get $b=\frac{37.5}{1.375}=27\frac{3}{11}\%$.

We can therefore with the assumed traffic in the present example reduce the number of secondary selectors by somewhat more than 27 %. This is of course under the assumption that the technical and practical requirements do not for some reason prevent the full utilisation of the arrangement of the devices.

The skewing arrangement thus appears to be very efficient for improvement of the access and thereby for savings in the supply of secondary selectors.

In the example that we have studied we are able to carry 5.5 telephone hours during the busy hour from 100 primary preselectors. This gives 3600 · 5.5 = 19800 telephone seconds. If each call is expected to have an average duration of 100 seconds, then we get 198 calls or per primary preselector 1.98 in this hour. Assuming a concentration of 14.3 %, ie. a concentration of approx. 1/7 of the whole day traffic in the busy hour, then this gives 1.98 · 7 = 13.86, or close to 14 calls or conversations per day per preselector or subscriber. As repeatedly emphasised, the calculations are carried out in a way that one wants to be on the safe side when utilising the results, so that most likely a somewhat larger traffic than calculated can be carried. With a greater traffic than $y = 5.5$ for $x = 10$, one must, however, be prepared that the degree of hindrance very soon will increase above the prescribed 1 ‰. One could for the present case further improve the system somewhat by making a skewing arrangement of more than 4 groups. With the derived formula it is possible to calculate the gain. However, any substantial increase will make the calculation rather difficult, while there is hardly any reason to assume that an increase of the number of skewed groups in excess of 4 or 5 will bring about particularly noticeable changes. In order to meet a substantial traffic increase in excess of 5.5 one must probably add more secondary selectors (or remove some of the primary ones, to place them into new groups), or one could furnish some spare arrangement.

X Calculation of the degree of hindrance in systems with spare arrangements

50. A system with spare circuits with attached selectors is characterised by adding, for a certain limited area, a number of circuits leading to a corresponding set of selectors that are common to several primary groups. These primary groups of subscribers or selectors each have access to a certain number of fully co-ordinate common circuits. The common spare circuits

serve to receive and carry on calls that are blocked from being carried by the ordinary circuits. Behind these spare sections additional spare selectors may supply outlets for calls that are blocked in the first spare section. We will, however, only consider one single row of spare selectors as outlets for calls blocked in the ordinary row of selectors.

51. Let us denote the number of common ordinary circuits for each limited group by x_0. These circuits lead to x_0 selectors in the ordinary row of selectors. The number of spare circuits for each group is denoted x_1. These lead to x_1 selectors in the spare row. We will assume that r groups have one such row or section in *common*. Thus the r groups share altogether only x_1 spare selectors under the condition that no skewing is introduced. The traffic intensity of each group is denoted y. We will assume that the subscribers by connection to the exchange are so evenly mixed that the average traffic is approximately equally large for each group.

We will now try to calculate the final hindrances and the corresponding degree of hindrance in such a system.

The traffic y will divide in two parts, one passes unhindered through the ordinary bundle of circuits, the other, being blocked in this bundle, search for free circuits in the spare section, where it will be carried or finally blocked. We will denote the first part of the traffic by y_0 and the second by y_1, so that we have $y_0 + y_1 = y$.

According to what has been derived earlier we have:

$$y_1 = \frac{e^{-y} y^{x_0+1}}{(x_0+1)!}\left(1 + 2\frac{y}{(x_0+2)} + 3\frac{y^2}{(x_0+2)(x_0+3)} + \dots\right)$$

$$< \frac{e^{-y} y^{x_0+1}}{(x_0+1)!\left(1-\frac{y}{x_0+2}\right)^2}$$

When we load the ordinary bundle only moderately, ie. we let x_0 be substantially larger than y, then we may consider the expression behind the inequality sign as approximately equal to y_1. The traffic that resort to the spare section from all r groups is equal to ry_1. The number of circuits that are for common use by this traffic is as mentioned x_1. The amount that is hindered because of busy or lacking circuits in the spare section is therefore:

$$Q_{r(x_1 y_1)} = \frac{e^{-ry_1}(ry_1)^{x_1+1}}{(x_1+1)!} \frac{1}{\left(1-\frac{ry_1}{x_1+2}\right)^2} \tag{58}$$

The degree of hindrance becomes:

$$\psi_{r(x_1 y_1)} = \frac{Q_{r(x_1 y_1)}}{ry} = \frac{e^{-ry_1}(ry_1)^{x_1+1}}{ry(x_1+1)!} \frac{1}{\left(1-\frac{ry_1}{x_1+2}\right)^2} \tag{59}$$

52. On this basis we will calculate the degree of hindrance for a system with $x_0 = 8$, $x_1 = 2$, $y = 6$, $r = 2$. This gives:

$$y_1 = \frac{e^{-6}6^9}{9!}\frac{10^2}{4^2} = \frac{e^{-6}}{9!}\frac{3^9 \cdot 2^9 \cdot 5^2 \cdot 2^2}{2^4}$$

$$= \frac{e^{-6}}{9!}\frac{3^9 2^7 5^2}{1}$$

$\log y_1 = 9 \log 3 + \log 128 + \log 25 - 6 \log e - \log 9!$
$= 9 \cdot 0.4771213 + 2.1072100 + 1.3979400$
$- 6 \cdot 0.4342945 - 5.5597631 =$

$$\begin{array}{lll} & 4.2940917 & \\ + & 2.1072100 & \\ \underline{+} & \underline{1.3979400} & \quad 7.7992417 \\ & & \quad + 1. \qquad\quad -1 \\ - & 2.6057670 & \\ \underline{-} & \underline{5.5597631} & \quad \underline{-\ 8.1655301} \\ & & \quad = 0.6337116 - 1 \end{array}$$

Num. $\log y_1 = 0.43024 = y_1$.

One must remark here that because of the small difference between x_0 and y we get a somewhat too large value of y_1 caused by setting the previously discussed quantity k equal to 2 in the fraction $\frac{1}{\left(1-\frac{y}{x+k}\right)^2}$. It should correctly be somewhat greater than 2, how much depends on x and y, but is otherwise a question that is not easily answered. It may be noted that a record of traffic hindrances based on Bortkewitsch' table shows a total hindrance of 0.3192 for $x = 8$ and $y = 6$. But this number is determined by summation of only a limited number of terms of a series that is not strongly converging, and should only be used with care. If we take the mean value of the two numbers 0.43024 and 0.31290 we get a value that does not differ much from the likely one. Thus for further treatment of the matter we set $y_1 = \frac{0.43024 + 0.31290}{2} = 0.37157$. Thus we have have $ry_1 = 2 \cdot 0.37157 = 0.74314$. We round this number upwards to 0.75 and set $ry_1 = 3/4$. Furthermore we will for the rest of the calculation set $k = 3$, which is more in agreement with the mentioned table. Thus we write

$$Q_{r(x_1 y_1)} = \frac{e^{-ry_1}(ry_1)^{x_1+1}}{(x_1+1)!} \frac{1}{\left(1-\frac{ry_1}{x_1+3}\right)^2}$$

$$= \frac{e^{-\frac{3}{4}(\frac{3}{4})^3}}{3!} \frac{1}{\left(1-\frac{\frac{3}{4}}{5}\right)^2}$$

$$= \frac{e^{-\frac{3}{4}} \cdot 3^3 \cdot 20^2}{3! \cdot 4^3 \cdot 17^2} = \frac{e^{-\frac{3}{4}} \cdot 27 \cdot 400}{6 \cdot 64 \cdot 289}$$

$$y_1 = \frac{e^{-y} y^{x_0+1}}{(x_0+1)!}\left(1+2\frac{y}{(x_0+2)}+3\frac{y^2}{(x_0+2)(x_0+3)}+K\right)$$

$$< \frac{e^{-y} y^{x_0+1}}{(x_0+1)!\left(1-\frac{y}{x_0+2}\right)^2}$$

When we load the ordinary bundle only moderately, ie. we let x_0 be substantially larger than y, then we may consider the expression behind the inequality sign as approximately equal to y_1. The traffic that resort to the spare section from all r groups is equal to ry_1. The number of circuits that are for common use by this traffic is as mentioned x_1. The amount that is hindered because of busy or lacking circuits in the spare section is therefore:

$$Q_{r(x_1 y_1)} = \frac{e^{-ry_1}(ry_1)^{x_1+1}}{(x_1+1)!}\frac{1}{\left(1-\frac{ry_1}{x_1+2}\right)^2} \tag{58}$$

The degree of hindrance becomes:

$$\psi_{r(x_1 y_1)} = \frac{Q_{r(x_1 y_1)}}{ry} = \frac{e^{-ry_1}(ry_1)^{x_1+1}}{ry(x_1+1)!}\frac{1}{\left(1-\frac{ry_1}{x_1+2}\right)^2} \tag{59}$$

52. On this basis we will calculate the degree of hindrance for a system with $x_0 = 8$, $x_1 = 2$, $y = 6$, $r = 2$. This gives:

$$y_1 = \frac{e^{-6}6^9}{9!}\frac{10^2}{4^2} = \frac{e^{-6}}{9!}\frac{3^9\cdot 2^9\cdot 5^2\cdot 2^2}{2^4}$$

$$= \frac{e^{-6}}{9!}\frac{3^9 2^7 5^2}{1}$$

$\log y_1 = 9 \log 3 + \log 128 + \log 25 - 6 \log e - \log 9!$
$= 9 \cdot 0.4771213 + 2.1072100 + 1.3979400$
$- 6 \cdot 0.4342945 - 5.5597631 =$

$$\begin{array}{lll} & 4.2940917 & \\ + & 2.1072100 & \\ + & \underline{1.3979400} & \quad 7.7992417 \\ & & + \ 1. \qquad\quad -1 \\ - & 2.6057670 & \\ - & \underline{5.5597631} & \quad \underline{-\ 8.1655301} \\ & & = \ 0.6337116 - 1 \end{array}$$

Num. $\log y_1 = 0.43024 = y_1$.

One must remark here that because of the small difference between x_0 and y we get a somewhat too large value of y_1 caused by setting the previously discussed quantity k equal to 2 in the fraction $\frac{1}{\left(1-\frac{y}{x+k}\right)^2}$. It should correctly be somewhat greater than 2, how much depends on x and y, but is otherwise a question that is not easily answered. It may be noted that a record of traffic hindrances based on Bortkewitsch' table shows a total hindrance of 0.3192 for $x = 8$ and $y = 6$. But this number is determined by summation of only a limited number of terms of a series that is not strongly converging, and should only be used with care. If we take the mean value of the two numbers 0.43024 and 0.31290 we get a value that does not differ much from the likely one. Thus for further treatment of the matter we set $y_1 = \frac{0.43024+0.31290}{2} = 0.37157$. Thus we have have $ry_1 = 2 \cdot 0.37157 = 0.74314$. We round this number upwards to 0.75 and set $ry_1 = 3/4$. Furthermore we will for the rest of the calculation set $k = 3$, which is more in agreement with the mentioned table. Thus we write

$$Q_{r(x_1 y_1)} = \frac{e^{-ry_1}(ry_1)^{x_1+1}}{(x_1+1)!}\frac{1}{\left(1-\frac{ry_1}{x_1+3}\right)^2}$$

$$= \frac{e^{-\frac{3}{4}\left(\frac{3}{4}\right)^3}}{3!}\frac{1}{\left(1-\frac{\frac{3}{4}}{5}\right)^2}$$

$$= \frac{e^{-\frac{3}{4}}\cdot 3^3\cdot 20^2}{3!\cdot 4^3\cdot 17^2} = \frac{e^{-\frac{3}{4}}\cdot 27\cdot 400}{6\cdot 64\cdot 289}$$

$\log Q_{r(x_1 y_1)} = \log\ 27 + \log\ 400 - 3/4 \log\ e$
$- \log 6 - \log 64 - \log 289 =$

$$\begin{array}{lll} & 1.4313638 & \\ + & \underline{2.6020600} & \\ & & \quad 4.0334238 \\ & & + \ 2. \qquad\quad -2 \\ - & 0.3257209 & \\ - & 0.7781513 & \\ - & 1.8061800 & \\ - & \underline{2.4608978} & \\ & & \underline{-\ 5.3709500} \\ & & = \ 0.6624738 \ -2 \end{array}$$

Num. $\log\ Q_{r(x_1 y_1)} = 0.04597 = Q_{r(x_1 y_1)}$

We thus get $\frac{Q_{r(x_1 y_1)}}{ry} = \frac{0.04597}{12} = 0.00383.$

The degree of hindrance for the system in question with a traffic intensity of $y = 6$ thus is near 4 ‰, which is more than recommended.

If we allow for $y = 5.5$ we get for the same system, when we in both cases set $k = 3$:

$$y_1 = \frac{e^{-\frac{11}{2}}\left(\frac{11}{2}\right)^9}{9!}\cdot\frac{22^2}{11^2} = \frac{e^{-\frac{11}{2}}11^9\cdot 2^2\cdot 11^2}{9!\cdot 2^9\cdot 11^2}$$

$$= \frac{e^{-\frac{11}{2}}11^9}{9!\cdot 2^7}$$

try to examine the case of a traffic intensity of $y = 6$. We now possess all necessary information for our calculation.

We will calculate $Q_{\substack{r\\x_0 y}}$ as we set the expression $\frac{1}{\binom{rx_0}{x_0}}$ on the form $\frac{(rx_0 - x_0)!x_0!}{(rx_0)!}$, and the previously discussed quantity k is set to $k = 3$. This is permissible and required, as the difference between x_0 and y is so small. We have, by a closer study of the infinite series in question

$$1+\frac{2y}{x+2}+\frac{3y^2}{(x+2)(x+3)}+\frac{4y^3}{(x+2)(x+3)(x+4)}+\ldots=\frac{1}{\left(1-\frac{y}{x+k}\right)^2},$$

found that k with sufficient exactness and required security may be set equal to 3 when

$$\frac{8y^3}{(x+3)(x+5)(x+3-y)^2} > \frac{2}{(x+2)}+\frac{3y}{(x+2)(x+3)}+\frac{4y^2}{(x+2)(x+3)^2(x+4)} \quad (60)$$

Even if this condition is not satisfied we may of course also set $k = 3$ when the sum of the terms containing the factor $\frac{1}{\left(1-\frac{y}{x+k}\right)}$ or $\frac{1}{\left(1-\frac{y}{x+k}\right)^2}$ is small compared to term or terms that do not contain this factor.

After this we get:

$$Q_{\substack{r\\x_0 y}} = \frac{e^{-ry}(ry)^{x_0+1}(rx_0-x_0)!}{(rx_0)!(x_0+1)}\Big(1+\frac{ry}{1!}+\frac{(ry)^2}{2!}+\frac{(ry)^3}{3!}+\ldots$$

$$+\frac{(ry)^{rx_0-x_0-1}}{(rx_0-x_0-1)!}\Big)$$

$$+\frac{e^{-ry}(ry)^{rx_0+1}(rx_0-x_0)(rx_0+3)}{(rx_0+1)!(x_0+1)(rx_0+3-ry)}$$

$$+\frac{e^{-ry}(ry)^{rx_0+1}(rx_0+3)^2}{(rx_0+1)!(rx_0+3-ry)^2}$$

$$=\frac{e^{-18}18^9 16!}{24!\cdot 9}\left(1+\frac{18}{1!}+\frac{18^2}{2!}+\frac{18^3}{3!}+\ldots+\frac{18^{15}}{15!}\right)$$

$$+\frac{e^{-18}18^{25}16\cdot 27}{25!\cdot 9\cdot 9}+\frac{e^{-18}18^{25}27^2}{25!\cdot 9^2}$$

For calculation of the sum that is contained in the great parenthesis of the first term we will not utilise the simplification discussed earlier, as it here, with the small difference between x_0 and y, would imply an unacceptable inaccuracy. Therefore, we calculate this sum and find by means of logarithms approximately:

$$1+\frac{18}{1!}+\frac{18^2}{2!}+\ldots+\frac{18^{15}}{15!}=18821628$$

Hereby we get, as we combine the two last terms in $Q_{\substack{r\\x_0 y}}$:

$$Q_{\substack{r\\x_0 y}} = \frac{e^{-18}18^9 16!}{24!\cdot 9}18821628+\frac{e^{-18}18^{25}\cdot 1161}{25!\cdot 81}$$

$$=A+B$$

$\log A$	$=$	$9 \log 18$	11.2974525	
	$+$	$\log 16!$	13.3206196	
	$+$	$\log 18821628$	7.2746572	31.8927293
				$+\ 1 \qquad -1$
	$-$	$\log 24!$	$-$ 23.7927056	
	$-$	$\log 9$	$-$ 0.9542425	
	$-$	$18 \log e$	$-$ 7.8173010	
				$-$ 32.5642491
				$=$ 0.3284802 $-$ 1

Num. log $(0.3284802 - 1) = 0.21305 = A$.

$\log B$	$=$	$25 \log 18$	31.3818125	
	$+$	$\log 1161$	3.0648322	
				34.4466447
				$+\ 1 \qquad -1$
	$-$	$\log 25!$	$-$ 25.1906456	
	$-$	$\log 81$	$-$ 1.9084850	
	$-$	$18 \log e$	$-$ 7.8173010	
				$-$ 34.9164316
				$=$ 0.5302131 $-$ 1

Num. log $(0.5302131 - 1) = 0.33901 = B$.

$$Q_{\substack{r\\x_0 y}} = Q_{\substack{3\\8,6}} = A+B = 0.55206$$

The part of the traffic $ry = 18$ that is blocked under the main selection and therefore flows over to the spare row is thus equal to 0.552, as we omit the last two decimals. Here we get: $x_1 = 2$, $y = 0.552$, $r = 3$, and by the same formula as above we get:

$$Q_{\substack{r\\x_1 y}} = \frac{e^{-1.656}4!\cdot 1.656^3}{6!\cdot 3}\left(1+\frac{1.656}{1!}+\frac{1.656^2}{2!}+\frac{1.656^3}{3!}\right)$$

$$+\frac{e^{-1.656}1.656^7\cdot 4\cdot 9}{7!\cdot 3\cdot 7.344}+\frac{e^{-1.656}1.656^7\cdot 9^2}{7!\cdot 7.344^2}$$

$$=\frac{e^{-1.656}1.656^3\cdot 3.4129}{90}+\frac{e^{-1.656}1.656^7\cdot 169.128}{7!\cdot 7.344^2}$$

$$=A+B$$

$$
\begin{array}{llrrr}
\log A & = 3 \log 1.656 & 0.6571809 & & \\
 & + \log 3.4129 & \underline{0.5331236} & 1.1903045 & \\
 & & & + \; 2 & -2 \\
 & - \log 90 & -\; 1.9542425 & & \\
 & - 1.656 \log e & -\; \underline{0.7191917} & & \\
 & & & -\; \underline{2.6734342} & \\
 & & & =\; 0.5168703 & -2
\end{array}
$$

Num. log $(0.5168703 - 2) = 0.0328753 = A$.

$$
\begin{array}{llrrr}
\log B & = 7 \log 1.656 & 1.5334221 & & \\
 & + \log 169.128 & \underline{2.2282155} & & \\
 & & & 3.7616376 & \\
 & & & + \; 3 & -3 \\
 & - \log 7! & -\; 3.7024305 & & \\
 & - 1.656 \log e & -\; 0.7191917 & & \\
 & - 2 \log 7.344 & -\; \underline{1.7318654} & & \\
 & & & -\; \underline{6.1534876} & \\
 & & & =\; 0.6081500 & -3
\end{array}
$$

Num.log $(0.6081500 - 3) = 0.0040565 = B$.

$$Q_{\substack{r\\x_1 y}} = Q_{\substack{3\\2;0,552}} = A + B = 0.0369318$$

This is thus the total traffic amount that suffers final blocking. It will appear from the previous that the 3 spare groups include the whole system of 9 groups of primary selectors. From these 9 groups comes an aggregate traffic of $9y = 54$. When we divide the whole finally blocked traffic by the whole original traffic (without hindrances) we get the degree of hindrance. In the present case it will therefore be:

$$\psi_{\substack{r\\xy}} = \frac{Q_{\substack{r\\x_1 y}}}{9y} = \frac{0.0369318}{54} = 0.0006839,$$

so we get approx. 7/10 ‰.

This gives a fairly good illustration of what can be achieved by the combined circuit arrangements as described: While a 10-outlet selector system with an ordinary bundle of circuits alone can carry a traffic of $y = 4$ via 10 secondary selectors with a degree of hindrance of a little more than 1 ‰, one can via 78 secondary selectors carry a traffic of $9y = 54$, and accordingly an average a traffic of $y = 6$ via $78/9 = 8\,{}^2/_3$ secondary selectors. Simultaneous to making possible a traffic increase of 50 % in each group we thus have a reduction of secondary selectors of $13\,{}^1/_3$ %. With a given traffic we can, therefore, after having reduced the number of secondary selectors by $13\,{}^1/_3$ % by means of the combined arrangement for an exchange section of suitable size, further reduce the remaining number by

$b = \dfrac{a}{1+\frac{a}{100}} = \dfrac{50}{1\frac{1}{2}} = 33\frac{1}{3}\%$. The total reduction may in this way become $100 - \left(100 - 13\frac{1}{3}\right)\frac{2}{3} = 42\frac{2}{9}\%$, given that the conditions for full utilisation of the calculated possibility of saving are fulfilled.

We want to see by how many percent we may increase the traffic when the secondary selectors are retained and the selectors that are made superfluous by the spare arrangement are utilised by new groups for the increased traffic. Then we may reverse the given equation for the relation between saving and efficiency given in percent and insert $b = 42\,{}^2/_9\,\% = 380/9\,\%$. Then we get:

$$a = \frac{b}{a - \frac{b}{100}} = \frac{\frac{380}{9}}{1 - \frac{\frac{380}{9}}{100}} = \frac{3800}{90-38} = \frac{3800}{52} = 73\frac{10}{13}\%$$

The total efficiency of a given number of secondary selectors is thus increased by approx. 74 %, of course still under the assumption that the suggested conditions for the feasibility of the theoretical calculations are fulfilled.

54. For comparison with the example treated above we will now consider an arrangement with the first 7 circuits from the contact arc for the main selection and the last 3 for the spare selection. We will here combine 4 and 4 groups of 7^2 primary selectors and let 3 such sections of $4 \cdot 7^2$ primary selectors be allotted a common spare system. The number of primary selectors thus will be $3 \cdot 4 \cdot 7^2 = 588$. We will here not let each single spare circuit be fully accessible for a whole section of 4 groups, as a calculation that is carried out shows that to be unsatisfactory, but we will let each spare circuit only be accessible for calls at 12 points, ie. each spare circuit encompasses $1\,{}^5/_7$ groups. As the calls from 12 primary groups flow in on $7 \cdot 12$ points, we get located in the longitudinal direction a sequence of 7 spare circuits, and since there are 3 rows of such circuits beside one another, we get 21 spare circuits. We will now let the first 3 groups of 3 spare circuits form a closed system and the last 4 groups of 3 spare circuits likewise form a closed system. Each skewing step for the spare circuits encompasses ${}^4/_7$, ie. 4 access points. We will now stay with a traffic intensity of $y = 6$. According to the foregoing we have in the main selection system $x_0 = 7$, $y = 6$, $r = 4$. In the spare part we have $x_1 = 3$ and r equal to 3 and 4 respectively. We could let the 7 spare groups form one single combination with $r = 7$, but we have preferred the indicated arrangement in order to simplify the calculation and to have the best possible opportunity to compare with the previous system.

The result of the main selection that takes place in an ordinary skewing system is as follows:

$$
\begin{aligned}
Q_{\substack{r\\x_0 y}} &= \frac{e^{-24}\,24^8\,21!}{28!\cdot 8}\left(1 + \frac{24}{1!} + \frac{24^2}{2!} + \frac{24^3}{3!} + \ldots + \frac{24^{20}}{20!}\right) \\
&\quad + \frac{e^{-24}\,24^{29}\cdot 21\cdot 31}{29!\cdot 8\cdot 7} + \frac{e^{-24}\,24^{29}\cdot 31^2}{29!\cdot 7^2} \\
&= \frac{e^{-24}\,24^8\cdot 21!\cdot 6428040000}{28!\cdot 8} \qquad {}^{9)} \\
&= \frac{e^{-24}\,24^{29}\cdot 651}{29!\cdot 56} + \frac{e^{-24}\,24^{29}\cdot 961}{29!\cdot 49} \\
&= A + B + C
\end{aligned}
$$

9) *approximately.*

$$\begin{array}{llrrr}
\log A & = 8\log 24 & 11.0416896 & & \\
 & + \log 21! & 19.7083439 & & \\
 & + \log 642804 & 5.8080786 & & \\
 & + \log 10000 & \underline{4.} & & \\
 & & & 40.5581121 & \\
 & & & +\ 1 & -1 \\
 & - \log 28! & -29.4841407 & & \\
 & - \log 8 & -\ 0.9030900 & & \\
 & - 24\log e & \underline{-10.4230680} & & \\
 & & & \underline{-40.8102987} & \\
 & & & =\ 0.7478294 & -1
\end{array}$$

Num. log $(0.7478294 - 1) = 0.55953 = A$.

$$\begin{array}{llrrr}
\log B & = 29\log 24 & 40.0261248 & & \\
 & + \log 651 & \underline{2.8135810} & & \\
 & & & 42.8397058 & \\
 & & & +\ 1 & -1 \\
 & - \log 29! & -30.9465378 & & \\
 & - \log 56 & -\ 1.7481880 & & \\
 & - 24\log e & \underline{-10.4230680} & & \\
 & & & \underline{-43.1177938} & \\
 & & & =\ 0.7219120 & -1
\end{array}$$

Num. log $(0.7219120 - 1) = 0.52712 = B$.

$$\begin{array}{llrrr}
\log C & 29\log 24 & 40.0261248 & & \\
 & + \log 961 & \underline{2.9827234} & & \\
 & & & 43.0088482 & \\
 & & & +\ 1 & -1 \\
 & - \log 29! & -30.9465387 & & \\
 & - \log 49 & -\ 1.6901961 & & \\
 & - 24\log e & \underline{-10.4230680} & & \\
 & & & \underline{-43.0598028} & \\
 & & & =\ 0.9490454 & -1
\end{array}$$

Num. log $(0.9490454 - 1) = 0.88930 = C$.

Accordingly we have:

$$\begin{aligned}
Q_{\substack{r\\x_0 y}} = Q_{\substack{4\\7{,}6}} = A+B+C = {} & 0.55953 \\
& +0.52712 \\
& \underline{+0.88930} \\
& = 1.97595
\end{aligned}$$

This is the traffic that is hindered within a section of 4 groups of primary selectors.

That leaves for each group:

$1.97595 : 4 = 0.49399 =$ approx. 0.494.

We shall now calculate the result of the spare row of the first section. Here we have $x_1 = 3$, $r = 3$ and $y = 0.494 \cdot {}^{12}/_{7} = 0.849$, as ${}^{12}/_{7}$ groups have unlimited access to one spare circuit. We round off the number for y to $y = 0.85$ and $ry = 2.55$.

Thus we get:

$$Q_{\substack{r\\x_1 y}} = \frac{e^{-2.55}}{9!}\frac{2.55^4\, 6!}{4}\left(1+\frac{2.55}{1!}+\frac{2.55^2}{2!}+\ldots+\frac{2.55^5}{5!}\right)$$

$$+\frac{e^{-2.55}\,2.55^{10}\cdot 6\cdot 12}{10!\cdot 4\cdot 9.45}+\frac{e^{-2.55}\,2.55^{10}\cdot 12^2}{10!\cdot 9.45^2}$$

$$=\frac{e^{-2.55}\,6!\cdot 42.2825\cdot 12.2251}{9!\cdot 4}$$

$$+\frac{e^{-2.55}\,2.55^{10}\cdot 18}{10!\cdot 9.45}+\frac{e^{-2.55}\,2.55^{10}\cdot 144}{10!\cdot 89.3025}$$

$$= A+B+C$$

$$\begin{array}{llrrr}
\log A & = \log 6! & 2.8573325 & & \\
 & + \log 42.2825 & 1.6260580 & & \\
 & + \log 12.2251 & \underline{1.0872525} & & \\
 & & & 5.5706430 & \\
 & & & +\ 2 & -2 \\
 & - \log 9! & -\ 5.5597630 & & \\
 & - \log 4 & -\ 0.6020600 & & \\
 & - 2.55\log e & \underline{-\ 1.1074510} & & \\
 & & & \underline{-\ 7.2692740} & \\
 & & & =\ 0.3013690 & -2
\end{array}$$

Num. log $(0.3013690 - 2) = 0.0200156 = A$.

$$\begin{array}{llrrr}
\log B & = 10\log 2.55 & 4.0654020 & & \\
 & + \log 18 & \underline{1.2552725} & & \\
 & & & 5.3206745 & \\
 & & & +\ 4 & -4 \\
 & - \log 10! & -\ 6.5597630 & & \\
 & - \log 9.45 & -\ 0.9754318 & & \\
 & - 2.55\log e & \underline{-\ 1.1074510} & & \\
 & & & \underline{-\ 8.6426458} & \\
 & & & =\ 0.6780287 & -4
\end{array}$$

Num. log $(0.6780287 - 4) = 0.00047646 = B$.

$$\begin{array}{llrrr}
\log C & = 10\log 2.55 & 4.0654020 & & \\
 & + \log 144 & \underline{2.1583625} & & \\
 & & & 6.2237645 & \\
 & & & +\ 4 & -4 \\
 & - \log 10! & -\ 6.5597630 & & \\
 & - 2.55\log e & -\ 1.1074510 & & \\
 & - \log 89.3025 & \underline{-\ 1.9508636} & & \\
 & & & \underline{-\ 9.6180776} & \\
 & & & =\ 0.6056869 & -4
\end{array}$$

Num. log $(0.6056869 - 4) = 0.00040335 = C$.

Thus we have:

$$\begin{aligned}
Q_{\substack{r\\x_1 y}} = Q_{\substack{3\\3;0.85}} = {} & A\ldots\ldots 0.0200156 \\
& +\text{B}\ldots\ldots\ 0.0004765 \\
& \underline{+\text{C}\ldots\ldots\ 0.0004034} \\
= {} & \qquad\quad 0.0208955
\end{aligned}$$

This is the traffic that suffers final hindrance in a section of 3 spare groups. As these 3 groups covers $3 \cdot {}^{12}/_{7}$ primary groups and as each primary group has a traffic of $y = 6$, then the traffic that has had to suffer the mentioned hindrance is equal to $3 \cdot {}^{12}/_{7} \cdot 6 = {}^{216}/_{7}$. The hindered traffic must be divided by this number to get the degree of hindrance. Accordingly we have:

$$\frac{\frac{Q_{r \atop x_1 y}}{216}}{7} = \frac{0.0208955 \cdot 7}{216} = 0.000677 = \psi_{r \atop xy}$$

We see that the degree of hindrance here is pretty much equal to the one calculated for the previously treated system. It may in both cases with sufficient accuracy be set to $^7/_{10}$ ‰.

The second spare section of 4 groups will according to our formula get a somewhat smaller degree of hindrance. The difference, however, is rather insignificant in the present case, and a calculation for this section may therefore be omitted.

As for the saving of secondary selectors, this is a bit smaller than that of the previous case. Here we have for the main selection $7 \cdot 12 = 84$ selectors and for the spare selection $3 \cdot 7 = 21$ selectors, altogether 105 selectors in the secondary groups. These 105 selectors give satisfactory outlet for a traffic of $12\,y = 72$, or on average a traffic of $y = 6$ over $^{105}/_{12} = 8\ ^3/_4$ secondary selectors. The saving of secondary selectors compared to an ordinary system is thus $12\ ^1/_2$ %, not considering the very significant saving that is obtained by the 50 % improved traffic access that amounts to $33\ ^1/_3$ %. The total improvement of the traffic possibilities, when calculated in the same way as earlier, will be $71\ ^3/_7$ %.

55. With the application of selectors with more than 10 contacts in the arc the improvement by the arrangements in question will be less in percent than calculated here, which will also be evident from the relevant formulas and from our 1 ‰ curve.

56. We will not fail to point out that the conditions for the validity of the simplified calculations, as formulated in section IV (12) and in section V (17) are not satisfied in the last case. According to the former condition we ought to have:

$$\left(x_0+1\right)\left(1-\sqrt{\frac{n-x_0-1}{\left(x_0+1\right)(n-1)}}\right) \geq y, \text{ or}$$

$$8\left(1-\sqrt{\frac{7^2-8}{8\left(7^2-1\right)}}\right)=\left(8-\sqrt{\frac{41}{6}}\right)$$

$= 5.3739 \geq 6$, which is not an insignificant disagreement. According to the latter condition we ought to have:

$$x_0+1-\sqrt{x_0+1}>y, \text{ or}$$

$8 - 2.8284 = 5.1716 > 6$, which indicates an even greater inaccuracy. However, after the calculations that gave cause for the mentioned conditions, there have also been made some other simplifications. For one thing we have in the skewing system not credited new calls the failure of previous calls, to which in the present case could be attached great significance for the main selection. Thus, without offering any mathematical proof of the matter, we may assume that altogether there is sufficient compensation for the excess made of the limits discussed in IV and V.

57. With regard to the comparison of the two assumed systems, of which the degree of hindrance is calculated in the preceding, it ought to be remarked that the traffic from each primary selector is quite different for the two cases. If these are denoted s_1 and s_2 we have:

$$s_1=\frac{6}{8^2}=0.09375 \text{ telephone hours in busiest hour}$$

$$s_2=\frac{6}{7^2}=0.12245 \text{ telephone hours in busiest hour}$$

Assuming each call or conversation to last 2 minutes = $^1/_{30}$ hour, we get:

$s_1 = 0.09375 \cdot 30 =$ approx. 2.81 conversations in the busiest hour

$s_2 = 0.12245 \cdot 30 =$ approx. 3.67 conversations in the busiest hour

With a concentration of $^1/_7$ or 14.3 % this leads, respectively, to an average of approx. $19\ ^2/_3$ and approx. $25\ ^2/_3$ conversations from each, which is an exceptionally large traffic. However, one will come quite close to practical requirements in larger cities by presuming approximately half that traffic. In that case one ought to allocate twice as many primary selectors in each group in order to fully utilise the secondary selectors. This is done by extending the groups in length, as the width is given by the numbers 8 and 7, respectively. But then the skewing also has to be changed, as *each* group has to form a *complete cycle*, in order to obtain the advantage as calculated. Therefore we double the skewing step for each circuit in the main selection system as well as the spare system, as we let the calls from the primary selectors have access to the double number of points and let the number of secondary selectors stay unchanged.

However, if we let the primary groups be extended by only 50 %, we may in the main system let a skewing step covering two access points be followed by a step covering only one point and carry out a corresponding change in the spare system to obtain a complete skewing for each group.

In a similar way one can find modifications corresponding to an increase of for instance 150 %, 200 % (ie. triple size) etc., without deviation from the conditions assumed for our calculations concerning the skewing.

If the changes suggested here in some occurring cases do not comply with the practical-technical requirements, then one can apply an *overcomplete* skewing, ie. the last access point on the last common circuit of a group is brought *past* the first access point of the first circuit of the following group. Such an overcomplete skewing will be *more favourable* for the forwarding of calls than the previously described solution with a complete cycle for each group. One will therefore still be on the safe side by using our formula for a such skewing. But this skewing type has the disadvantage of being less surveyable. It is of course possible to find some special calculation method even for an overcomplete skewing. However, at present that might be less needed, partly because of the drawback as mentioned and partly because of the circumstance that we already, for the sake of comparison, by the simpler systems have had to let the values of x and y come so close that we to some extent have exceeded the derived conditions of *always* being on the safe side in our calculations. A discussion of a more far-going, and in that connection more complicated skewing arrangements, must largely be considered as being outside the frame of the present work, and prework, and presumably ought to be left for closer investigation by means of pure experience. *Experience* is also what offers the necessary guidance with respect to the application of the theoretically based results, and what will serve to deepen the understanding of, and thereby also provide an increasingly exact mathematical treatment of the technical problems.

XII Summary

58. On the basis of pure probability considerations and through a series of simplifications of the derived general mathematical expressions we have found that the likely degree of hindrance ψ_{xy} of an automatic telephone system for a certain number x of fully co-ordinate and co-operating central exchange circuits that are supplied to carry a traffic of intensity y in the selected time (a busy hour or a busy quarter-hour), with sufficient accuracy can be expressed by:

$$\psi_{xy} = \frac{e^{-y} y^x}{(x+1)!}\left(1 + \frac{2y}{(x+2)} + \frac{3y^2}{(x+2)(x+3)} + \frac{4y^3}{(x+2)(x+3)(x+4)} + \text{K}\right)$$

$$< \frac{e^{-y} y^x}{(x+1)!} \cdot \frac{1}{\left(1 - \frac{y}{x+2}\right)^2}$$

By the degree of hindrance is here understood the ratio between the part of telephone traffic that suffers hindrance because of the insufficiency of the x circuits, and the total traffic that searches for access over the same circuits. By traffic intensity is understood the average number of simultaneous telephone connections over the selected time. This number is equal to the average number of telephone calls that search for access over the x circuits in the course of the average duration of a connection in the selected time.

We have shown that the simplifications done are always allowable under certain further stated conditions, given in mathematical form, and we have thereby been able to make certain that we are on the safe side with our calculations.

We have shown that, by including a limited number of terms from the given infinite series, we get a result very close to that of the expression behind the inequality sign when doing calculations for values of x and y that lead to small hindrances, corresponding to those that will be allowed in practice. As an allowable degree of hindrance we have in accordance with current practice set forth one per mill, and we have constructed a curve for a such degree of hindrance on the basis of calculations where we have set $\psi_{xy} = \frac{e^{-y} y^x}{(x+1)!} \cdot \frac{(x+2)^2}{(x+2-y)^2}$. The curve shows the value of the traffic y that is allowed for a given x or the number x needed to carry a traffic y. This curve is presented beside two other curves that have been utilised by telephone engineers as guidance for determining x in relation to y or of y in relation to x, and which have been constructed on the basis of calculation of the probability of blocking, ie. the probability of all x circuits being busy. These curves are denoted, respectively, 1 ‰ and 4 ‰ curves for full load.

59. Furthermore, we have carried out calculation of the relation of the average daily degree of hindrance to that of the busy hour, and we have found that with the concentration of traffic in the busy hour such as usually found in larger cities one may state that a degree of hindrance of 1 ‰ in the busiest hour corresponds to an average degree of hindrance of $^4/_{10}$ ‰ for the daily traffic.

60. Next we have calculated the degree of hindrance for what is named a skewing system (slip multiple system), where the co-operating central exchange circuits are not fully co-ordinate, but skewed in relation to one another. As we combine in the longitudinal direction of the system r groups of such skewed central exchange circuits, each group consisting of x circuits with attached selectors, and make sure that the combination is arranged as a complete cycle and that in each group there are equally many skewing steps as there are skewed circuits, then for a traffic of y for each group we can determine the likely degree of hindrance for the skewing system by:

$$\psi_{r\,xy} =$$

$$\frac{e^{-ry}(ry)^x (rx-x)!}{(rx)!(x+1)}\left(1 + \frac{ry}{1!} + \frac{(ry)^2}{2!} + \frac{(ry)^3}{3!} + \text{K} + \frac{(ry)^{rx-x-1}}{(rx-x-1)!}\right)$$

$$+ \frac{e^{-ry}(ry)^{rx}(rx-x)}{(rx+1)!(x+1)} \cdot \frac{1}{\left(1 - \frac{ry}{rx+2}\right)} + \frac{e^{-ry}(ry)^{rx}}{(rx+1)!} \cdot \frac{1}{\left(1 - \frac{ry}{rx+2}\right)^2}$$

By a detailed calculation we have by means of this shown that a skewing system consisting of 4 groups of ordinary 10-contact selectors result in an increased accessibility of at least 37 $^1/_2$ % compared to a system of the same selectors without skewing.

61. Furthermore we have calculated the likely degree of hindrance for a non-skewing system with spare circuits, ie. a system where a certain number of fully co-ordinate central exchange circuits with attached secondary selectors are common for several groups of primary selectors. The calculation is based on the expression in this report for the likely hindered part of the traffic, which is the degree of hindrance multiplied by the total or non-hindered traffic. We have found that by letting 2 secondary selector groups with 8 circuits or selectors in each group be supplemented by 2 spare circuits with attached selectors, which can receive for forwarding the calls that are hindered in the 2 groups, we can also obtain an improvement of accessibility of 37 $^1/_2$ % (or a saving in the number of concerned selectors of 27 $^3/_{11}$ %), not considering the saving of secondary selectors immediately resulting from the spare arrangement. We have under this instead of the factor $\frac{1}{\left(1 - \frac{y}{x+2}\right)}$ in the relevant expressions applied $\frac{1}{\left(1 - \frac{y}{x+3}\right)}$, as we have justified this modification when the ratio y/x by a reduction of x reaches a certain magnitude, which in a system with spare selectors will be the case for the secondary selector groups that receive the majority of calls.

62. Finally, we have by application of the derived equations with corresponding use of the last mentioned factor calculated the degree of hindrance for a couple of combined skewing and spare systems consisting of 10-contact selectors arranged in a simplest possible manner. In this way we have found an im-

provement of the access possibilities of 50 % beside a saving in the number of secondary selectors of 12 $^{1}/_{2}$ to 13 $^{1}/_{3}$ % obtained by the spare arrangement. This results in a total increase in the efficiency of a certain larger number of secondary selectors of 73 $^{10}/_{13}$ % (or with a given traffic a total reduction of 42 $^{2}/_{9}$ % of the number of these selectors), all under the assumption that the conditions for the application of the calculations are satisfied.

63. The last mentioned arrangements allow such a reduction of the *x* circuits needed for the traffic *y*, while retaining the permitted degree of hindrance, that we in part exceed the conditions that guarantee the calculations to always be on the safe side. Possible calculations of what can be obtained by extended skewing arrangements supplemented by spare selectors for further increasing the accessibility seem, therefore, to be linked with special difficulties on the present basis, when the calculations are required to give sufficiently certain results. Consequently, they probably ought to be carried out by application of empirical formulas to be derived by means of experiments or experience, in case more complicated arrangements of the types mentioned can be regarded as useful in spite of the more complicated circuitry that will be involved.

Annex I

Table showing the *degree of hindrance* for different values of *x* = the number of circuits and *y* = the traffic intensity according to calculations based on *Bortkewitsch'* tables.

y = / x =	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0
0																				
1																				
2	0.0328	0.1037																		
3	.0040	.0234	.0570																	
4	.0004	.0044	.0159	.0375																
5		.0007	.0036	.0112	.0246															
6		.0001	.0007	.0029	.0078	.0164														
7			.0001	.0007	.0022	.0053	.0118													
8				.0001	.0005	.0015	.0042	.0084												
9					.0001	.0003	.0013	.0041	.0062											
10						-	.0004	.0010	.0024	.0044										
11							.0001	.0003	.0008	.0017	.0032	.0057								
12							-	.0001	.0003	.0006	.0012	.0023	.0042							
13								-	.0001	.0002	.0004	.0009	.0018	.0031						
14									-	-	.0001	.0003	.0007	.0013	.0024					
15											-	.0001	.0002	.0005	.0010	.0018				
16												-	.0001	.0002	.0004	.0008	.0014	.0023		
17													-	.0001	.0002	.0003	.0006	.0010	.0017	
18														-	.	.0001	.0002	.0004	.0008	.0013
19															-		.0001	.0002	.0003	.0006
20																-	-		.0001	.0003
21																	-	-	-	.0001
22																			-	-
23																				-
24																				

Annex II

Table showing the *degree of hindrance*

$$\psi_{xy} = \frac{e^{-y} y^x}{(x+1)!\left(1-\frac{y}{x+2}\right)^2}$$

for the values x = 4, 7, 13, 16, 19 and 22, and

$$\psi_{xy} = \frac{e^{-y} y^x e^{x+1}}{\sqrt{2\pi(x+1)}(x+1)^{x+1}\left(1-\frac{y}{x+2}\right)^2}$$

for the values x = 30, 50, 70 and 100, applying various suitable values of y, where x denotes the number of circuits and y the traffic intensity.

y =	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
x =										
1										
2										
3										
4	.000376	.004415	.016734							
5										
6										
7			.000136	.000710	.002382					
8										
9										
10							.000416	.001083	.002423	

y =	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0
x =										
11										
12										
13	.000493	.001032	.001986							
14										
15										
16					.000458	.000860	.001526			
17										
18										

y = / x =	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5	15.0
18											
19	.000680	.001144	.001852								
20											
21											
22						.000871	.001338	.002041			
25											
30											
40											
50											
60											
70											
80											
90											
100											

y = / x =	19.0	19.5	20.0	35.0	36.0	37.0	52.0	53.0	54.0	79.0	80.0	81.0
18												
19												
20												
21												
22												
25												
30	.000954	.001369	.001919									
40												
50				.000609	.001036	.001703						
60												
70							.000525	.000812	.001232			
80												
90												
100										.000593	.000839	.001173

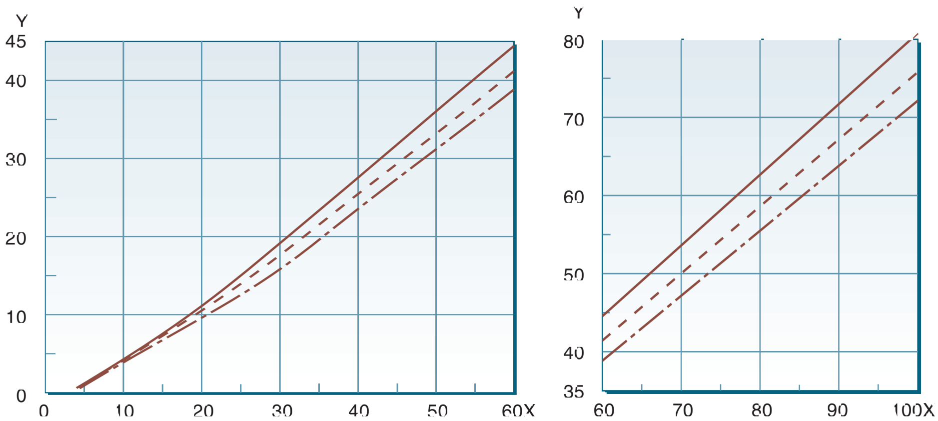
Annex III

Presentation of values of traffic intensity *y* and efficiency α with a degree of hindrance $\psi_{xy} = 0.001$ corresponding to circuit numbers of $x = 4, 5, 6, ..., 22, 30, 40, ..., 100$. (For the underlined values of *x* the calculation of *y* is done by means of the formulas given in the report. For the others interpolation is applied.)

x	y	α	x	y	α
4	0.70	0.175	18	9.62	0.534
5	1.12	0.224	19	10.37	0.546
6	1.60	0.266	20	11.13	0.556
7	2.12	0.303	21	11.89	0.567
8	2.71	0.339	22	12.67	0.576
9	3.33	0.370	30	19.07	0.636
10	3.96	0.396	40	27.40	0.685
11	4.62	0.420	50	35.92	0.718
12	5.29	0.441	60	44.61	0.743
13	5.98	0.460	70	53.46	0.764
14	6.69	0.478	80	62.40	0.780
15	7.41	0.494	90	71.42	0.794
16	8.13	0.508	100	80.53	0.805
17	8.87	0.522			

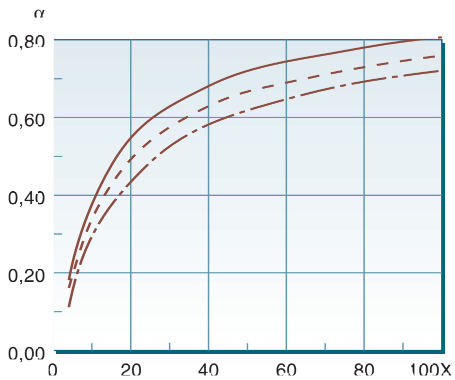
Annex IV

Diagrams showing the likely relation between traffic and circuits in an ordinary automatic telephone system.



Annex V

Diagram showing the likely efficiency of circuits in an ordinary automatic telephone sys-



Remarks on the translation

General

The present document is a translation from Norwegian to English of a report written in 1915. The time span from the time of writing to the time of translation is thus more than 80 years. This fact has naturally influenced the translation on several counts and in several ways. To the general development comes, possibly, personal traits of author and translator as well.

As the translator in case I want to point out the following:

- As most languages of modern societies undergo continuous change, this has probably been the case to a larger extent for Norwegian than for most other languages because of its particular history of the last century.
- The author appears to have been a very accurate person who all along wanted to make sure that all relevant conditions and reservations at any point were clearly expressed. That seems to have led to a meticulous accumulation of characterising words and many subordinate clauses.
- The development of underlying technology during the given time span has been formidable, and even a present day translation should not presume this development.
- The concepts and terminology of teletraffic theory and that of probability theory in general has undergone a similarly formidable development, and should not be fully anticipated in the translation of a text originated in 1915.
- In veneration of the author and with a wish to keep the time colouring of the original report, not much effort has been done to modernise the text, but for some cases when a long single period has been cut in two or more, while avoiding any alteration of the meaning.

Specifics

In the report there are some quite central terms that may be translated in different ways, and a choice has to be made in order of correct understanding and in view of the statements above. Some of these terms have even been replaced by new terms in modern Norwegian. Some examples will be listed and commented.

- *velger* (orig.: vælger) The most direct translation is *selector*, which is also used in chapters IX – XI about gradings. In these chapters the pre-selection function (in two stages) is treated. As a generic term in the rest of the report we have used *switch*. That includes different functional devices, like selector, preselector, finder, connector, etc.
- *hindring, hindringsgrad*: *hindrance, degree of hindrance*. This is the one most central term in the report. Today the commonly used Norwegian term is 'sperr' = blocking or congestion. In the main traffic model of this report a combined delay and blocking concept is more appropriate, and the proposed terms seem very suitable. It is worth mentioning that in the only known publication by Engset, written in German, the terms 'Hinderung' and 'Hinderungsgrad' are used.
- *blokkering, overfylling: blocking, overfill.* Those are descriptive terms for the probabilities of at least x (= number of servers) and at least $x + 1$ positions occupied. Here there is no weighting by the number in queue.
- *forskyvning, reserveledninger*: *skewing, spare circuits*. The term *grading* (Norw.: 'gradering') is not applied and was probably not in common use. The report treats two types of grading, one of a slip multiple type and the other progressive. The first has a skewed structure and the other straight.

A R N E   M Y S K J A

The Engset report of 1915

Summary and comments

ARNE MYSKJA

A typed report of 130 pages written by Tore Olaus Engset in 1915 indicates a much more extensive pioneer work in the teletraffic field by Engset than has been publicly known. The well known Engset formula is solely based on the only published teletraffic work by Engset, a two-page article from 1918.

The 1915 report was discovered by Villy Bæk Iversen in the files of Copenhagen Telephone Company, and was first presented in a guest lecture at the University of Trondheim in 1995 and in a short paper at the 13th Nordic Teletraffic Seminar (NTS-13) in Trondheim in 1996. The assumed original of the report was later found at Norsk Telemuseum, Oslo.

The Engset report contains original modelling of teletraffic according to a model that was taken up by Conny Palm in a more general form more than twenty years later. It is a queue system with departure from the queue. Engset even allows an arbitrary number of individual traffic sources. He includes aspects that are only treated much later by Palm, Cohen, Dartois and others. Even daily variations are modelled, and simple cases of progressive and slipped multiple gradings and combinations of both are treated.

Throughout, Engset aims at practical useful modelling. He therefore introduces simple approximations, and he is keen to assure that his approximations are on the safe side. With the same objective he even presents dimensioning diagrams for practical use.

Photo: Norwegian Telecommunications Museum

Tore Olaus Engset, 1865 – 1943

1 Introduction

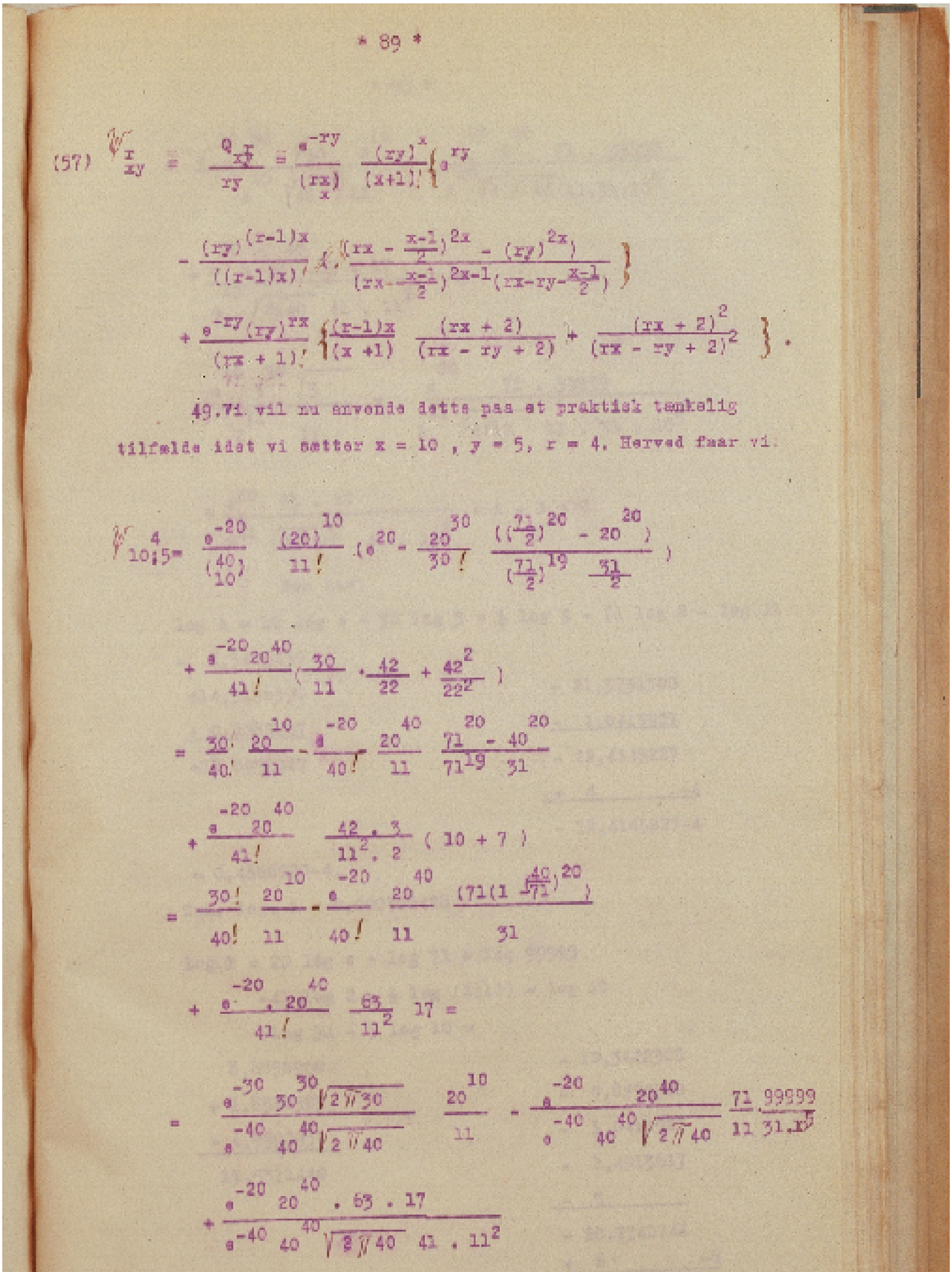
In the teletraffic community Engset is a household name connected with the mathematical formula that carries his name. Until recently the only known teletraffic work by Engset is the article that presents his formula: 'Die Wahrscheinlichkeitsrechnung zur Bestimmung der Wählerzahl in automatischen Fernsprechämtern'. The article, which is a concentrated work of merely two pages, appeared in *Elektrotechnische Zeitschrift*, 1918, Heft 31 [1] and in a French translation [2] in 1921. An English translation by Eliot Jensen appeared in *Telektronikk* No. 1, 1992 [3], with the title 'The Probability Calculation to Determine the Number of Switches in Automatic Telephone Exchanges'. In the same issue of *Telektronikk* Eliot Jensen also presented a commentary [4], discussing Engset's paper in the context of later developments.

In 1995 a copy of a report from 1915 by Engset [5] was discovered in the files of Copenhagen Telephone Company (KTAS). A survey of the report was presented in a guest lecture at the Norwegian Institute of Technology by Villy Bæk Iversen the same year, and a brief written presentation appeared in the records of The Thirteenth Nordic Teletraffic Seminar of August 1996 [6]. In the abstract of this paper Iversen states: "Based on an unknown report by Engset from 1915 and a review of the history of traffic engineering in the period 1900 – 1925 it is shown that Engset's contributions to traffic theory are more fundamental than so far acknowledged".

On this background, and after getting acquainted with a copy of the report, the question was brought up with editor Ola Espvik of the Telenor (Norwegian Telecom) technical journal *Telektronikk,* whether there was an interest in publishing information about Engset's more extensive work. First a search in the Telenor files was initiated, and the presumed original of the KTAS report copy was found.[1] *Telektronikk*'s editor encouraged a translation into English [7] of the report, as it was assumed that there might be some interest outside Scandinavia, and as it was in line with the present general policy of *Telektronikk* to address itself to the international telecommunication community.

The main interest in the report today will be from a historical point of view, as there are no aspects of the report that have not been studied and reported during the ensuing 83 years. It is worth noting, though, that the report was finished in 1915, while Erlang's B-formula [8] came in 1917 and Engset's formula [1] in 1918. Since also the modelling procedure is different from that of Erlang, and several aspects of the report were first treated in publications many years later, there is good reason for the above quote by Iversen. It is fair to say, though, that Engset in his 1918 paper gives due credit to Erlang and his colleagues in Copenhagen, and it is my opinion that even if the 1915 report had been published, it would not have reduced the original value of Erlang's pioneering work.

[1] *There is a number (some 25 noted) of minor differences between the two versions, mostly due to manual insertions in the typewritten text. Generally, the Telenor version is the more complete/correct one.*



* 89 *

$$(57)\quad \mathcal{W}^{r}_{xy} = \frac{Q^{r}_{xy}}{ry} = \frac{e^{-ry}}{\binom{rx}{x}}\,\frac{(ry)^{x}}{(x+1)!}\Bigg\{e^{ry}$$

$$-\frac{(ry)^{(r-1)x}}{((r-1)x)!}\,\frac{\left(rx-\frac{x-1}{2}\right)^{2x}-(ry)^{2x}}{\left(rx-\frac{x-1}{2}\right)^{2x-1}\left(rx-ry-\frac{x-1}{2}\right)}\Bigg\}$$

$$+\frac{e^{-ry}(ry)^{rx}}{(rx+1)!}\left\{\frac{(r-1)x}{(x+1)}\,\frac{(rx+2)}{(rx-ry+2)}+\frac{(rx+2)^{2}}{(rx-ry+2)^{2}}\right\}.$$

49. Vi vil nu anvende dette paa et praktisk tænkelig tilfælde idet vi sætter $x = 10$, $y = 5$, $r = 4$. Herved faar vi:

$$\mathcal{W}^{4}_{10;5} = \frac{e^{-20}}{\binom{40}{10}}\,\frac{(20)^{10}}{11!}\left(e^{20}-\frac{20^{30}}{30!}\,\frac{\left(\frac{71}{2}\right)^{20}-20^{20}}{\left(\frac{71}{2}\right)^{19}\frac{31}{2}}\right)$$

$$+\frac{e^{-20}20^{40}}{41!}\left(\frac{30}{11}\cdot\frac{42}{22}+\frac{42^{2}}{22^{2}}\right)$$

$$=\frac{30!}{40!}\,\frac{20^{10}}{11}-\frac{e^{-20}}{40!}\,\frac{20^{40}}{11}\,\frac{71^{20}-40^{20}}{71^{19}\;31}$$

$$+\frac{e^{-20}20^{40}}{41!}\,\frac{42\cdot 3}{11^{2}\cdot 2}\,(10+7)$$

$$=\frac{30!}{40!}\,\frac{20^{10}}{11}-\frac{e^{-20}}{40!}\,\frac{20^{40}}{11}\,\frac{71\left(1-\left(\frac{40}{71}\right)^{20}\right)}{31}$$

$$+\frac{e^{-20}\cdot 20^{40}}{41!}\,\frac{63}{11^{2}}\,17 =$$

$$=\frac{e^{-30}\,30^{30}\sqrt{2\pi 30}}{e^{-40}\,40^{40}\sqrt{2\pi 40}}\,\frac{20^{10}}{11}-\frac{e^{-20}\,20^{40}}{e^{-40}\,40^{40}\sqrt{2\pi 40}}\,\frac{71}{11}\,\frac{99999}{31\cdot 10^{5}}$$

$$+\frac{e^{-20}\,20^{40}\cdot 63\cdot 17}{e^{-40}\,40^{40}\sqrt{2\pi 40}\;41\cdot 11^{2}}$$

Photo: Svein E Fasteng/SAMFOTO

A page from the original Engset report of 1915

One can only speculate why Engset did not go on to publish his findings in the 1915 report. First of all, there was no tradition for scientific publishing in his working environment. In connection with system studies for a possible introduction of an automatic telephone system in Norway a national committee had been established, with Engset as one of the committee members. The work led to close contact with Danish teletraffic experts, and the complete 1915 report was made available in Copenhagen (and probably in Stockholm). The report contains references to Erlang and Christensen (Denmark), Grinsted (England) and Spiecker (Germany). The Grinsted/Spiecker reference is applied by Engset in one of three different approaches to obtain the binomial (Bernoulli) formula with the Poisson formula as a limit case. The Erlang/Christensen reference is to a 1913-publication with the so-called Christensen formula, that applies the Gaussian error curve for dimensioning.

Engset's approach seems to be his own original one, based on individual traffic demands. It permits waiting for service, and the subscriber holding time is independent of the fate of the call. Engset may have been discouraged from further pursuit of his model, simply by becoming convinced that Erlang's limited server group truncation was a better model. This truncation is adopted in the 1918 paper, which otherwise applies the general approach of the 1915 report.

In the following we shall mainly stay with the 1915 report and only by particular mention refer to the 1918 paper.

Engset's formal education may look rather modest. After his junior high school exam and telegraph course in 1883 he never carried out any full time studies. However, based on spare time work he graduated as a student in 1892 and as M.Sc. in mathematics and physics at university in 1894. In itself this indicates a very great capacity. His teletraffic work shows that he was well versed in mathematics at a medium to high level, and his works later in theoretical physics (not to be discussed further here) indicate a notorious capability in advanced mathematics. Engset's approach to teletraffic analysis is clearly coloured by the deep insight in the practical aspects that he had acquired as a telephone engineer with many years of experience. The main reference in the report is Bortkewitsch: 'Das Gesetz der kleinen Zahlen' [9].[2]

Engset is very conscious of making his assumptions clear. He applies mainly two basic theorems of probability calculation, that of the probability sum of disjunct events and that of the probability product of mutually independent events. Furthermore, he starts with the basic assumption of a limited number of individual traffic sources, each of which asks for a certain service time per time unit. This traffic demand may be different for each source. No particular holding time distribution is assumed, and neither is any interarrival time distribution prescribed, only full independence between sources.

Engset applies the concept of 'statistical equilibrium'. His main model assumes a limited number of n sources and a limited number of x servers. There is no mention of a queue. However, he specifies $n + 1$ different states, which in fact is descriptive of a system with x servers and $q \geq n - x$ queue positions. The implication of this is a queuing system where departure from the system is independent of whether it is from a server or from a waiting position. From the outside the system looks like that of an infinite server group with no ordering of positions, with arrivals and departures according to some given distributions. From the point of view of an arbitrary customer arriving he may find a free server and immediately occupy it. Otherwise, he may stay waiting until a server becomes free (he may have to compete), or he loses patience and leaves. Only in the latter case is the call lost.

The outcome of these assumptions is that within a given time period ($T = T_s + T_w$) there will be two distinct traffic volumes, one given by the total time (T_s) spent in the server group and the other given by the time (T_w) spent waiting. The calls will consist of three categories, those only being served, those only waiting, and those first waiting and then being served. Engset defines two relative quantities: the 'degree of efficiency' $\varphi = T_s / (T_s + T_w)$ and the 'degree of hindrance' $\psi = T_w / (T_s + T_w)$. Consequently, $\varphi + \psi = 1$. There is thus no defined 'loss' or 'congestion'. There are of course lost calls, namely those that leave the system (for loss of patience) without being served. This part is not calculated, and neither are waiting times.

One may note here that this Engset model is neither the 'lost calls cleared' (Erlang) nor the 'lost calls held' (Molina) model. A comprehensive treatment of what in this context may be termed an extended Engset model by Conny Palm is given in [10] and [11] (*K. Telegr.-Styr. Tekn. Med.* No 7-9, 1937 and *Tele* No. 1, 1957). The Palm model, which naturally has no reference to Engset's unpublished work, treats a system with unlimited queue with exponential service and waiting, however with separate service and waiting parameters. Engset's case is a special case of Palm's general model, with a common holding time parameter.

Apart from this preliminary discussion of the main model we shall point out a few particulars of the report:

1 A limited number of identical traffic devices serve a larger, still limited, number of individual traffic sources. A calling source gets immediate service if there is at least one free device, otherwise it is allowed to wait until it gets delayed service, or it leaves the system voluntarily.

2 All traffic sources are mutually independent, calls arrive at random and there are no dependencies between call arrivals and holding times.

3 No assumption is made about holding time distributions. Individual traffic characteristics are according to the generalised Engset formula as later introduced by Cohen (1957) [12].

4 Engset presents a proof that identical sources is the least favourable case. A publication on this matter was first presented in 1970 by Dartois [14].

5 Several simplifications are made in order to be able to calculate particular cases. Foremost is the assumption of equal offered traffic from all sources, the assumption of an unlimited number of sources (Erlang), and introduction of approximations. Engset is consistent in proving that approximations are on the safe side, ie. the approximation is less favourable than the exact solution.

[2] *The name of Bortkewitsch is according to his book in German, edited in Leipzig 1898. In some library references the name is spelled Bortkiewicz.*

6 The independence condition between arrivals is shown to lead to the binomial (Bernoulli) distribution and in the limit to Poisson distribution (referred to as Bortkewitsch' formula).

7 Engset discusses at length the highest one hour concentration (busy hour) of the full 24 hour traffic, and introduces a geometrical model for this, in order to find the total 24 hour loss, given the busy hour loss.

8 Engset carries out detailed studies of slip multiple gradings, progressive gradings and combinations of both, with example calculations. Like other early authors on the matter he seems not to be aware of the special overflow characteristics of random traffic.

9 Calculations according to two approximations are carried out and the results tabulated. Diagrams for dimensioning purposes are drawn.

2 Preliminaries

In the introductory Chapter I Engset introduces the classical model of random drawing of black and white balls from an urn, along with the two 'main theorems of probability': 1) The sum theorem of mutually disjunct events, and 2) The product theorem of mutually independent events.

In Chapter II a detailed discussion of practical telephone traffic aspects is carried out. A telephone exchange with a set of x identical connecting devices (lines, selectors) is assumed. These devices serve a set of n individual traffic sources (subscribers). An observation time T_c is assumed to be a period of homogeneous traffic. A number r of calls in T_c with an average duration of t defines an average traffic intensity of $y = rt / T_c$. The traffic is individualised by assuming a subscriber A_i being busy with outgoing calls an accumulated time T_i out of total observation time $T_t = D \cdot T_c$, where D is a (great) number of days in order to get enough observations for a single subscriber. An alternative to measuring the single intervals that make up T_i is to assume that T_t consists of a very large number N of very short intervals Δt and counting the number of intervals m_i when A_i is busy. Then the expression $T_i / T_t = m_i / N = p_i$ may be interpreted as the relative time that A_i is busy = the probability that A_i is busy = connection probability of A_i = the traffic load of A_i. (Throughout, only outgoing calls are considered.)

3 State probabilities, efficiency and hindrance

In Chapter III the combined states of the n subscribers are studied. Only the number r of busy subscribers is of interest, and any state is completely described by this number. Thus, with n subscribers there can be $n + 1$ different states: 0, 1, 2, ..., r, ..., n. For the $x \le n$ serving devices there can be only $x + 1$ different states: 0, 1, 2, ..., x. The state x includes $n - x + 1$ total states with 0, 1, ..., $n - x$ subscribers that are busy, but not being served. For the total number $n + 1$ of states there are as many probabilities defined: $P_0, P_1, P_2, ..., P_x, P_{x+1}, ..., P_n$.

This system is identical to a queuing system with x servers and $n - x$ (or an unlimited number of) queue positions. It is implicitly assumed that the subscriber holding time is independent of its fate, if it is immediately served, if it is served after waiting in queue, or if it leaves the queue without service.

An arbitrary state r is defined by r subscribers being busy and the remaining $n - r$ subscribers free. There are $\binom{n}{r} = \binom{n}{n-r} = \frac{n!}{r!(n-r)!}$ ways of picking r out of n individuals. Each subscriber i has its individual busy probability p_i and the complementary free probability $(1 - p_i)$. The probability of each of the $\binom{n}{r}$ combinations with r busy and $n-r$ free subscribers can be expressed on the form $\overset{(r)}{\Pi}_j\, p_i \cdot \overset{(n-r)}{\Pi}_j \left(1-p_k\right)_{k\neq i}$. Here $\overset{(r)}{\Pi}_j\, p_i$ means the j^{th} product of r individual probabilities p_i. Similarly, $\overset{(n-r)}{\Pi}_j \left(1-p_k\right)_{k\neq i}$ is the j^{th} product of $(n - r)$ individual probabilities $(1 - p_k)$. The total probability of state r will then be:

$$P_r = \sum_{j=1}^{j=\binom{n}{r}} \overset{(r)}{\Pi}_j\, p_i \cdot \overset{(n-r)}{\Pi}_j \left(1-p_k\right)_{k\neq i} \tag{1}$$

The normalising condition is:

$$P_0 + P_1 + P_2 + \cdots + P_x + \cdots + P_n = 1 \tag{2}$$

It should be noted that P_r is the probability of r busy subscribers, whereas the probability of all x servers being busy is

$$P_x + P_{x+1} + P_{x+2} + \cdots + P_n$$

At this stage Engset introduces the concept of 'state probability', and he notes that the above calculation is valid irrespective of any hypothesis concerning the duration of the single connections or the average duration of connections.

With these assumptions the total traffic intensity can be expressed in two ways:

$$y = \sum_{i=1}^{n} p_i = \sum_{r=0}^{n} r \cdot P_r \tag{3}$$

where each P_r is a function of $p_1, p_2, \ldots, p_n$, as given above.

The traffic intensity y consists of two parts, one effective part that can be measured directly on the x servers, and one waiting part that can be measured on the calling subscribers while they are waiting for service. There is no traffic rejected by the system. All subscribers waiting for service either leave the system voluntarily without being served, or they get delayed service when finding a free device. The same call may contribute to both parts. Engset considers the first part as the *effective* traffic and the second part as the *hindered* traffic.

In order to operate with relative quantities he defines a *degree of efficiency* φ_{nx} and a *degree of hindrance* ψ_{nx}. These are in a compact way defined by

$$\varphi_{nx} = \frac{1}{y} \cdot \left[\sum_{r=1}^{x-1} r \cdot P_r + x \cdot \sum_{r=x}^{n} P_r \right] \tag{4}$$

$$\Psi_{nx} = \frac{1}{y} \cdot \sum_{r=x+1}^{n} (r-x) \cdot P_r, \text{ where} \quad (5)$$

$$\varphi_{nx} + \psi_{nx} = 1 \quad (6)$$

Engset states that these are exact expressions, given that all p_i can be considered unaffected by the degree of hindrance. He points out, however, that that may only be the case for small degrees of hindrance.

4 The first simplification of calculations

It may be noted that the assumption of individual traffic demands, expressed by p_i for any subscriber A_i, is in fact the assumption made for what is later called the generalised Engset formula. The solution given above is thus in fact the generalised solution, however with the difference due to the different handling of calls that are not given immediate service.

It is pointed out by Engset that this general assumption leads to very complicated calculation even for limited numbers of subscribers of, say, $n = 100$. Because of this in Chapter IV he makes the simplifying assumption of setting:

$$p_1 = p_2 = p_3 = \ldots = p_n = p \quad (7)$$

This condition leads to the simple expressions for the state probabilities of

$$\overline{P}_r = \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r}, \quad (8)$$

which is in fact the binomial (Bernoulli) distribution.

By introducing equation (8) in (4) and (5) and setting $y = np$ we get the special expressions for the degree of efficiency and the degree of hindrance (using the dashed symbols to indicate the difference):

$$\overline{\varphi}_{nx} = \frac{1}{np} \cdot \left(\sum_{r=1}^{x-1} r \cdot \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r} + x \cdot \sum_{r=x}^{n} \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r} \right) \quad (9)$$

$$\overline{\psi}_{nx} = \frac{1}{np} \cdot \left(\sum_{r=x+1}^{n} (r-x) \cdot \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r} \right) \quad (10)$$

Engset investigates in his report how the condition (7) influences the values of $\overline{P}_r$ as compared to P_r, and in particular the values for $r > x$. He shows that if $\overline{P}_r > P_r$ for $r = x + 1$, then the relation applies to all values $r > x + 1$, and then also $\overline{\psi}_{nx} > \psi_{nx}$. Essentially, this is the same conclusion as that reached by Dartois [13] in 1970. Thus, under those conditions, it is safe to use the simplifying assumption of (7). The critical values of $y = np$ is tabulated as a function of n and r (= x + 1) for some small values. The conclusion is that the simplification of assuming all traffic sources to be identical leads to maximum hindrance and thus is justified under normal circumstances, ie. for practical values of traffic and degree of hindrance.

5 The second simplification of calculations

A further simplification of expressions is obtained in Chapter V by letting the number of traffic sources approach infinity, while the total traffic is kept constant. This is the basic Erlang assumption. Engset applies the elementary mathematical manipulation leading to

$$\lim_{n\to\infty} \overline{P}_{nr} = \lim_{n\to\infty} \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r} = \frac{y^r}{r!} e^{-y} \quad (11)$$

ie. the Poisson formula, which he denotes Bortkewitsch' formula.

Since Engset wants to stick to the assumption of a limited number $n = n_0$ of sources, he intends to prove that above a certain value of r all the Poisson terms are greater than the corresponding Bernoulli terms. This is done by taking the derivative of (the log of) the Bernoulli terms with respect to n, and finding the condition that makes this derivative positive. This leads to the condition

$$r > y + \frac{1}{2} + \sqrt{y + \frac{1}{4}} \quad (12)$$

When condition (12) is satisfied, the Poisson terms are greater than the corresponding Bernoulli terms, and the calculation is on the safe side when the Bernoulli terms in (10) are replaced by Poisson terms from $r = x + 1$ to $r = n = n_0$:

$$\overline{\psi}_{\infty x} = \frac{e^{-y}}{y} \cdot \sum_{r=x+1}^{n_0} \frac{(r-x) \cdot y^r}{r!} \quad (13)$$

The corresponding non-truncated, weighted sum is

$$\overline{\psi}_{\underline{\infty} x} = \frac{e^{-y}}{y} \cdot \sum_{r=x+1}^{\infty} \frac{(r-x) \cdot y^r}{r!} = \frac{y^x}{(x+1)!} \cdot e^{-y} \cdot \frac{1}{\left(1 - \frac{y}{x+k}\right)^2} \quad (14)$$

where k can be chosen to satisfy the equation.

After some manipulation a more compact expression of the truncated sum in (13) is obtained:

$$\overline{\psi}_{xy} = \frac{y^x}{(x+1)!} \cdot e^{-y} \cdot \frac{1}{\left(1 - \frac{y}{x+h}\right)^2}, \text{ where } 2 < h < \frac{n_0 - x}{2} + 3 \quad (15)$$

Again, in order to be on the safe side, the value $h = 2$ is chosen and the notation is then replaced by $\overline{\psi}_{xy} > \overline{\psi}_{\infty x}$.

(Later in the report, under the study of gradings, $h = 3$ is assumed to be a better choice.)

For practical calculations Stirling's formula for the factorial function is proposed, to give:

$$\psi_{xy} = \frac{1}{\sqrt{2\pi(x+1)}} \cdot \frac{y^x}{(x+1)^{x+1}} \cdot \frac{e^{x+1-y}}{\left(1 - \frac{y}{x+2}\right)^2} > \overline{\psi}_{xy} \quad (16)$$

(The indexing and use of – (dash) for marking of y is not quite systematic.)

Engset has calculated tables with a selected set of values from formulas (13) and (15)/(16), in order to compare the two approximations. For formula (13) he uses values from a four decimal Poisson table (found in the Bortkewitsch book), whereas formula (15)/(16) is calculated with 7 digit accuracy. By means of interpolation he has further determined table values for traffic y and utilisation α for a selection of integer x values, given a fixed degree of hindrance of $\psi_{xy} = 0.001$. Similarly, curves for the values 0.002, 0.003, ..., 0.00r can be drawn for dimensioning purposes.

6 Other approaches to the evaluation of efficiency

The efficiency measure in the previous is given by the ratio of the waiting traffic to the total offered traffic, given that the traffic demand of each subscriber is unaffected by the fate of the calls. Thus the state probability of each state has to be multiplied by the number of waiting subscribers to give the waiting traffic.

Engset in Chapter VI points out two other possible efficiency measures, namely the probability $P_{(xy)}$ that all x devices are busy and the probability P_{xy} that there are subscribers waiting. Here he applies the assumption of infinitely many subscribers by using the Poisson distribution:

$$P_{(xy)} = e^{-y} \sum_{r=x}^{\infty} \frac{y^r}{r!} \tag{17}$$

$$P_{xy} = e^{-y} \sum_{r=x+1}^{\infty} \frac{y^r}{r!} = P_{(xy)} - \frac{y^x}{x!} \cdot e^{-y} \tag{18}$$

In addition to the above mentioned curve for $\psi_{xy} = 0.001$, curves are also drawn for $P_{(xy)} = 0.001$ and $P_{(xy)} = 0.004$, illustrating the similar character of the two functions.

With reference to a representative of Western Electric Co. as a source, Engset also presents an integral form with $p_{xy} = \frac{y^x}{x!} \cdot e^{-y}$ as a generative function:

$$P_{xy} = 1 - e^{-y} \sum_{r=0}^{x} \frac{y^r}{r!} = \int_0^y \frac{t^x}{x!} \cdot e^{-t} dt \tag{19}$$

7 Objections and alternative approaches

Engset in Chapter VII carries out a discussion where he looks more closely at the assumptions concerning the fate of calls that are not immediately served, related to the technical possibilities of the system. The crucial question is whether the technical system rejects a call when busy, or allows the call to be served after waiting, without having to be repeated.

During the initial part of the discussion the reasoning leads to the result that the distribution of holding times is of no significance, so one may as well assume that they are constant. It is worth noting that waiting times are not considered. Also here he starts with the assumption of individual traffic demands per subscriber, proceeding to the assumption of identical properties (Bernoulli) and finally to an unlimited number of subscribers (Poisson).

At this point there is no clear distinction between the number of calls arriving during a mean holding time t_m and the number of busy devices. The weakness of the reasoning is that the birth-and-death process is not described, as it is in the Erlang model of lost calls cleared. Engset considers only the calls arriving during an average holding time, irrespective of previous calls already being served. (He does mention these calls, but thinks they are of no significance to the argument.) Thus he assumes that the first x calls in t_m obtains service, whereas calls no. $x + 1$, $x + 2$, $x + 3$, ... are rejected. Because of this he remains with the non-truncated Poisson distribution, which is of course valid for the arrival process. Thus, even with the assumption of calls being rejected, the same degree of hindrance is obtained as in the case of waiting calls according to the original model.

8 Calculation of the degree of hindrance for the daily traffic

Up to this point only the traffic of an arbitrary *stationary* period has been considered, and all the formulas developed are based on the stationarity assumption. As Engset is fully aware of the real traffic fluctuations over time, and thus also the periodic character from day to day, in Chapter VIII he goes on to studying the consequences of this on traffic dimensioning.

Given the degree of hindrance ψ_{xy} for a traffic intensity of y offered to a device group of x, the hindered traffic intensity = the hindered traffic volume per time unit is

$$H_c = y \cdot \psi_{xy} = e^{-y} \cdot \sum_{r=x+1}^{\infty} \frac{(r-x) \cdot y^r}{r!} \approx \frac{y^{x+1}}{(x+1)!} \cdot e^{-y} \cdot \frac{1}{\left(1 - \frac{y}{x+2}\right)^2} \tag{20}$$

(Here ψ_{xy} replaces $\overline{\psi_{\underline{\infty} x}}$ in equation (14), and the form is that of $\overline{\psi}_{xy}$ in equation (15) with $h = 2$.)

Assuming now that the offered traffic intensity varies with time, $y = y(t)$, then the mean value during a period h can be expressed by

$$y_h = \frac{1}{h} \int_0^h y(t) dt, \tag{21}$$

with a corresponding traffic volume of $h \cdot y_h$ over h.

The mean hindered traffic intensity over h will be

$$\overline{H_h} = \frac{1}{h} \int_0^h H_c(t) dt \tag{22}$$

Thus, the degree of hindrance over the period h becomes

$$\psi_{xy_h} = \frac{\overline{H_h}}{y_h} = \frac{1}{h \cdot y_h} \int_0^h H_c(t)dt$$

$$= \frac{1}{h \cdot y_h} \int_0^h \left(e^{-y(t)} \cdot \sum_{r=x+1}^{\infty} \frac{(r-x)y(t)^r}{r!} \right) dt \qquad (23)$$

It is considered very complicated and in practice impossible to solve this problem in a general way. A Gaussian based function of the form $y = \bar{y}e^{-p^2t^2}$ as well as a Poisson based function of the form $y = \bar{y} \cdot qe^{-t}\frac{t^h}{h!}$ are considered.

The latter is found the most favourable option, after introducing Stirling's factorial approximation and a concentration factor $c = k/h$, leading to

$$y = \bar{y} \cdot e^{-ct} t^{ch} \left(\frac{e}{h}\right)^{ch} = \bar{y}\left(\frac{e}{h}\right)^k \cdot e^{-\frac{k}{h}t} \cdot t^k \qquad (24)$$

The factor k is denoted the concentration coefficient, as it can be used to adjust the traffic concentration without changing the basic parameters $\bar{y}$ and h. Characteristic values are $y = 0$ for $t = 0$ and $y = \bar{y} = y_{\max}$ for $t = h$. Further adjustments are made by replacing h by ph in (24) in order to cover a normal full day traffic. The expression thus obtained is

$$y = \bar{y}\left(\frac{e}{ph}\right)^k \cdot e^{-\frac{k}{ph}t} \cdot t^k \qquad (25)$$

The full day average degree of hindrance is obtained by introducing y as a function of t according to (25) in the integral expression of (23).

Engset realises that the exact calculation according to this is rather complicated, and he introduces a linear approximation to obtain the simplified relation $dt = \frac{a-r}{\bar{y}}dy$, where a and r are constants in the approximation. Furthermore, he introduces arithmetic approximations for the infinite series of integrated terms from (23), and as always he makes sure that his approximations are on the safe side. The calculations are very detailed with many practical assumptions. The main outcome of the given model, with a concentration to the busy hour of 14.2 % of the daily traffic, is that a busy hour degree of hindrance of ψ_b leads to a full day degree of hindrance ψ_d expressed by

$$\psi_d = \psi_b \cdot \left(1.385 \frac{\left(x+2-\bar{y}\right)^2}{\left(x+3-\bar{y}\right)^3} \cdot \frac{(x+3)^3}{(x+2)^3} + 0.307 \right) \qquad (26)$$

For a practical case of $x = 17$ and $\bar{y} = 9$ we get $\psi_b = 1$ ‰ and $\psi_d = 0.43$ ‰.

That is of course only a single example based on specific parameters. Several examples indicate values in the range ψ_d = 0.3 – 0.45 ‰ when $\psi_b = 1$ ‰ with the given model.

9 Gradings. Slip multiple

Engset's basic model is in Chapter IX extended to the case of slip multiple gradings. It is applied to a system where each subscriber is connected to a primary pre-selector that hunts over a set of secondary preselectors or group selectors in a multiple. A primary group consists of x^2 primary preselectors that are served by x secondary preselectors. The value of x is of course a fixed one, as it is determined by the very construction of the selector devices. (The ratio $x^2/x = x$ seems rather arbitrary and may be suitable only for a limited range of x and of traffic per subscriber.)

The total traffic intensity per group is y, which is served by x devices. The x devices must carry the traffic y with a prescribed degree of hindrance. A better utilisation can be obtained by combining two or more, say r, similar primary groups in a slip multiple that makes up a complete ring where $r \cdot x^2$ subscribers share $r \cdot x$ devices, but with limited access. Full symmetry among the groups and equal traffic y per group are assumed. In the simplest case $r = 2$ the connections should be:

Preselectors				Secondary selectors		
1	to	x	are connected to	1	to	x
x+1	to	2x	are connected to	2	to	x+1
2x+1	to	3x	are connected to	3	to	x+2
-						
-						
-						
(x–1)x+1	to	x^2	are connected to	x	to	2x–1
x^2+1	to	x^2+x	are connected to	x+1	to	2x
x^2+x+1	to	x^2+2x	are connected to	x+2	to	2x and 1
x^2+2x+1	to	x^2+3x	are connected to	x+3	to	2x and 1 to 2
-						
-						
-						
2x^2–x+1	to	2x^2	are connected to	2x	and	1 to x–1

(Engset described the same connection pattern in a slightly different way). A similar cyclic pattern is obtained for an arbitrary integer value of r. The total traffic offered to such a graded system is $r \cdot y$.

Assuming first that $r = 2$ and that s devices (secondary selectors) out of $2x$ are free, then hindrance occurs only when $s \le x$. Only the cases s = 1, 2, 3, ..., x need to be considered.

The probability that a call does not immediately find a free outlet is found as the ratio between all possible ways of finding s among x and all possible ways of finding s among $2x$:

$$R_s = R_{x-i} = \frac{\binom{x}{s}}{\binom{2x}{s}} = \frac{\binom{x}{x-i}}{\binom{2x}{x-i}} = \frac{\binom{x+i}{x}}{\binom{2x}{x}} = \frac{\binom{x+i}{i}}{\binom{2x}{x}} \qquad (27)$$

Engset obtained the last of the three variants in a somewhat different way. He further defined a set of combined probabilities:

$$S_i = \sum_{j=0}^{i-1} R_{x-j} = \frac{\binom{x+i}{i-1}}{\binom{2x}{x}} \qquad (28)$$

$$S_i = \sum_{j=0}^{i-1} R_{x-j} = \frac{\binom{x+i}{i-1}}{\binom{2x}{x}} \tag{28}$$

The combinations of occupation states in this grading system become rather complicated, whereas according to the assumptions it is possible to define occupation states based on busy subscribers only, expressed by the Poisson distribution:

$$\overline{P}_k = \frac{e^{-2y}(2y)^k}{k!} \tag{29}$$

Hindrance of calls (calls held in waiting) only occurs for $k > x$, and the probability that a call does not get access, given that there are $x + 1$ calls in the system is given by $R_x = S_1$.

The contribution to the hindrance thus becomes $q_1 = \overline{P}_{x+1} \cdot S_1$.

Similarly for $x+2$ we get $q_2 = \overline{P}_{x+2} \cdot S_2$, $q_3 = \overline{P}_{x+3} \cdot S_3$ etc.

(See remarks below for the general case).

For the more general case of r groups (27) and (28) take the forms of

$$R_s = R_{(r-1)x-i} = \frac{\binom{x+i}{x}}{\binom{rx}{x}} = \frac{\binom{x+i}{i}}{\binom{rx}{x}}, \quad i = 0,1,2,\ldots,(r-1)x-1 \tag{27a}$$

and

$$S_i = \sum_{j=0}^{i-1} R_{(r-1)x-j} = \frac{\binom{x+i}{i-1}}{\binom{rx}{x}} \tag{28a}$$

Under Engset's general assumption that no calls are lost, only hindered, and here also the source number is infinite, the state probabilities for a total r-group system can be expressed exactly by

$$\overline{P}_k = \frac{e^{-ry}(ry)^k}{k!} \tag{29a}$$

Only the states of $k > x$ will imply hindered calls, and by similar reasoning as in the case of $r = 2$ we get the values of hindered traffic for each state from $k = x + 1$ to $k = rx$:

$$q_i = \overline{P}_{x+1} \cdot S_i = \frac{e^{-ry}(ry)^{x+i}}{(x+i)!} \cdot \frac{\binom{x+i}{i-1}}{\binom{rx}{x}}, \quad i = 1,2,3,\ldots,(r-1)x \tag{30}$$

For states above rx all new calls are put on waiting, and we get correspondingly:

$$q_{(r-1)x+h} = \overline{P}_{rx+h}\left(S_{(r-1)x} + h\right) = \frac{e^{-ry}(ry)^{rx+h}}{(rx+h)!} \cdot \left(\frac{(r-1)x}{x+1} + h\right), \quad h = 1,2,3,\ldots, \tag{31}$$

The total hindrance can be expressed by

$$Q_{r\,xy} = \sum_{i=1}^{\infty} q_i \tag{32}$$

and the degree of hindrance by

$$\psi_{r\,xy} = \frac{Q_{r\,xy}}{ry} \tag{33}$$

In order to be able to do practical calculations some approximations are introduced like before. Two examples are calculated with $r = 4$ groups, $x = 10$ selector outlets and a traffic per group of $y = 5$ and $y = 5.5$ respectively. The calculation indicates that $y = 5.5$ is permitted within the requirement of 1 ‰ hindrance. The non-graded case permits only $y = 4$, and that means a traffic increase of 37.5 %, or alternatively, a saving of ≈ 27 % in the amount of selectors for the same grade of service.

Remark: The assumption of equations (28) and (28a) of the S-parameter given as a sum of R-parameters is a simplification, as only first order probabilities are included. This implies that events are mutually disjunct, which is not the case. For example, when $r = 2$, the first case, $S_1 = R_x$ is correct, whereas the second, $S_2 = R_x + R_{x-1}$ should be replaced by $S_2 = R_x + R_{x-1} - R_x \cdot R_{x-1}$. Engset seems to be aware that his solution is not exact. However, as elsewhere, he is satisfied that the error is on the safe side, and that the error is small for low values of the degree of hindrance.

10 Progressive and combined gradings

As a first step Engset in Chapter X describes a straight grading where each of r groups is served by a set of secondary devices x_0 that is offered a traffic y. The traffic carried on x_0 is y_0, and the remaining total traffic of $r \cdot y_1 = r \cdot (y - y_0)$ is offered to a common group of x_1 devices. The calculation is straightforward according to his previous method for non-graded systems. Two example calculations are carried out for $r = 2$, $x_0 = 8$, $x_1 = 2$ and $y = 6$ and $y = 5.5$ respectively. The case of $y = 5.5$ leads to slightly less than 1 ‰ hindrance. The obtained advantage of the grading is thus very close to that found for the slip multiple grading, 37.5 % traffic increase or ≈ 27 % equipment saving.

Engset also considers the possibility of having several stages in the grading, as a more general progressive grading, without carrying out any calculations. The method is, however, easily applied to such cases.

At the time there was no understanding of the effect of the overflow character of the traffic, and Engset does not take that into account. He only considers the first moment of the traffic distribution, which of course leads to too optimistic results.

Finally, in Chapter XI a combination of a slip multiple grading as a first stage, followed by progression to another slip multiple stage, is studied. The calculation method does not differ from that used for each of the two separate methods. According to the calculation a saving in the amount of selectors of ≈ 42 % is obtained, while maintaining the grade of service. As an alternative the equipment could be kept non-reduced in order to carry more traffic. It is concluded that, while maintaining the grade of service, the grading would permit a traffic increase of ≈ 74 %.

A discussion with some coarse estimates is carried out with the aim of adapting the slip multiple grading to varying traffic requirements. Among other the multiple slips may be modified to double steps and to overcomplete cycles.

11 Concluding remarks

The above discussion attempts to give a brief survey of Engset's quite voluminous report, without missing any of his major points. The report goes a long way in showing details of the mathematical development as well as the approximations felt necessary to obtain practical usable results. Also detailed numerical calculations are carried out. Much of this is of course of less interest today, given the tools available for numerical calculations.

The report demonstrates very clearly that the primary objective was to find ways of dimensioning the new automatic telephone systems that were being studied for introduction in the Norwegian telephone network. Engset had a leading position in these efforts, as he was member of a committee established in 1910 to give advice on the matter for the General Director of Telegrafverket (Norwegian Telecom Administration).

This committee was doing extensive studies by travelling in Europe and America as well, and came out with a report in 1913. The implementation of the committee recommendations was delayed because of the war, and the system offers were ready for evaluation in 1916. For the purpose of this evaluation the committee was supplemented by Johannsen and Christensen from Copenhagen and Hultman from Stockholm, all very notable people in the business. This is very interesting in view of the Engset report being completed in 1915.

One may also note that from 1917 on there were regular meetings between the Nordic telecom administrations. Engset participated in ten such meetings from 1917 to 1935, whereas the teletraffic community in Denmark was primarily within the private company KTAS, and thus not participants in these meetings. (Also in Norway there were many private telephone companies at that time.) It is this author's impression that there was a lively exchange of information between the Nordic telecommunication communities during the transition from manual to automatic systems. Traffic matters were of great operational and economic importance.

Engset's 1915 model fell in a way between Erlang's two models of 1917, that of loss (congestion) system and that of waiting system. Erlang applied the birth-and-death process in an explicit way, with emphasis on the distribution types. Engset, in a more implicit way, showed that the arrival independence lead to Bernoulli and Poisson distributions, and argued that his model set no conditions on holding time distributions. However, with a limited number of traffic sources a bias would exist for uneven traffic demand per source.

References

3 Engset, T. The Probability Calculation to Determine the Number of Switches in Automatic Telephone Exchanges. Translation by E Jensen. *Telektronikk,* 88, (1), 90–93, 1992.

4 Jensen, E. On the Engset Loss Formula. *Telektronikk,* 88, (1), 93–95, 1992.

5 Engset, T. *Om beregningen av vælgere i et automatisk telefonsystem.* Kristiania, 1915. Unpublished report.

6 Iversen, V B. Thore O. Engset : a pioneer in teletraffic engineering. *NTS13,* Trondheim, 1996.

7 Engset, T. On the Calculation of Switches in an Automatic Telephone System. Translation of [5] by A Myskja. *Telektronikk,* 94, (2), 1998 (this issue).

8 Erlang, A K. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. In: Brockmeyer, Halstrøm, Jensen. *The Life and Works of A.K. Erlang.* Copenhagen, 1948. First published in Danish 1917.

9 Bortkewitsch. *Das Gesetz der kleinen Zahlen.* Leipzig. B.G. Teubner, 1898.

10 Palm, C. Some investigations into waiting times in telephone plants. *Tekniska Meddelanden från Kungl. Telegrafstyrelsen,* 7–9, 1937. (In Swedish.)

11 Palm, C. Research on telephone traffic carried by full availability groups. *TELE,* 1, 1957. Edited by E. Malmgren.

12 Cohen, J W. The generalised Engset formula. *Philips Telecommunication Review,* 18, 158–170, 1957.

13 Dartois, J P. Lost call cleared systems with unbalanced traffic sources. *Proc. of the 6th International Teletraffic Congress, Munich,* 215/1-7, 1970.

Appendix

Queue with departures. Palm's and Engset's models

1 Introduction

Conny Palm published in 1937 a study of a queueing system where customers waiting in a queue may leave the queue without being served.

The study was first published in Swedish [10] and later in French. A revised version appeared in a collection of articles by Palm published in 1957 [11].

It turns out that Engset's main model in his report from 1915 [5] under certain conditions is identical to this model by Palm. Palm's model assumes the following:

- The number n of servers is limited
- The number N of customers is unlimited

- Offered traffic $\underline{a}$ per subscriber tends to zero
- Total offered traffic A is given ($A = N \cdot a$)
- Arrivals are Poissonian
- Service times are exponential
- The number of queue positions is unlimited
- Waiting times in queue, provided no service, are exponential (patience function).

Engset's model has no explicit mention of a queue, but the customers are allowed to wait actively for service. The numbers of servers and of customers are limited. Each customer $\underline{i}$ generates an offered traffic load p_i, $0 < p_i < 1$.

Customers operate independently, but no condition is put on holding time distributions. At any instant the servers are either free or they are partly or wholly occupied, and in the latter case there may or may not be waiting customers. The holding time of a call is independent of its fate, and it may be spent fully in waiting, partly in waiting and partly in service, or fully in service. Thus, the total offered load will consist of two parts: the service load and the waiting load. By integrating over time there will be two traffic volumes, that of service and that of waiting. The two corresponding ratios of the offered traffic are termed the *degree of efficiency*, φ, and the *degree of hindrance*, ψ, where φ + ψ = 1.

It should be pointed out that even if Engset has the more general assumptions concerning the number of sources, and even allowing for individual differences between traffic demands per source, Palm's study takes a much more general approach concerning time parameters and distributions, and he also carries the analysis much further.

2 The Palm model

Definitions:

Number of sources:	$N = \infty$
Number of servers:	n
Mean call interval:	$1/y$
Mean service time:	s
Mean departure time (expressing patience):	b
Offered traffic:	$A = y \cdot s$
Parameter	$\alpha = y \cdot s / n = A/n$
Parameter	$\beta = n \cdot b / s$
Number of servers occupied:	$0 \le p \le n$
Number of queue positions occupied:	$0 \le q \le \infty$
Probability of state x:	$[x]$

Statistical equilibrium requires:

$$[0]^{-1} = \sum_{p=0}^{n-1} \frac{(s \cdot y)^p}{p!} + \frac{(s \cdot y)^n}{n!} \cdot W(\alpha,\beta) \tag{1}$$

with

$$W(\alpha,\beta) = \sum_{q=0}^{\infty} \frac{(\alpha \cdot \beta)^q}{(\beta+1)(\beta+2)\ldots(\beta+q)} = \beta! \sum_{q=0}^{\infty} \frac{(\alpha \cdot \beta)^q}{(\beta+q)!} \tag{2}$$

where the term for $q = 0$ is set equal to 1.

With the notation $E_{1,n}$ for Erlang's loss formula we get

$$[0]^{-1} = \frac{A^n}{n!}\left(W - 1 + \frac{1}{E_{1,n}}\right) \tag{3}$$

For $p \le n$ and $q \ge 0$ respectively, we get the state probabilities

$$[p] = \frac{E_{1,n}}{1+(W-1)E_{1,n}} \cdot \frac{n!}{p! \cdot A^{n-p}} \tag{4}$$

$$[n,q] = \frac{E_{1,n}}{1+(W-1)E_{1,n}} \cdot \frac{(\alpha \cdot \beta)^q}{(\beta+1)(\beta+2)\ldots(\beta+q)} \tag{5}$$

Two limit cases are

1) $\beta = 0 \rightarrow W(\alpha,0) = 1$, leading to $[n] = E_{1,n}$

2) $\beta = \infty \rightarrow W(\alpha,\infty) = \;= \frac{1}{1-\alpha}$, leading to

$$\sum_{q=0}^{\infty}[n,q] = E_{2,n} = \frac{E_{1,n}}{1-\alpha\left(1-E_{1,n}\right)}$$

corresponding to Erlang's two formulas for loss and waiting, respectively.

The probability of congestion, defined as the probability of all servers busy, in the general case is given by

$$C_n = \sum_{q=0}^{\infty}[n,q] = \frac{W \cdot E_{1,n}}{1+(W-1)E_{1,n}} \tag{6}$$

The mean waiting time for all calls waiting can be expressed in a simple way by

$$T_w = \frac{\sum_{q=1}^{\infty} q \cdot [n,q]}{y \sum_{q=0}^{\infty}[n,q]} \tag{7}$$

where the numerator gives cumulated waiting time per time unit and the denominator gives the number of calls being exposed to waiting per time unit.

By means of equation (5) equation (7) can be transformed to

$$T_w = \frac{\beta}{y}\left(\frac{1}{W} + \alpha - 1\right) \tag{8}$$

leading to mean waiting time per call for all calls

$$\tau_w = C_n \cdot \frac{\beta}{y}\left(\frac{1}{W} + \alpha - 1\right) \tag{9}$$

As there are y calls per time unit, the waiting traffic can be expressed by

$$A_w = \tau_w \cdot y = C_n \cdot \beta\left(\frac{1}{W} + \alpha - 1\right) \tag{10}$$

3 The special case of *b* = *s*. Comparison with Engset's result

In the special case of $b = s$, which is in fact Engset's assumption, we get the congestion

$$C_n(A,n) = \sum_{v=n}^{\infty} \frac{A^v}{v!} e^{-A} \tag{11}$$

Engset's definition of 'hindrance' or 'hindered traffic' is identical to that of equation (10). The degree of hindrance, ψ, is given by the ratio between the hindered traffic and the offered traffic. In the *general* case we get from (6) and (10)

$$\psi = \frac{A_w}{A} = \frac{W \cdot E_{1,n}}{1 + (W-1)E_{1,n}} \cdot \frac{n \cdot b}{A \cdot s}\left(\frac{1}{W} + \alpha - 1\right) \tag{12}$$

In the given case with $b = s$ we have from (10) and (11)

$$\psi = \frac{1}{\alpha}\left(\frac{1}{W} + \alpha - 1\right) \cdot \sum_{v=n}^{\infty} \frac{A^v}{v!} e^{-A} \tag{13}$$

In this case $\alpha \cdot \beta = A$, $\beta = n$ and $\alpha = A/n$. From equation (2) we get

$$W = \frac{n!}{A^n} \sum_{v=n}^{\infty} \frac{A^v}{v!} \tag{14}$$

With substitution in (13) we obtain

$$\psi = \frac{A^{n-1}}{(n-1)!} \cdot e^{-A}\left[1 - W\left(1 - \frac{A}{n}\right)\right] \tag{15}$$

A simple manipulation leads to

$$\psi = \frac{A^n}{(n+1)!} \cdot e^{-A}\left[1 + \frac{2A}{n+2} + \frac{3A^2}{(n+2)(n+3)} + \ldots\right]$$

$$= \sum_{v=n+1}^{\infty} \frac{(v-n)A^{v-1}}{v!} e^{-A} \tag{16}$$

This is identical to equation (14) in Engset's report [5], except that Engset truncated the infinite series at N = the limited number of traffic sources. (In Engset's notation $N = n_0$.) The last term in the parenthesis thus becomes

$$\frac{(N-n)A^{N-n-1}}{(n+2)(n+3)\ldots N}$$

By omitting this truncation it is demonstrated that Engset's model leads to a result identical to the Palm model with an infinite number of sources, a number n of servers and an unlimited queue, where customers leave the queue according to an exponential 'patience' function with a waiting time parameter equal to that of service time.

The relative value ψ is the ratio between the traffic load in the queue and the offered traffic A.

Engset developed the expression (16) by weighting the Poisson terms $\frac{A^x}{x!}e^{-a}$ from $x = n+1$ upwards by 1, 2, 3 etc., since these are the numbers of subscribers in queue for the corresponding terms.

In order to justify using the Poisson distribution instead of the Bernoulli (binomial) distribution for a limited number of sources, he shows by taking the derivative of the logarithm of the binomial terms with respect to the number of terms, that increasing the number of terms towards Poisson leads to an increasing queue traffic when $A < n + 1 - \sqrt{n+1}$. For practical dimensioning that will normally be the case. By applying Poisson instead of Bernoulli when there is a limited number of sources, the calculation will thus be on the safe side, ie. waiting in queue will be less than that calculated.

Under the given conditions one will have

$$\sum_{v=n+1}^{N} (v-n)\binom{N}{v}\left(\frac{A}{n}\right)^v \left(1 - \frac{A}{n}\right)^{N-v}$$

$$< \sum_{v=n+1}^{N} \frac{(v-n)A^{v-1}}{v!} e^{-A}$$

$$< \sum_{v=n+1}^{\infty} \frac{(v-n)A^{v-1}}{v!} e^{-A} \tag{17}$$

The common factor $(v - n)$ represents the weighting by the number in queue, while the three expressions without weighting represent, respectively, the sum of the Bernoulli terms from $n + 1$ to N, the truncated sum of the Poisson terms from $n + 1$ to N, and the sum of the Poisson terms from $n + 1$ to ∞.

Engset has calculated values with four decimals according to the middle expression of (17). The choice of N-value is not stated and most value sets are such that the value of the expression will be < 0.01. In those cases the differences, if any, between corresponding values only turn up in the last (fourth) decimal.

Arne Myskja is Professor Emeritus at the Norwegian University of Science and Technology, where he has been employed since 1960. He was formerly with the Norwegian Telecommunications Administration, and has had teaching/research assignments in the US, Denmark, Sweden, Australia and France. His main areas have been switching systems and teletraffic theory. Most recently he has devoted his time to studies related to T. Engset.

e-mail: arne.myskja@item.ntnu.no

The man behind the formula

Biographical notes on Tore Olaus Engset[1]

ARNE MYSKJA

Tore Olaus Engset is known in the teletraffic community for the formula that carries his name. To most others he is rather anonymous, or he is known as the Director General of the Norwegian Telecommunications Administration from 1930 to 1935, after having served the same institution his whole adult life from 1883 on.

New light is cast on Engset by a newly discovered extensive report on teletraffic modelling, a report that was never published, but which contains original aspects, some of which were first treated much later by other authors.

Even less has been known of the impressive scientific work by Engset on atomic theory, as demonstrated by a report in German finished in the last year of his life, 1943. This again led to the discovery of a 50 page series of articles on the same subject in the reputed journal *Annalen der Physik* in 1926 – 1927, which – among other things – discusses the then recently published Schrödinger wave equation.

Who was this man? Very little has been publicly known, and very few people now can remember him. The time is due to supplement this first publishing of some of his unknown written work by a broader personal presentation. Maybe his very modesty limits the possibilities of following his track, but there is something that is much more than nothing.

1 Introduction

Engset's formula is well known and appears in most present day textbooks on teletraffic theory. His only publication [1] on a traffic theory topic is still being referred to 80 years after its appearance in the German journal *Elektrotechnische Zeitschrift* in August 1918.

The Danish scientist and engineer A.K. Erlang is rightly called 'the father of teletraffic theory'. His publications on the subject and related matters span from 1909 until his death in 1929. His principal paper [2] containing the B- and C(D)-formulas came in 1917.

Engset's formula has been – and may be – considered as an appendix to or an extension of Erlang's B-formula. The difference in principle is that while Erlang assumed an unlimited number of traffic sources the Engset model permits any arbitrary number of sources. Thus, one may as well turn the matter around and say that Erlang's formula is a special case of Engset's formula, obtained by letting the number of sources increase to infinity. However, the sequence of the publications should not be neglected.

In spite of the naming of the formula, Engset as a teletraffic pioneer has not been a focus of interest within the scientific community, which may be seen as natural in view of his very limited known production. The reason for a renewed interest is the discovery of a voluminous report by Engset of 130 typewritten pages from 1915 – two years before Erlang's main paper – on teletraffic matters [3]. The report is an original work with many aspects, some of which were taken up by others only many years later.

[1] *In the baptism protocol the name is Thore Olaus Engeseth. Himself, he always wrote Tore Engset. Locally, the normal form today in writing is Engeset, while pronounced more like Enset.*

When Engset's 1915 report now appears in an English translation [4], along with a summary article [5] on the report, it is an opportune occasion to supplement all this by some biographical notes.

T.O. Engset is a publicly known name on two counts: One is that of the teletraffic expert as focused here, and the other is that of the man serving Telegrafverket (later: Televerket/Telenor, the public telecommunication operator in Norway) for 52 years, the final five years as its Director General.

There is no extensive biography of Engset written up till now. However, there are bits and pieces several places. In T. Rafto's *Telegrafverkets historie 1855–1955* [6] there are many references to Engset.

The best separate presentation is offered by L.A. Joys in Televerket's internal periodical *Verk og Virke* no. 4, 1967 [7]. The presentation contains the main biographical data, some citations from ref. [6], and it has also an anecdotal character. Joys, by the way, was probably the one person who devoted most effort in studying Engset's formula and its significance. He wrote several papers on the formula, and at the age of 70 earned his doctorate on the subject. Joys' recursion formula has been applied for calculation of Engset tables [8]. In the introduction to his Engset presentation Joys regrets the lack of sources. The 1915 report was not known to him, and would certainly have supplemented his knowledge and elevated his already high esteem of Engset.

An interesting short biography is presented by Eliot Jensen in a commentary article [9] in 1992 to his English translation [10] of Engset's 1918 paper. E. Jensen states the following:

> *"Engset was very interested in mathematics. This, in addition to more than 20 years of experience in practical teletraffic matters, forms the background for the theoretical work he published in 1918. The main material of the paper had been established already by 1915, but Engset at that time did not consider it of sufficient scientific substance to have it published. However, subsequent contact with prominent people of the Copenhagen Telephone Company, bringing him impulses from the traffic research there, appears to have changed his mind".*

It is not clear how E. Jensen knew that "the main material had been established already by 1915". There is no other indication that he had any closer knowledge of the 1915 report. The impulses from the CTC people to publish his 1918 paper are very likely, but he may as well have been discouraged from publishing his somewhat different model in the 1915 report.

The editor of *Telektronikk,* Ola Espvik, requested me to collect information for a more comprehensive biographic presentation. For reasons of time and other limitations it has not been possible to write a full biography. That would require extensive search in public and private archives, and would also require a different frame of publication. It is even uncertain if much more can be found than what is presented here.

The intent is not to draw a picture of the professional man only, but rather to give a broader picture, including the human aspects, in a narrative of an in many ways remarkable man. Since it is to appear in a technical journal, it is only fair to advise those readers who are solely interested in the scientific aspects to leave at this point and rather go to references [1], [4] and [5].

2 The background

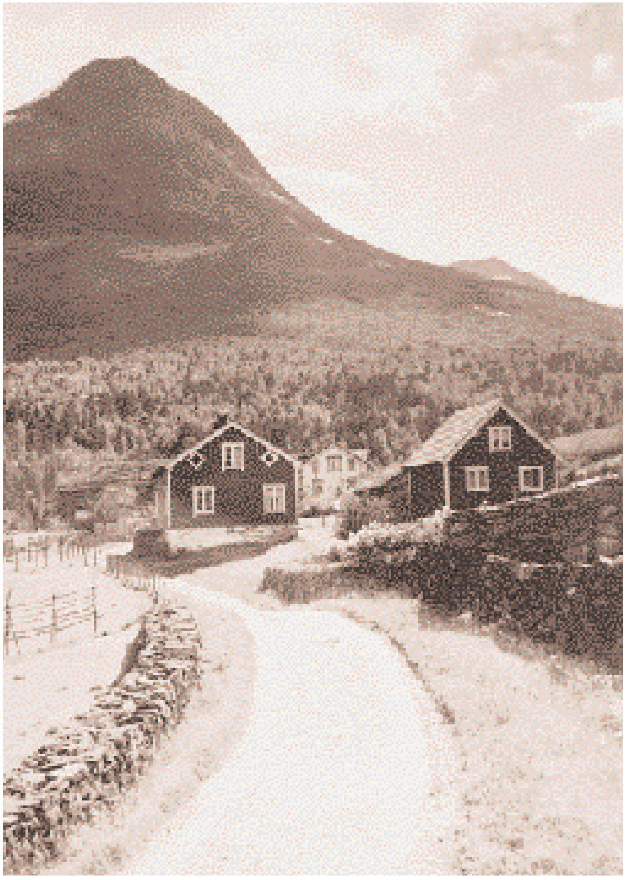
Tore Olaus Engset was born 8 May 1865 at Stranda in the county of Sunnmøre in western Norway. Stranda lies in the heart of the fjord country that is famous for its breathtaking nature views, with nearby Geiranger, for decades the target of large tourist cruisers. T. Engset, in retrospect after retirement, reviews some of the great panoramas that he had seen in his life, notably the Grand Canyon and Niagara Falls, "But of all I have seen of great and beautiful from the Creator's hand nothing has for me been like the view from the top of Roaldshorn that Sunday in my 19th year".

Stranda in 1865 was a mainly agricultural community with very modest – many would say poor – living conditions, but has since then developed into a prosperous industrial town, manufacturing furniture, foodstuffs, clothing, mechanics, electronics etc., with foodstuffs having taken the lead over furniture. As an illustration the industrial turnover increased from 95 million kroner in 1967 to 750 in 1987, a factor of eight in nominal value. Since then the figure has more than doubled. Tourism is important, and even national championships in alpine skiing has been hosted by the town. The present unemployment rate is less than one percent.

Engset (or rather Engeset) is a small community with a little group of small farms along a rather steep slope. Tore grew up on the farm named after his great-grandfather Kornelius (today, the 'Karneles farm'). The owners of the farm are known in an unbroken line since 1540. The earliest of Tore's forefathers on the farm was born in 1655.

Tore was the second youngest among ten children, four of whom died at one year or less. There were also two older half-brothers and one half-sister. Tore's mother died when he was nine, and her last words to her then 22 year old daughter was: "Don't forget Tore". And she did not. She stepped into her mother's place and delayed her own marriage for several years. Elen Regine was the only of Tore's sisters to grow up. In spite of her own ten children, she told in a letter that he always kept a special room in her heart. She admired her younger brother very much, but there were limits: When he told her that in the future she would be able to speak over long distances without telephone wires, she replied that he might tell her whatever he wanted, she did not believe a word.

The school records from 1875 show that 10-year old Tore never missed one day. He was later sent to junior high school in the town of Aalesund, from where he graduated at the age of 18, receiving a prize from a local legacy in the form of two book volumes with golden letters on the back, with a hand-written note by the school principal, for "diligence, progress and good conduct".



The 'Karneles farm' at Engset, Stranda, where Tore Olaus Engset grew up

3 Professional career

T. Engset was admitted to the telegraph school in Stavanger in 1883, and that turned out to be decisive for his future career. He got his exam at the end of the year, and from January 1884 he was employed as a telegraph assistant at Lødingen, Arendal and Trondheim, until in January 1890 he got a position as junior secretary at the Director General's office in Kristiania (Oslo).

Since he had only his junior high school exam from 1883, he went on with studies in his spare time. He graduated as a student in 1892 and went on with university studies in mathematics and physics and got his M.Sc. degree 1894.

Gradually he advanced to secretary, office manager and head of the traffic section of Telegrafverket. As such he was also the deputy for the DG.

From 1897 there were two departments, the administrative and the technical, with the administrative department as the more 'political' branch, which became Engset's responsibility. In 1920 a further subdivision was done, with five departments and with Engset as head of the traffic department. Then in 1926 there was a reversal back to two divisions, one traffic division and one technical division. In fact it was a grouping of departments on a higher level, according to the growth of the institution. Engset kept his position as second in command.

When Rasmussen retired from the position of Director General in 1905 Engset was one of five applicants for the vacant position. However, the new DG, Heftye, was chosen from outside ranks, and he kept the position until he died in an accident in

1921. Engset acted in the position for parts of 1921 and 1922 and applied a second time with no more success than before. The new man was Nickelsen, also an outsider, who at the end of 1929 left the position because of political differences with the ministry about the independent authority of his institution.

Engset was now 65 years old, and he admits in a private letter that he was very much in doubt whether to apply for the top position a third time. A couple of colleagues suggested that he might simply neglect to apply this time, as he would probably be asked anyway. Realising that such procedure would be too risky he sent his application at the latest possible time. There were nine applicants. Engset was appointed on 22 August 1930 and received a telegram of congratulations from the Minister.

Before this Engset had to go through a hard fight with the all-male union within the organisation. In his interim position as acting DG he had appointed a female head of staff. The union leader was also editor of the newsletter of his organisation, and from this position he fought for the male candidate, claiming that the female candidate was given her previous position in the staff office in order to take care of women's interests. He even paid a visit to the Minister trying to influence him on the matter, and to get a speed-up of the process of appointing the DG. Engset claimed that his candidate had responsibility for the whole staff, and was clearly best qualified. He put the matter forward to the Minister, who admitted that he generally preferred men for such high positions, but he did not want to reverse the appointment. Even formally, Engset claimed that he was on safe ground. A 1924 decision by the National Assembly had given equal rights of higher positions, given proper educational background.

Engset feared that the episode might weaken his possibility to obtain the top position, so he felt a great relief when he received the telegram from the Minister. The particular background of having been second in command for so many years, and now at the age of 65 was acting DG for a second time, made him particularly sensitive to the risk of being bypassed once more. He was certainly aware that there had never been an insider in the position. He turned out to become the first. He had asked the Minister to avoid a long interregnum like that in 1921–22, "when they let me hang in the position a full year, and then cut me down to let me lie in the gutter (that was exactly what I said)"[2]. Those were strong words by a man of Engset's character.

Engset's lifelong career in telecommunication administration coincides with the first fifty years of telephone in Norway. From the beginning there was competition between many local private companies that were established for the promotion of the new medium and its basic technology on one side, and the publicly owned and operated telegraph company on the other. The private companies had the edge in most cities because of local business initiatives and a more flexible operation. There were, however, problems of parallel network construction and lacking interconnection. The companies were solely locally based, whereas the public telegraph company had a national coverage.

The matter became one of national political interest. The rural representation in the National Assembly was strong, and the idea of equal opportunities for all citizens, irrespective of location and operational cost, became decisive. After much delay a new law was passed by the National Assembly in 1899, giving exclusive rights of telephone establishment and operation to the state, and the public telegraph company was authorised to carry out the task.

The monopoly law of 1899 remained virtually unchanged for more than eighty years, and almost as long as it took to carry out the process of transferring all private telephones to the state(!). Much less time was needed to reintroduce full competition 99 years after the initial law enactment. The long delay in enforcing the law was due to the lack of public funding and to local resistance. In concentrated local communities it was possible to establish and operate a telephone service at cheaper rates than could be done nationwide. Financing of expansion according to demand by loans and stock emission was straightforward in private companies, whereas the authorities did not permit that for the public state owned system.

Increasingly, though, also local telephone came under state ownership, but not without fierce fights in many places. A very informative recount is found in [6]. According to this source even notable people like Fridtjof Nansen and Erik Werenskiold were involved in the 'telephone war'. In the 'violent dispute' there were demands that "the director general as well as his expert advisor, head of office Engset, must be removed". (Erik Werenskiold, by the way, later made the official portrait painting of Director General T. Engset, 53 years after he made the corresponding portrait of Tank-Nilsen, the first DG of Telegrafverket).

No doubt Engset was in the front line of the telephone development. He had to stand against the public dissatisfaction with supply being far behind demand, and to fight for better funding from the government. In fact, even in the National Assembly lengthy discussions were carried out on what would today be considered petty details. At the same time his primary task was the technical planning and accomplishment.

A particularly challenging task was that of introducing a new generation of telephone exchanges. In spite of the first automatic exchange having been put in operation in 1892 in USA, manual operation was predominant for another 20–30 years. In 1910 a three-man expert committee was established to evaluate the development of telephone switching and give recommendations for the choice of a future system. An immediate cause was the need for expansion of telephones in the capital city.

The committee took its task very seriously. They visited nine cities in Sweden, Denmark, Germany, Austria and Holland, and eight cities, from New York to Los Angeles and San Francisco, in the United States. They spent 48 days travelling in Europe and 71 days on the US trip. The final committee recommendation [11] was put forward in March/April 1913, and contained 67 printed pages and 13 appendices. The committee proposed conversion to a fully automatic system with primary and secondary exchanges. The imminent world war led to changes and delays, and the first automatic exchange was not put in operation until 1921.[3] As this committee work represents a very

[2] *In a private letter September 1930.*

[3] *The first public automatic telephone exchange in Norway was put in operation in 1920 by the private telephone company in the town of Skien.*

important part of Engset's career, we shall quote from ref. [6] in translation:

> *"It was a great, all new and difficult task the study committee of 1910 had embarked upon. It should make a judgement on the three main solutions for telephone exchanges: the manual, the semi-automatic, and the fully automatic. It should further estimate the actual and future demand and invite offers from the telephone manufacturers. Chief engineer Abild, office head Engset and telephone director Iversen each had insight, experience and detailed knowledge in telephone matters. They now investigated thoroughly everything concerning modern telephone operation and carried out excursions in European countries and through America all the way to the Pacific coast. The result was a thorough report that recommended transition to a fully automatic system with primary and secondary exchanges.*
>
> *The fully automatic solution would among other things solve the difficult problem of large exchanges, as it permitted a subdivision in several satellite exchanges and establishment of the single exchanges according to subscriber density. In design the automatic system represented a decentralisation, in operation a centralisation. The new City Centre exchange in Kristiania (Oslo) was intended to assemble the large business subscribers, that were the greatest users of telephone and had the most mutual correspondence. For connection to the main exchange were planned four sub-exchanges in the city for private and small business telephones. All exchanges should have connection to four suburban exchanges belonging to the Kristiania telephone area. The system was aimed at 30,000 subscribers, but could be adapted to as many as 90,000 numbers. The construction period was set to four years, and the cost estimated at 8.5 million kroner.*
>
> *On the basis of the report an initial grant of 980,000 kroner was given by the National Assembly for the start of reconstruction works. A disappointment was that 'for budgetary reasons' the 4-year plan could not be kept and construction time had to be extended to 6 years. A factor that nobody could have foreseen came up: the war. As the war erupted, all the ongoing works had to cease for a while. The delays that followed could not be regained. On the contrary, new interruptions arose incessantly. The bids on system installations could not be brought to a decision until 1916. For examination of the bids the committee was enforced by three experts from abroad in order to guarantee best knowledge and fair choice: Telephone director Fr. Johannsen had headed the reconstruction and reorganisation of the Copenhagen telephone, so that it was made 'a gold mine for experts wanting to study modern telephone operation'. Telephone director Axel Hultman from Stockholm was an outstanding telephone authority with international reputation, and he had himself designed a telephone selector; while the Danish telephone engineer P.V. Christensen, as a specialist on the mathematical side of the matter, assisted with calculation of the traffic equipment."*

Engset had already at this stage had some contact with the Copenhagen people, and that may even have instigated the more active co-operation around the evaluation of system proposals. What is clear, however, is that he must have been working on his traffic modelling for some substantial time before the 1916 co-operation. His comprehensive 130 page report, dated 1915, must have taken considerable time, at least months, more probably years. It was developed from scratch as a left-hand activity beside his work in a demanding administrative position as second in command in a big national public service organisation. This point will be considered closer in the next chapter. It is mentioned here in order to highlight the very close relation between Engset's traffic studies and the great task of introducing a radically new technical solution for large-scale telephone switching in the national network. Practical solutions on a safe theoretical basis is the governing idea of the report.

World War I was a time of great expansion in telegraph and telephone traffic, but also of scarcity of material and equipment. For the first two years the available funding of expansion was in fact reduced, and serious crises arose in traffic performance, as demands far exceeded capacity. It led to outcries, and an emergency plan was issued to improve the situation. It did help, but inflation sapped some of its intended improvement. However, the post-war period was a constructive one with substantial expansion, until the 1921 crisis hit hard. Engset was in charge, until the new DG Nickelsen took over in 1922. The following years were marked by struggle for more independence from direct government control with more freedom of financing and operation. The limitations were felt as a straitjacket that prevented expansion according to market needs. When Nickelsen left in protest and Engset was finally appointed, the new worldwide economic crisis was already in progress.

In a review of Engset's career at the time he was appointed, T. Rafto in [6] writes: *"From 1894 he worked as secretary at the 'Main office' in Kristiania. From now on and all through the years he was head of traffic and operational matters. This man of administration 'par excellence' was absorbed in his work with ardent interest. Prudent, knowledgeable and with impressive energy he became the best support of the director. He expressed openly his gratitude of having been given a 'place in the sun', and he enjoyed his work."*

The newly appointed Engset, with his experience from the crisis in 1921–22 had to handle a similar situation from 1930 onwards. There were restrictions instead of expansion in most areas. *"As Engset in 1930 got the full responsibility, a similar situation emerged as that of ten years earlier when he acted as director during the vacancy. While then the post-war depression reigned, now a new world crisis was under development. The boom in world markets in the latter half of the twenties gave the illusion of continuous progress of prosperity. But during summer and autumn 1929 the radiant hope of better times was crushed. ... It was helpful that Engset was accustomed to dealing with tight budgets and wage reductions."*

National coverage by long distance telephone had been obtained by 1920. While landline telegraph gradually stagnated, long distance telephone increased continually. Radio telegraph and – from the twenties – radio telephone were also growing. Engset, with his radio expert colleague Hermod Petersen, worked on expansion of the radio network and on creating a national broadcast system. Also picture telegraph came on the agenda, and the first unit was ordered in 1930.

During Engset's five year period on top, the fight for more independence continued. A committee was established to take a new look at the question. A favourable recommendation got part support and part opposition from different ministries. It was supported by the parliament committee, but the proposal fell in the plenary session.

The economic crisis gradually eased during the five year period and the matter of integrating the whole national telephone system under one administration emerged again. Engset's political views in general are not known, but already in 1930 he expressed the opinion that "if the country got a labour government, it would probably bring in order the telephones in Norway like in many other countries." He was confident that his institution would be capable of taking over, but it must be done in a lenient way, should the transition be for the better. When the labour government took office in 1935, an initiative in the direction was taken. Limited funding, however, kept the process at a slow pace, and the take-over was not finalised until 1972(!).

Engset's leading position in the organisation of the national communication system, with an ever increasing interconnection with a global system, naturally brought him in touch with international work.

Already in 1901–1902 he stayed with his wife for an extended period in Copenhagen. The main reason was poor health. His doctor ordered him to quit work for a while to recover. Although it is not known, it is fair to assume that he got in touch with colleagues in Copenhagen already at this stage. The committee work from 1910 onwards gave him a broad view of the whole western world, and later in close co-operation with Danish, and to a lesser extent, Swedish colleagues.

From 1917 on a more regular form of contact was established between the telecom administrations in the Nordic countries. A survey of 80 years of Nordic telecom co-operation is given in [12]. Altogether 96 conferences are listed. Out of 14 such conferences in the period 1917 – 1934 Engset took part in ten. He was also a delegate to the very important international telecommunication congress held in Madrid in 1932. This was the congress where the International Telecommunication Union (UIT) was founded on the basis of the three separate committees CCIF, CCIT and CCIR.

A Norwegian Encyclopaedia from around 1950 (Norsk Allkunnebok) [13] has a mention of Engset, where his main career data are given. The further information is as follows: "He had little schooling, but made a lot of studying on his own, collecting great knowledge, even outside his own profession. He was Norwegian delegate to several international conferences on telegraph and telephone matters and gained much honour at home as well as abroad."

Tore Engset retired at the age of 70 in 1935. An outstanding man of duty and independence had finished his 52 year career. At that point the international and national economic conditions were on the rise. Optimism was returning, until – alas – World War II erupted. But that is another story.

4 Science for service

There is this dual picture of Tore Engset: The lifelong professional career man – modest and ambitious at the same time – at the front of a team creating a many-faceted national telecommunications system; and the introvert amateur academic who quietly developed his mathematical traffic model. Which one of the two was the most influential? In Rafto's words [6]: "Beside his primary function, all through the years he was occupied with scientific investigations".

In his own country and among non-specialists of traffic theory he is the professional man who only at the end of his career reached the top, and therefore was less visible as a man behind the scenes. In the international teletraffic community he is the man behind the Engset formula. However, his basis is only the two-page article from 1918, appearing in the wake of Erlang's more famous work of 1917.

The newly discovered report from 1915 changes the picture radically. Engset created his own traffic model from basics as an original work. His only assumption is that of mutually independent sources that leads to Bernoulli/Poisson distribution. Otherwise he does not make any assumptions of distributions. From this basic model he arrives at results that are only published much later: Departure from queues (Palm 1937), the generalised Engset formula (Cohen 1957), most unfavourable case when identical sources (Dartois 1970), total loss with day-variations of traffic, calculation of gradings, etc.

Engset was not satisfied with only theoretical solutions. For him the application of scientific methods was not a target in itself, but rather a tool to obtain improved telephone service. His very clear intention was to create practical dimensioning tools that could be applied within the new automatic telephone systems that were being planned by himself and his colleagues. With the limited means that were available for numerical calculations he was very conscious of two conditions: Approximations had to be introduced, and these approximations had to be on the safe side. The second condition implied that the real service should be better than that calculated. Even with the simplified calculations available, he found it important to establish dimensioning diagrams that allowed direct reading of value sets, and he concluded his report with calculated tables and diagrams.

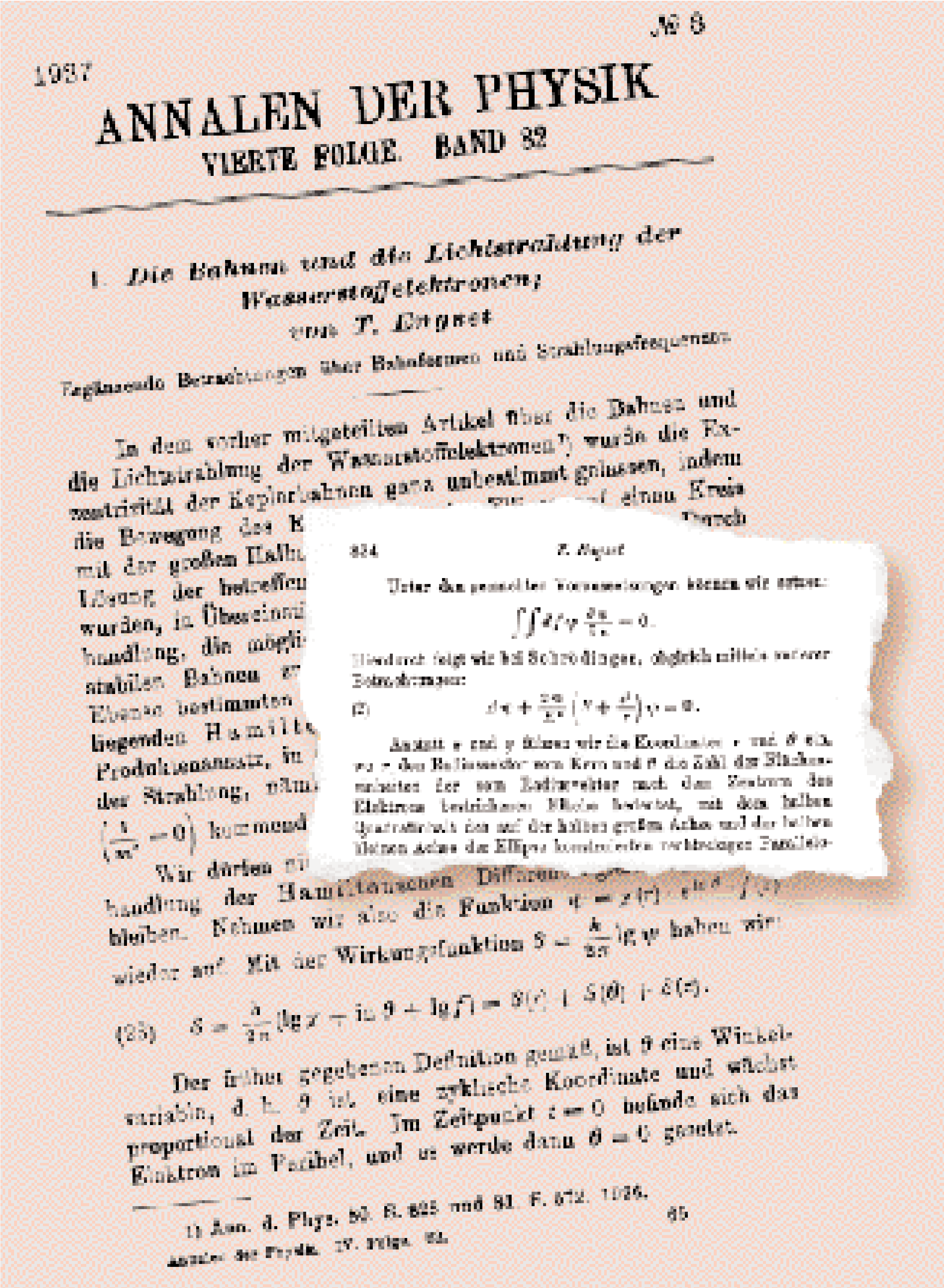
As the 1915 report was never published, and the Erlang model was accepted as a better proposal, also by Engset himself, the report was put aside and kept in the files without being utilised. The Engset formula of 1918, however, applied the Erlang truncation (1917) and was accepted as the solution for loss systems with limited number of sources. The Erlang and Engset formulas are both formulas for congestion probabilities due to service system limitation, the former given by a truncation of the Poisson distribution and the latter by a truncation of the Bernoulli distribution.

As the scientific teletraffic work of Engset is discussed in an accompanying article [5], we shall simply refer to those articles and omit any further treatment at this point.

This, in the present context of teletraffic, ends the story of Engset's scientific work. However, in the final section of this article a quote from Joys' biographical notes indicates that Engset after retirement worked on issues of atomic theory, and that he was very interested in the utilisation of nuclear power.

5 The physicist

The information in Joys' notes led me to a further search that brought me in contact with Engset's step-grandson in Oslo, Willi Urne Fonahn. It turned out that he kept some papers of Engset's in a safe deposit box. By a recent inspection of the contents such a dissertation was in fact found. It is dated spring and summer 1943, a few months before Engset's death in Octo-



1927

№ 8

ANNALEN DER PHYSIK

VIERTE FOLGE. BAND 82

Page from Annalen der Physik, *1927 [16]* ber the same year. The dissertation is written in German and is of 147 typewritten pages [14]. It is a heavily mathematical work, with the mathematics expressions mainly written in by hand. It takes special knowledge and probably a lot of time to evaluate the quality of the work. This is not done so far, as several physics professors, after a brief look, have indicated that it would take more time than they have available to come up with a serious evaluation.

However, professor Iver Brevik of NTNU took time off over a weekend to look into the matter, and he has permitted me to quote his comments after that limited survey:

> *"This is a very extensive work, almost 150 typewritten pages in all. The author focuses on the fundamental building blocks of the hydrogen atom, namely the proton and the electron, and tries to describe the physical properties of these 'Korpuskeln' and their interaction within the framework of the classical theory, and partly also the elder version of quantum mechanical theory.*
>
> *The work may roughly be divided into three different parts. The first part describes classical models of the proton and the electron, as particles that have finite extensions and are kept in mechanical equilibrium by means of mechanical stresses. This physical picture of the elementary particles appears outdated today, although to a certain extent it has become revived again in recent years in connection with the so-called Casimir effect (ie. the quantum mechanical theory of zero point oscillations). The second part of the work considers the electron in its orbit around the proton in the hydrogen atom (elliptic orbit), according to the first version of the quantum theory as constructed by Bohr, Sommerfeld and others around 1920. The last part treats the theory of electrically charged particles moving at high velocities, and one may consider this as a study of the scientific discipline which is now called relativistic electrodynamics. The special theory of relativity here plays an important role.*
>
> *On the whole, the present work seems to be very carefully done. It is clear that Engset has made a thorough study of the background material belonging to the classical theory, and partly also to the quantum theory in its elder version. It is quite striking how Engset goes a long way to test his theory with experiment, by carrying out detailed numerical calculations of the theoretical predictions.*
>
> *One may ask oneself: is this large work of scientific importance today, or is it mainly of historical interest? I would think that the second answer is the more correct. At the end of the 1930s the new version of quantum mechanics was developed, implying that the motion of the electron orbit in the hydrogen atom ought properly to be described in terms of wave theory, constructed in accordance with the Schrödinger equation. On the basis of this, it lies at hand to conclude that the physical picture of this work was essentially outdated already at the time it was written. However, the situation is actually more complicated: In fact, Engset published some years earlier several short papers in the well known journal* Annalen der Physik *(Leipzig), showing that he was well aware of Schrödinger's work. The first of these papers was even published almost immediately after Schrödinger's paper, in the same year (1926). Engset developed a wave equation which, however, is stated by him to contain some differences from Schrödinger's equation in the concrete situation discussed.*
>
> *There are some strong statements in Engset's treatise which can be criticised on the basis of present day's knowledge of physics. However, leaving such points aside, one cannot help being impressed by this major work made by a man in his retiring age. After all, this field of research lay outside Engset's main activity in his younger years. One may of course wonder about what could have been Engset's motivation for undertaking such a 'tour de force'. Was it to continue his earlier studies, aiming at further publications in* Annalen der Physik*? Or was the motivation to submit the work as a doctoral thesis? It seems very natural to suggest that the latter option is correct, although apparently we do not have written sources helping us to draw a more definite conclusion at this point."*

Brevik, by his browsing through the report, discovered references to the publications by Engset in *Annalen der Physik* in 1926 and 1927. A further search led to two publications [16, 17] in seven parts, altogether about 50 journal pages. Those articles put the 1943 report in a new light, as they indicate a serious interest and capability in fundamental physics already in the twenties, while Engset was still in his most active period of managing and expanding the national telecommunications network. It is interesting to note that the starting point of the first part of [15], dated June 1926, is a reference to Schrödinger's publication of his famous wave equation in the same journal earlier the same year(!).

Knowing Engset's seriousness in all matters, it is hard to believe that the dissertation is not also a serious piece of work. He is not known to have done any physical experiments, so his sources of knowledge in the field must have been confined to literature sources, but, as Brevik points out, he did detailed numerical calculations to check his results. The report, however impressive, will most probably be of historic interest today. The value of the 1926/27 publications in *Annalen der Physik* is not further assessed so far.

There is a straightforward explanation to the eight year long empty interval from 1927 to 1935 of Engset's active research in physics. In 1927 his wife became seriously ill and died later in the year after her husband had taken leave from his duties to tend to her. It meant a serious break in the private life of a man in his sixties. Then there was a turbulent time with political implications that led to the resignation of DG Nickelsen and the take-over of the top position by Engset, until he resigned at 70 in 1935.

Engset's two reports, the 1915 teletraffic report and the 1943 atomic physics report, are now safely deposited at Norsk Telemuseum, Oslo, where also other information, like contemporary newspaper clips from Engset's time, is available.

6 The private man

In his biographical notes [7] Joys recounts the following anecdote:

> *"I never met Engset myself. True enough I had read about Erlang and Engset as a young engineer while employed in the Standard corporation, first in Antwerp and then in Madrid. Since 'our' traffic theoretician at the time was M. Merker, who possessed the absolute truth in the best of all concerns, I remember that I pitied my slightly confused compatriot. The*

fact that Engset was Director General was unknown to me until the author and 'tramp' Erling Winsnes turned up in Madrid in 1932. He said that it would ease his personal, and thereby his friends', financial situation, if he were able to send to Dagbladet *an interview with Engset, who just then attended an international telegraph and radio congress in Madrid.*[4]

Neither Winsnes nor any of us other Norwegians knew what Engset looked like, but Winsnes found him. Winsnes explained that it happened in the way that he went to the congress building at the end of a session. People poured out, chatting and gesticulating. Finally, a lonely fellow with perplexed looks came out. Winsnes went straight over to him, saying: 'How do you do, Mr. Engset?', getting the answer: 'Just fine, and how are you?'."

As Joys' recount of Engset is probably the most complete one up till now, it is worth continuing the quote:

"Engset was undoubtedly introvert and absentminded at times, as his close colleagues could confirm. Sleepless he was also. Even as office head in the central branch he might turn up in the long distance operator hall in the middle of the night for inspection. He was not always welcome, as I understand.

Engset's grandchild (in fact step-grandchild)[5] *recounts that he was easily allured to take him to the movies, as he would fall asleep immediately.*

He seems to have been something of a pedant and very much a perfectionist. Among his colleagues Engset was said to have acquired enough knowledge for a law degree at the University, but the fear of not getting top grades held him back. The truth of this is of course doubtful, but it does say something about Engset's character.

When the editor of Telegrafbladet *(the newsletter of the male union)*[6] *asked for a picture for the newsletter on the occasion of Engset's appointment to Director General, Engset replied that he would have a photograph taken the same day. A suggestion of using an already available picture was rejected with a remark that he wanted his staff newsletter to have a picture of the Director General, not of the office manager.*

Engset was, however, a complex character. He was sober and objective as administrator and scientist, friendly and obliging among colleagues and at home. At the same time he was emotional and poetic. He gave elegant versified talks, once even in German at a major event with German participants.

He was deeply religious, and even wrote a hymn for his own funeral.

It is said that Engset and some of his colleagues once were chatting at random. One of them brought up the question of what they ought to bring out with them if a fire broke out in the building. Engset immediately declared that the most important things to save were his personal papers.

As I heard about those papers, I thought that something of interest might be found in the main files of the institution or elsewhere. It turned out that kept in the family there were papers containing an extensive dissertation about the theory of nuclear fission, by which Engset had occupied himself for years. He assumed already before World War II that nuclear power would be the future form of energy to the advantage of mankind. The subject engaged him very strongly. His dissertation is dated 1943, the year of his death, two years before Hiroshima, which he certainly could not have dreamt of.

With the rapid development of nuclear theory after Engset's time one must assume that his dissertation is obsolete. Maybe at his time it represented pioneer work, maybe not. Anyway, Engset had obviously found a new hobby to replace traffic theory.

It is said that Engset in his retirement had his daily walk to a place where there was good sand for writing with his stick. Most people have their hobbies. Some have sports, a few have mathematics as their pastime. Hobbies are as a rule a form of relaxation from daily toil. Its goal is limited and most often related to private life. Whether the hobby even happens to leave a mark in society is 'pure chance' – to stay within the terminology of traffic theory.

Sometimes the foremost of amateurs reach the top of professional level. This, however, presupposes a particular environment. Engset was such an amateur. The group around Telephone Director Johannsen was the environment."

So far L.A. Joys. It is fair to point out that the only teletraffic work by Engset available to Joys was the two-page article from 1918. That the 1915 report, which actually existed in the central files, was not found, is of course a pity, considering Joys' keen interest in Engset's teletraffic work.

The recount by Joys on Engset's life as a private man can be supplemented as an extension to the description of his background as given in the beginning of this article.

Tore Engset had a career of more than 13 years in Telegrafverket behind him when he married in 1897 at the age of 32. His wife was Marie Amalie Nord from Kristiania who was one

Tore Olaus Engset with his wife Marie Amalie

---

[4] *See previous mention of the 1932 Madrid congress.*

[5] *Author's remark.*

[6] *Author's remark.*

Photo: Norwegian Telecommunications Museum

Tore Engset (seated at the head of the table) with heads of department in 1930

of a family of 12 children. She was then a widow with two children, a boy, Fritz Gude, and a girl, Julie Kathrine (Lilli), after her first husband Anton Julius Bolivar Kreutz, a ship-owner and captain who died on a trip to Canada. She was 41 when she married Tore. They had no children, but Tore considered his stepchildren as his own. The boy went to America quite early, and the family lost track of him, in spite of many efforts to find him. Lilli was married to Adolf Fonahn, later professor at the University of Kristiania. They had one boy, Willi Urne, who always considered Tore Engset as his grandfather.

Willi Urne Fonahn (81) is probably the only person today with any closeness to Tore Engset, who still remembers him. He had and still has great admiration for his grandfather, whom he remembers as very hardworking, honest and at the same time an affectionate family man. He describes "an intelligent, endearing person and a wonderful grandfather". The boy Willi had once avoided to pay his fare on the tramway, and he was kind of proud of his achievement when he told his grandfather, who immediately gave him a coin and sent him back to pay and excuse himself.

His working day was often extended far into the night, and sometimes he had two bowls, one with hot and one with cold water to put his feet into in order to stay awake.

Engset was a shy person, in particular towards women in his youth. How he got the courage to propose to the nine year older widow seems a mystery, unless she helped him very much on the way. Anyway, it turned out to be a very happy marriage. In a letter to his niece Anna Saxegaard Gjerding, with whom Marie had had regular correspondence for years, Tore paints a very touching picture of his loving wife for 30 years after she died in 1927. "My office occupied me at all times, and in my scarce leisure time I jumped to studies and problems that stole her husband away from her. And when I broke down from overexertion or for other reasons was thrown on the sickbed, she completely gave herself to nursing me. Such was her love, sincere, rich, great, with no thought of herself. ... We always loved to be together. She was, as you know, so nice and cheerful and sociable. So youthful and brave". She also, in her private notes, expresses her sincere love for her husband. He, on the other hand, took leave from his work to stay with her when her illness became more severe.

Tore Engset's generosity is another quality which is clear from private correspondence between two of his nieces, talking about the Christmas presents received from their uncle in Kristiania in 1900. "What a joy there was in the home that evening. Such a joy was never seen in our crowd of children, my mother said." There were ten children. Later, during the war with food scarcity in the capital, when the retired widower Engset was living with his stepdaughter, the generosity was returned with farm products being sent from his relatives at Stranda, for which he expressed great gratitude.

When Engset was finally appointed to the top position after having applied for it on three occasions, he writes to Anna, in reply to her congratulations, about all the received attention: "... and all this has pleased me much, not least the kind lines from you, dean Saxegaard's daughter, who always maintained the correspondence with my beloved Marie! Oh, if only she and Elen had lived to see that their dear Tore received an open and honest recognition!" His sister Elen died only a short time before his appointment.

Tore Engset's poetic nature is demonstrated in his funeral hymn, but even more in an invited memorial article in the Christmas edition 1941 of the regional newspaper *Sunnmørsposten* from his home region. In this article he writes about a glorious Sunday in August 1883. He had just received a telegram granting his admittance to the telegraph school in Stavanger.

"I was jubilant to send a telegram of acceptance, and on my way home I saw the mountain Roaldshornet flooded with sunlight." He decided to climb to the top, as he had heard of the fascinating view, and he was well used to mountain hiking. "I was surprised and thrilled – yes, overwhelmed". He goes on to describe the view in poetic terms, naming more than fifteen well known mountain tops, then likewise the many islands lining the coast.

"The great contrasts in the environment of nature and, in spite of these, the wonderful harmony of the panorama under the glorious sun of the summer day had a fascinating effect on my sensitive soul: I lay down on my back on the highest point, looking towards the sky, and wept like a child".

From this he goes on to describe other places in Norway, where he had travelled extensively during his work, continuing with famous views of Germany, Switzerland and Italy, and the blue, blue Mediterranean, to end up with the Grand Canyon and Niagara falls. At Niagara Falls he claims to have spoken to the man who had jumped the Falls in a steel barrel, Bob Leach. Arms and legs had been crushed, but he was now all right. Worse was the fate of the woman who wanted to repeat the deed. She turned mad, proving that women should not always try to be as capable as men(!).

He reflects on the contrast between the extreme happiness that Sunday on the home mountain, when he had got admittance to the national telecom organisation, and his time on the top of that organisation, "... it was rather stormy, candidly spoken a rather disgusting weather, with little sun ...".

In an illustrated biographic dictionary [17] from 1916 Tore Engset, then department head, is presented with family relations and personal vita. The editor had put questions to all persons presented in the dictionary: 1) Which persons, institutions, events have had the greatest influence on your development and life? 2) Tell some little anecdote that characterises the environment of your adolescence and yourself.

Engset's replies were: 1) Dean Saxegaard of Stranda, with whom I went to confirmation in 1880. 2) I have never felt more proud and happy and more amply rewarded than when, as a little boy, I received my first pay of 2.00 kroner (20 pence) for eight days as a shepherd boy in the mountains of Sunnmøre.

Dean Saxegaard, by the way, happened to become Engset's brother-in-law 17 years after that confirmation.

The complexity of Tore Engset's character also implied a general scepticism towards humankind. Maybe he felt that his experience called for some carefulness. He is quoted to have said: Watch out for complaisant people!

The event of Engset's retirement from his position as Director General on his seventieth birthday was of course one of general public interest. After a broad newspaper presentation of the achievements of his career he was asked for an interview. He refused, saying: "It is not that I have done anything that should be concealed for the world. However, it pleases me more to disappear quietly." And the newspaper continues: "Quietly and without great gestures Engset has worked during his 52 years of service in Telegrafverket. He is, however, a man of insight, determination and the courage of his convictions, faithfully carrying out his work within his administration, where he has been of great influence in the inner circles. And thus even for the whole country".

One might reflect further on Tore Engset's dual life: That of a very demanding position, normally requiring the full mental and physical capacity of a very robust man, which he was definitely not; and that of very capable scientific work, carried out mainly during evenings and nights, and in two very separate areas. It brings to mind the duality of Robert Louis Stevenson's 'Dr. Jekyll and Mr. Hyde', though certainly without the sinister traits of the latter.

His achievements are the more admirable, as his health was generally not very good. He suffered from asthma and was prone to catching colds. He had a full year sick leave in 1901–1902, when he stayed with his wife for several months in a pension in Copenhagen. In private letters his wife expresses concern about his physical health as well as his nerves and his sleeplessness. Himself he refers to periods of overexertion. However, after his sick leave the situation improved, and he was capable of carrying the full burden of his responsibilities for many years to come.

Engset was made *Commander of the Second Order of Dannebrog* and *Knight of the Legion of Honour*.

After his retirement Engset moved to stay with the family of his step-daughter Lilli Fonahn, and it seems that they were on very good terms, as indicated in private letters and also according to recounts by his step-grandson. This was the time when he worked on his dissertation on atomic physics.

We now write 1998, 115 years after Tore Engset began his career as a boy of 18 and 55 years after his death. With the increased knowledge that we now have of this outstanding man it is appropriate to give him the honour he deserves by issuing the present publications. In addition to this, The Norwegian Society of Chartered Engineers (NIF) has decided to establish the *Engset Award* for young candidates in telecommunications. The award is supported by Telenor and some major telecom companies in Norway and will be presented at an annual conference arranged in co-operation by NIF and NTNU (Norwegian University of Science and Technology), the first time in January 1999.

Acknowledgement

In addition to the written sources of information, as given in the bibliography below, several people have contributed with bits and pieces to the mosaic of Tore Engset's life.

I was received with enthusiasm at Stranda by members of The Historical Society of Stranda (Stranda Sogelag), Harald Berge and Signy Mork (granddaughter of Anna Saxegaard Gjerding, as mentioned above), together with the mayor of Stranda, Anne Lise Lunde, and the editor of the local newspaper, Helge Søvik. They have furnished valuable information. Later information has been added by Lina and Ole Gjerde, who are writing a memorial article on Tore Engset for the 1998 Annual of Stranda Sogelag. They all admit to having practically no previous knowledge of the scientific side of Engset's career.

Willi Urne Fonahn and his daughter Wenche Fonahn were extremely helpful in adding to the information, and by transferring the atomic physics dissertation to Norsk Telemuseum, where Arve Nordsveen and Anne Solberg took care of the transfer, and supplied further information, in particular newspaper clippings.

Other sources of information that I have utilised are Statsarkivet (the National Archives) in Trondheim and Riksarkivet (the National Archives of Norway) in Oslo.

References

1 Engset, T. Die Wahrscheinlichkeitsrechnung zur Bestimmung der Wählerzahl in automatischen Fernsprechämtern. *Elektrotechnische Zeitschrift,* Heft 31, 1918.

2 Erlang, A K. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. In: Brockmeyer, Halstrøm, Jensen. *The Life and Works of A.K. Erlang.* Copenhagen, 1948. First published in Danish 1917.

3 Engset, T. *Om beregningen av vælgere i et automatisk telefonsystem.* Kristiania, 1915. Unpublished report.

4 Engset, T. On the calculation of switches in an automatic telephone system. Translation of [3] by A. Myskja. *Telektronikk,* 94, (2), 1998 (this issue).

5 Myskja, A. The Engset report of 1915 : summary and comments. *Telektronikk,* 94, (2), 1998 (this issue).

6 Rafto, T. *Telegrafverkets Historie, 1855 – 1955.*

7 Joys, L A. Til minne om T. Engset, Kristiania. *Verk og Virke,* 4, 1967.

8 Rubas, J. *Table of Engset Loss Formula.* Melbourne, Postmaster-General's Department Headquarters, 1969.

9 Jensen, E. On the Engset Loss Formula. *Telektronikk,* 88, (1), 93–95, 1992.

10 Engset, T. The Probability Calculation to Determine the Number of Switches in Automatic Telephone Exchanges. Translation of [1] by E. Jensen. *Telektronikk,* 88, (1), 90–93, 1992.

11 Iversen, Abild, Engset. *Indstilling fra Komiteen for telefoncentralstationsordningen i Kristiania.* Kristiania, 1913.

12 Toft, Utvik. *80 års nordisk telesamarbejde 1917 – 1997.* NTOB/NORDTEL. København, Fihl-Jensens Bogtrykkeri/Offset AS, 1997.

13 *Norsk Allkunnebok.* Fonna Forlag 1950 –.

14 Engset, T. *Die Wasserstoffkorpuskeln, ihre Strahlung und die fundamentalen physikalischen Grössen.* Unpublished report, 1943. (German)

15 Engset, T. Die Bahnen und die Lichtstrahlung der Wasserstoffelektronen (4 parts). *Annalen der Physik,* 80 (1926) 823; 81 (1926) 572; 82 (1927) 143 and 184.

16 Engset, T. Die Bahnen und die Lichtstrahlung der Wasserstoffelektronen. Ergänzende Betrachtungen über Bahnformen und Strahlungsfrequenzen (3 parts). *Annalen der Physik,* 82 (1927) 1017; 83 (1927) 903; 84 (1927) 880.

17 With, N. *Illustrert Biografisk Leksikon.* Kristiania, A/S With & Co.s Trykkeri, 1916.

Arne Myskja is Professor Emeritus at the Norwegian University of Science and Technology, where he has been employed since 1960. He was formerly with the Norwegian Telecommunications Administration, and has had teaching/research assignments in the US, Denmark, Sweden, Australia and France. His main areas have been switching systems and teletraffic theory. Most recently he has devoted his time to studies related to T. Engset.

e-mail: arne.myskja@item.ntnu.no