KUNGL TEKNISKA HÖGSKOLAN

An Accesspoint for Mobile Terminals

Support for Simulation and Evaluation

and MEDIA Evaluation Reports D1.3 and D1.5

Andreas Schmidt

aschmidt@it.kth.se

Computer Communication Systems Laboratory
Department of Teleinformatics
KTH Stockholm, Sweden

KUNGL TEKNISKA HÖGSKOLAN

Technische Universität Graz

An Accesspoint for Mobile Terminals

Support for Simulation and Evaluation

and MEDIA Evaluation Reports D1.3 and D1.5

Master of Science Thesis

Andreas Schmidt

Summer 1998

aschmidt@it.kth.se
Computer Communication Systems Laboratory
Department of Teleinformatics
KTH Stockholm, Sweden

aschmidt@iaik.tu-graz.ac.at
Institute for Applied Information Processing
and Communications – IAIK
TU Graz, Austria

Abstract

This *Master of Science Thesis* evaluates the system architecture of a network for mobile communication.

The work is carried out within the framework of the European Union's ESPRIT Frame IV long term research project MEDIA – Strategic Research in Network Systems for Multimedia Mobile Computing.

The topic of research is how to efficiently support terminal mobility in a packet switched network environment, and selecting or providing the technologies that fit best to accomplish that goal and combining them to a system is the aim of the MEDIA Project.

The ongoing standardization process has already defined the basic protocols necessary to support terminal mobility. Based on the network architecture design efforts presented earlier by the project, an elaborated model will be presented that takes the mobility specific messages and additional messages due to the distributed nature of our system as well as statistical user data traffic into account to simulate the entire system on the network level.

Full advantage of state-of-the-art simulation tools is taken to verify the functionality of that model on this level. Simulation is also used to provide testvector data for subsequent detail-levels in this iterative design process.

As a result of these similations, we are in a position to answer questions about the utilization of channels, message delays, buffer space usage, energy consumption, and scalability of our network architecture.

As a result the network topology is enhanced and finalized, meeting the economical design goal of a minimum price for the access point – the network node that has to be available in large quantities in a picocellular environment as ours.

All our results indicate the technical feasibility and the performance gain of our resulting system architecture.

Acknowledgments

The work for this Master of Science thesis has been carried out during the period February till September 1998 at the Computer Communication Systems Laboratory (CCS-Lab) at the Department of Teleinformatics (IT) at the Royal Institute of Technology (KTH) in Stockholm, Sweden.

This thesis is submitted to the Institute for Applied Information Processing and Communications (IAIK) at the University of Technology, Graz, Austria, as *Diplomarbeit*, in partial fulfillment of the requirement for the degree *Dipl.-Ing.* (corresponds to *master*).

I have had the pleasure to work with Professor Gerald Q. "Chip" Maguire Jr. as my thesis supervisor at KTH, to whom I hereby wish to express my deep respect and gratitude. Moreover I would like to thank my supervisor at the IAIK Graz, Prof. Reinhard Posch.

This work has been accomplished in the framework of, and funded by the European Union's ESPRIT Frame IV long term research project "21929 – MEDIA, Strategic Research in Network Systems for Multimedia Mobile Computing".

Furthermore the generous financial support I have received from the SOKRATES/ERASMUS Student Exchange Agency is thankfully acknowledged. I also want to thank KTH's External Relations Office, especially Ann-Charlotte Kohut and Ingeborg Löfgren, for their dedicated effort to make my stay in Stockholm pleasant and successful.

I highly think of IAIK's Dipl.-Ing. Dr. Karl Christian Posch as he was the one paving the way for my semester abroad and, as a reviewer of my thesis, provided me with comments and feedback which I am holding in high esteem.

To my colleagues and the staff at the Teleinformatics department I am grateful for their cooperation and assistance.

I want to thank Rita Johnsson, for carrying out all the administrative work smoothly as I have not seen it anywhere else before. I furthermore appreciate Jon-Olov Vatn's help – he spent several hours discussing economical aspects with me.

Last but not least several big 'Thank you's shall go to my friends in Stockholm for making my stay besides my study-tasks, more pleasant, eventful, exciting, and also cosy – just more perfect than I could have possibly imagined before.

Stockholm, September 1998
Andreas Schmidt

Table Of Contents

Table Of Contents

Table Of Contents

Objectives and Outline of the Thesis

The primary objective of this thesis is to evaluate a proposed communication network architecture by simulation and elaborate on protocol- and topologic enhancements.

The secondary objective of the present thesis is to introduce the background information considered during the preparation of this thesis. With a relatively long introductory part in the chapters 1.1. – 1.7. this has been accomplished.

Based on Juntong Liu's thesis work and his suggested model of the network, the model will be presented and explained in detail in chapter 2.2. and 2.7.

We will estimate the additional backbone-network load caused only by the control traffic in chapter 2.3. and 2.4., based on extended models and extensive simulation within the COMNET III environment. This is part of the D1.3 deliverable to the MEDIA Project.
From the MEDIA Project specification:

D1.3 Evaluation report. Derive a network architectural model based on the above investigations [D1.2 First version of the base station functional specification], supported by extensive simulation results, and document the network architecture strategies and guidelines that can be used to develop the next generation picocellular infrastructure. The deliverables would be a network architecture model.

An investigation of the scalability of our design is carried out in chapter 2.5. and the finalized topology is presented in chapter 2.6.

Guidelines for the crucial task of partitioning the functionality between access point and access point server will be presented in chapter 3. This is part of the D1.5 deliverable to the MEDIA Project.
From the MEDIA Project specification:

D1.5 Second version of the basestation functional specification. Investigate how the control and management functionality should be partitioned between the basestations and the computational servers attached to the broadband network, with consideration to the network architecture strategy being studied in the first part of the project plan. The deliverables would be a strategy for partitioning control and management functionality between basestations and broadband network.

Chapter 4 briefly discusses the task of providing testvector data for other models. Especially the representation of "traffic" suitable to verify the functionality of the IEEE 802.11 MAC layer as SDL-model and an 802.3 (Ethernet) model, both developed at KTH's ESD-Lab, is of interest to this project.

Chapter 5 points out future work and with Chapter 6 "Conclusions" this thesis ends.

1. Introductory Chapters

By the time of writing this thesis the necessary networking-, radiocommunication-, chip- and system-design technology is available, for providing *mobility* to computer terminals. Thus giving Internet users with their *mobile terminals* the freedom to seamlessly roam between domains and maintain a stable and reliable connection to the network.

Selecting or providing the technologies that fit best to accomplish that goal and combining them to a system is the purpose of the MEDIA Project.

To provide connectivity to an area, the approach of the *picocellular network* has been chosen. A *picocell* is considered to be an area of 20m radius, comparatively tiny to existing systems, as GSM. Such a cell is served by an transceiver with intelligent network access, known as an *access point*, and logically a lot of such access points will be needed in order to cover a whole city.

Thus, from an economical point of view, access points for a picocellular system have to be foremost *cheap* in order to be successful on this highly competitive market.

The advanced technologies of radio transmission have a substantial impact on the system design, affecting the lower (physical) layers up to the higher (transport) levels.

The physical layer e.g., and especially the radio, is critical when designing a system that is expected to be economically successful. The traditional approach is expensive, but soft-radio technology is available and can be integrated on-chip as well as most of the physical's of the wired network part, and thereby keeping the mass production cheap.

A single chip, integrating the functionality of the *access point* is the declared goal[1].

The process of protocol standardizing has already brought up the support for mobility based on Internet technology. However, the implementation thereof is behind and there is no mobility support at all for most operating system platforms so far.

Of particular interest are the demands on the underlying communication network that connects the access points and other nodes of the system. The design of an network architecture and the partitioning of functionality among the network nodes are crucial.

During the design process special care must be taken to ensure the proper functionality from on early stages, and to be aware of all effects of design decisions.

Network simulation not only ensures this two general demands, but furthermore provides valuable data to subsequent and lower detail levels of the system design enabling simulation on different levels, consistently linked by automated tools.

[1] Any implementation however lies outside of the scope of the MEDIA project.

This first section with its six subsections listed below, gives an introduction, which is essential for understanding the material presented in chapters 2–6 thereafter. It also forms the background and basis of this thesis and serves as reference for future work.

- **Teleeconomic aspects**
  An economical motivation for a technical project like this from a provider's point of view is presented in the following chapter. Due to its similarity to this project, Internet telephony has been chosen as an example to discuss cost, pricing and policy consideration.

- **The *new* Internet protocol IPv6**
  Every future implementation for the Internet will be based on this protocol – the successor of the IP version four (IPv4) in use hitherto. The approach the designers of IPv6 have taken in particular to the node mobility topic is studied in the next chapter.

- **The mobility protocol for the Internet**
  Besides a general introduction to mobility, the different approaches taken so far to handle node mobility on the internet are presented. In particular Mobile IP and its concept is discussed in detail in this third chapter. The specific terms used in the following chapters are first introduced here.

- **The GSM system as an example**
  Undoubtedly *the* large scale system providing mobility for the user is the GSM mobile telephony system in all its variants. Following the history of design decisions of this example and understanding its strength and weaknesses was the preparatory work for this project and still provides valuable input to our system design process.

- **Our proposed network architecture**
  This is the immediate basis for the presented thesis – the work performed so far in the MEDIA Project group, summarized from reports and preceding thesis work. Here all concepts which are important to our proposed network architecture and terms going along with it are presented.

- **The tool we use for network simulation**
  Comnet III is the simulation tool used throughout this thesis. A short overview to the system is given in this last introductory chapter and together with the description of the model in chapter 2.7. it serves as a quick start manual.

1.1. Teleeconomic Aspects

This chapter discusses communications policy models with a special focus on Internet telephony. Furthermore pricing and cost issues arising from such services are addressed, based on [MCKN98]. I present this material so that the reader can understand the economical context of my work.

This chapter concludes that an *Open Communications Policy framework*, as described in [NMKS97], is consistent with the technology and economics of the Internet, which relies heavily on *ad hoc, corporatist, intermediary organizations* for standards-setting and other governance functions. Both new pricing strategies and a supportive policy framework are needed for Internet telephony and other services to recover costs and to integrate successfully the Internet and telecommunications industries for the benefit of customers and suppliers.

It would be desirable if government policymakers could study and understand these issues and develop appropriate policies without introducing economic and technical distortions into the rapidly developing Internet market.

1.1.1. From Internet Telephony to IP Telephony

Initially, Internet telephony has been developed to provide interactive voice communications over the existing IP-based Internet. In practice only a personal computer with compatible application software running on both the originating and receiving computers is needed. Both computers need to be connected to the Internet (for example via an ISP) and equipped with a sound card, a microphone, and speakers. As multimedia PCs come equipped with just such hardware, and even notebook computers are so equipped — essentially all users have access to the necessary technology.

This form of connectivity was the primary focus for the first stage of development of Internet telephony technology. The extension of connectivity from a single network across the Internet was the advance which made Internet telephony possible.

The development of voice communication capabilities, using the Internet Protocol (IP) over the Internet, has progressed to where these capabilities may be viewed more accurately as *IP Telephony* rather than as Internet telephony. IP telephony is now becoming widespread. The limitations of the circuit-switching techniques of conventional telephony may now be overcome through integration with IP technologies.

The areas which IP telephony addressed include: delivery of incoming calls from the Internet to existing call centers, provision of virtual office capabilities for at-home workers (telecommuters), provision of mobile desktop capabilities for traveling employees, and multimedia-enabling of enterprise WANs.

A key element in making effective use of the emerging IP telephony technology is the development and use of *gateways* that bridge the IP environment of the Internet and the circuit-switched environment of the public telecommunications network. Gateways thus enable the interoperability between the public telecommunications network and the Internet. It also permits telephone to telephone calls to be carried by IP networks, as well as calls between telephone and computer.

The public Internet is not a controlled network environment. The IP technology currently employed on the Internet uses non-deterministic switching (transport of datagrams), which in version 4 (IPv4) offers little potential for approximating real-time voice connectivity in such an environment, except through overprovisioning of bandwidth.

As a consequence, Internet telephony over the public Internet is not, as yet, a significant direct competitor in the field of voice telephony. However, using IP networks, either conceived as Intranets or Virtual Private Networks, to provide voice services across a well-managed, over-provisioned, best-effort network, is technically feasible today. In fact, new operators are coming into existance which have built their network using IP as its basis.

Typically, the end user or customer may not know or care that her voice was carried across a packet network rather than across a circuit-switched network along a portion of its transmission path.

1.1.2. Policies and Regulations

The Internet is a network of networks using public protocols to share resources and economic benefits. Having arisen outside the realm of broadcast, telecommunications, or print media, the Internet has effectively resisted classification and regulation within the models of other media and services.

National regulators, political authorities, international agencies, and multinational firms from a variety of industries have all sought to control the Internet — All have failed. Seemingly, the Internet is a new species of technical-economic-political activity, which is self-organizing and self-governing. Truly existing only in the minds of its users due to their choice to receive and send messages, using its language and methods.

The Internet's seemingly mysterious, self-governing and fluidly reorganizing capabilities are not unique, however. Other entities, such as expert commissions, standards organizations and public-private research consortia, exhibit similar qualities and possess comparable strengths and weaknesses. A difference may be, that they rarely have a common simple basis for communication.

The international implications of the communications policy transformation that is attributable to the growth of the Internet is profound. The notion that the Internet is intrinsically borderless and outside the control of national governments is false. Especially with the growth of new naming and numbering schemes for intellectual property protection, taxation, and directory services, as well as provisions for varying qualities of service, there are several levels of influence that governments may exert over Internet users.

However, if their control becomes excessive the internet will simply route around them. As Internet pioneer John Gilmore famously put it, *"The Net interprets censorship as damage and routes around it."* Thus a country could become a third world internet entity if their restrictions become greater — unless *all* other countries acted the same way.

Communications policy for information infrastructures that ignores the lessons of the Internet model are unlikely to succeed. Internet technology was created by agile government technology policy programs, while the Internet, as a communications and interaction environment, is a key infrastructure for *virtual governance* [NMKS95].

The Internet model is used increasingly by standards organizations, and by distributed organizations (often, communities of interest). The consensual nature of most interactions on the Internet is important in facilitating *ad hoc, corporatist* governance of the information infrastructure.

The principles are compatible with and supportive of an information or networked society. The features of *ad hoc, corporatist organizations* allow them to process, analyze, and sort an exploding information base which is beyond the ability of any individual, firm, or government body to understand in all its facets.

They operate in a sea of public discourse and debate, and rely increasingly on the Internet as a communications backbone. The International Engineering Task Force (IETF) may be conceptualized as an ad hoc, corporatist organization whose mission is to support and integrate the results of pre-competitive, generic research and development of Internet protocols.

The principles of *ad hoc corporatism* are:
· Create impermanent, publicly sanctioned organizations to develop and implement public policy;
· Cross institutional boundaries; and
· Employ interdisciplinary teams to analyze problems, identify opportunities, and assess outcomes.

An open communications policy combines the diversity and dynamism of the American cultural values with a coordination and planning mechanism that serves to strengthen industry and improve government performance. Communications policies should take cognizance of the attributes of *ad hoc, corporatist* policy processes to improve flexibility, accelerate decision-making, reduce risks, and share costs.

Any technology and industrial policy, whether for communications markets or other sectors, that does not possess these attributes is ultimately doomed to failure.

The U.S. has benefited from years of accumulated experience with the Internet. New industries which conduct business on the Internet are beginning to emerge. For example, electronic publishing is providing ample information at low cost and in digital format. The cost of information lies *not only* in the *creation* of content, but in the *storage* and efficient *delivery*. In essence, the cost of paper, printing, trucks, warehouses, and other physical distribution mechanisms, plus the cost of personnel. The Internet demonstrates that electronic networks can reduce reproduction and delivery costs by orders of magnitude.

An Internet-centric model of policy development, i.e., one that relies on voluntary cooperation of autonomous agents to achieve goals defined through interaction, is admittedly imperfect. However, the alternatives, fetishizing or embracing proprietary models of network design, are fatally flawed. Communications policymakers must recognize the critical features of the global information economy and fashion policy instruments suited to networked firms and the principles of ad hoc, corporatist policymaking.

1.1.3. Pricing of Internet Telephony

Costs for the provision of Internet telephony services are not only concentrated in the modem banks and transport services of the ISP, but in addition to these technical items, items such as customer service, sales, and marketing represent a substantial portion of an ISP's costs of providing Internet telephony.

The bottom line for an ISP is that revenues will increase slightly, while costs will almost double with only a moderate use of Internet telephony (see also [MKBL98]). Hence, Internet Telephony Service Providers (ITSPs) need to consider how to minimize the cost impact of Internet telephony and/or how to gain additional revenue to operate at profitable levels.

For ISPs to remain in business, they eventually will need to recover these increased costs through tiered pricing, a higher flat price, or other mechanisms such as advertising or transaction fees or they have to figure out how to lower their cost. The Resource ReSerVation Protocol (RSVP) developed by the Internet Engineering Task Force (IETF) could be used as one mechanism for implementation of usage-sensitive pricing to recover those costs.

But usage-sensitive pricing will not be an option until protocols and systems that monitor the use of Internet telephony are developed and deployed.

When developing pricing schemes, service providers will have to look *beyond* the Internet telephony service and consider how to price integrated Internet services, *one* of which is Internet telephony. Additionally, an integrated regulatory framework will be required to permit the provision of such integrated services.

1.1.4. Costs for the ISPs

The Internet telephony scenario described below represents a case in which moderate use of Internet telephony creates a large cost increase to an ISP.

For comparative purposes, one might wonder how other changing factors would affect an ISP's costs. One such factor is the Federal Communications Commission's (FCC) recent revamping of access charge rules. Under the new rules, ISPs will see an increase in cost of their analog dial-in lines. Two principal changes cause this increase: an increased Subscriber Line Charge (SLC) and the Presubscribed Interexchange Carrier Charge (PICC).

The SLC rose from a cap of $5.60 per-line, per-month to $9.00 on January 1, 1998 (although few local exchange carriers (LECs) will be able to charge as high as the cap, the average has been calculated to be $7.61).

The PICC increases from $0.53 to $2.75 per-line, per-month. Using the average charges, the impact on ISPs (or any multi-line business) will be a $4.23 per-month increase for each analog line.[1]

Inserting these updated costs into the ISP cost model, previously developed by the Internet Telephony Consortium, yields an increase for the analog dial-in subscribers' cost for both the baseline and Internet telephony scenarios. The following table shows the providers cost in US$ per subscriber and indicates the throughout higher cost and the cost increase (in brackets) for transport of IP telephony after the access reform.

[1] ($7.61 - $5.60) + ($2.75 - $0.53) = $4.23

	Baseline	Baseline with Access Reform	IP Telephony	IP Telephony with Access Reform
Capital Equipment:	$2.70	$2.70	$3.90	$3.90
Transport:	$2.98	$3.44 (+0.46)	$7.72	$8.37 (+0.65)
Customer Service:	$7.50	$7.50	$10.80	$10.80
Operations:	$3.07	$3.07	$3.26	$3.26
Other Expenses:	$6.52	$6.57 (+0.05)	$7.73	$7.78 (+0.05)
Total:	$22.77	$23.27	$33.42	$34.12

As shown in the table above, the ITSP cost model, developed by the Internet Telephony Consortium (ITC), of Internet telephony service providers places the ITSP's costs into five categories:

- Capital Equipment — the hardware and software of the network,
- Transport — the leased-lines of the network and interconnection costs,
- Customer Service — staff and facilities for supporting the customers,
- Operations — billing, equipment and facilities maintenance, and operations personnel, and
- Other expenses — sales, marketing, general and administrative.

[MKBL98] shows that even a moderate use of Internet telephony causes total ISP costs to almost double (an increase of 42%) while revenues increase only slightly, by 19%.

Increasing from 24% to 28% of total costs, transport costs become the largest cost category in the Internet telephony scenario. The implication for ISPs based on this result is that an ISP that operates its network most efficiently will have a competitive advantage over other ISPs if the Internet telephony scenario takes place. Such efficiencies could come from scale economies, facilities-based networks or network optimization techniques. However, if one believes that the market for transport is already efficient and that transport is essentially a commodity, then there would be fewer opportunities for competitive advantage resulting from owning a network. Even so, network optimization techniques would prove advantageous whether or not the ISP owns or leases its network.

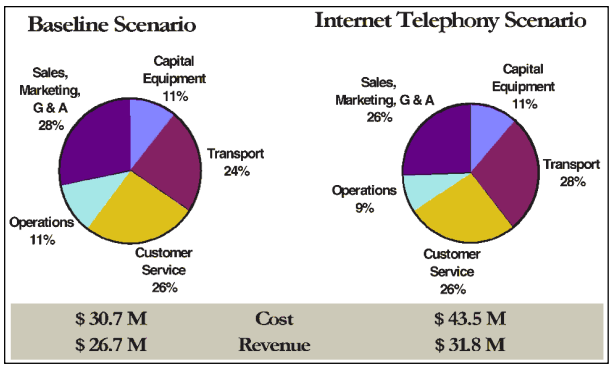


Figure 1.1.1. Comparative Cost Results

While no cost increase is advantageous to ISPs, the recent FCC actions should be considered much less threatening than the potential impact of Internet telephony, as of the broader impact of, for example, the imposition of universal service charges on ISPs in addition to on the lines they purchase from Local Exchange Carriers (LECs).

A principal conclusion that one reaches based on these cost results is that ISPs need either to *prevent* widespread use of Internet telephony, or to *change* the current pricing structure of Internet access services in order to recover the increased costs. What other actions the FCC may take with regard to Internet telephony are unclear, though a ruling with regard to the ACTA petition of 1996 (see [KATS97]) is said to be imminent.

1.1.5. Policymaking in the European Union

As we have noted above, perhaps the greatest challenge for Internet telephony is how it will be treated by governments. The Internet is indeed growing in importance in the United States, and therefore has been focused on at the highest levels of the U.S. government—more than in most other nations.

However, it would be a mistake to ignore regulatory dilemmas and proposed approaches arising elsewhere. In particular, the European Commission's approach to determine policy for Internet telephony merits attention, because of the obvious impact such policies may have in enabling or inhibiting the continued growth of a worldwide market for advanced Internet services.

Heterogeneity has been a key characteristic of the Internet from its beginning. The question of how much heterogeneity in Internet policy is tolerable for various classes of service will soon be answered in practice by policymakers and Internet users. In 1997 the European Commission has established several criteria that Internet telephony must meet before it will be subject to regulation:

- Such communications are the subject of a commercial offer;
- Such communications are provided for the public;
- Such communications are to and from the public switched network termination points on a fixed telephony network; and
- Such communications involve direct transport and switching of speech in real-time.

Based on these criteria, Internet telephony is not as yet considered voice telephony because Internet telephony does not meet the (last) criterion of "real-time" communication, due to the current, high levels of delay experienced by Internet telephony users on the public Internet. Hence, Internet telephony services in Europe are not subject to regulation at this time. How long this will last is open to speculation.

1.1.6. Conclusions

Internet telephony service providers confront a variety of challenges. The costs, the technologies, the pricing models, and the policy environment for Internet telephony are all unsettled and in a state of rapid evolution.

To date, a relatively hands-off policy approach has been taken by the FCC and the White House in the U.S. A similar policy position has been taken by the European Union. In spite of misguided efforts in some countries to ban Internet telephony, the expert's opinion is, that the real challenge is how to align the costs, technologies, prices, and policies to enable a rich new class of integrated and differentiated Internet services to flourish, subsequently bringing substantial benefits to consumers. Premature regulation of Internet telephony would be an ineffective and inappropriate response to this developing technology and market.

An Open Communications Policy model, supported by *ad hoc, corporatist* technology and policy intermediaries, is best suited to this task.

The lesson we learn for our project is clearly to keep the cost of transport (i.e., interconnection) as low as possible. In contrast to the GSM system (introduced in chapter 1.4.) we will eliminate the cost for an overlaid operation and maintenance network, and (comparing GSM's BTS to our AP), we will greatly reduce the cost of the access points.

1.2. Introduction to IPv6

The new Internet Protocol IPv6 is based on a rather simple philosophy: The Internet could not have been so successful in the past years if IPv4 had contained any major flaw. IPv4 was a good design, and IPv6 was designed to keep most of its characteristics. In fact, it could have been sufficient to simply increase the size of addresses and to keep everything else unchanged. However, ten years of experience brought lessons and IPv6 is built on this additional knowledge. It is not a simple derivative of IPv4, but a definitive improvement.

This chapter gives a summary of IPv6's design, based on [HUIT98] with particular attention to addressing, auto configuration, and mobility support features of IPv6.

1.2.1. Similarities and Differences in the IPv4 and IPv6 Headers

The IPv4 header is based on the state of the art in 1975. Twenty-three years later the new design has the following appearance: The IPv6 header is composed of 64 bits of fields followed by two 128 bit addresses summing up to a total header length of 40 bytes .
The initial 64 bit are partitioned as follows:

- 4 bit Version
- 8 bit Class
- 20 bit Flow Label
- 16 bit Length of Payload
- 8 bit Type of Next Header
- 8 bit Hop Limit

The new header is in fact much simpler than that of the classic IPv4. It has only six fixed fields and two addresses, while the old version has 10 fields, two addresses, and some options.

Only the first field retained its function: If the *Version* code is '4' it is a IPv4 packet, a version code of '6' marks an IPv6 packet. Nevertheless, packet demultiplexing is accomplished at the media layer, using distinct content types, not at the network layer (as initially intended).

Six IPv4 fields were dropped: *Header Length, Type of Service, Identification, Flags, Fragment Offset,* and the *Header Checksum*.

The definition of three fields changed: *Length*, *Protocol*, and *Time to Live (TTL)*. The *Length* field still limits the packet size to 64kB, however the *jumbogram* option allows for much larger datagrams. The renaming of "Time to Live" to "Hop Limit" takes into account the practice of decrementing this field once per hop and not according to the seconds of delay, as specified for IPv4.

Three fields are new: *Class, Next Header,* and *Flow Label*. The fields *Class* and *Flow Label* are mostly designed to facilitate the handling of real-time traffic. *Class* describes the priority and *Flow Label* is used to distinguish packets which "require the same treatment". The *Next Header* field is used to structure a daisy chain of an IPv6- and one or more extension-headers, such as *hop-by-hop options-, routing-, fragment-, authentication-, encrypted security payload-,* and *destination options-headers*. Chained in that recommended order, followed by the upper layer header (for example, TCP or UDP).

To summarize, there are tree major simplifications:

- The headers have a fixed format
- No header checksum
- Removal of the hop-by-hop segmentation procedures

IPv4 provides a segmentation procedure so that senders could send large packets without worrying about the capacities of relays. The rule with IPv6 is that hosts should learn about the maximum acceptable segment size through a procedure known as *Path MTU Discovery*. A host sends a packet of the desired size (usually as large as the local interface allows) and if this is larger than one hop's MTU, an ICMP message (type 2, "*Packet to Big*") containing this link's MTU as a parameter is issued and this allows the host to adapt its MTU setting.

Future extensions of the functionality of IPv6 on the header level can be added in two ways: By allocating additional header-type numbers (which are a scarce resource) and by using the *Destination Options* header.

The change in address length has some impact on the upper layers. Transport protocols, such as TCP or UDP, attach a checksum to their packets. This checksum is computed over an imaginary packet, which is the concatenation of the actually transmitted packet and a pseudoheader. The pseudoheader contains the source- and destination-address and the payload length. The definition of the pseudoheader is an integral part of the TCP and UDP specification and, in fact, of any upper-layer protocol that includes addresses in its checksum computation. But a new version of the TCP and UDP pseudoheaders is provided in the IPv6 specification. It includes the source and destination address, the next header type and the payload length. Since IPv6 does not contain any header checksum the use of checksums at the upper layer is mandatory, even in case of UDP.

IPv6 has an impact on the DNS. An address is normally obtained through a DNS lookup. The DNS database stores various resource records for each Internet domain. These resource records are identified by a type. For example, an IPv4 address is stored in a record of type A. Each A record contains one 32 bit address. A new resource record type has been defined for IPv6 that contains the 128 bit address and is therefore named AAAA. Applications programs should normally manipulate domain names rather than numeric addresses.

1.2.2. Routing and Addressing in IPv6

The most salient feature of IPv6 is undoubtedly the enlarged address format. Going from 32 to 128 not only guarantees that it will be possible to number an abundance of hosts, but it also provides room to insert many more degrees of hierachy than the basic three layers of network, subnet, and host offered by IPv4.

Since IPv6 is based on the same architecture principles as the classic Internet Protocol IPv4, one could very well expect IPv6 addresses to be larger versions of the IPv4 addresses, and one would basically be right. However, one big difference is that IPv6 routinely allows each interface to be identified by several addresses to facilitate routing or management.

IPv6 addresses belong to one of three categories:

- Unicast (also known as point-to-point) addresses identify exactly one interface
- Multicast addresses define a group of stations
- Anycast identifies a group too, but the packet is only delivered to the *nearest* member

IPv6 128 bit addresses are written as 8 colon separated 4 digit hex (16 bit) numbers. With possible abbreviations such as omitting leading zeros and denoting one further sequence of zeros by a double colon.

1.2.2.1. Initial Address Assignment

Implementations are not supposed to have full knowledge of the various address allocation prefixes and formats. In most cases, it is perfectly legitimate for a host to treat addresses as opaque string of 128 bits. Similarly, entries in router tables will be simple prefixes between 1 and 128 bits in length. The only exception concerns special addresses. Hosts and routers must indeed recognize multicast addresses, which cannot be processed the same way as unicast or anycast addresses.

The initial address prefix allocation within IPv6:

- 010... Aggregatable Global Unicast Addresses
- 100... Reserved for Geographic based Unicast Addresses
- 0000 001... Reserved for NSAP Allocation
- 0000 010... Reserved for IPX Allocation
- 1111 1111... Multicast Addresses
- 1111 1110 10... Link Local Use Addresses
- 1111 1110 11... Site Local Use Addresses

One of the most significant shares is that of the global unicast addresses, but that covers only 12.5% of the total space. All in all, more than 70% of the space remains unassigned, which should provide ample opportunity for trying new assignments in the future.

1.2.2.2. The Aggregatable Global Unicast Address Format

Aggregatable global unicast addresses are composed of the following 5 fields of fixed length:

- 3 bit Prefix “010”
- 13 bit Top Level Aggregator (TLA)
- 32 bit Next Level Aggregator (NLA)
- 16 bit Site Local Aggregator (SLA)
- 64 bit Interface ID

The IANA is expected to allocate ranges of TLAs to the various continental registries, which could possibly further delegate the allocation of sub-ranges to national or regional registries. The TLA does not necessarily identify a provider.

The NLA field will be structured by long-haul providers and exchanges, allocating, for example, the N top-most bits to identify a second-tier provider and the remaining 32-N bits to identify a subscriber to that provider.

The SLA identifier is normally allocated to a link within a site, and the site itself is an atom in the addressing hierachy. In the case of renumbering a site, only the TLA and NLA need to be changed.

The Interface ID, after all, uniquely identifies a particular interface on that link. It is furthermore assured, that this ID is globally unique, to make renumbering easier.

1.2.2.3. Multicasting and Anycasting

Multicasting capabilities were formally added to IPv4 in 1988 with the definition of class D addresses and the Internet Group Management Protocol (IGMP). The deployment of these capabilities was sped up by the arrival of the MBONE in 1992, but this deployment is still very far from being universal. IPv4's anycasting capabilities were even less advanced and their incorporation into IPv6 will offer a lot of flexibility to network managers.

IPv6 is defined such that multicast addresses are handled by each router and IPv4's IGMP is incorporated in IPv6's ICMP by a set of additional messages.

A multicast address is composed out of the following 4 fields of fixed length:

- 8 bit Prefix "1111 1111"
- 4 bit Flags, only one bit is defined as *Transient* flag (set for temporary addresses)
- 4 bit limit the scope (node local, link local, site local, organization local, global)
- 112 bit Group ID

A range of Group IDs will be registered by the IANA, another portion will be used by the network discovery procedures and several ranges of addresses corresponding to the various scopes of transient addresses will be assigned.

The principle of anycasting on the other hand is simple: Instead of sending one packet to a specific server, one sends the packet to a generic address that will be recognized by all the servers of a given type, and one trusts the routing system to deliver the packet to the nearest of these servers. One could use anycasting to find out e.g., the nearest DNS or timeserver.

Hosts treat anycast addresses the same way as unicast addresses. The load is on the routing system, which has to maintain one route for each anycast address that is active in a given site.

1.2.3. Autoconfiguration

Through autoconfiguration the host will automatically discover and register the parameters necessary to connect to the Internet. This problem is sometimes referred to as "The dentists office and the thousand computers on the dock". Networking specialists imagine that dentists are rich enough to buy several computers, but that they have been educated in dentistry, not computer networks, so they can do little else other than take the machine out of the box, plug in various connectors, switch it on, and expect it to work. A requirement of IPv6 is that this should indeed be sufficient, even if the dentist is not connected to the Internet and even if there is no router in the office's network.

But just setting up a machine once and forever is not sufficient. IPv6 is capable to support *dynamical address change* (to facilitate renumbering when the dentist connects via an ISP or switches ISPs) and *multiple addresses* for an interface. The address configuration is now an integral part of IPv6 and comes in both a *stateless mode* and a *stateful mode*.

1.2.3.1. Link Local Addresses

Such an unique address can be formed by concatenating the
- 10 bit Prefix "1111 1110 10",
- 54 bit zeros,
- 24 bit, the MSBs from the ethernet address,
- 16 bit fixed as "FF FE", and
- 24 bit, the LSBs from the ethernet address.

An address formed as described is unique, since the 48 bit ethernet address can be considered worldwide unique. This link local address can only be used on the local link, and it may well solve the dentist's problem, but it will not be sufficient for organizing a large network.

1.2.3.2. Stateless and Stateful Autoconfiguration of temporary Addresses

An IPv6 node starts initializing its *stateless configuration* behaviour by joining the multicast group "All Nodes" (address FF02::1). Then it sends a *solicitation message* (ICMP, type 133) to the multicast group "All Routers" (address FF02::2) containing its link layer address in the message's option field.

Periodically or on receipt of a solicitation message, the router responds with a *router advertisement message* (ICMP, type 134), sent to "All Nodes" or directly towards the link layer address obtained from the solicitation message respectively. From that message the node learns about the router's and therefore it's own prefix.

Stateful configuration on the other hand is the IPv6 version of IPv4's Dynamic Host Configuration Protocol (DHCP). It has the advantage that it does not demand any such server to accomplish the address configuration. But to be operational, a station must also discover the location of other useful servers, notably a DNS. Stateful configuration is capable of that and can even provide further options to the clients.

Internet users usually own their IPv4 address, which is no longer the case with IPv6 addresses. To be able to aggregate network numbers one has to accept the dependency between address and network topology. Any change of the latter (i.e., by changing provider, reorganisation of the backbone, ...) might imply a change of the address.

Thus, addresses obtained through either stateless or stateful configuration will have a limited lifetime. In stateful configuration, the lifetime will be indicated by the address server and in stateless configuration, the lifetime of the address will be deducted from the lifetime of the prefix, as indicated in the router's advertisement.

An address whose lifetime has expired becomes invalid and cannot be used as a source- or destination-address any more. To prevent a break in a TCP connection when an address becomes invalid, actually two lifetimes are issued (the valid- and preferred lifetime). Hosts are expected to continuously receive router advertisements and to update the lifetime of their address or the address as necessary.

Hosts know that they don't own their addresses; they merely lease them. To ensure continuous ownership they must repeat their DHCP request before the lifetime expires. If they fail to do so, servers can reasonably assume that the hosts don't need the addresses any more, and may re-allocate them to other clients.

1.2.3.3. Address Resolution

To transmit packets on a network, Internet stations must determine the prefix or the MAC address of the target station. IPv6's ICMP encompasses the functions of ARP as well as router- and neighbor discovery procedures.

An ICMP message, as with any other message, can only be transmitted if the host knows the MAC address of the destination. In IPv6 this is solved by the use of multicast transmission. As long as the MAC address remains unknown, messages are sent to multicast addresses.

1.2.3.4. Neighbor Discovery

The neighbor discovery procedure maintains four separate caches:

- The *destinations cache* has an entry for each destination address toward which the host recently sent packets. It associates the IPv6 address of these destinations with that of the neighbor toward which the packets were sent.
- The *neighbor's cache* has an entry for the immediately adjacent neighbor to which packets were recently relayed. It associates the IPv6 address of that neighbor with the corresponding MAC address.
- The *prefix list* includes the prefixes that have been recently learned from router advertisements.
- The *router list* includes the IPv6 addresses of all routers from which advertisements have recently been received.

To transmit a packet successfully the host must find out the MAC address of the *next hop* for the packet's destination. In many cases this address will be found to be already in the *destinations cache*. If the address is found in the *prefix list*, the destination is local and therefore constitutes the next hop. Otherwise the host selects a router from the *router list* and sends the packet to it.

In case the MAC address cannot be found in the *neighbor's cache* the host sends an ICMP *neighbour solicitation message* and updates the cache upon receiving a response. This message is sent to the *solicited node multicast address*, an address derived from the node's IPv6 address. Nodes listen to their multicast address and when detecting a solicitation message addressed to them, they respond with a *neighbor advertisement message*. This procedure, with modifications, can also be used to check whether a target address is already in use on a link.

Several routers can send advertisements to a link so that several entries will be in the *router list*. The host has to choose one and this choice can be wrong. The router that receives such a misdirected packet will retransmit it to the correct next hop which eventually ensures delivery for the cost of double transmission on the local link. Furthermore the router will issue an ICMP *redirect message* which informs the host about the correct next hop and optionally including its MAC address. When a host receives such a message it should consider this suggestion and update the respective entries in its *destination cache* and *neighbor's cache*.

The neighbor discovery algorithm clearly relies on the presence of routers. However the hosts should also be capable of operating in routerless environments, such as dentists' offices. This situation is indicated by permanently empty *router-* and *prefix lists*. In this case the hosts simply consider all destination addresses as local.

It is important to check that the information present in the various caches is still valid. This is the role of the *neighbor unreachability detection* procedure. That a destination is indeed reachable can be learned from acknowledged TCP packets or any progress on higher protocol layers, which is communicated to the IPv6 layer. But for strictly unidirectional connections the cache entries age and by default after 30 seconds are tagged *dubious*. The host can still use such information, but should actively seek confirmation of the neighbor's reachability by sending a solicitation message.

A potential risk is an attacker from outside sending forged ICMP advertisement or redirect messages. However, setting the hop count of every outgoing message to 255 defeats this kind of attack, since any arriving message with an hop count less than 255 apparently has an off-link origin and is thus bogus.

1.2.3.5. IPv6 over Circuit Switched Nonbroadcast Media

Media capable of broadcasting, in particular packet switching networks such as ethernet or FDDI are better fit for the IP as we know it than circuit switched networks such as telephone networks, ISDN, X.25, or ATM, which have usually no broadcast capability. IPv6 can indeed be run over non-broadcast switched circuits, but this will always require some manual configuration.

But IPv6 makes sure that the neighbor discovery procedure still works in this environment, whose main characteristic is the absence of any multicast capability. ATM, i.e., belongs to the *NonBroadcast Multiple Access* category (NBMA).

To ensure a host works in such environment it should be configured with the MAC address of at least one router. One of these routers will always considered as next hop, because in the absence of multicast the host will not be able to send neighbor solicitation messages.

Unless ATM is just used as an interconnect between IP devices, a problem specific to ATM is the obvious contradiction between ATM's global scope and IPv6's inherently hierachical structure, which can only be overcome by breaking the ATM "large cloud" into multiple "logical groups" connected by routers and providing a "cut through" procedure for the establishment of direct connections between stations that belong to different groups.

1.2.3.6. Node Mobility with IPv6

The general idea of IPv6 mobile nodes follows the "Mobile IP" concept presented in the next chapter. The *home agent* is informed about the current *care-of address* of the mobile node by sending a packet with a *Binding Update* end-to-end option to it. Upon receipt of such a packet other hosts send the packets thereafter directly to the stated (updated) address. Because the binding update option will in effect change the destination of packets bound to the mobile, it must be carried securely. The specification mandates that the packets carrying such options must be properly authenticated.

1.3. Introduction to Mobile IP

In offices we currently expect to have permanent access to global network resources. But as people move around from place to place with their laptop, keeping connected to the network is challenging and sometimes frustrating or expensive. However in the very near future, communicating via laptop, PDA, or any other mobile computing device should become as natural as using ones GSM-phone.

We are focusing on wireless means of accessing the network, most notably (diffuse) infrared (IR) and radio frequency (RF) links. Currently cellular telephone technology is of interest to the mobile user, relying on the telephone company to maintain connectivity and paying a considerable price for it. Point-to-point infrared LAN attachments, e.g. HP's NetBeam IR, are in use too, but their short range rather limits the user's mobility.

The following introduction is based on the first four chapters of [PERK98], which continues to give insight on the protocol level of Mobile IP.

Evidently *the* internet protocol TCP/IP plays a central role in current internetworking and therefore has to be considered when mobility becomes important. IP network address allocation and administration have historically assumed that there is a fixed relationship between a computer's IP address and its network location. So the IP address inherently contains routing information *and* uniquely identifies one node in the global net.

Until the advent of mobile IP, TCP/IP could not cope with these two conflicting requirements: The network address should be changeable for routability and mobility purposes but was expected to be stable and hence was used for identification.

Applications use IP addresses to identify routes by which datagrams may be exchanged between two network nodes. In addition, the IP address used by the application was also used to identify the endpoints themselves. This dual use of the IP address by the application endpoints causes problems when trying to use the application while changing one's point of attachment to the Internet. Clearly, applications need an unchanging way to identify the communication endpoints, but just as clearly the routes between the endpoints must change as they move from place to place within the Internet.

The internet is far to big to be administered by using a single flat address space. Therefore each 32bit IP address is partitioned in two parts, the boundary of which is determined by the netmask: The *routing prefix* and the *host address*. The former, also known as the *network* part, represents an administrative allocation or a subspace of the address space.

From the point of view of routing, the problem with mobility is that mobile computers move from one IP subnet to another, but then have the wrong subnet prefix for their current destination. Of course no packet destined for the old address will be routed to the new subnet, unless we introduce some new routing mechanism

1.3.1. Care-Of Address and Home-Address

Mobile IP solves the problem by maintaining *two* addresses. One for *locating* the mobile computer and one for *identifying* a communication endpoint on the mobile computer. Thus there is a need for some kind of *directory* to store the associations between these addresses and keep them up to date. Indexing by the address used to identify the mobile node to the Internet seems obvious. Each entry therefore contains the *home address* for identification purposes and an associated address for locating the mobile node, known as *care-of address*.

Since for the rest of the Internet the mobile node appears as if it were actually located on its home network, the name *home address* makes sense. If the source of a packet is assumed to be an Internet node without any special mobility support, the source will be unaware whether anything special happens when sending packets to the mobile computer. Indeed, nothing special needs to happen, if the mobile node is "at home", i.e., located at its home network. However, if the mobile node is not attached to its home network, then the datagram somehow needs to follow it to its care-of address. Since only the existing Internet infrastructure should be utilized, it seems clear, that the address of the datagram needs to be changed, in order to follow the mobile node when it is away from its home network.

This operation of changing the address of the datagram for further routing is known as *readdressing*. The destination address of the datagram (the home address) is replaced by the new destination address (the care-of address).

The inverse operation is required if the higher level protocols in the mobile node and the node with which it corresponds are to operate in a symmetrical manner. Any reasonable mobile IP architecture must be built with the intention of eliminating any modification to existing higher level protocols. The *inverse readdressing* function transforms the datagram so that the care-of address is replaced by the home address.

The exact ways in which the functions are applied in different nodes distinguish the various approaches. A common feature is the need to update the *location directory* when the mobile node attaches to a new point within the Internet. The update message sent to the directory has the effect of directing traffic from the home network to the mobile node's new location, and is therefore known as *remote redirection*. Authentication of such messages is mandatory and strong cryptography is used to solve this problem.

1.3.2. Agents and Example Architectures

The idea behind the **IETF Mobile IP Protocol** [RFC2002] is to position a *home agent* at the mobile node's home network and assign a long-term IP address to it. The *home agent* performs the *readdressing* and *location directory* functionality and uses a technique hereafter described as *tunnelling* to forward the datagrams to the roaming mobile computer.

The *inverse readressing* function and delivery of the resulting datagram to the mobile node is performed by the *foreign agent* located at the mobile node's visited network using the *care-of address*. Provided a sufficiently capable mobile node, it can take over this functionality from the *foreign agent,* i.e., be its own foreign agent. Furthermore the *foreign agent* is usually the mobile's *default router*. Thus when a mobile node sends packets, their first link destination is generally the foreign agent.

The mobile node must have at least the capability to acquire a valid IP address within the foreign network, for example via a *DHCP-Server* [RFC2131], that assigns unique addresses to every mobile node. However, there is an advantage in centralizing the *foreign agent's* functionality since it can server several mobile nodes using only one *care-of address*.

A different approach worth mentioning is known as the **Columbia Protocol** [IOAN91]. The Columbia University's mobile IP relies on the configuration of a collection of *mobile support routers* (MSRs) that conspire to create a virtual subnet – the *mobile subnet* of IP addresses administered for use by the mobile nodes. The MSRs communicate by way of a multicast address.

As mobile nodes move, they inform their current MSR about their needs and request that the current MSR informs their previous MSR of their movement. Thereby the network of MSRs performs *forward-* and *reverse address translation* and the *location directory* service in a distributed manner.

1.3.3. The IETF Mobile IP Protocol – Overview

Mobile IP is, in essence, a way of doing three relatively separate functions:

1. *Agent discovery* - Home agent and foreign agents may advertise their presence on the network for which they provide service, to let the mobile nodes determine, whether they are in their home network or away. A newly arrived mobile node can take the initiative and send a solicitation message on the link to discover available agents, but this is optional. For agent discovery Mobile IP extends the existing router advertisement and router solicitation messages defined for ICMP router discovery.

2. *Registration* - When the mobile node is located on its home network it operates without needing mobility service. But when the mobile node roams away from its home network, it has, after acquiring an IP address, to register its care-of address with its home agent. Depending on the method of attachment (see below), the mobile node will register either directly with its home agent or through a foreign agent, which forwards the registration message to the home agent. UDP is used for these messages.

3. *Tunneling* - In order for datagrams that arrive at the home agent to be forwarded to the mobile node, the home agent has to *tunnel* the datagrams to the care-of address by *encapsulating* and re-sending them, relying on conventional IP routing. Outgoing packets from the mobile node are sent conventionally, i.e. without needing to be encapsulated.

And there are two ways of acquiring a care-of address:

1. A *foreign agent care-of address* is owned by a foreign agent and published through its *agent advertisement* messages. In this mode, the foreign agent is the endpoint of the tunnel and on receiving tunnelled datagrams, decapsulates them and delivers the inner datagram to the mobile node. This mode is advantageous because it allows many mobile nodes to benefit from a shared care-of address and therefore does not place unnecessary demands on the already limited IPv4 address space.

2. A *colocated care-of address* is a care-of address acquired by the mobile node itself as a local IP address. The address may be dynamically acquired, such as through DHCP, or it may be owned by the mobile node as a long-term address for its use only while visiting a particular foreign network. When using a colocated care-of address, the mobile node is the endpoint of the tunnel and performs the decapsulation itself.
   An additional advantage of a colocated address is that the MT can use this address directly without the need of any agent intervention, if it has no intention to move during the lifetime of such a connection. Thereby the long delay path via its home agent is bypassed.

The home agent can be located on a separate system on the home network, integrated into a router there, or located on a virtual network. In the latter case there is no physical link to attach to, thus the mobile computer is always treated as being away.

Possibilities for placing the foreign agent are even greater, since they need not rely on TCP/IP networks, the only requirement is a suitable link-layer connection between the homeagent and the foreign agent.

1.4. Introduction to GSM

The Global System for Mobile communications (GSM) is a digital cellular communications system initially developed in an European context which has rapidly gained acceptance and market share worldwide. It was designed to be compatible with ISDN systems and the services provided by GSM are a subset of the standard ISDN services (speech is the most basic service).

The functional architecture of a GSM system is depicted in Figure 1.4.2. and can be divided into the Mobile Station (MS), the Base Station (BS), and the Network Subsystem (NS). The MS is carried by the subscriber, the BS subsystem controls the radio link with the MS, and the NS performs the switching of calls between the mobile and other fixed or mobile network users as well as mobility management. The MS and the BS subsystem communicate across the Um interface also known as radio link.

1.4.1. Economical Background

At present there are six analogue systems operational in Europe, and a mobile designed for one cannot be used with another.

In 1981 a joint Franco-German study was initiated to develop a common approach which, it was hoped, would become a standard for Europe. Soon after, the main governing body of the European PTTs (CEPT) set up a committee known as the Groupe Special Mobile (GSM) under the auspices of its Committee on Harmonisation (CCH). GSM was charged with defining a system which could be introduced across Europe in the 1990s. The stage was now reached where equipment could be installed and services bought in any European country and used anywhere in Europe.

For the first time in the history of the European electronics industry, the political, commercial and industrial forces had come together to generate a home market for the European mobile telephony industry, matching that of the United States.

That this initiative is in the rapidly expanding and technologically challenging field of cellular telephony makes the opportunity all the more exciting. The enterprise offers European industry the chance to supply in equipment to many rapidly growing operators. Two UK cellular operators, Cellnet and Vodafone, have shown the most rapid growth world-wide of any cellular network operators.

One of the major features of GSM is the open specification of all the interfaces. It should also be noted that the specifications were driven by the *operators* who wanted open competition by supplieres.

1.4.2. The Creation of the Specification

The early years of GSM were devoted primarily to the selection of the radio techniques for the air interface. By 1986 GSM was ready to undertake trials of the different candidate systems proposed for this interface and later that year six different systems were trialled in Paris. GSM had drawn up a list of seven criteria ranked in order of importance (shown in Table 1.4.1), to be used in assessing these candidates. At the top of the list was spectral efficiency, measured as the number of simultaneous conversations per megahertz per square kilometer.

At the same time the UK, which did not have a candidate in the field, undertook a study of the different systems against the criteria shown in Table 1.4.1 and started development of a test bed to assess the critical features. In the event none of the candidates was selected. The information gleaned could be used to generate the outline specifications of a system which could capitalize on ideas and approaches from several of the candidates.

Table 1.4.1 - Criteria to be met by the GSM System

① Spectrum efficiency
② Subjective voice quality
③ Mobile cost
④ Hand-portable feasibility
⑤ Base-station costs
⑥ Ability to support new services
⑦ Co-existence with existing systems

On one subject there was no serious debate. The performance of cellular radio is restricted primarily by co-channel interference and a given quality of telephony can be achieved at much higher levels of co-channel interference if digital transmission is used. This allows the cells to be reused more frequently, and it has been estimated that this factor alone offers the GSM system a three-fold improvement in spectral efficiency over the baseline Nordic Mobile Telephone (NMT) system.

Another factor that encouraged the trend to digital transmission was that the telecommunications industry world-wide had been converting rapidly to digital methods and the advent of ISDN demanded a much higher level of digital signaling than had hitherto been the case [HERR86].

Equally important is the evolutionary path that is offered by digital transmission. If the system was to stand the test of time it was essential that it be able to evolve to accommodate the improvements that future systems might introduce. The most important of those foreseen was the introduction of an 8 kbps speech coder. As a digital system is more easily configured to change channel characteristics than was an analogue system.

Considerable debate took place, however, over the most suitable transmission method. A variety of approaches were represented by the candidate systems (FDMA, TDMA, CDMA - see [GOOD97], Chapter 9.1 for a summary) and the final decision to adopt a TDMA structure was made in April 1987.

Table 1.4.2 - GSM Recommendations [ETSI90]

00 Preamble
01 General
02 Service aspects
03 Network aspects
04 MS-BS interface and protocols
05 Physical layer on the radio path
06 Audio aspects
07 Terminal adapters for mobile stations
08 BTS/BSC and BSC/MSC interfaces*
09 Network interworking
10 Service interworking
11 Equipment specification and type approval specification
12 Operations and Maintenance

*A list of GSM acronyms and abbreviations is given in Figure 1.4.2.

The period 1987 to 1990 saw a tremendous effort by the working parties and their supporting expert groups to create and document a complete mobile telecommunications system. The specifications and their explanatory notes were substantially complete by 1990 and run to 138 documents, some of which are several hundred pages long. They are divided into 13 Sets of Recommendations [ETSI90], covering different aspects of the system, as shown in

Table 1.4.2. above. These make extensive reference to relevant CCITT, CEPT, and ISO Standards and provide system definitions, standardization aspects, mandatory and optional features, as well as guidance on system design.

1.4.3. System Description

1.4.3.1. Open Interfaces

The main philosophy underlying the GSM approach is that the major interfaces in the system should be open and in the public domain. This has the twin benefits of allowing supplier competition for each network node and encouraging the evolution of ideas, albeit within the constraints of the interface definitions.

Considerable time and energy went into the development and standardisation of all the interfaces shown in bold in Figure 1.4.2. In terms of the standard ISO seven-layer model, GSM specifies layers 1, 2, and 3 (i.e., the physical, data link and network layers) [WAKI87]. This approach represents a significant departure from previous practice, and it is one of the major aspects in which the GSM system differed from the existing analogue systems.

1.4.3.2. System Components

In a chapter of this nature it is only possible to give an overview of the major system elements and briefly describe their functions. In the following, each of the system components shown in Figure 1.4.2 will be described as it comes into operation when a mobile station makes a call.

1.4.3.2.1. Mobile Station

The mobile station comes in a number of different forms, ranging from the traditional vehicle-mounted phone operating at 20W, through transportables operating at 8W and 5W, to the increasingly popular hand-portable units which typically radiate less than 2W. A fifth class for hand-portables operating at 0.8W has been specified for Micro Cellular versions of the network.

One of the main factors governing the hand-portable size and weight is the battery pack. Several features of the system are designed to allow this either to be smaller or to give a substantially longer life between charges. Chief among these is Discontinuous Receive (DRX). This allows the mobile when not engaged in a call to sleep by synchronizing its listening period to a known paging cycle of the network. This can typically reduce the standby power requirements by over 95% (as can be derived from [GOOD97] in the section “Paging Channel Operation, Sleep Mode”, p.180).

1.4.3.2.2. The Radio Sub-system

When the mobile user initiates a call, the MS will search for a local base station’s control channel. A base station sub-system (BSS) comprises a base station controller (BSC) and several base transceiver stations (BTS), each of which provides a radio cell of one or more channels. The BTS is responsible for providing layers 1 and 2 of the radio interface, that is, an partially error-corrected data path. Each BTS has at least one of its radio channels assigned to carry control signals in addition to traffic. The BSC is responsible for the management of the radio resource within a region. Its main functions are to allocate and control traffic channels, control frequency hopping, undertake handovers (between cells inside its region) and provide radio performance measurements. Once the mobile has accessed, and synchronized with, a BTS the BSC will allocate it a dedicated bi-directional signaling channel and will set up a route to the Mobile services Switching Center (MSC).

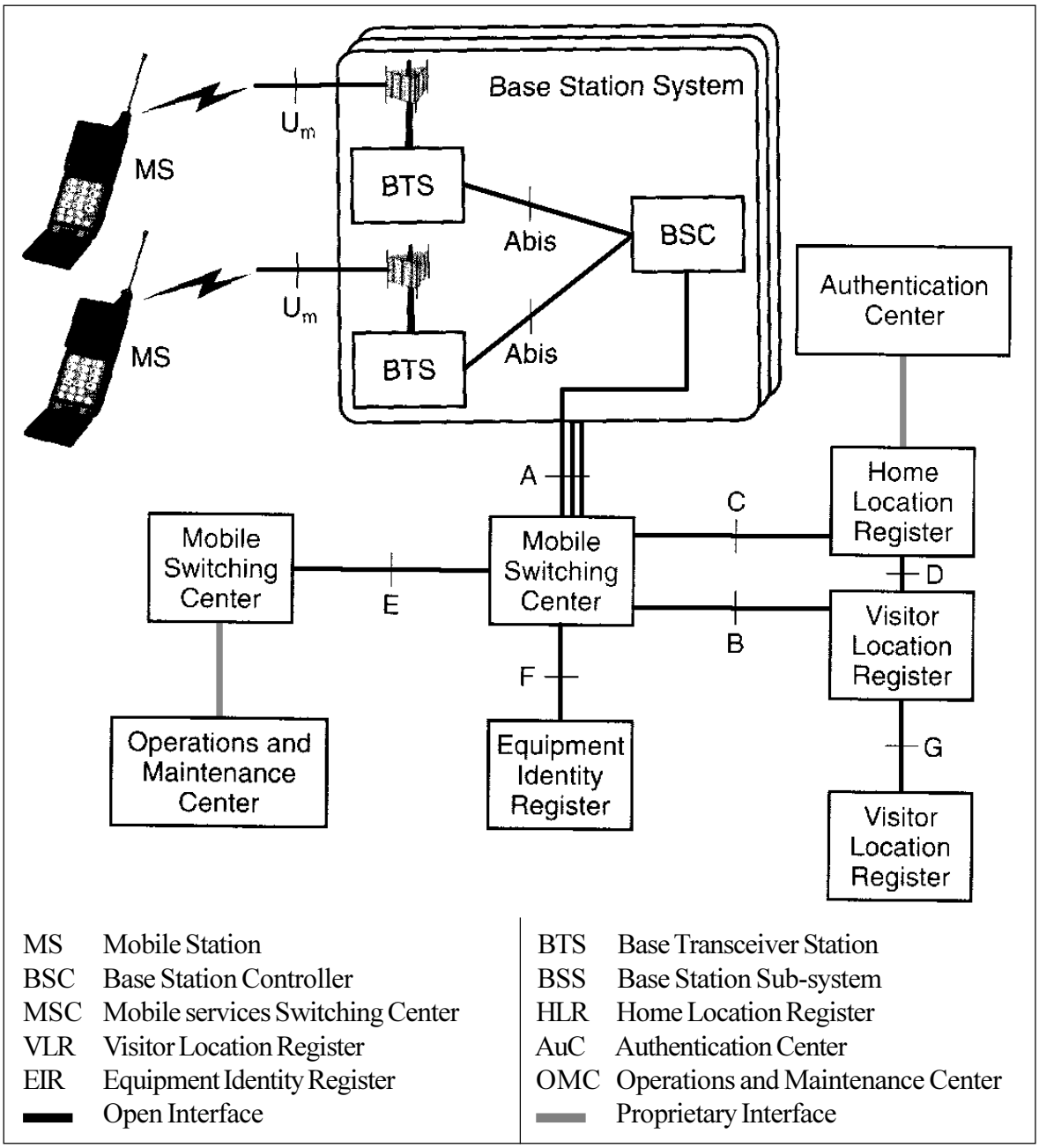


Figure 1.4.2. GSM Network Components and Standardized Interfaces

1.4.3.2.3. The Switching Sub-system

The MSC routes traffic and signaling within the network and interworks with other networks. It comprises a trunk ISDN exchange with additional functionality and interfaces to support the mobility. When a mobile requests access to the system it has to supply its International Mobile Subscriber Identity (IMSI). This is a unique number which will allow the system to initiate a process to confirm that the hereby authenticated subscriber (not necessarily the equipment!) is allowed to access it. Before it can do this, however, it has to find where the subscriber is based. Every subscriber is allocated to a home network, associated with an MSC within that network. Each subscriber has an entry in the Home Location Register (HLR) of their home system, which contains information about the services the subscriber is allowed. The HLR also contains a unique authentication key and associated challenge/response generators. Optionally there can be one key for each subscriber (i.e., per system key or optionally per user keys)

1.4.3.2.4. Mobility Management and Security

Whenever a mobile is switched on, and at intervals thereafter, it will register with the system; this allows its location in the network to be established and its location area to be updated in the HLR. A location area is a (geographically) defined group of cells. On first registering, the local MSC will use the IMSI to interrogate the subscriber's HLR and will add the subscriber data to its associated Visitor Location Register (VLR). The VLR now contains the address of the subscriber's HLR and the authentication request is routed back through the HLR to the subscriber's Authentication Center (AuC). This generates a challenge/response pair which is used by the local network to challenge the mobile.

Alternatively, and practically used regularily because of the resulting bandwith saving on the backbone network, a temporary mobile subscriber identity (TMSI) is assigned when first registering and used hereafter *without* the need of a HLR query. In addition, some operators also plan to check the mobile equipment against an Equipment Identity Register (EIR), in order to control stolen, fraudulent, or faulty equipment.

The authentication process was intended to be very powerful. It primarily protects the network operators from fraudulent use of their services. It does not however protect the user from eavesdropping. The TDMA nature of GSM coupled with its frequency hopping facility will make it very difficult for an eavesdropper to lock onto the correct signal however and thus there is a much higher degree of inherent security in the system than is found in today's analogue systems. Nevertheless for users who need assurance of a secure transmission, GSM offers encryption over the air interface. This is based on a public key encryption principle and is thought to provide very high security. But since the algorithmn is secret it has not been widely evaluated.

1.4.3.2.5. Call Set-up

Once the user (and possibly his equipment) is accepted by the network, the mobile must define the type of service it requires (voice, data, supplementary services, etc.) and the destination number. At this point a traffic channel with the relevant capacity will be allocated and the MSC will route the call to the destination. Note that the network may delay assigning the traffic channel until the connection is made with the called number. This is known as off-air call set-up, and it can reduce the radio channel occupancy of any one call, thus increasing the system traffic capacity.

1.4.3.2.6. Handover

GSM employs mobile-assisted handover. In this technique the mobile continuously monitors other base stations in its vicinity, measuring signal strength and error rate. These measurements are combined in a single function and the identities of the best six base stations are transmitted back to the system. The network can then decide when to initiate handover. The use of bit error rate, in addition to signal strength, adds considerably to the ability of the network to make informed handover decisions and is another example of the advantage of digital over analogue transmission. The BSC can initiate and execute handover if both BTS's are under its own control. In this case, the BSC can be considered as the manager of a specific group of radio frequencies for a geographic region and can control that resource to maximize its utilization. Alternatively, and whenever handover must take place to a cell outside the control of the BSC, the MSC controls and executes the handover.

1.4.3.2.7. Mobile Terminated Calls

When setting up a call from the fixed network to a mobile the procedure is much the same. First, however, the mobile must be found which initiates a query of the HLR for the user location. The locating is achieved by means of a paging signal which covers the location area in which the mobile has registered. Mobiles intermitally monitor the paging channel and, on detecting a call to them, undertake the access procedure described. It shall be noted that authentication is not mandatory for received calls, but it is expected that most operators will demand it.

The paging procedure has been designed to facilitate significant battery-saving potential in the hand-portable. Unless a hand-portable is used extensively the biggest drain on its battery comes not from the time spent using it, but from the standby cycle as it monitors the paging channel, just in case it is being called. In the GSM system the DRX mode, (discussed in 1.4.3.2.1.) allows the mobile, once it has located the paging signal, to synchronize with a clock knowing that it will not get another paging signal until a specified time has elapsed. It can thus power down its circuits for most of the time hence standing by.

1.4.3.3. Network Management

So far we have discussed the network elements used directly when making a call. A network of this complexity, however, needs to be managed and maintained. This is the function of the Operations and Maintenance Center (OMC). GSM leaves decisions on O&M to the individual operators, but general guidelines are given which reinforce the move towards an overlaid telecommunications management network. In this approach five separate management functions are identified, as shown in Table 1.4.4. The activities covered by each major area are further defined in Tables 1.4.4.1 to 1.4.4.5.

Table 1.4.4. - Network Management Functionality

① Operations and Performance Management
② Maintenance
③ System Change Control
④ Security Management
⑤ Administration and Commercial Functionality

Table 1.4.4.1. - Operations and Performance Management Functions

- *Network Status Information* (circuits, nodes, signaling)
- *Operations* (interworking, network configuration, control of system elements, node configuration)
- *Performance Data* (traffic measurements, quality of service observations, availability performance data, throughput measurements, handover statistics)
- *Performance Management* (radio network management, switching network management, routing control)

Table 1.4.4.2. - Maintenance Functions

- *BSS Maintenance* (BTS, BSC, mobile integrity)
- *MSC Maintenance* (hardware, software)
- *Transmission Maintenance* (lines, microwave, multiplexing)
- *DataBase Maintenance* (HLR, AuC, EIR)

Table 1.4.4.3. - System Change Control Functions

- *Enhancements* (provision of: new features, new functions)
- *Extensions* (addition of: existing equipment, existing services)
- *Reconfigurations* (reorganization of: existing equipment)

Table 1.4.4.4. - Security Management Functions

- *Network Access* (management of: authorisation, access control, authentication facilities)
- *Security Reporting* (reporting on: feature status, access status, management status)
- *Data Security Management* (management of: keys and encryption, security routing, subscriber identity module)

Table 1.4.4.5. - Administration Functions

- Administration and Commercial Tariff
- Charging and Accounting
- Subscriber and Mobile Equipment Management
- Customer Services

The telecommunications management network is hierarchical in structure. It starts with the O&M functionality of individual network elements, integrates them with operations and maintenance centers (OMC) and provides a network management center (NMC).

1.4.4. The Air Interface

The air interface is the radio link over which the mobile stations communicate with the fixed infrastructure. The technology for information transport across this interface is the most prominent feature of each practical system. The nature of the air interface reflects the tradeoffs among a large number of design goals including costs, coverage area, transmission quality, spectrum efficiency, and user mobility.

1.4.4.1. Frequency Allocation

Throughout Europe, GSM has been allocated a specific 50 MHz of spectrum divided into transmit and receive bands separated as shown in Figure 1.4.3. It should be noted however that not all countries are able to use the full allocation at present. This is due to existing commitments, often military in nature. Each radio channel is 200 kHz wide and thus there are a total of 125 paired channels available. In practice, where there is more than one operator and to prevent interference, guard bands have to be provided between the frequency bands allocated to them, and thus the actual number of channels available is less than the maximum.

The salient features of the air interface are shown in Table 1.4.6.

Table 1.4.6. - Basic Air Interface Parameters

Channel spacing ..... 200 kHz
Modulation ..... GMSK
Modulation depth ..... BT = 0.3
Data transmission rate ..... 270.833 kbps
Number of channels/band ..... 8 (16, with half rate codec)
User data rate (nominal) ..... 16 kbps (8, with half rate codec)
TDMA frame period ..... 4.62 ms
Time-slot duration ..... 0.58 ms

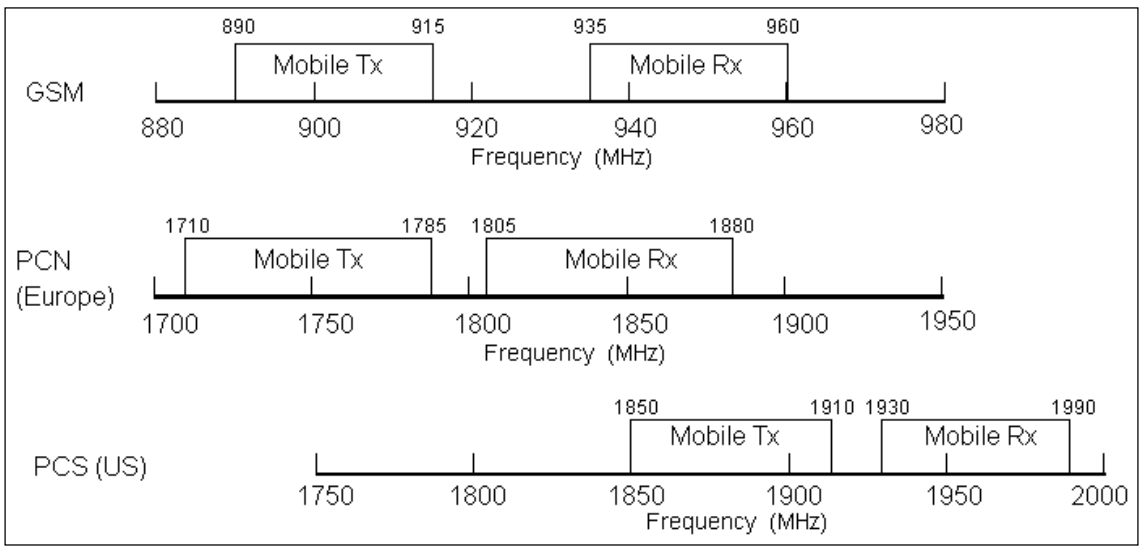


Figure 1.4.3. GSM Spectrum Allocation

1.4.4.2. Speech Coding

The spectral efficiency target set for the GSM system demands a speech codec which can provide toll quality speech at 16 kbps or less. The solution adopted is based on a Residually Excited Linear Predictive coder (RELP) enhanced by the inclusion of a long-term predictor (LTP) [NATV88]. This improves the speech quality by removing the structure from the vowel sounds prior to coding the residual data. It has the effect of removing the coarseness often associated with linear predictive coding, especially on female voices. The basic data rate from the coder is 13 kbps and speech is processed in 20 ms blocks, as shown in Figure 1.4.4.

The resulting code is split into two parts, the most critical bits being put first. This first part has a half-rate convolutional code applied to it and when recombined with the second part the total block length is 456 bits. As we will see later this block length can be fitted into four time slots, but in practice it is spread over eight. This process is called *diagonal interleaving* and it allows the convolutional code more chance to recover if a sequence of TDMA frames is badly corrupted during radio transmission. But it is also obvious that not all bits are equally protected since some have *no* error correction!

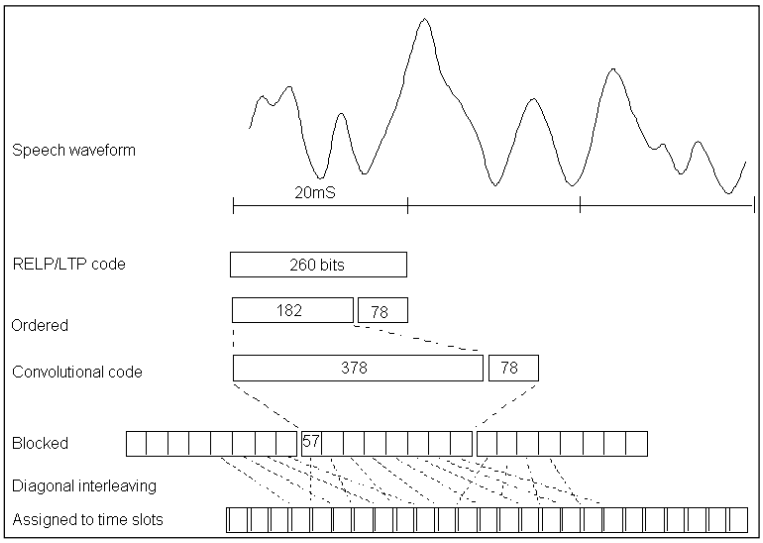


Figure 1.4.4. The Speech Coding Process

1.4.4.3. Data Structure

The specific parameters selected for the GSM air interface were shown above in Table 1.4.6. The data transmission rate has been set as high as possible commensurate with cost-effective equalization of the expected multipath effects. The precise figure was then chosen to allow major system clocks to be derived from a common source.

The basic traffic data rate allows eight channels to be accommodated on a single RF carrier. With an eye to the future, however, the specification allows two separate channels to be interleaved within the same frame. This facility will effectively double the traffic capacity once a half rate speech codec became available. Figure 1.4.5. shows the basic frame structure and the time slot organization for a traffic or signaling channel.

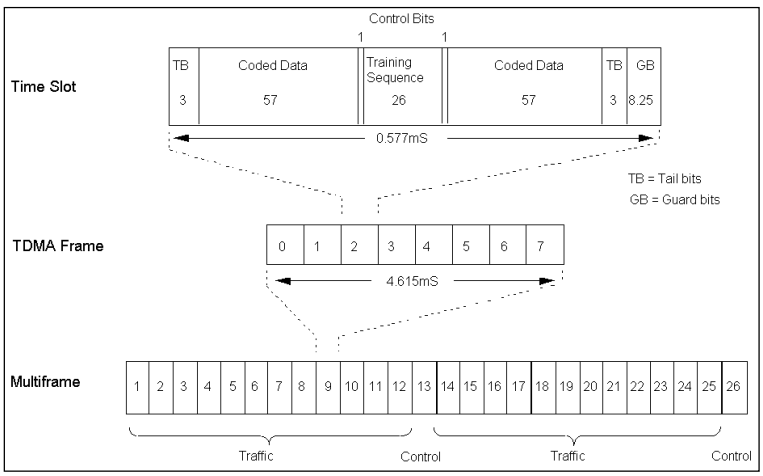


Figure 1.4.5. GSM Frame Structure

The 26 frame multiframe provides 24 frames allocated to traffic and two (the 13th and the last) allocated to control and supervisory signals associated with the traffic channels. Each traffic channel is assigned one of the 16 time slots in these two frames. (For 13 kbps speech, where only eight traffic channels can be carried, the last TDMA frame is not used.)

In addition to these two slow associated control channels (SACCh) allocated in the 13th and 26th frame, the system provides:

- Fast Associated Control Channels (FACCh) – which steal time slots from the traffic channel and are used for irregular control requirements such as handover.
- Dedicated Control Channels (DCh) – which are multiplexed onto a standard traffic channel and are used for registration, location updating, authentication, and call set-up.
- Broadcast Control Channels (BCh) (down-link only) – which provide mobiles with base station identity and information pertaining to the cell.
- Random Access Channel (RACh) (up-link only) – which is used by the mobile to request access to the network.
- Access Grant Channel (AGCh) (down-link only) – which replies to a random access and assigns a dedicated control channel for subsequent signaling.
- Paging Channel (PCh) (down-link only) – which informs the mobile that the network wants to signal to it.

Apart from the random access channel all these control channels have the same structure as the traffic channel shown in Figure 1.4.5. The random access channel has a different equalizer training sequence because no system timing information is available at this stage, see section 1.4.4.4.

Each time slot lasts 0.577 ms and comprises 148 bits, with an 8.25 bit guard period between slots. The traffic carried by the slot is divided into two separate 57 bit blocks and each block is assigned data from separate speech coding frames. Thus eight such slots are needed to convey the 20ms of speech data, but each slot is actually carrying data from two speech blocks simultaneously. Thus four time slots in consecutive frames provide 456 traffic bits in 185 ms. This accommodates the 456 speech bits created every 20ms and the additional 15 ms adds up over 24 frames to provide the additional two control frames in the multiframe.

The control bit shown associated with each data block is used to flag whether the block is carrying normal traffic, or has been stolen by the fast associated control channel.

In the center of each slot is a sequence of 26 bits which are used by the receiver to set the parameters of its equalizer/demodulator functions. This is necessary to overcome the multipath problem described in Section 1.4.4.6.

1.4.4.4. Timing Advance

TDMA requires that the signals from all the mobiles using a single channel must reach the base station at the right time. They must not overlap each other. If the base station provides a reference signal, those mobiles nearest it will respond earlier than those towards the perimeter of the cell. GSM has been specified to allow cells to extend up to 35 km from the base station.

The time taken for a radio signal to travel the 70 km to the perimeter and back is 0.23 ms and thus a guard period of this length would have to be provided on each time slot. This is clearly inefficient, and GSM overcomes this by informing the mobile how much in advance of the reference it should transmit in order to be correctly synchronized at the base station. This allows the guard period to be reduced to 0.03 ms (8.25 bits).

1.4.4.5. Modulation

The modulation scheme chosen by GSM is GMSK with a bandwidth data-rate product (BT) of 0.3 [MURO81]. Gaussian Minimum Shift Keying modulation produces a better defined spectral occupancy than Frequency Shift Keying (FSK) or Differential Phase Modulation (DPM) and its resilience to co-channel interference, while not as good as DPM, is adequate for the GSM requirements.

1.4.4.6. Multipath and Equalization

At the frequency band occupied by GSM, radio waves do not refract very well and thus there are potentially many shadow areas created when a mobile or a base station transmits. This is compensated for by the tendency for the signals to reflect from buildings, hills, high-sided vehicles, etc., and these reflections help to fill in the shadows. Many different reflections can reach the same point, however, and even when there is a direct path it is not unknown for strong reflections to be received as well. The radio paths taken by the reflections must, of course, be longer than the direct path and at the bit rates chosen for GSM the difference in path length can be equivalent to several bit periods. Figure 1.4.6 demonstrates this effect, and it can be seen that the combined signal received at the mobile's antenna can be severely corrupted.

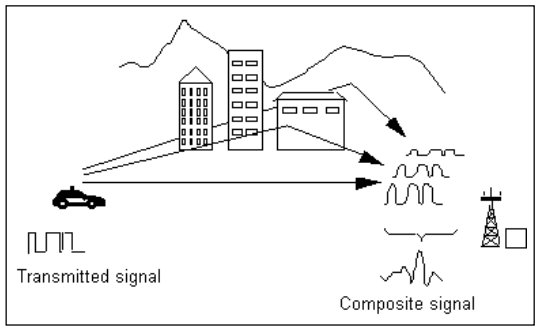


Figure 1.4.6 - The Multipath Problem

To date, radio systems have substantially avoided this multipath effect by choosing bit rates which are long compared with the expected multipath delays. Equalization is a technique however, which allows us to recover the wanted signal despite severe multipath corruption. Until recently the cost and complexity of applying equalization was prohibitive. Now, however, 50000 gate VLSI devices are not unusual and equalization can be reduced to a single, albeit complex, chip. This means that higher data rates can be used with consequent improvements in spectral efficiency. Nowadays equalization is done by a 50 MIPS DSP used in GSM phones.

Equalization works by making an estimate of the impulse response of the transmission medium and then constructing an inverse filter through which the received signal is passed. There are several methods for estimating the transfer function of the transmission path and several variations of algorithm associated with each, but whatever the method they all rely on receiving a known sequence of data. This is the training sequence which is transmitted in the middle of each time slot. The receiver detects this sequence, and, knowing what bit pattern it represents, is able to estimate the transfer function most likely to have produced the signal received. The calculation of the coefficients of the filter required to compensate for the response is then relatively straightforward.

The multipath effects can change very rapidly in practice. The wavelength at 900 MHz is only 30 cm and thus a change in the differential path length of only 15 cm between two signals received at an antenna can change their interference from constructive to destructive. The GSM specifications are designed to accommodate vehicles moving at up to 250 kph and thus the mobile could have moved up to 32 cm in the 46 ms between successive traffic channel timeslots. Add to this the problems of reflections from other moving vehicles and it is clear that each time slot has to be treated independently. It is also important to provide the best possible estimate of the path characteristics and it can be seen that placing the training sequence in the middle of the slot reduces the time between it and the data bit most distant from it.

1.4.5. Network features

1.4.5.1. GSM as an Intelligent Network

1.4.5.1.1. Intelligent Network (IN) Architecture

Intelligent networks are identified as a network architecture that relocates specific services and data bases from switches to one or more network control and decision points. The principal driving force behind IN's is the inability of current network architectures to support the rapid development and deployment of advanced services due to the need to specify, develop, test, and deploy software in each switch in a public network for each new service. The move towards IN results in networks containing:

- Switches with bearer and basic service control capability known as Service Switching Points (SSPs).
- Elements with advanced service control capability known as Service Control Points (SCPs).
- A Service Management System (SMS) that controls the deployment of services and the associated service data.

The link between the SSPs (the switches) and their associated SCP(s) (the service logic/databases) will use the Transaction Capability (TC) part of the Signalling System No. 7 (SS7) with the Signalling Connection Control Part (SCCP) and the Message Transfer Part (MTP) providing the basic mechanism on which to build the query portion of these new services, though for SCP functions co-located with an SSP, a local high speed link may be substituted.

1.4.5.1.2. GSM Network Architecture

GSM has been designed with close reference to the IN model and can be considered to be the first true instantiation of an Intelligent Network. It exhibits:

- An open distributed architecture
- Separation of service control and switching functions
- Full use of SS7 as the signaling communications infrastructure
- Clearly defined and specified interfaces
- Intelligent network structure

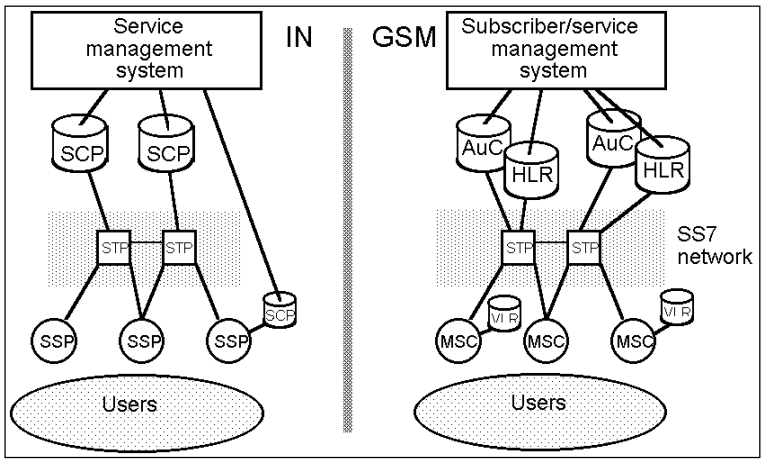


Figure 1.4.7. GSM and IN Physical Architectures

With particular reference to the latter point, examination of the GSM network architecture in terms of SSP and SCP units indicates close alignment with the general IN architecture. This is illustrated in Figure 1.4.7. above.

The SSP, or service switching point, is responsible for the interface with the service users. It provides a number of service components. The SSP also provides the bearer capability for telecommunications traffic and generates service triggers that cause service control requests to be directed at the SCP, which may be remotely located for infrequently used services, or connected locally for frequently used services. In GSM terms the SSP appears as the Mobile Switching Center (MSC) and the associated radio subsystem responsible for providing service access to mobiles.

The SCP is the service control point. It manages service triggers and controls the progress of a call based upon the nature of the trigger and the service programs running on its internal processing platform. Referring to Figure 1.4.7. again, the SCP functionality occurs in the the GSM database entities such as HLR, AuC, and VLR.

1.4.5.2. Services

GSM offers a wide range of services to its users. Not all these services were expected to be available from day 1, hence a group formed to coordinate the views and commercial positioning of the Operators (this is the GSM MoU - which simply stands for the GSM Memorandum of Understanding). They defined targets for the introduction of specific services. The main reasons for this are to provide guidance on priorities to the system suppliers and to ensure as far as possible that when subscribers roam they will be able to receive the same set of services as they are used to on their home network.

The GSM MoU has defined four categories for the introduction of service; three of these, E1, E2, and Eh are time related, the fourth, A, means the introduction is optional at the operator's discretion. The categories are:

E1 ..... Start of Service
E2 ..... End of 1994
Eh ..... On availability of half-rate channels
A ..... Optional

The asynchronous (transparent and non-transparent) data services at rates at and above 2.4 kbps were assigned in E2 and subsequently Eh. All other bearer services are assigned category A.

Teleservices are assigned categories as shown in Table 1.4.7.

Table 1.4.7. Introduction of Teleservices [1]

Telephony	E1 then Eh
Emergency Calls	E1 then Eh
Transparent fax	E2
SMS Point-to-Point Mobile terminated	E2
SMS Point-to-Point Mobile originated	A
Non-transparent fax	A
SMS Cell Broadcast	A

Supplementary services are assigned categories as shown in Table 1.4.8.

[1] Tables 1.4.7 is provided by kind permission of the GSM MoU Group.

Table 1.4.8. Introduction of Supplementary Services [1]

Code	Supplementary Service	Introduction
CLIP	Calling Line Identification Presentation	A
CLIR	Calling Line Identification Restriction	A
CoLP	Connected Line Identification Presentation	A
CoLR	Connected Line Identification Restriction	A
CFU	Call Forward Unconditional	E1
CFB	Call Forward on Busy	E1
CFNRy	Call Forward on no Reply	E1
CFNRc	Call Forward on not Reachable	E1
CW	Call Waiting	E2
HOLD	Call Hold	E2
MPTY	Multi Party	E2
CUG	Closed User Group	A
AoC	Advice of Charge	E2
BAOC	Barring of All Outgoing Calls	E1
BOIC	Barring of Outgoing International Calls	E1
BOIC-exHC	Barring of Outgoing International Calls except to Home Country	A
BAIC	Barring of All Incoming Calls	E1
BIC-Roam	Barring of Incoming Calls when Roaming	A

1.4.5.3. The Subscriber Identity Module

An important innovation introduced by the GSM committee is the idea of using a Smart Card in conjunction with a mobile phone, thus decoupling the subscriber from the equipment. The credit sized card, containing a microprocessor and a small amount of memory, is being introduced in a number of fields, such as banking and security. Associating a cellular subscription with a card instead of a mobile phone introduces considerable flexibility. This is the concept of the Subscriber Identity Module (SIM). It means that the subscriber can use any phone to receive incoming calls and have any outgoing calls made charged to his own account. In effect, while his card is in the phone, it becomes his personal phone. All his personal data: short-code dialing, services subscribed to, authentication key, IMSI, etc. are stored in the smart card.

GSM allowings manufacturers to offer a semi-permanent SIM that is plugged inside the equipment. This module is identical to the standard SIM except it is mounted on a cut-down (reduced size) card with modified connectors. This option is particularly attractive to the hand-portable manufacturers.

Security conscious subscribers also want to incorporate a PIN (Personal Identity Number) in their SIM so that nobody else could use it without their authority. GSM offers this facility. Indeed the PIN is standard, but it is expected that most operators will choose to allow the user to disable the function after the first registration although they highly recommend its continued use. If it is not disabled the PIN has to be entered every time the card is inserted in the phone and/or when it is switched on.

The introduction of a second PIN (PIN2) is planned to extend the use of some services, such as Advice of Charge (AoC), in Phase 2.

A multiple International Mobile Subscriber Identity (IMSI) SIM card allows ‘semi-automatic’ roaming. It can store up to 32 network identities. The user retains a subscriber number, but also keys in a two-digit code to access the local service provider.

[1] Tables 1.4.8 is provided by kind permission of the GSM MoU Group.

1.4.5.4. Short Message Service

As in the already well established paging services a feature of GSM, not found in first generation cellular networks, is the ability to transmit short data messages, each up to 160 alphanumeric characters long, over the signaling channels. This connectionless service operates rather like a pager function with the added advantages that the messages can pass in either direction and confirmation is provided that the message has been received.

Two types of Short Message Service (SMS) are planned, Cell Broadcast and Point-to-Point. In the Cell Broadcast case a message is transmitted to all those operating mobiles present in a cell that have the capability to receive SMS. This service is clearly only one way and no confirmation of receipt is obtained. It is expected that Cell Broadcast, which will be the first service to be offered, will be used to transmit messages about traffic conditions, sports results, stock exchange information, etc. The subscriber can arrange to filter out unwanted cell broadcast message types. The Point-to-Point service, as its name implies, allows a message to be sent to a particular mobile or lets the mobile send a message to a specific addressee in a service centre.

Mobiles used for SMS services must contain special software to enable the messages to be decoded and stored. Typically messages will be stored on the SIM and read using the standard display of the mobile.

The Short Message Service requires a Service Center to receive incoming messages, check and organize them, send them to the Operator and receive and pass on any confirmation message. The Operators themselves will no doubt set up their own Service Centers and possibly offer Service Providers the opportunity to connect their own service centers to the GSM network. In most cases the subscribers will have to register with the SMS Service Provider and pay a monthly fee to receive the service.

Furthermore the innovative SMS services are based on using the SMS message to manipulate the SIMs contents.

1.4.5.5. Data

GSM recognized that data over cellular was going to play an increasingly important role in the future and the digital nature of the standard makes it well suited to supporting such a service. It was perhaps one of the few failings of the committee however that the specification for data services encompasses all the existing services offered by the PSTN's throughout Europe. It seems it was not possible for the committee to cut through the variety and provide a service geared to the 1990's. Thus GSM offers synchronous and asynchronous, transparent and non-transparent service, at data rates of 300 bps, 1.2 kbps, 2.4 kbps, 4.8 kbps and 9.6 kbps. It also supports Group 3 FAX.

1.4.6. The Past

By September 1992, 27 operators representing 18 countries, had committed themselves to GSM and joined the MoU Group:

Australia (TELECOM Australia), Austria (PTA), Belgium (RTT Belgacom Belgium), Denmark (TELE Danmark Mobile, Dansk Mobil Telefon DMT), Finland (Telecom Finland, OY Radiolinja AB), France (France Telecom, SFR), Germany (Deutsche Bundespost Telekom, Mannesmann Mobilfunk), Ireland (Telecom Ireland), Italy (SIP Italy), Luxembourg (P&T luxembourg), Netherlands (PTT Telecom), Norway (Norwegian Telecom, NetCom GSM A/S), Portugal (Telecomunicacoes Moveis Nacionais (TMN), TELECEL), Spain (Telefonica Spain), Sweden (Telia AB, Comvik GSM AB, AB Nordic Tel), Switzerland (Swiss PTT Telecom), Turkey (PTT Turkey), UK (Cellnet, Vodafone).

Of these, Denmark, Finland, France, Germany, Italy, Norway, Sweden, and the UK already had one or more networks in operation and offering service before the end of 1992.

GSM has created much interest world-wide and it was anticipated that it will be taken up by many more countries outside Europe. Hong Kong has indicated that it will license competitive service and the Far East in general is very interested in the standard. In total there were then 23 countries in Europe and a further 20 outside Europe who have agreed to use GSM. The forecast was that there would be about 13 million digital cellular subscribers world-wide by 1996; over half of these will be using GSM equipment. This would be on target to meet the 10 million subscriber base by the end of the decade, which was the forecast in 1987.

1.4.7. Current Status

At the beginning of 1998 there were over 65 million GSM subscribers, including pre-paid customers, in well over 100 countries. Including individual provincial and city networks in China and Russia, there were well over 350 operational networks, with a further 150 planned. [URL003] offers a thorough, factual and comprehensive review of developments, providing a solid and reliable picture of the status of GSM in a world cellular market still growing at rates in excess of 45% per annum.

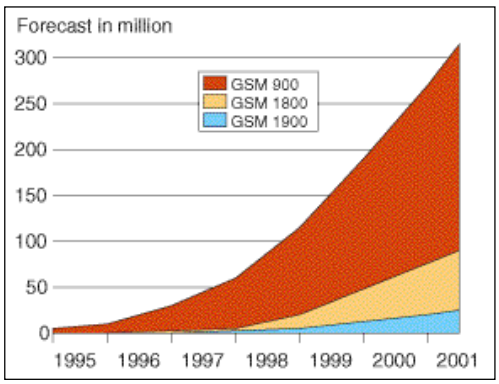


Figure 1.4.8. GSM Subscribers [1]

Countries with at least one GSM provider include the following 111: Albania, Andorra, Argentina, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Belgium, Bosnia-Herzegovina, Brunei, Bulgaria, Cambodia, Cameroon, Canada, Chile, China, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Estonia, Fiji, Finland, France, French West Indies, Georgia, Germany, Ghana, Gibraltar, Greece, Guernsey, Guinea Republic, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Ireland, Isle of Man, Israel, Italy, Ivory Coast, Jersey, Jordan, Kenya, Kuwait, Laos, Latvia, Lebanon, Lesotho, Libya, Lithuania, Luxembourg, Macau, Macedonia, Malawi, Malaysia, Malta, Martinique, Mauritius, Monaco, Mongolia, Montenegro, Morocco, Mozambique, Namibia, Netherlands, New Caledonia, New Zealand, Norway, Oman, Pakistan, Philippines, Poland, Portugal, Qatar, Réunion, Romania, Russia, Saudi Arabia, Senegal, Serbia, Seychelles, Singapore, Slovakia, Slovenia, South Africa, Spain, Sri Lanka, Sudan, Swaziland, Sweden, Switzerland, Tahiti, Taiwan, Tanzania, Thailand, Togo, Tunisia, Turkey, UAE, Uganda, UK, Ukraine, USA, Uzbekistan, Venezuela, Vietnam, and Zimbabwe.

[1] Figure 1.4.8. is taken from [URL001]

GSM infrastructure suppliers are companies as AirNet, Alcatel, Celcore, Ericsson, Italtel, Lucent, Matra, Motorola, Nokia, Nortel, Orbitel, Plexsys, Siemens, and Tecore.

Manufacturers of GSM terminals include: Acer, AEG/Matra, Alcatel, Ascom, Aselsan, Benefon, Blaupunkt, Bosch/Dancall, Dancall, Ericsson, GCS, Goldstar, Goldtron, Hagenuk/Cetelco, HB Electronics, Kokusai, Matra, Maxon, Mitsubishi, Motorola, Naewae, NEC, Nixxo, Nokia, Nortel, Orbitel, Panasonic, Philips, Sagem, Samsung, Siemens, Sony, Telital, Toshiba, Uniden, and Voxson.

1.4.8. The Future of GSM

Some areas of research and developement in this rapidly moving field are mentioned here together with references for further reading.

Radio spectrum allocation and assignment: The amount of radio spectrum available controls the number of competitors who can be licensed and the capacity of their networks, both of which have a significant impact on end user prices and services. National and international policies on allocation (the decision to set aside a particular band for cellular) and assignment (the decision as to which operators are given rights to that band) will have a major impact on the future growth of cellular radio. In [URL005] a look at some of the issues and trends in the short and long term for spectrum allocation and assignment is presented.

New and improved speech services: As in the case of the half rate speech codec, already foreseen in the original design, and later on specified by ETSI, the European Telecommunications Standards Institute, in January 1995 and in use since, future developements in coder technologiy will definitely lead to even less bandwith and higher quality.
For further reading [URL002] and [URL004] is given here.

New and improved other services: Higher user data bit-rates are achieved through multi-slot High Speed Circuit Switched Data (HSCSD), which will be available this year (1998) and General Packet Radio Services (GPRS), which will be available during 1999 and eventually offer a data rate of 115kbit/s.
As a packet-based technique, GPRS will enhance GSM data services significantly - especially for bursty Internet/intranet traffic - and make optimal use of available radio spectrum and channels. With the new GPRS nodes, direct IP access from the radio access network will be provided.
GPRS will allow many users to share the same channel and allow users to stay virtually “on line” all of the time: Radio resources will only be used when data is actually being transmitted or received. Call set-up will be almost instantaneous and users will be charged on the basis of actual data transmitted, rather than connection time.
Details to those and other enhancements can be found on [URL001] and on GPRS in particular in [HÄKA95].

1.4.9. Conclusions

For a detailed comparison of the GSM with our proposed system see chapter 1.7.!

1.5. Towards a Future Ubiquitous Wireless Mobile Network Architecture

This chapter summarizes [JLIU98] and [JLGM97] on which the present thesis is based.

Chapter 2 of the former covers mobility in general and gives an introduction to the *mobile agent* concept.

Chapter 3 describes the system architecture for an integrated multimedia mobile computing and communication environment consisting of a Mobility Support Server (MSS), Base Stations (BSs), and Mobile Terminals (MTs). The goals are formulated and the idea of micro- and macro-handover is introduced.

Chapter 4 presents an efficient mobile routing scheme which avoids triangle routing. The concept of a meta-network is introduced.

Chapter 5 focuses on efficient handover and route update schemes. The methods of pre-registration and pre-connection are used.

Chapter 6 presents the idea of a "location aware IP layer and a MSS with DNS extension" and improves thereby the current mobile IP by overcoming problems such as triangle routing, double encapsulation, multicast support, firewalls, and support for new protocols (RSVP).

Chapter 7 evaluates the system architecture and presents simulation results.

1.5.1. An Overview of the System Design Procedure

From the system designers point of view a *picocellular multimedia mobile computing infrastructure* with support for terminal mobility, and good scalability consists of the following:

- overall system architecture
- mobility management and control scheme
- mobile routing scheme
- definition, partition, and specification of mobility management and control functions
- mobility management and control protocol used among mobile agents
- selection of physical and link layer protocols and hardware for implementation

1.5.2. General

The essential elements of our architecture are the Mobile Terminals (MT), Access Points (AP, or Base Station, (BS)), the Mobility Support Server (MSS), router/switches/router-switches, and connections to other networks, as shown in Figure 1.5.1.

Each MSS provides service for a number of BSs. It is important to note that the MSS is not involved in processing each packet, but rather is only involved when a mobile terminal first comes to this cell, when it is about to leave, when it leaves, and just after it leaves.

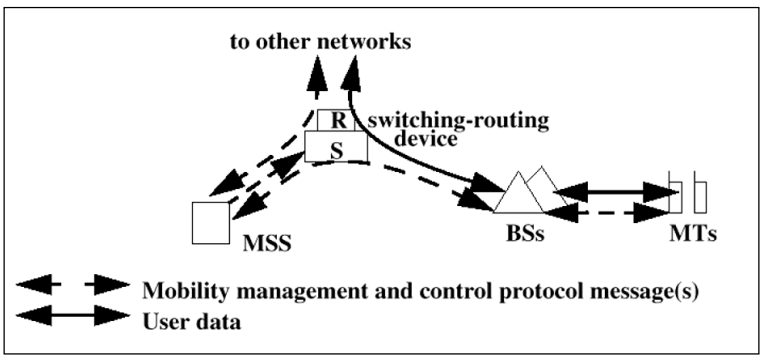


Figure 1.5.1. Anatomy of a Wireless Mobile System

In order for a future environment to support both high user data rates and high capacity, a lot of base stations will be needed for a picocellular system. Unless the cost of base stations is low, this would increase the cost of such an infrastructure to an unreasonable level as compared to a macrocell system. There are a number of possible ways to decrease the cost of an picocellular system infrastructure:

- Use of soft-radio or reconfigurable radio. So that we can make one kind of radio, therefore increasing manufacturing volume.
- Simplify base station structure by making them look like mobiles, but with a connection to the wired/fibered/... infrastructure.
- It all has to fit on one chip which is made in high volumes.
- Leverage something you already have to the greatest possible extent, thus re-use the existing facilities by attaching BSs to the LAN or in place of telephones.

For the *wireless physical layer* IEEE 802.11 [IEEE11] was been selected. It defines three possible physical layers: Direct Sequence Spread Spectrum (DSSS), Frequency hopping Spread Spectrum, or Diffuse IR.

For the *link layer* it is an open question what the optimal number of link layer addresses to have at each device would be. This will effect how easy it is to perform the transformation of frames which are received and transmitted by the access point.
The possible answers include:

- One per client and one for the device itself: De-Multiplexing is based on the link address. Thus, directly from the link layer address, we know which client to send which frame to. This requires that we have a pool of link layer addresses which we assign to the clients
  Or that we proxy with the clients' link address (since each client already has one link address). The one address for the device itself allows us to manage it.
- One for all clients and one for the device itself: We simply output onto the wireless interface all packets which arrive for the "clients" address. The remotes have to sort things out for themselves.

The second approach uses network layer addressing. The impacts, in particular on the MT's power consumption has to be examined further.

1.5.2.1. Micro and Macro Mobility

The efficient and reliable handover and route update scheme is based on the following idea: Handover is a potentially frequent event in a picocellular environment.
We have split handovers into two cases:

- In-LAN handover, or *micro handover*: This turns out to be relatively easy to handle, since all the entities are local and known to each other.
- Cross-LAN handover, or *macro handover*: This usually involves a routing update

To support macro mobility we assigned to the care-of address of each MT in our system an in-LAN multicast address. This kind of address allows the packet forwarding device to add branch routes to each of the target basestations, thus forming a logical multi-BS.

This has the following advantages:

- It facilitates handover: a branch route can be pre-established using the in-LAN multicast address before a handover, and branch routes can even be pre-established to several BSs.
- It enables the MSS to perform hierachical mobility management: using an in-LAN multicast address, we can limit the micro handover within a local area and perform partial route re-establishment while the MT experiences only a micro handover. This can greatly reduce the global route re-setup rate.
- Since the same in-LAN multicast address can be used within the local network, MTs do not have to re-register with its home MSS while performing micro handover within a local network. This can avoid a lot of globally exchanged messages.

1.5.2.2. Pre-Registration

To provide high performance macro handover, we propose an additional system threshold which extends the threshold concept used in cellular phone systems to two dimensions:

- One we call *min-threshold*, which has the same meaning as in cellular phone systems. It defines the outer boundary of a cell, i.e., the minimum received power sufficient for communication.
- *Max-threshold* defines an inner perimeter, at which the MT begins to look for other candidate BSs.

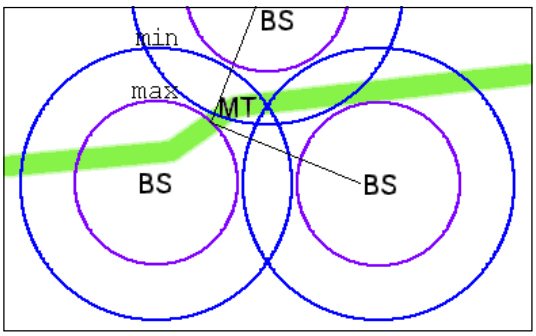


Figure 1.5.2. Min- and Max-Treshold

The strength of the radio signal decreases with the distance between the MT and the BS. In order to enable the MT to determine its location, every BS periodically transmits a beacon. If the MT can listen for different BSs beacons, this hint can be used to trigger the handover, before loosing connectivity.

An MT may use a mobile motion prediction algorithm which typically uses measurements of received signal strength indication (RSSI) or bit error rate (BER) and corresponding max-threshold and min-threshold to decide the likely candidate BS and selected BS. The system sets thresholds for RSSI and BER.

The establishment of pre-connection between an active BS and a candidate BS will eliminate the packet loss caused by the handover by avoiding an interruption of traffic. But on the other hand, the pre-connection imposes extra traffic burden on the local network. The extra traffic is mostly on the LAN since the over-the-air traffic need not be sent until the mobile actually is in the new cell. A reasonable choice of thresholds can minimize the costs.

The main purpose for introducing the pre-registration mechanism is to minimize the interruption and the packet loss caused by handover. The thresholds used to trigger this action must be chosen carefully to enable pre-register and pre-connection to be performed in time.

Since the distance between a MT and its communication partners may be very far, handover management messages as well as the usual packets may suffer a long round-trip delay. For the former, the delay will have a bad effect on handover performance, whereas a delayed user data packet is uncritical.

1.5.2.3. Network Management

For *network management* purposes the Simple Network Management Protocol (SNMP) is employed, and as a network management station we can just use a commercial system. We need to define the network management information base (MIB) for our devices and have to decide, where the network management agents should reside.

The network management system running on the wireless system should have additional functions to deal with these special features. The following wireless mobile system related states can be observed:

- general traffic statistics
- report the number of mobile terminals in a cell
- report the availability of access points
- report the event of service refusal due to lack of RF or IR resources
- report the (link) connection time
- report the event of service refusal due to security violation
- report the number of different type's of mobile terminals

1.5.3. The Mobility Support Server (MSS)

The main difference between a general purpose router and a Mobility Support Server is that the MSS concentrates on terminal mobility management related tasks. Such routers are expected to be near the network leaves rather than the core, so the routing should be simpler. An additional purpose is to help to simplify base stations as much as possible, as picocellular infrastructures will require a very large number of base stations.

The mobility management layer at the MSS processes mobile terminal registration messages, answers the domain query directed to it by a DNS server, and maintains three tables:

- *Visiting Register Table* (VRT): Stores the registration information related to the mobile hosts which belong to other networks, but have now roamed into this network.
- *Home Register Binding Table* (HRBT): Stores the revised domain name record information related to the mobile hosts which belong to this network.
- *Flow and Multicast Address Binding Table* (FMAT): Stores flows information which is associated with an in-LAN multicast address.

The size of those tables is proportional to the MTs active in the cells served by a particular MSS. Since this server posseses enough memory resources the table size does not limit the number of MTs.

1.5.4. The Access Point (AP)

An access point acts more like a bridge than a router, since there are at least, but usually only two interfaces for each access point, one radio interface and one fixed interface. Almost all frames received from the radio interface go to the fixed interface, and most of the frames received from the fixed interface are most likely going to the radio interface. Exceptions to this rule are SNMP-frames directed to the AP itself and broadcasts as mentioned hereafter.

For the access point, the basic data paths are:

- from radio interface to ethernet interface,
- from ethernet interface to radio interface, and
- from ethernet interface to UDP for management of the AP itself

However, it is important that the bridge filters out many of the broadcasts when they are irrelevant to the device(s) on the other side of the bridge. For example, the following filters should be supported: Eliminate specific nodes, Filter protocols (IP, IPX, NetBEUI, ...), Broadcast traffic filters (IP/ARP, SAP, and LSP).

The basic function that the operating software needs to perform in an access point is to bridge the frames between a radio and a fixed interface. Some limited data processing for some protocols such as SNMP is also needed (only, if the SNMP functions are implemented in the access point itself). It may be possible to derive the required information from other devices and hence avoid having to actually support an SNMP agent in the access point.

Since we do not need complex routing functions in an access point (because we have few choices of which interface to send a packet to and because we know that there *are* routers which we can send packets to if we need more complex routing decisions made), we can greatly simplify the network layer software's structure in our access point operating software. Here what we need is a very limited access point routing function to perform the above mentioned limited tasks.

So, we can use "frame in, frame out" to illustrate most of the data passing through an access point, except for the control and management protocol messages. For a two port device, the two input to output paths are largely independent.

Having independent paths means that the memories can be simpler, management of the queues becomes simpler, interrupts become simpler - in fact, one does not strictly need interrupts as the machines can run independently.

Thus, the access point's major operating software functions are:

- bridging frames between interfaces
- performing necessary data processing
- perform SNMP functions
- exchange the necessary protocol messages for supporting mobility

1.5.4.1. Optimizing and Partitioning the Access Point

The bridging algorithm (denoted as R⇒E in Figure 1.5.3.) itself can have one of two forms:

- *Transparent* (i.e., the identity transformation): Each of the bridges has the same apparent MAC address for the wireless side as the wired side. A transparent bridge must implement the spanning tree protocol.
- *Non-Transparent* (i.e., turning one sort of address into another): If we assume that we have a pool of ethernet MAC addresses for use with this access point, we can assign a new ethernet MAC for each new radio MAC address which we are going to serve.

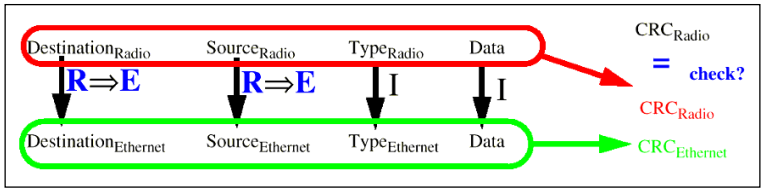


Figure 1.5.3. Frame from Radio- to Ethernet Interface

As can be seen from Figure 1.5.3. the important operations done on each frame are: field extract, associative lookup, field replacement, CRC computation, and filtering.

The link broadcast address maps to the broadcast address on the other link. It is used by the higher level processes to get a specific link layer address. In the case of IP, this is generally done via the ARP protocol.

New devices which are to be served by an access point initially use the broadcast MAC address to get service - as they don't necessarily know any other link layer address besides their own.

Since frames are only going either from the radio to the ethernet or from the ethernet to the radio, we can split these into two different machines. The only interaction in both cases will be at the MAC layer of each media, as otherwise the paths are independent (see Figure 1.5.4.).

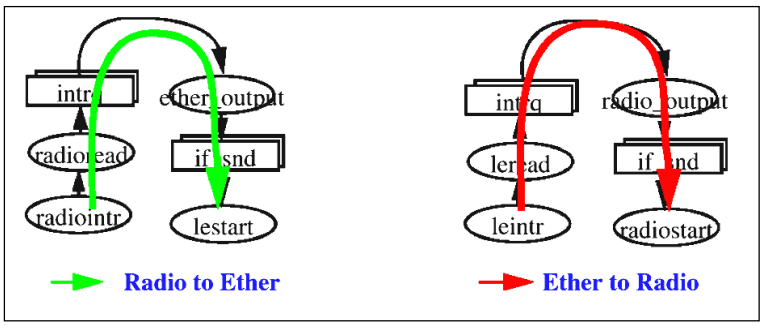


Figure 1.5.4. Reduced Bridging Paths

The access point contains three parts:

- radio network interface software and hardware
- fixed network interface software and hardware
- state machine which processes frames and passes data between the two interfaces.

It is suggested to split the functions between the access points and an access point server:

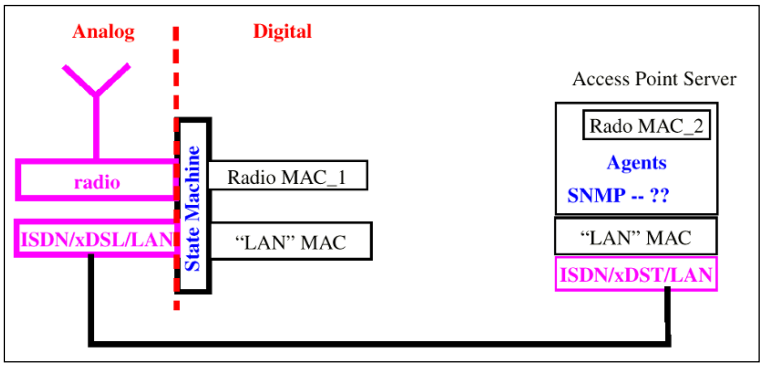


Figure 1.5.5. Access Point and Server Structure

How many transceivers should be used in the radio part MAC? We think it is best to have two radios. One is used as the current communication channel and the second is time multiplexed:

- to look for other access points,
- when bursts of extra bandwidth are needed,
- to acquire information which is used to determine your location and heading,
- as a secondary channel to listen to (and perhaps transmit) contents which is not available, via the main communication channel, and
- listen to non-traditional communication emissions in your environment, etc.

1.5.4.2. Access Point Implementation Issues

Implementation of the ethernet *LAN MAC*: Full custom VLSI or use of existing off-the-shelf MAC controllers is suggested. Full-custom VLSI chips or a macrocell in a library require 15K-17K gates, while designs for 10 and 100 megabit per second controllers require approximately 55K gates. This can further eliminate the unnecessary bus interface and external DRAM control of most chips. Alternative implementations using FPGAs or software implementations (as widely used for modems) with generic processors, use of DSPs, or RISC processors optimized for implementing communication protocols have to be investigated further.

Our interface will only send frames to an access point server, a correspondent host or another access point on this LAN segment, or to a switch or router to reach another station. Thus we might reduce the controller to handle only well formed packets, since this will substantially reduce the size of the wired LAN MAC.

An open question is the possibility of partitioning the *radio MAC*. IEEE 802.11 is likely going to become a widely used world wide standard for wireless LANs. It has the most complex MAC protocol of any of the 802.x standards. Is it possible to split the radio MAC functionality between the access point and the access point server? An open problem is to understand if and how this partitioning can be done.

The basic technology to support switching between multiple MACs is quite straight forward:

- FIPSOC supports two stored configurations on one chip (“Hardware Swap”)
  See URL `http://www.sidsa.es/fipsoc.htm` for details!
- Motorola offered an interesting FPAA technology on
  URL `http://www.mot-sps.com/fpaa/index.html`

It is not clear if either of these chips can support the necessary performance to do the analog processing which we would need for our physical interfaces. However, it represents an interesting possibility of putting even many of the analog functions into reprogrammable logic.

Implementation of the *wired physical layer*: LSI Logic has a set of cells to do almost all the analog processing required for 10baseT.

See URL `http://www.lsilogic.com/mediakit/unit3_1d.html` for further details!

1.6. Introduction to "Comnet III"

The **Comnet III** environment, release 1.3 from *CACI Products* was used extensively throughout this thesis. Therefore I will give a brief introduction to this system, and present it in the style of a manual, to provide an easy startup for future work. The comprehensive reference is the user's manual [COMN96].

It shall be noted that the model description- and programming language unterlying Comnet III is quite complex and constitutes an own product, namely Modsim II.

1.6.1. Introduction to Network Simulation

Comnet III lets you analyze and predict the performance of networks of any size quickly and easily. It supports a building-block approach where blocks are objects you are familiar with in the real world. These blocks are available from libraries and their parameters are adjustable to correspond with the *real* object.

As networks become larger and more complex and new technologies and types of applications are introduced, as in this thesis, the design and management of this system becomes an even more challenging task. Only simulation can help to assess alternatives at an early stage and adapt the design of the network with minimal risk. It would be expensive, impractical or even impossible to build three alternatives, pick the best, and throw away the other two.

1.6.2. Description, Applicability, Availability and Overall Approach of Comnet III

Comnet III is a performace analysis tool for computer and communication networks. Based on a description of a network, its control algorithms, and workload, Comnet III simulates the operation of the network and provides measures of network performance.

No programming is required. The single package supports all phases of model design, model execution, and presentation of the results. The data needed to describe the network comprises parameterized nodes and links, characteristics of the workload placed on the network, and selection of various protocols on different layers. Comnet III also supports the use of the non-standard topologies and routing algorithms which were needed in our proposed network architecture.

Comnet III is designed to accurately estimate the performance characteristics of computing and communication networks. Estimating means in particular a dynamic simulation of the network by using simulated traffic. During this execution phase data is gathered and various reports are generated. Comnet III also provides for displaying textual reports and diagrams and offers some standard statistical methods.

We are running Comnet III on a SUN Ultra-10 (SunOS 5.5.1) workstation equiped with 64MB of memory - since 32MB worked out to result in unacceptable slow execution speed.

1.6.3. Introduction to the Frontend - The Main Toolbar

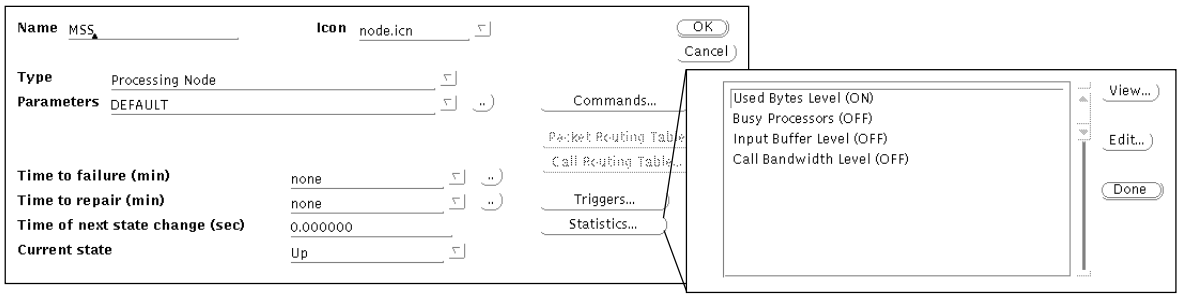
Most of the buttons in the main toolbar can be used quite intuitively. The first four buttons functions are: Create an *new* project, *open* an existing one, *save* the current one and *print* the current workspace. *Cut*, *copy*, *paste* buttons come next and work in the well known manner.

Clone duplicates a selected item and it's properties as many times as you want. The following button allows editing the *properties* of the the selected object. The same can be achieved by doubleclicking on the object. The cross *deletes* an object.

The three buttons on the very right control execution of the model. *Start* and *stop* a simulation run and take a *snapshot* while running.

1.6.4. Nodes in Comnet III

This is the node icon in Comnet III. It represents a single network node (i.e. workstation) in the model. Nodes can even have processing capabilities, memory and hard disc storage assigned to them. In the future this will enable us to associate a service time to messages, but so far we have not used these features. In our models, nodes are only sending, forwarding, and replying to messages. The properties of a node are usually left unchanged from their default values which are shown in the snapshot below.

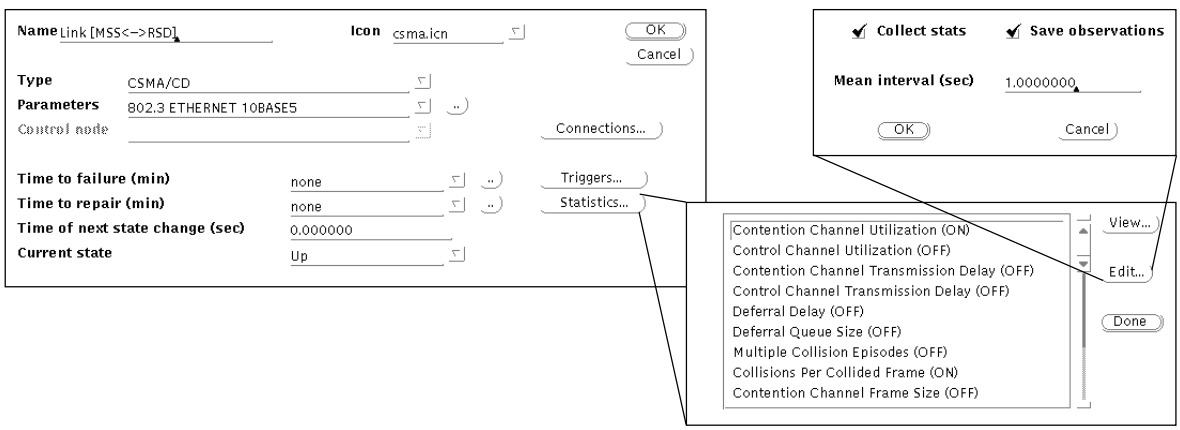


The Statistics button lets you enable the logging of parameters and the same window leads to the display section which is explained in section 1.6.7.

1.6.5. Networks in Comnet III

The link icon represents networks and links in Comnet III. The properties of this object let you select the type of the network among several standards. Ethernet or 100Base-T are used as it is, for radio links (wireless links) the Ethernet defaults were changed to a 1Mbps bandwith link. Regardless of the graphical representation of a network and its links as a hub, the topology can be a bus or ring as defined in the form.

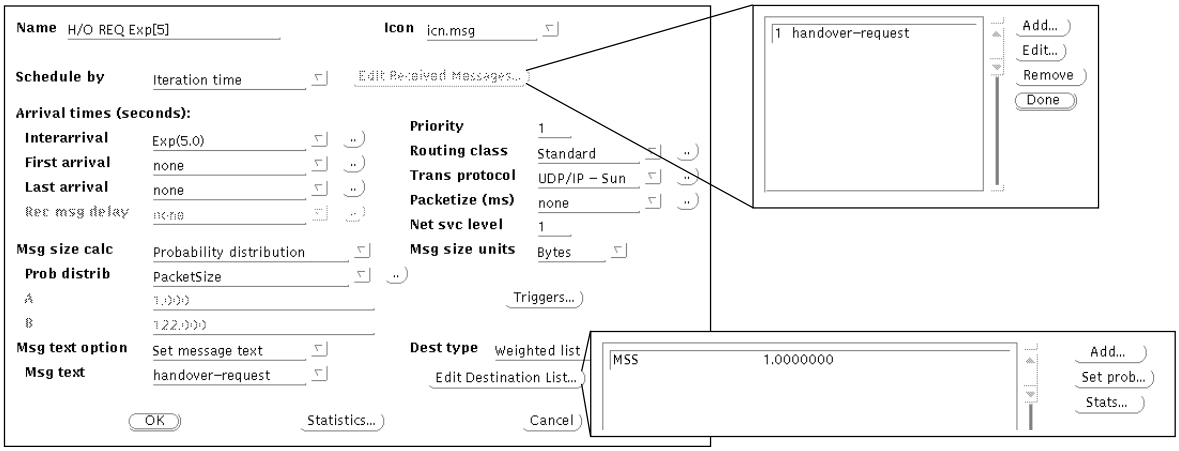
Various reports can be requested via the Statistics button; the "Contention Channel Utilization" is usually the most interesting one. Care must be taken when choosing the "Mean Interval" for the statistics, since events of interest could be missed or averaged away. One second is used as the default.



1.6.6. Messages in Comnet III

In Comnet III one or several message-sources are attached to a node and thereby define the set of messages a node sends, replies to, or forwards and how it does so. In the basic case a message with the string value specified in *Message text*, a certain size, and time-period is sent during the simulation run. The latter two can be an expression that describes a complex statistical distributions or a simple constant message size and interarrival time.

Usually there is a single destination of the message entered in the *Edit Destination List* dialog, but multiple destinations and ways to choose one of them can be specified .



If a node should also respond to a message sent to it, *Schedule by* must be set to *Received message*. That enables the *Edit* button next to it which allows the selection of messages this message should respond to.

For example to measure the delay of a particular message, a report of this statistics can be requested and after execution the report can be viewed via the statistics dialog.

1.6.7. Reports, Trace Files and Diagrams in Comnet III

Comnet III can generate an abundance of textual reports, showing events or statistics from every single node and link of the network, but we do not use this excessive amount of detail. Rather we present results graphically as diagrams or histograms. Comnet III offers a good interface to produce such output.

However the complete log of events can be dumped to a file via the *Simulate/Trace...* menu option. We will use this trace files to generate test vector data later on in this thesis.

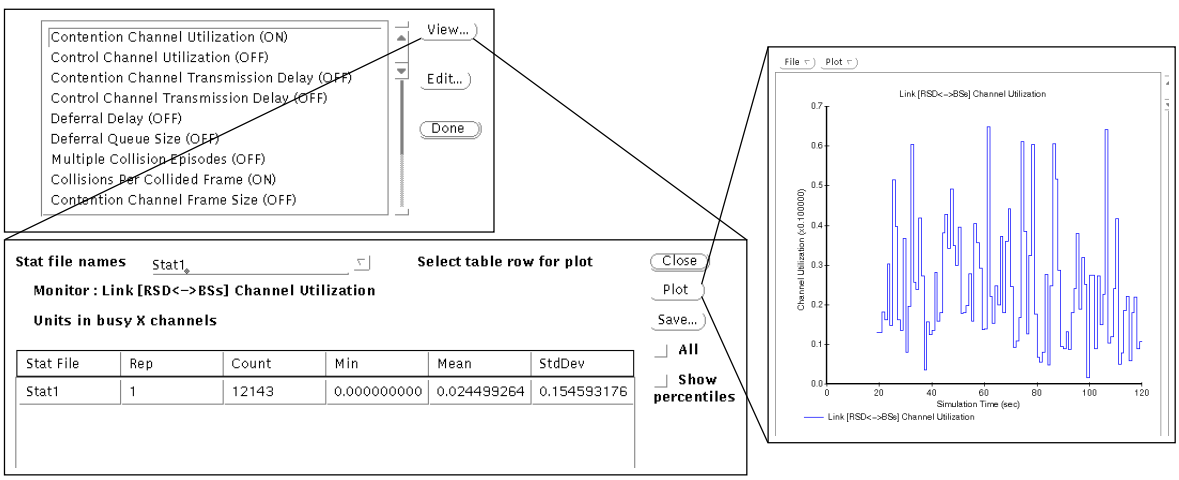


Figure 1.6.1. Plotting reports in Comnet III

After executing the model, the *View...* button in the *Statistics...* dialog is enabled and allows us to select a report that is switched *ON*. Clicking *View...* leads to a form where the gathered data is presented as rows in a table. A row has to be selected by clicking on it before the *Plot* button is enabled. This finally renders the diagram as shown in Figure 1.6.1.

Plots can be smothed, zoomed, or presented as a histogram via the *Plot* menu.

The *File* menu offers a *Print...* to PostScript function.

In chapter 2.7. the objects in our model are explained in detail and with the intention of making clear how the parameters known in the real world are entered and used in the model.

1.7. Lessons learnt from the GSM

Comparing a GSM network with our proposed network architecture on the system level there are certain analogies and differences that are pointed out hereafter.

Clearly the GSM's access-point, known as Base Transceiver Station (BTS), corresponds to our basestation (BS). The intermediate level of Base Station Controllers BSCs has no equivalent in our approach. GSM's Mobile services Switching Center (MSC), centralized in it's nature, finds it's analogy in the combination of the home- and the foreign-agent and the backbone router (RSD). The functionality is thereby carried out in a more distributed manner. The various centralized databases and some services of the GSM-system (HLR, VLR, AuC) are also distributed among the agents. The dedicated overlaid GSM administration/maintenance network (OMC, NMC, ADM) is not needed in our case, since all nodes of the network are already connected by a TCP/IP network and can be administered using SNMP.

Obviously our system takes a more distributed approach than GSM did. The reason is that decentralized systems scale better than centralized ones and GSM lacks this property when it comes to large geographic scale coverage. For example, inter-BSC handover is an compareably infrequent event in GSM, that demands for MSC intervention, but the analogous macro-handover occurs quite often in our mobility supporting network.

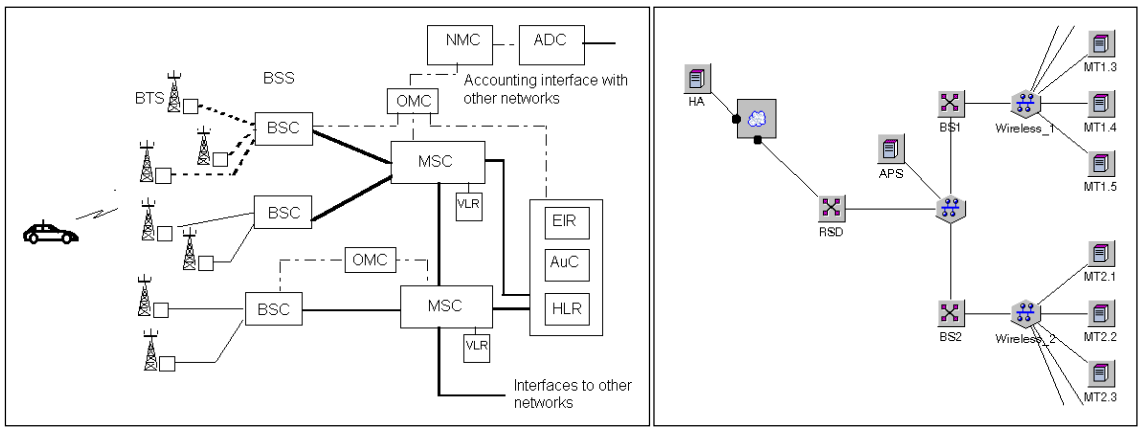


Figure 1.7. Comparison: GSM vs. Our Network with Mobility Support

2. Chapter II – and Deliverables to the MEDIA-Project

2.1. Link Utilization and Message Delay Simulation

2.1.1. Introduction to the Simulation and Parameters

To examine the link utilization and the message delay occurring in the proposed protocol architecture, as presented in [MD1297], we used the same example system parameters as in [MD1297] section 7.4, and stated below, and the resulting statistical models provided in [JLIU98] section 7.4, and constructed a simulation model using the COMNET III simulation environment.

System coverage area:	36 km^2
Population Density:	400 km^{-2}
MobileTerminal owner:	80 %
Power on percentage:	80 %
Velocity (average):	10 km/h
Cell radius:	20 m

Random user movement is assumed.

2.1.2. The Simulation Tool: COMNET III

COMNET III is a performance prediction and analysis tool for computer and communication networks. Other methods than simulation are inappropriate because of the stochastic nature of network traffic and the complexity of the total system. Based on a description of a network, its control algorithms and workload COMNET III simulates the operation of the network and provides measures of network performance. Network descriptions are created graphically and a single package performs all functions of model design, model execution, and presentation of the results. By using discrete event simulation methodology, COMNET III provides realistic and accurate results. The alternative to discrete event simulation is to use traditional mathematically based analytical methods which cannot cope with the effects of queuing, event interdependence, and random variance when analyzing complex communication networks. Furthermore COMNET III's logfile facilities constitute an excellent tool for testvector generation for subsequent models. For illustration and details of the various functions of COMNET III see [COMN96].

2.2. The 1st Simulation Model

In the protocol architecture proposed so far, the links connecting the Mobility Support Servers (MSSs) might become a bottleneck in our design [MD1297]. The sole purpose of these simulation efforts is to provide an estimation of the added load on the backbone-network by the control-messages *only*. The topology of our example-network is depicted in Figure 2.2.1.:

The MSS on the left side is connected with a set of 6 basestations (BSs) on the right via the link under examination "Link [MSS<->RSD]" and a router (RSD). We set the bandwith of the channel between RSD and MSS as 10Mbps.

The simulated scenario of "no user data - only control messages" already represents the worst case for link utilization of the link [MSS<->RSD], since any additional user-data-traffic is not routed over that link and thus by its utilization of the wireless network can only decrease the load on the link [MSS<->RSD]. We consider only 6 BSs connected to the MSS rather than the entire 28647 BSs to limit the simulation complexity. Note that given the results, that one MSS could serve more than 6 BSs, and the results scale accordingly.

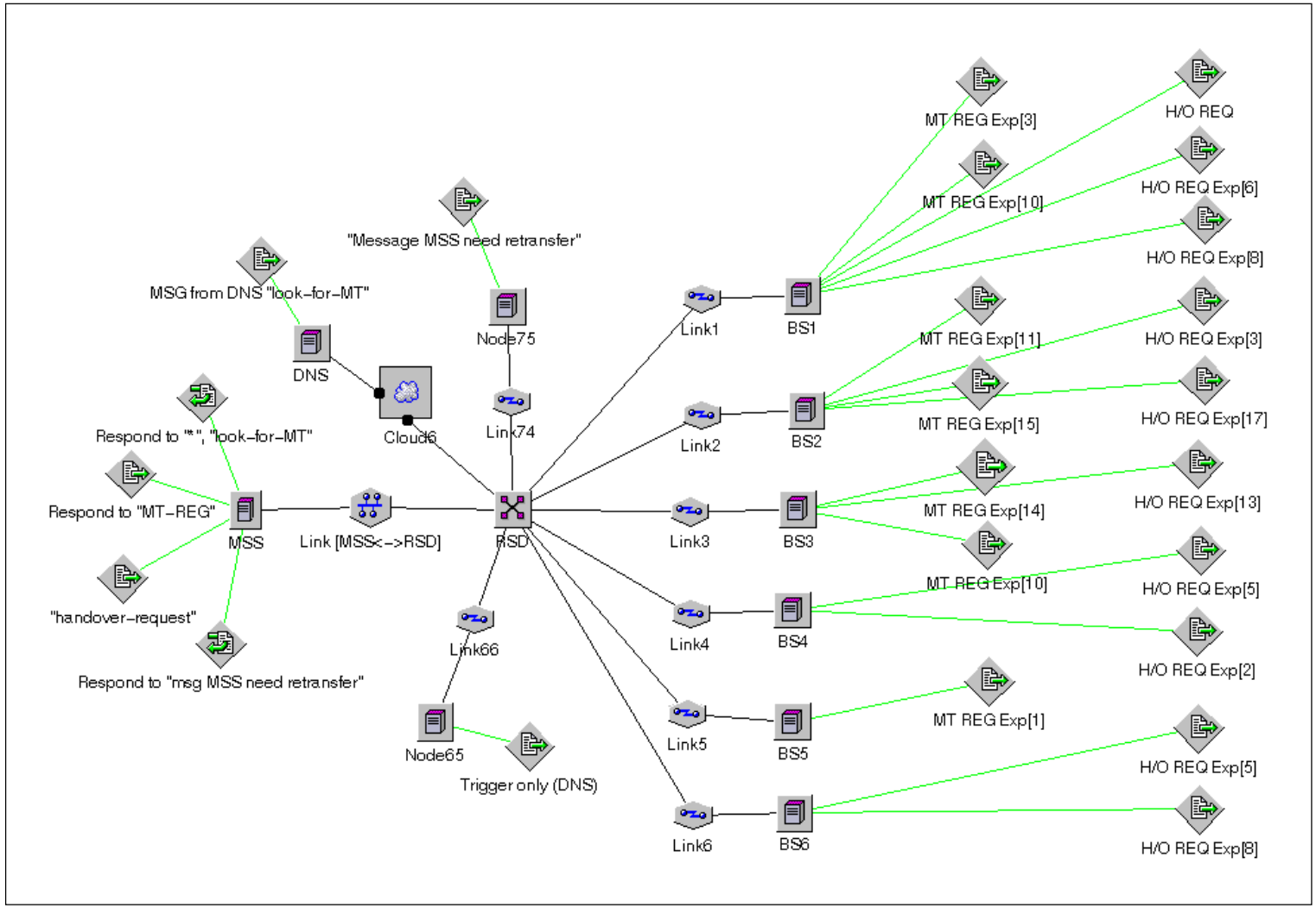
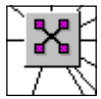


Figure 2.2.1. Network Topology

This is a “message source”
It describes the properties of the message, which is sent via a node.

This is a network “node”. It is the source or destination of messages.

This is a network “link”.
It connects nodes or routers. It describes the network properties.



This is a network “router” with the usual functionality.

Several messages, shown as “message source”-icons in Figure 2.2.1, are exchanged - in particular:

- A mobile-terminal-registration-message “MT REG” is sent from a MT to the MSS (as shown in Figure 2.2.2.) which sends a message to the DNS (as shown in Figure 2.2.3.). The DNS responds with a “look for MT” message to the MSS (as shown in Figure 2.2.4.).

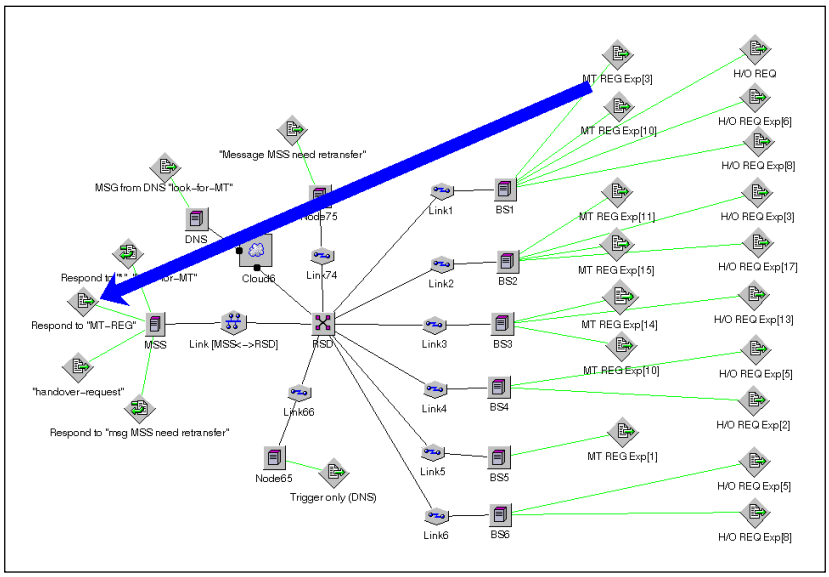


Figure 2.2.2.
Mobile Terminal registers via Basestation with its MSS.

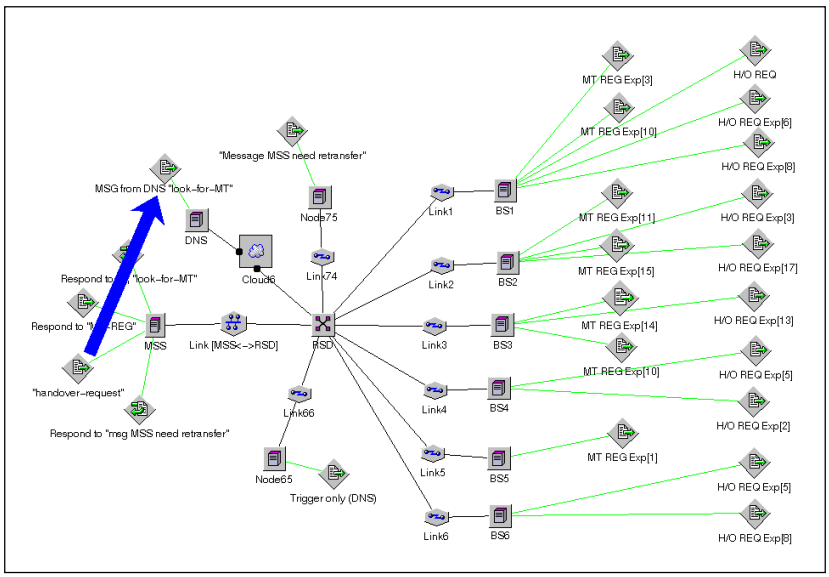


Figure 2.2.3.
The MSS updates the DNS.

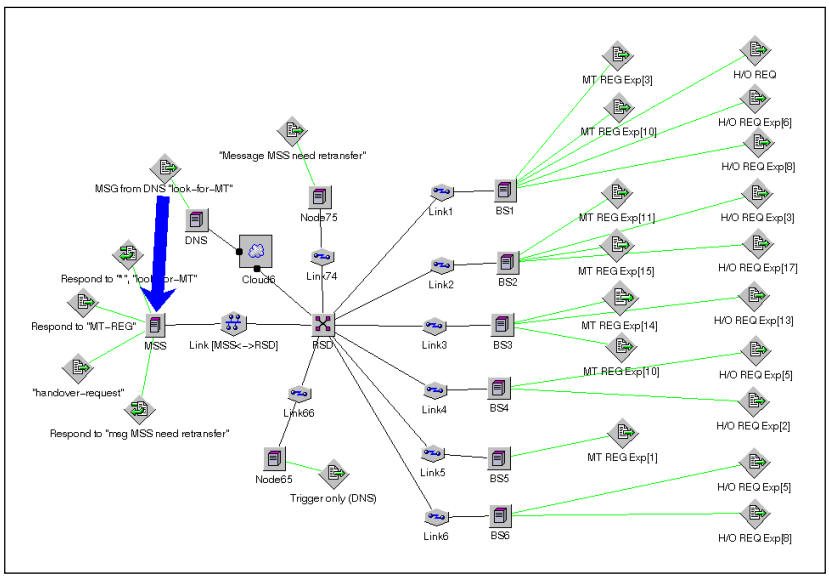


Figure 2.2.4.
The DNS sends its response to the MSS.

❖ A handover-request-message “H/O REQ” is sent from a MT to the MSS (as shown in Figure 2.2.5.) which sends this message via Node65 to the DNS (as shown in Figure 2.2.6.). The DNS responds with a “look for MT” message to the MSS (as shown in Figure 2.2.7.).

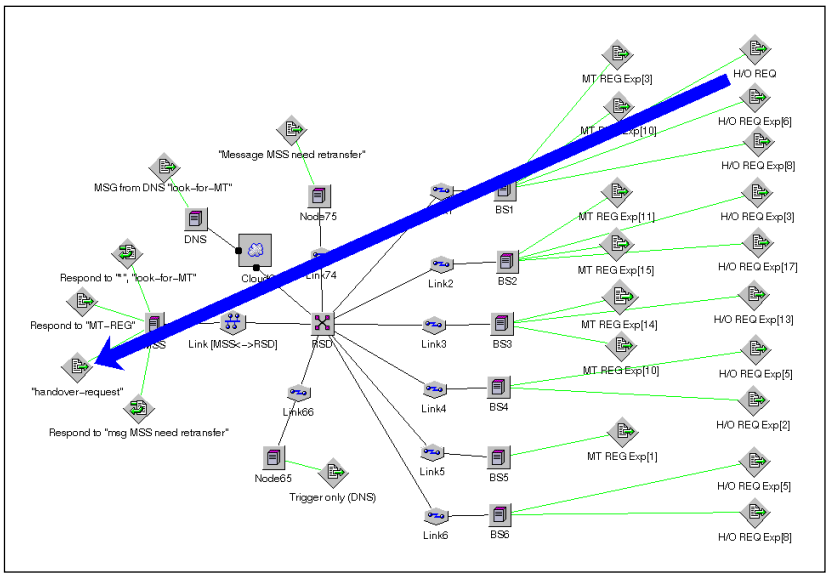


Figure 2.2.5.
Handover request sent by the Mobile Terminal to the MSS via a Basestation.

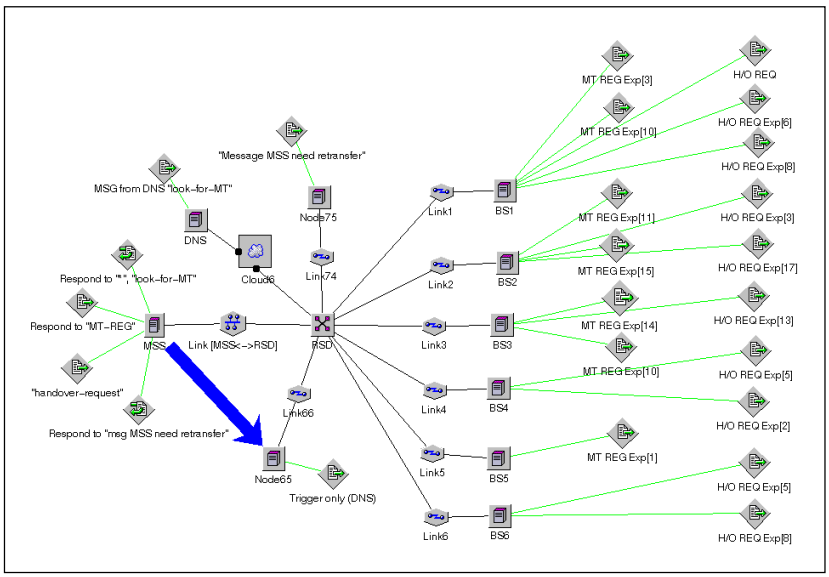


Figure 2.2.6.
The MSS updates the Router (Switch) to forward traffic to the Mobile Terminals new location.

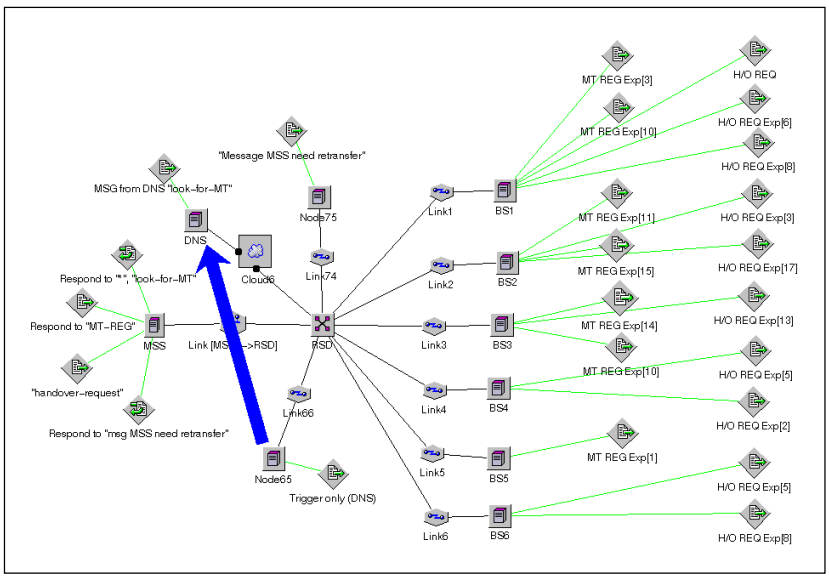


Figure 2.2.7.
The MSS update also triggers an update of the DNS.

❖ A third type “message MSS needs retransfer” occurs in a macro handover situation and is sent from Node 75 (which has the role of the former MSS serving the MT) to the MSS (as shown in Figure 2.2.8.) and the reply is shown in Figure 2.2.9.

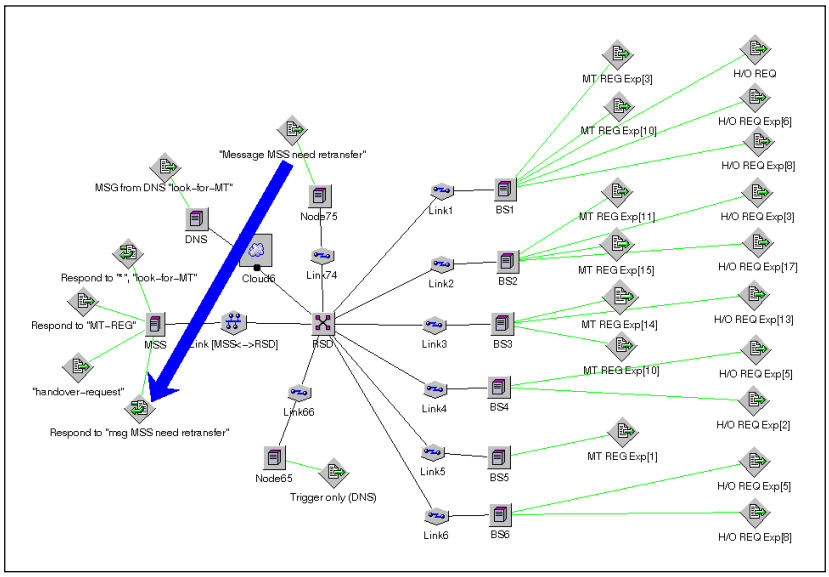


Figure 2.2.8.
The former MSS sends a message to the new MSS.

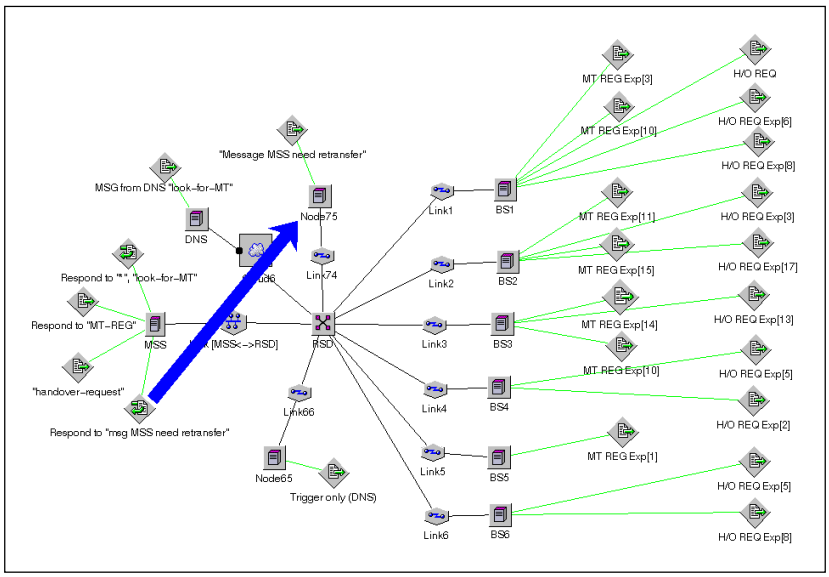


Figure 2.2.9.
A response is sent back.

2.2.1. Simulation Results

The model generates an average of 16 messages per second with 32 bytes per message. These numbers give an average utilization of 0.04% of the 10Mbps-link between MSS and RSD.

Using this simulation we examined the utilization of the link between MSS and RSD. Figure 2.2.10. shows the resulting link utilization and Figure 2.2.11. shows the distribution of utilization levels. As can be seen from the graph in Figure 2.2.10., the protocol messages take a limited bandwith while performing their function. The peak channel utilization is less than 0.14%. Assuming that, the 10Mbps-Channel can carry up to almost 11500 messages per second as the messages are each around 100 bytes in size.

Another important aspect is the protocol message delay while helping MTs performing their handover. Our simulation results also show the delay incurred by transferring protocol messages between the MSS and MTs. Figure 2.2.12. shows the delay of protocol messages exchanged between “Base Station” BSs and MSS. There are two kinds of protocol messages that are frequently exchanged between MSS and BSs. One is an MT registration request message and the other is a

handover request and control message. Only the latter one is time critical, since extensive delay in the handover process causes an apparent link failure as seen by the MT. As can be seen from Figures 2.2.12. and 2.2.13., our simulation results show that the maximum delay of this kind of message in our model system is less than 1ms. In a general system, this delay may cause some bad effects and worsen the handover situation. However, by using our pre-registration mechanism, shown in detail in [JLIU98] chapter 5.5, our system can easily eliminate the effect caused by this message delay.

A conclusion from this simulation work (6 BSs cause a peak channel utilization of 0.14%) is that the link connecting the MSS and the RSD is capable of carrying the aggregate load of 4285 BS's control traffic, hence considerations of data traffic via the switch will dominate the number of BSs attached to an MSS.

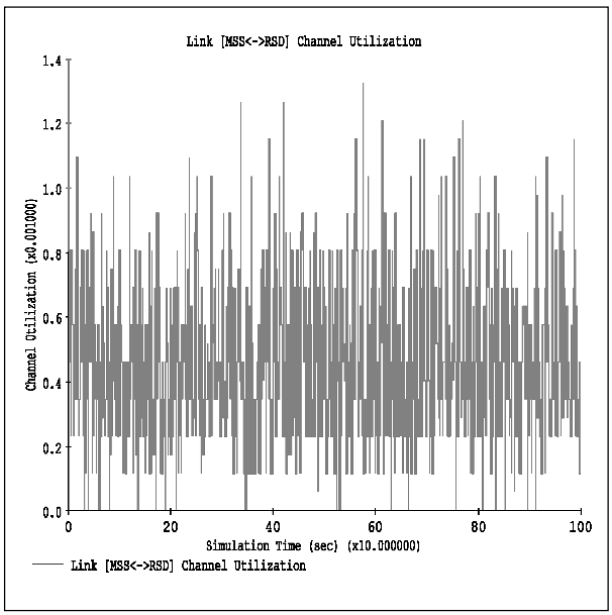


Figure 2.2.10. Channel Utilization

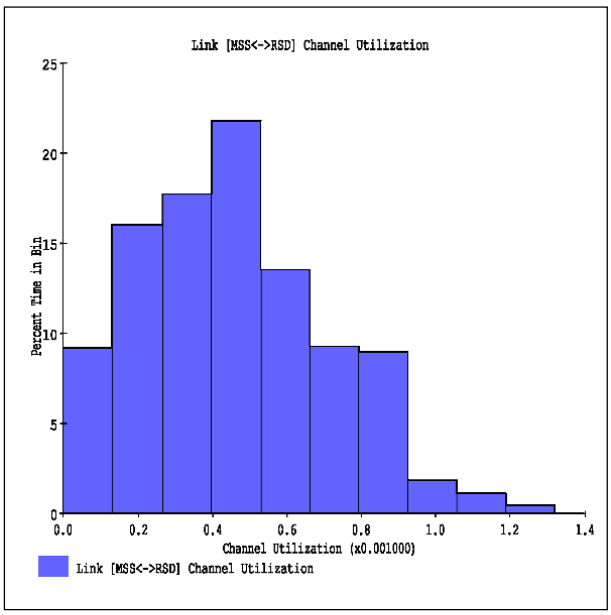


Figure 2.2.11. Channel Utilization - Histogram

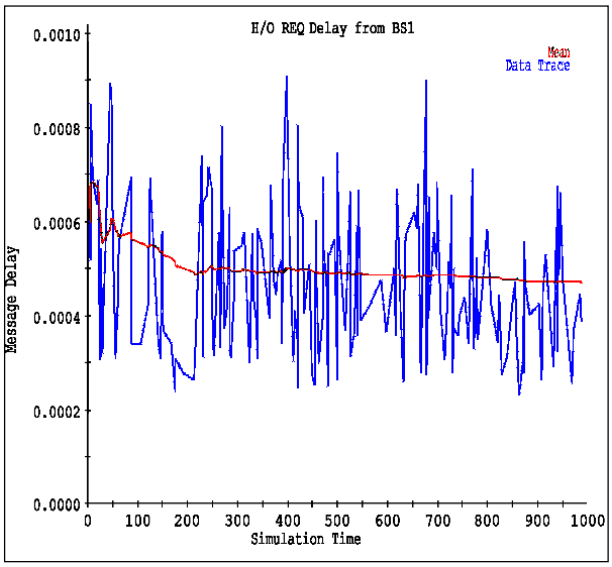


Figure 2.2.12. Message Delay

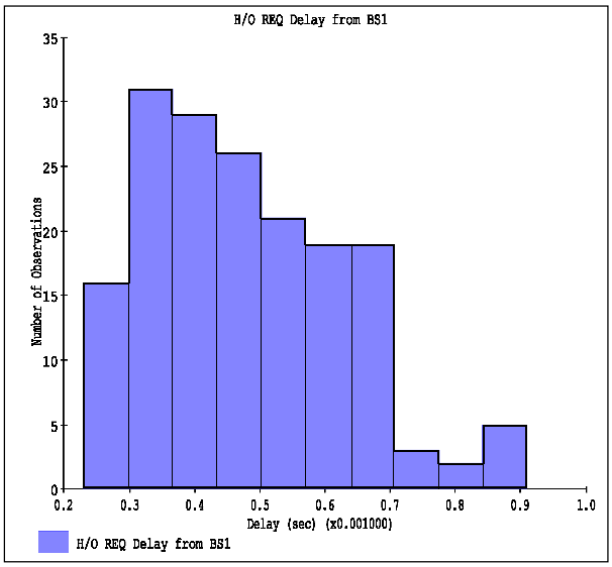


Figure 2.2.13. Message Delay - Histogram

2.3. A New Model – the 2nd

The extended model presented in this section and depicted in Figure 2.3.1. includes the following enhancements over the first modelling approach:

- Modelling of the wireless network part:
  Two 1Mbps CSMA/CD channels "Wireless_1" and "Wireless_2".
- Additional load on the backbone network:
  "Traffic Source" utilizing 10% or 30% of the network with simulated user data traffic.
- More realistic assumption of the packet size and protocol:
  30 bytes, UDP

The exchanged messages are the same as described in Section 2.2.

2.3.1. Considering the Amount of BSs per MSS

The maximum length of a network segment and the throughput of an ethernet network naturally limits the number of stations connected to a segment of the network and thereby the number of MTs per MSS. Assuming a maximum length of 186m and the hub sits in the center of a circle of that radius, a network with an area of 108.687 km^2 is spanned. Since in our simulation, each picocell has 20m radius, we need 100 BSs to cover the whole area.

Looking at Figure 2.3.2. it's also important to remember that we have assumed a 1Mbps bandwidth for each wireless link, shared between all the MTs assigned to a particular BS. Together with a reasonable backbone network utilization of 30%, the aggregate user traffic over the backbone should not exceed 3 or 30Mbps for ethernet or fast ethernet respectively. If we assume full utilization of the BSs and a fast ethernet backbone no more than 30 BSs can be served by one MSS.

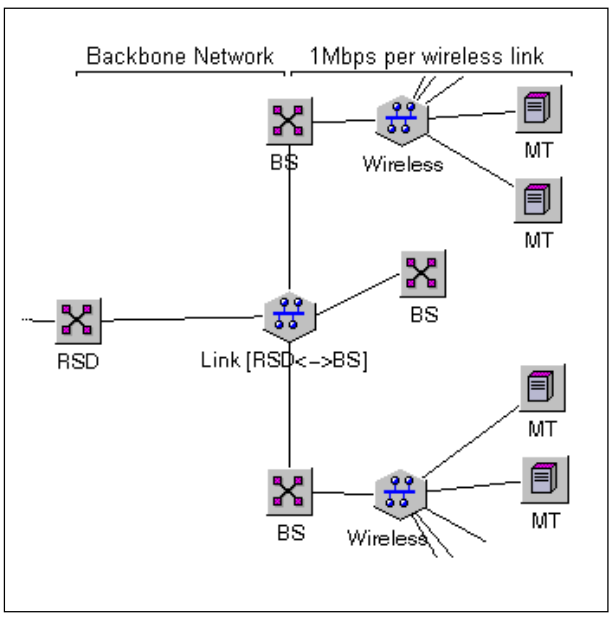


Figure 2.3.2. Bandwith Consideration

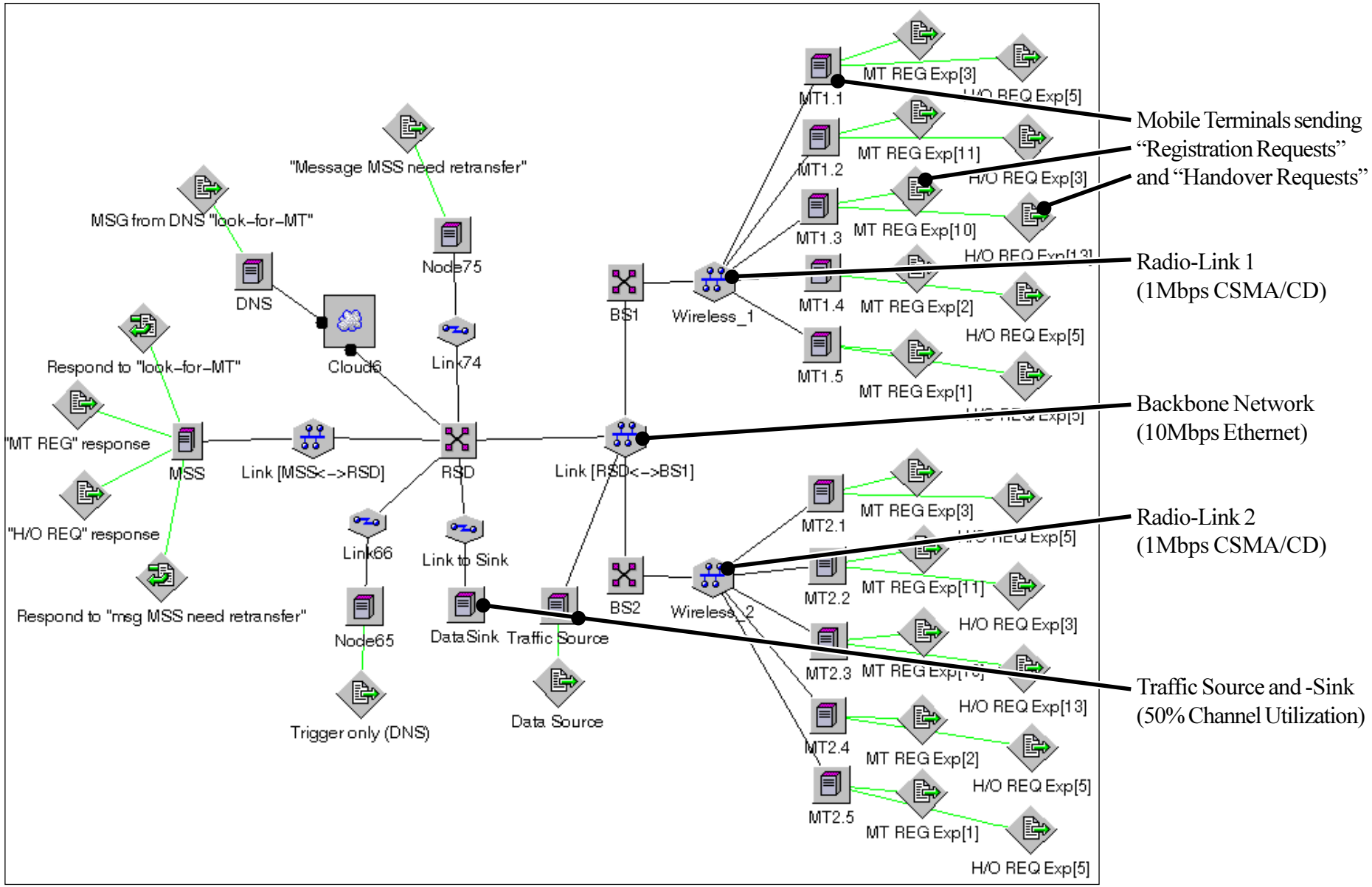


Figure 2.3.1. Network Topology

2.3.2. Simulation Results from the 2nd Model

Figure 2.3.3. shows the Link [MSS<->RSD] utilization as there is no extra traffic on the backbone Link [RSD<->BS]: The utilization is always less that 0.3% (median 0.13%). The message delay for this condition is 0.8ms at a maximum.

Next we consider a poorly utilized backbone network, as "Traffic Source" puts 10% extra load on it. The plot in Figure 2.3.5. indicates, that the peak channel utilization on the link between MSS and RSD is as well less than 0.3%. So this channel should be capable of taking the load of 3300 MTs, which is less than the earlier result (4285 BSs) and explainable by the more realistic assumption of the packet size and protocol as mentioned in the list above.

The "Message Delay"-plot in Figure 2.3.6. shows, that the worst case delay 0.92ms has slightly increased. 1ms can be taken as a safe maximum delay for a poor utilized network.

The situation does not even change significantly as the network experiences 30% heavy load (which can be considerd the maximum stable load for ethernet). The Link [MSS<->RSD] utilization is less than 0.35% as shown in Figure 2.3.7. and Figure 2.3.8 shows that the measured delay for handover requests still lies in the range less than 1ms.

2.4. The 3rd Model

Up to now only 'simulated' user data in the form of additional traffic on the backbone network has been considered. But in fact most of the backbone traffic consists of user data sent to and received from the MTs. A suitable model for user data traffic can be found in [BJUR98] and gives an Exp(2) distributed Inter Arrival Time (IAT) and roughly a message size with Normal(2.5k, 3k) distribution for an eMail application.

Another step towards a complete and realistic model is the consideration of the MT's sleep-cycle - which means: These nodes are powered down most of the time to reduce their energy consumption. COMNET III can cope with this behaviour by simulating a "node failure" and it's "repair" in appropriate cycles. To explorer the parameter space a cycle time of 1s and two sleep/awake-ratios are used: 50% and 90%.

The concept of an Access Point Server is introduced in the following section and a representation thereof is integrated in the described COMNET III model, which is depicted in Figure 2.4.1.

2.4.1. Buffers in the MTs or in an APS?

The model of our MTs demands for a considerable amount of memory used for buffering messages and packets even on the MAC layer. Having the cost of memory on chip in mind, a cheaper solution would be to move this buffer memory from the MT to an Access Point Server (APS) as shown in Figure 2.4.2. The details of this partitioning and the necessary protocol are yet to be defined, but the necessary buffer space in the APS and the additional backbone load due to some message overhead are simulated in the following section.

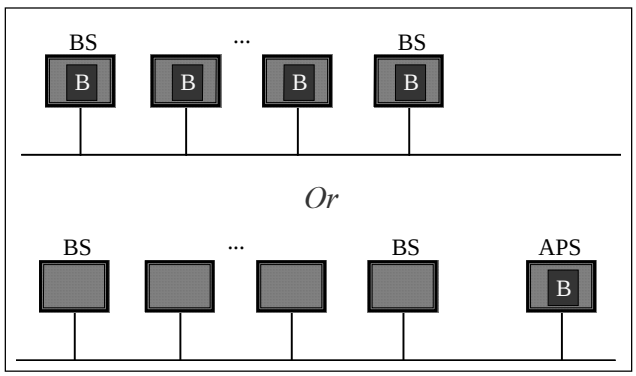


Figure 2.4.2. Buffer Memory

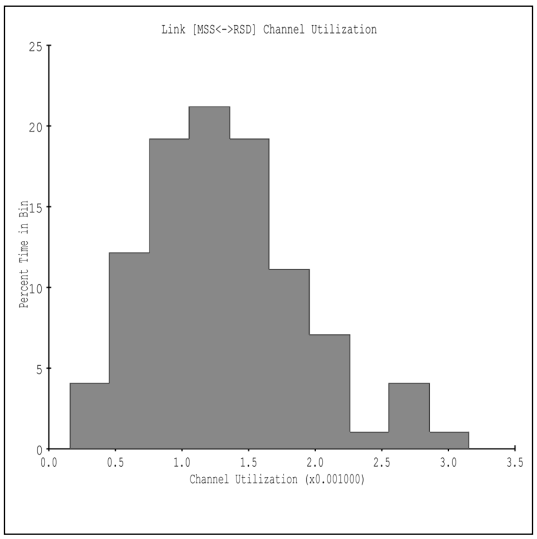


Figure 2.3.3. Channel Utilization @no extra Utilization

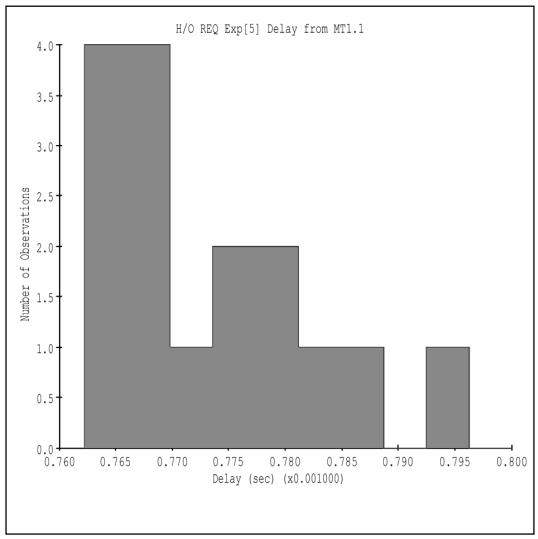


Figure 2.3.4. Message Delay @no extra Utilization

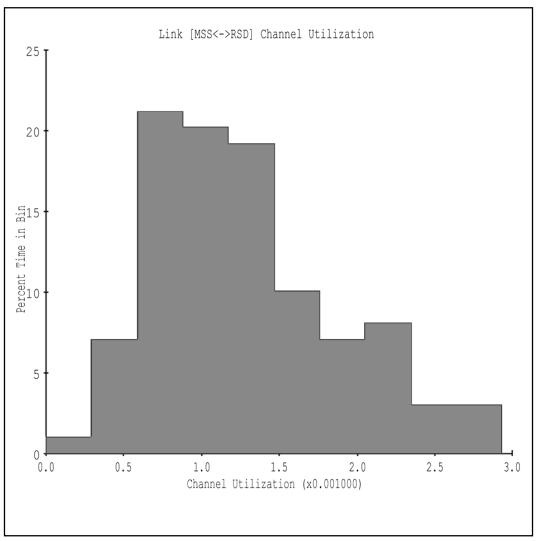


Figure 2.3.5. Channel Utilization @10% Utilization

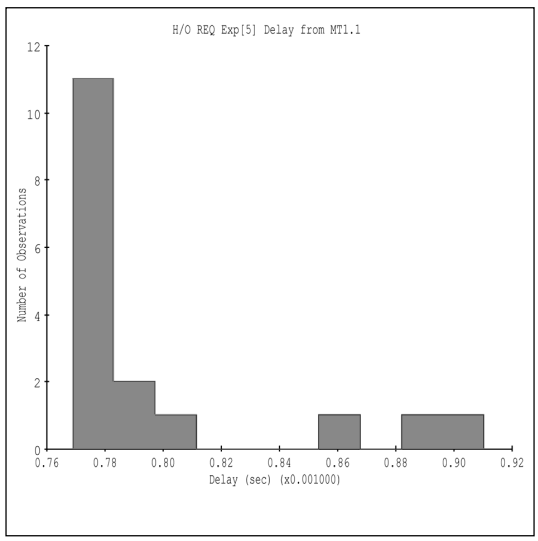


Figure 2.3.6. Message Delay @10% Utilization

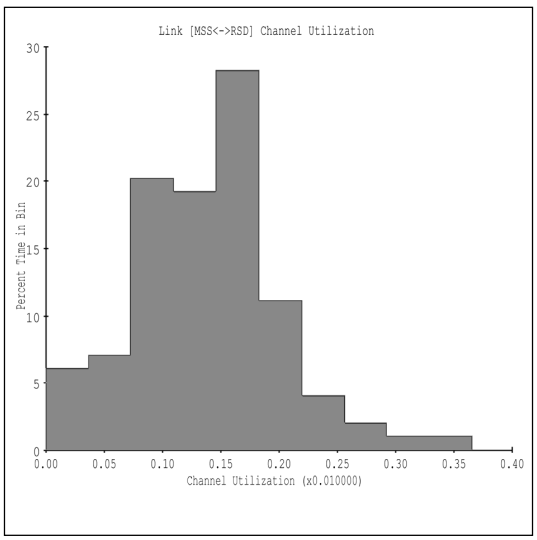


Figure 2.3.7. Channel Utilization @30% Utilization

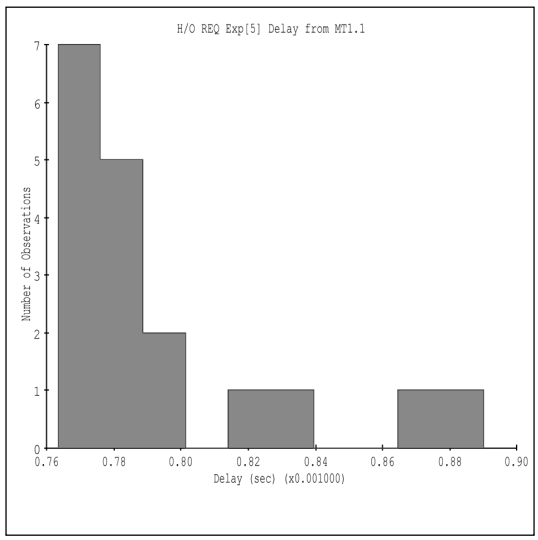


Figure 2.3.8. Message Delay @30% Utilization

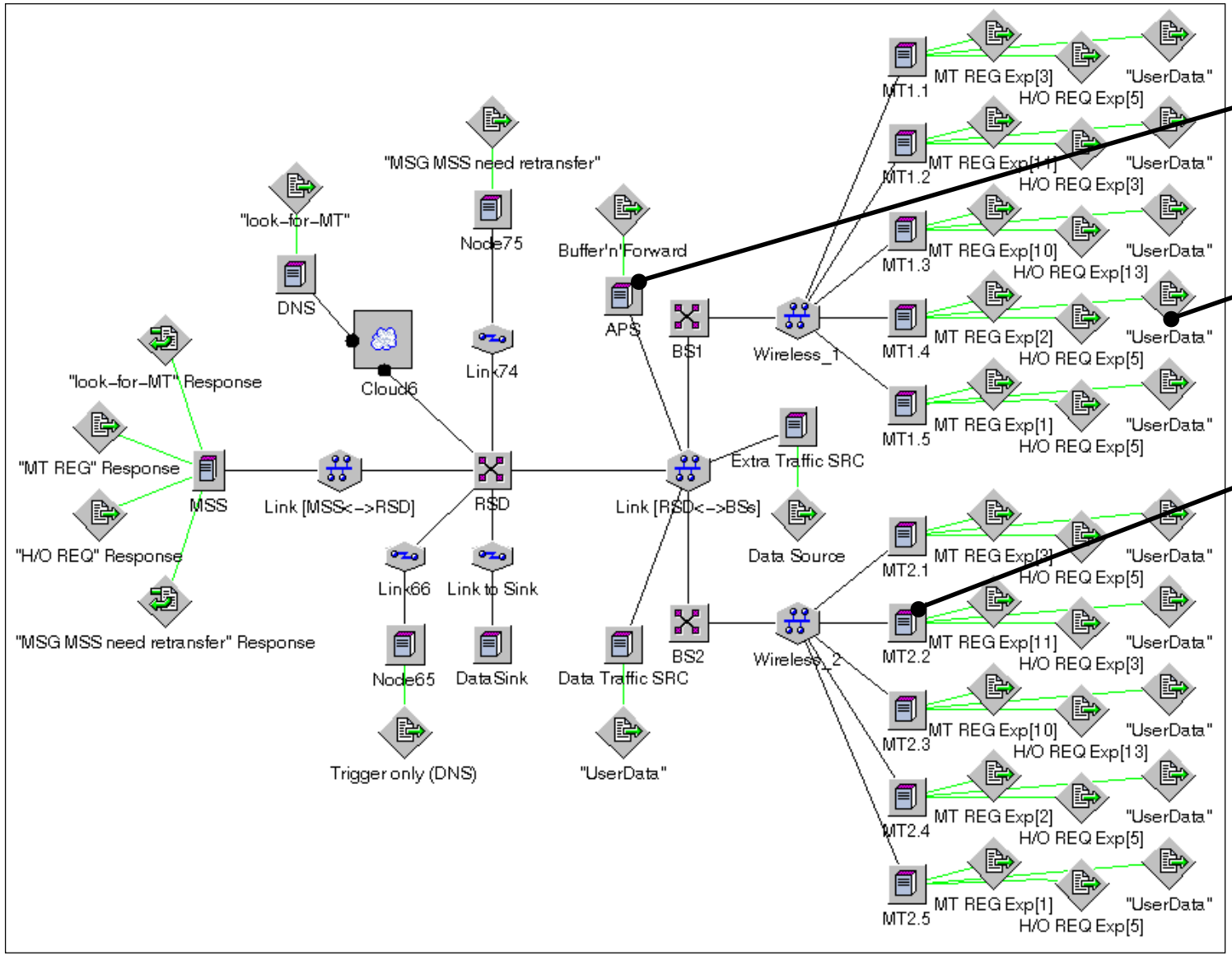


Figure 2.4.1. Network Topology

The Acces Point Server (APS) receives messages and distributes them randomly among the MT's.

MTs transmit User Data to Data Sink (90%) and to the APS and thereby to other MTs (10%)

The MTs "power down" and wake up in a cycle. The sleeptime is either 50% or 90%.

2.4.2. Simulation Results from the 3rd Model

The primary question was about the necessary buffer space needed in the APS to operate. The amount of space in numbers of packets stored to be forwarded, when the destination MT wakes up again is obviously dependant on the sleep-time to wake-time ratio of the MTs. In the following Figures the situation for 50% and 90% sleep time is examined.

One second sleep and 1s activity of all 5 MTs in the simulation demand for a maximum of 9 packets to be buffered at the APS. The simulation run is shown in Figure 2.4.3. and the corresponding histogram of buffer space used is depicted in Figure 2.4.4.

The assumption of 90% sleep, 1s activity and 9s sleep mode, is more realistic and examined in Figure 2.4.5. and Figure 2.4.6. Roughly 50 packets have to be buffered at the APS as a maximum. And the system is *stable* over several thousand simulated seconds, in the sense that the average used buffer space is bounded and less than 30.

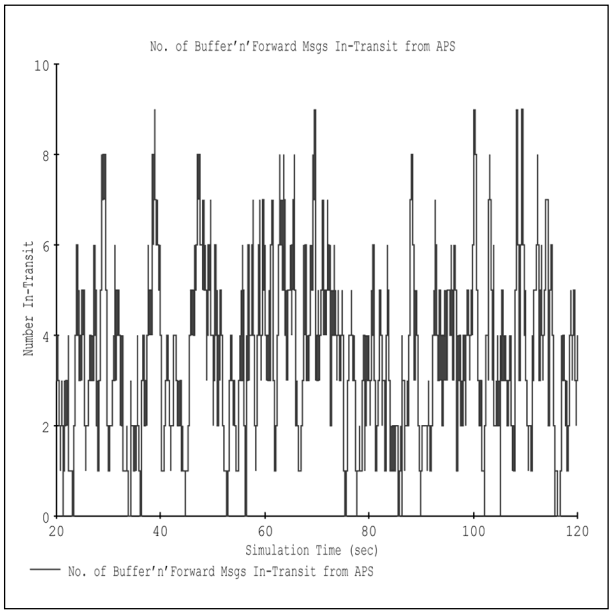


Figure 2.4.3. Buffer Utilization 50% sleep

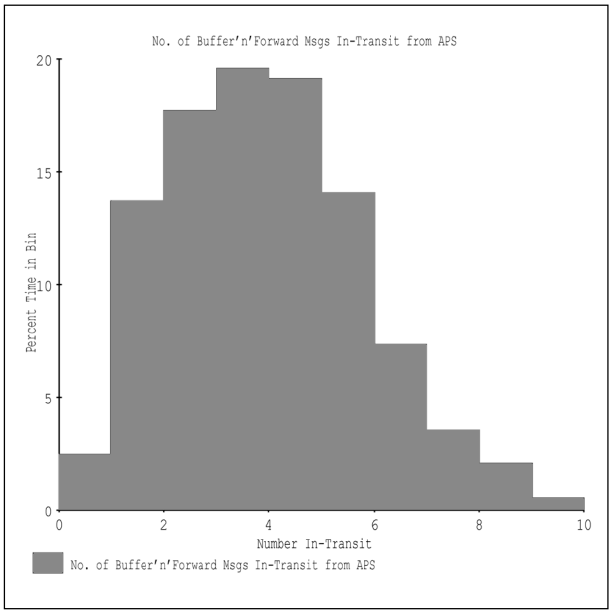


Figure 2.4.4. Buffer Utilization - Histogram 50% sleep

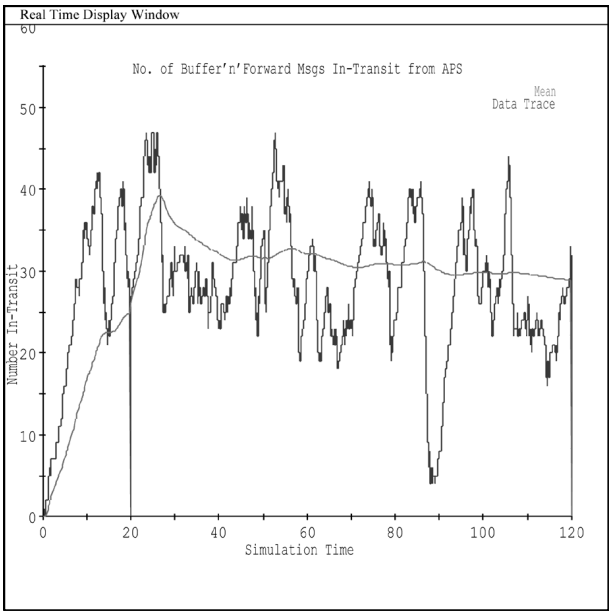


Figure 2.4.5. Buffer Utilization 90% sleep

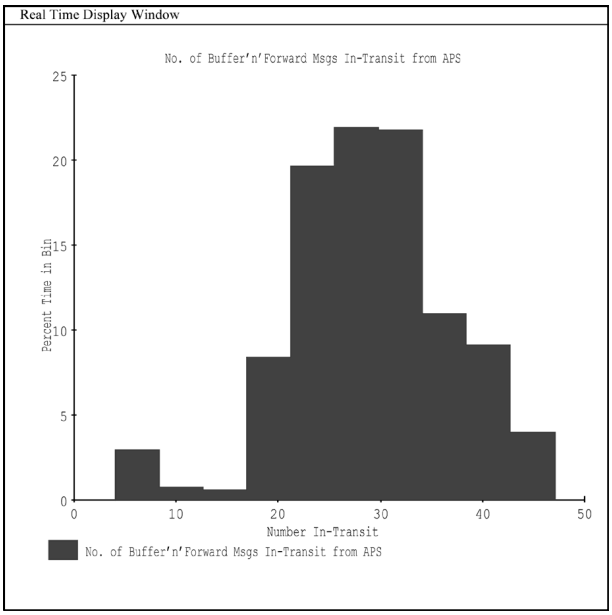


Figure 2.4.6. Buffer Utilization - Histogram 90% sleep

Next we take a look at several sections of the network and do some examination of the channel utilization. Remember these results have to scaled for the actual number of MTs and BSs used!

- The wireless networks 1Mbps CSMA/CA channel utilization is shown in Figure 2.4.7. and Figure 2.4.8. The average is around 6% with peaks up to 50%.
- The backbone network is an 10Mbps ethernet and it's mean level of utilization is less than 3%. Peak utilization goes up to 7% (see Figure 2.4.9. and 2.4.10.).
- The 10Mbps link between MSS and RSD is utilized only with 0.06% of load.

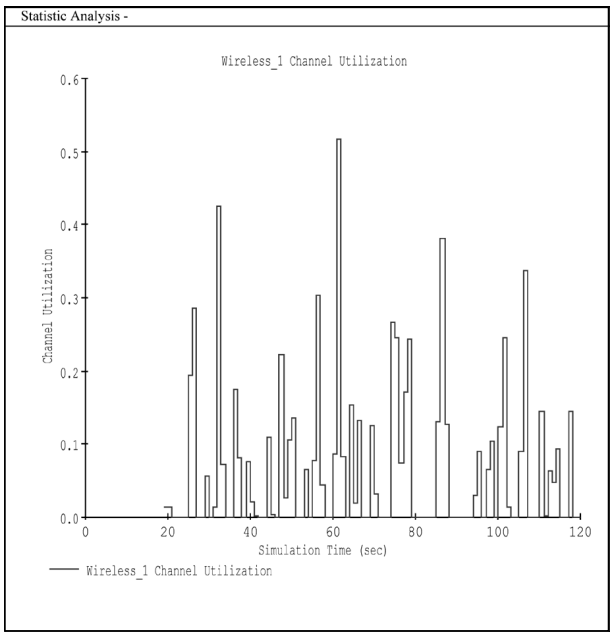


Figure 2.4.7. Wireless Channel Utilization

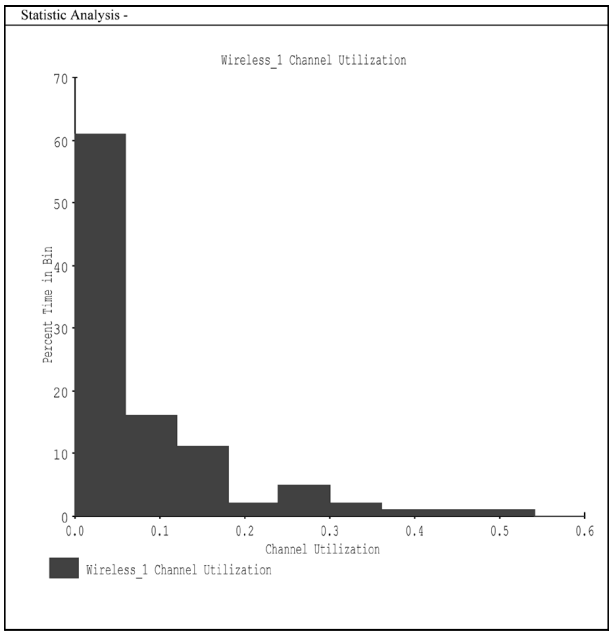


Figure 2.4.8. Wireless Channel Utilization

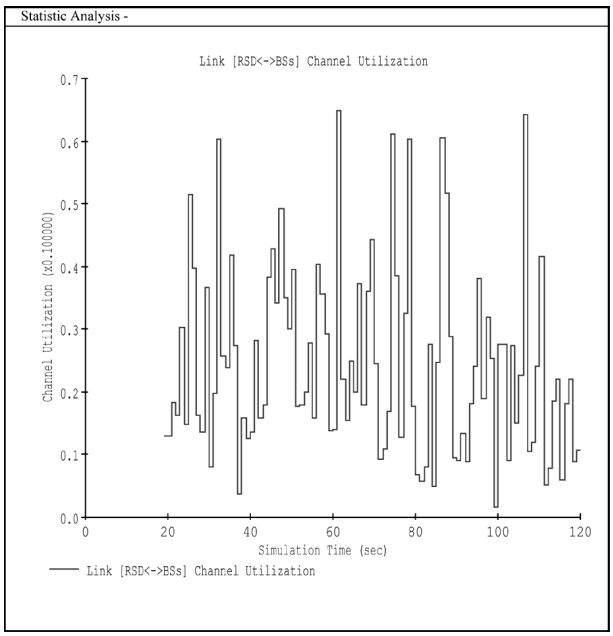


Figure 2.4.9. Backbone Channel Utilization

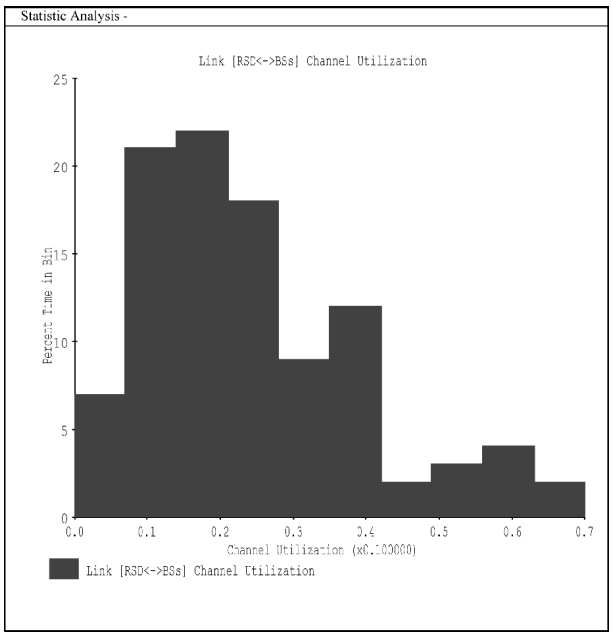


Figure 2.4.10. Backbone Channel Utilization

2.5. A Gigabit Backbone

The *gigabit ethernet* technology is described in [SEIF98] is available and it's impact on the presented system will be discussed in this subsection. Three properties of the 1000BASE-X technology make it interesting for our system design:

- Fullduplex mode operation
- Segment length up to 3000m (without repeater)
- 1Gbps bandwith

Since the medium is no longer shared as in halfduplex-mode, there is no need for collision detect and loopback circuitry as sketched in Figure 2.5.1. The same reasoning leads to a freedom from the CSMA/CD maximum length restriction. Furthermore the sustained channel-utilization is not limited as it is with CSMA/CD. Together with higher bandwith than 100BASE-T these properties together result in enhanced throughput and decreased delay.

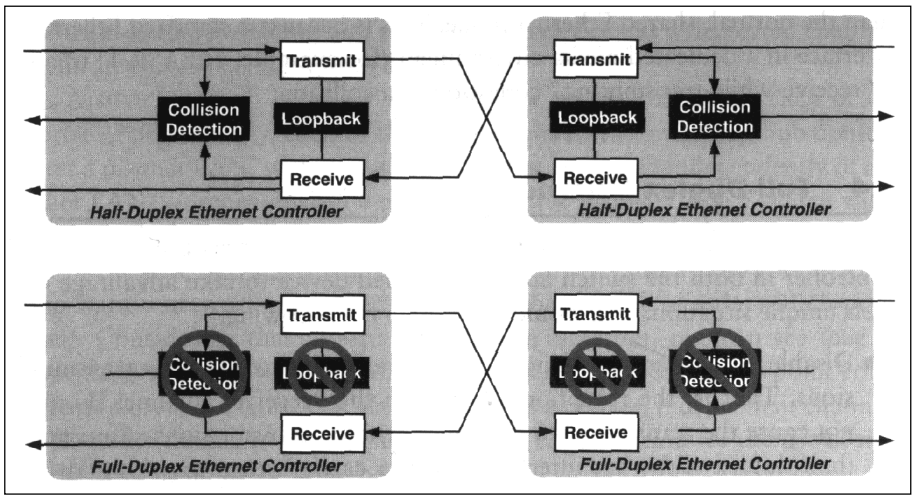


Figure 2.5.1. Half-Duplex vs. Full-Duplex

The extended segment length enlarges one MSS coverage area by a factor of 900. With 30 times more bandwith than 100BASE-T and based on the simulation results in the last chapter (2 BSs cause 3% utilization) thus making it possible to join up to 600 BSs together to a single backbone network served by one MSS and one APS.

Comparing the bandwith of the medium (1Gbps) and the traffic sources (600 BSs @ 1Mbps and the APS) this is a feasible solution, iff the worst case message delay was within reasonable bounds. This case occurs, when one particular BS requests a packet and all 599 other BSs are served first:

Nevertheless the network delay for this first packet is less than 10ms ($599 \cdot 1500$ bytes$\cdot 8/10^9$ bps) and zero for all subsequent ones, due to the fact that the transmission of a single packet over the wireless 1Mbps channel takes longer than 10ms.

A faster backbone is mandatory if in the future the bandwith of the wireless links increases, thus the proposed system architecture can grow as well with the technology.

2.6. Merging the MSS and APS

It is only reasonable to simplify the system design by merging the functionality of nodes and thereby reducing their number. The "Mobility Support Server" and the "Access Point Server" can clearly be combined and located anywhere on the backbone network which provides the necessary bandwith needed by the APS. The additional load of control traffic to/from and to the MSS module is negligible low as we have shown in Section 2.3.2.

One could even think about sharing the MSS functionality among several backbone networks.

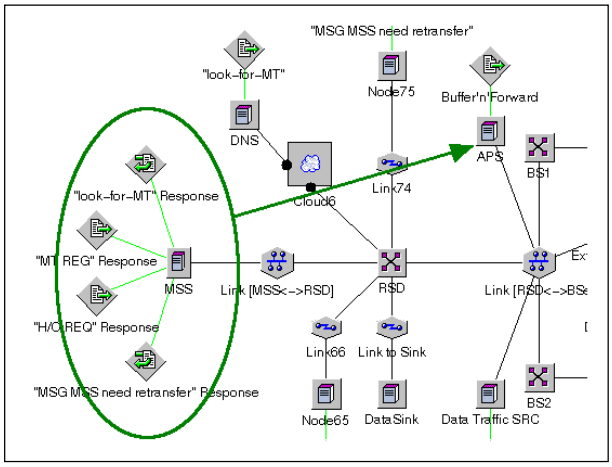


Figure 2.6.1. APS includes MSS functionality

After merging MSS and APS the network has a topology as depicted in Figure 2.6.2. below.

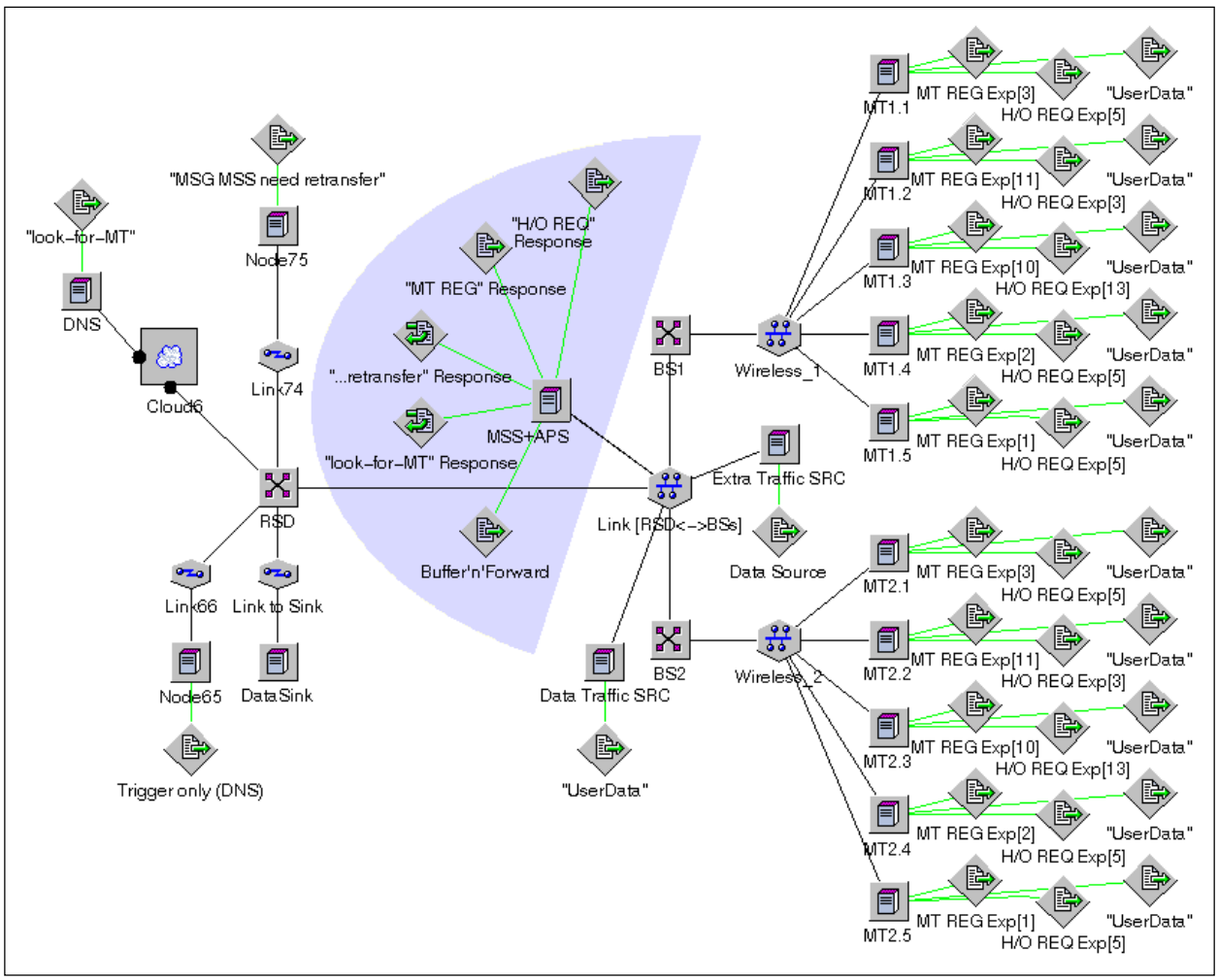


Figure 2.6.2. Merged MSS and APS

2.7. Properties of the Objects in the Simulation

This chapter describes all objects used in the latest model of the simulation to provide a better understanding of the model and the simulation software. Mainly messages and links are described here since all other objects have no properties relevant to the objective of this chapter. In the figures arrows mark items of interest.

The messages themselves and their interdependency in the simulation model is explained in detail in Section 2.2.

2.7.1. Messages of the Mobile Terminal

The *handover request* message ❶ originates from every MT with a certain inter-arrival time ❷ (IAT) - 5 second exponential distributed, in the case shown in Figure 2.7.1 - and size. The size is defined in a global variable *PacketSize* ❸. The protocol is UDP ❹ and the destination ❺ is always the APS.

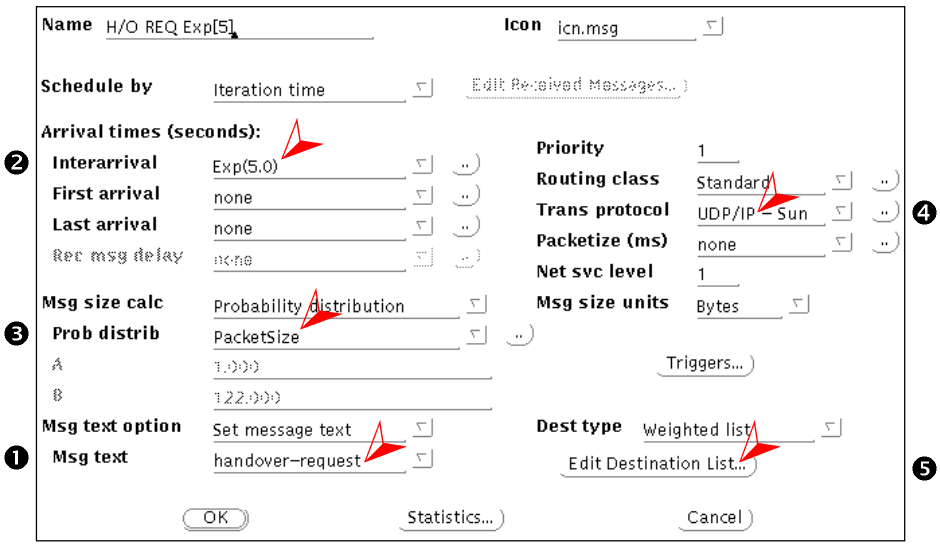


Figure 2.7.1. Handover Request Message

Figure 2.7.2.: The *mobile terminal registration request message* is similar and differs from the *H/O REQ Message* only in IAT ❶ and message text ❷.

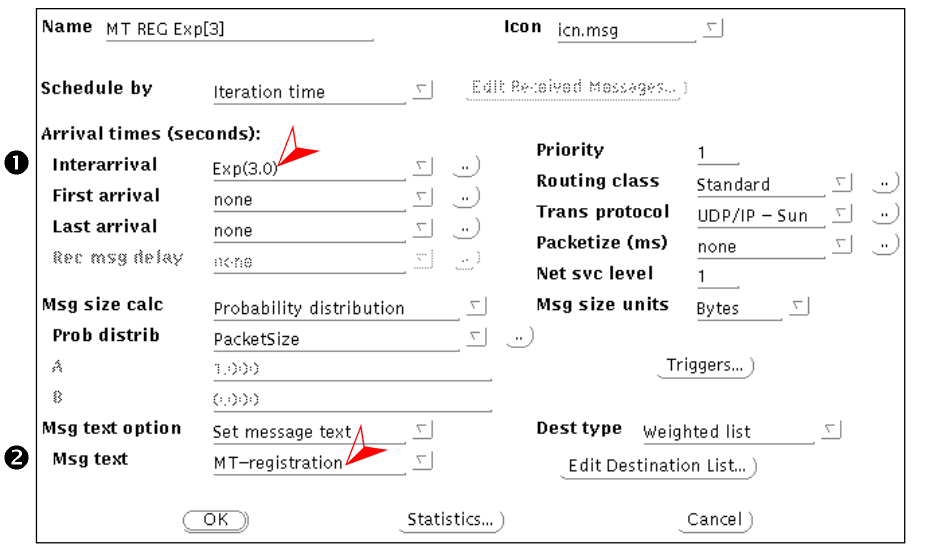


Figure 2.7.2. Mobile Terminal Registration Request Message

Figure 2.7.3.: The *user data message* ❶ sent from the MTs uses two global variables to allow the *IAT* ❷ and *message size* ❸ to be easily changed for all MTs. This message is sent to the *APS* or to the *DataSink* (a node on the backbone that absorbs all messages addressed to it) according to the probabilities given in the (weighted) *destination list* ❹. It is an UDP message too ❺.

Thereby traffic from one MT to another (p=20%) and to the Internet (p=80%) is simulated.

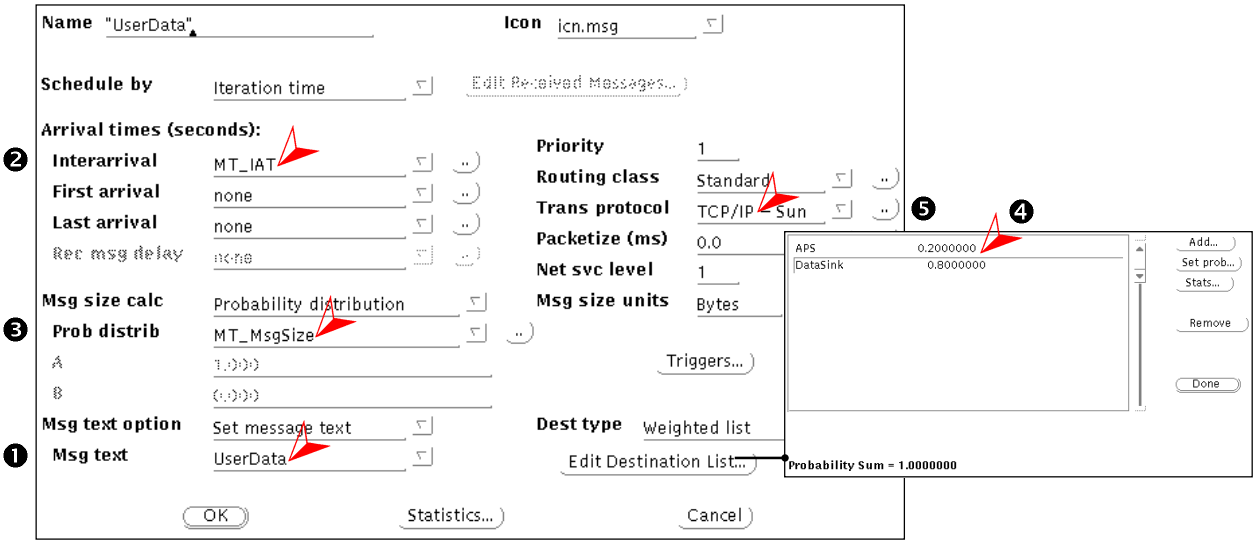


Figure 2.7.3. Mobile Terminal User Data Message - and Destination List

Figure 2.7.4.: The *mobile terminals* node dialog has only a few options. In order to include in our model the limited availability of the MTs due to their power down cycle, a node failure and automatic repair ❶ is simulated.

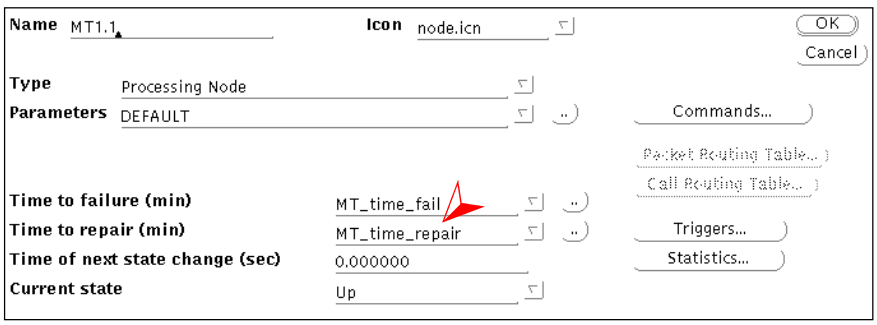


Figure 2.7.4. Mobile Terminal Node

2.7.2. Messages of the Access Point Server

Figure 2.7.5.: The *handover request response* message ❶ is sent from the APS to Node65 ❷ (which subsequently triggers a message to the DNS) whenever the message *handover-request* is received ❸. The response is a 72 byte UDP message ❹❺.

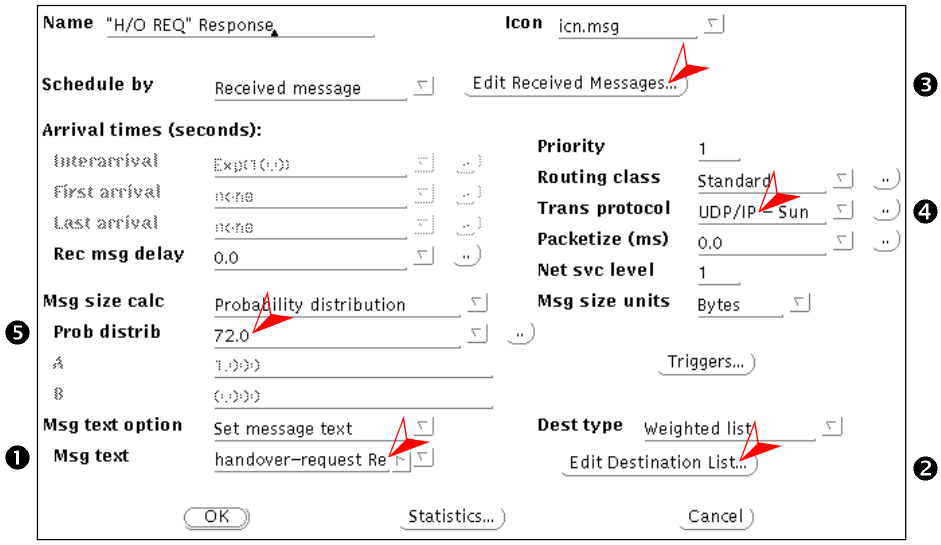


Figure 2.7.5. Handover Request Response Message

Figure 2.7.6.: The *MT registration response message* is similar. As can be seen by the zero *message delay* ❶, no processing time for these requests is included, as reasonable data is not available yet.

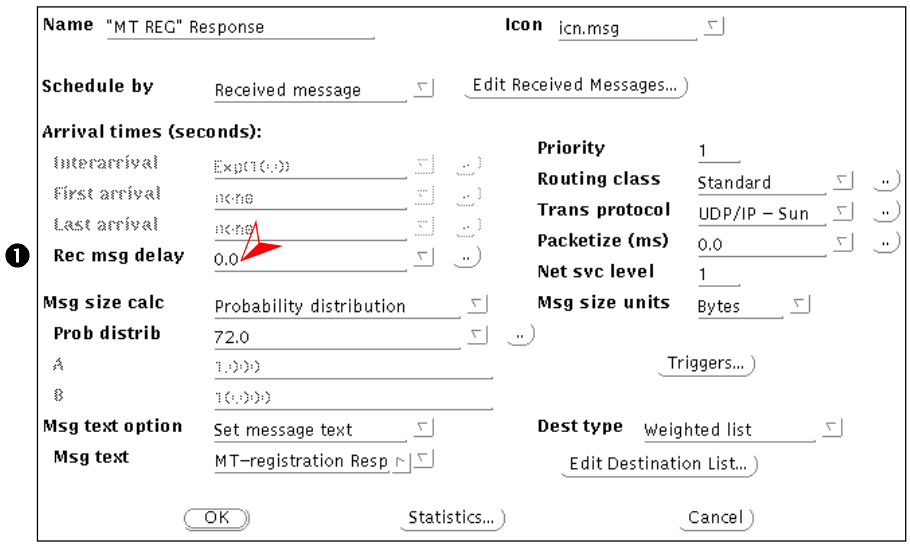


Figure 2.7.6. MT Registration Response Message

Figure 2.7.7.: The *buffer and forward* message ❶ dialog controls the APSs user data storage and distribution functionality. On receipt of a *UserData* message ❷ a TCP/IP message ❸ of the same size ❹ is sent to one of the MTs according to the *destination list* ❺.

At this point it is obvious, that a UserData-message is not directed to a specific MT, but is always sent to the APS. In this simulation it is the APS's responsibility to distribute these messages randomly among the MTs. Nevertheless the network traffic as seen externally behaves as if the messages were directed to specific MTs.

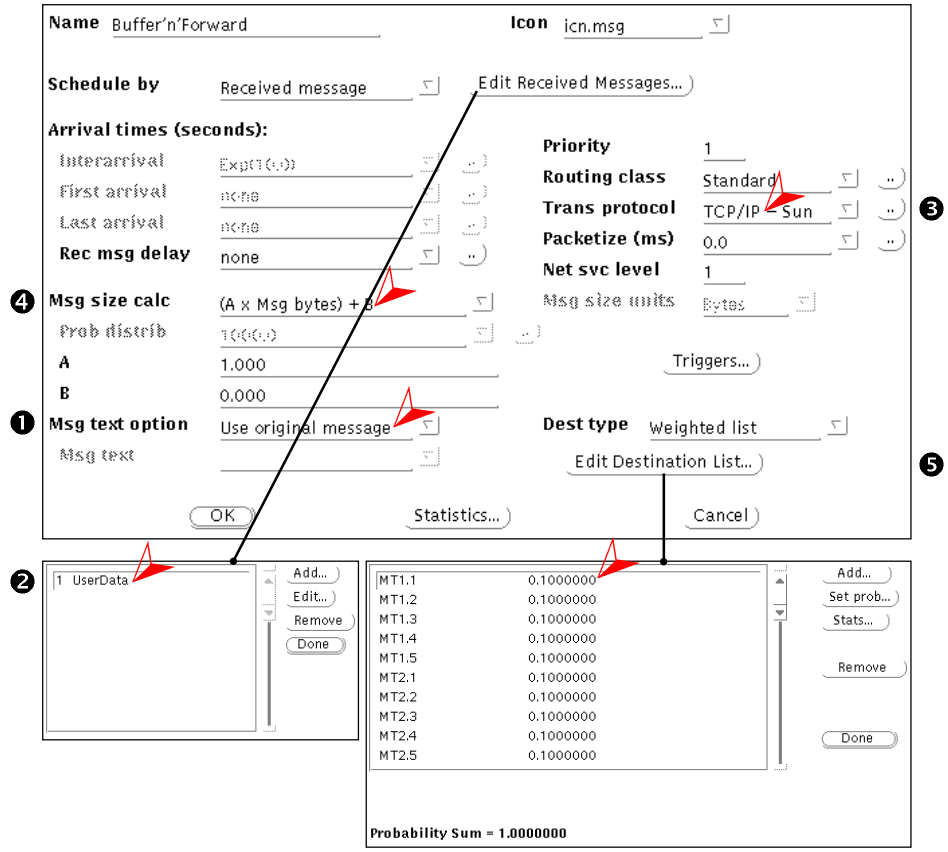


Figure 2.7.7. Buffer and Forward Message

The *Retransfer* message dialog shown in Figure 2.7.8. and the almost identical ❶ *Look for MT response* message dialog simply consists of an echo of the original message ❷ back to the respective original sender.

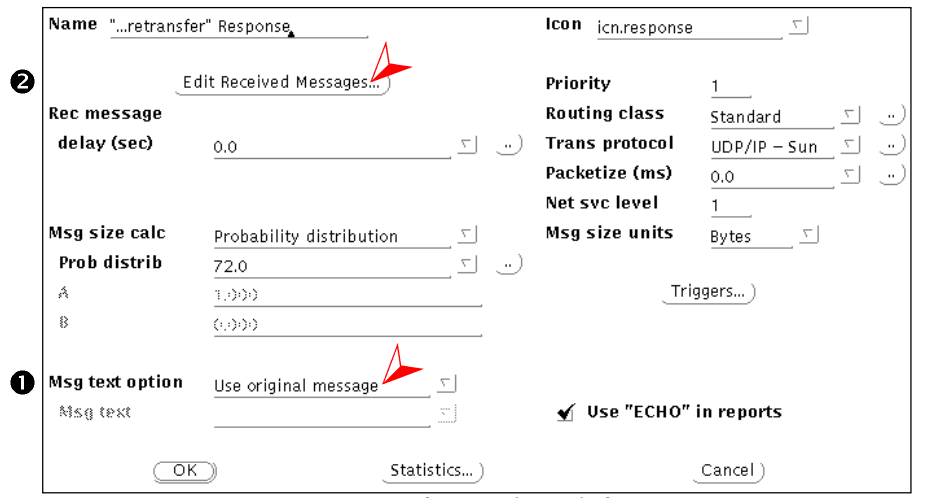


Figure 2.7.8. Retransfer- and Look for MT Response

2.7.3. Other Messages

Figure 2.7.9.: This message is triggered on receipt of a *handover request response* message ❶ and sends actually the same message (*copy message name* ❷) to the destination node DNS ❸. Protocol and size is fixed as UDP, 72 byte ❹❺.

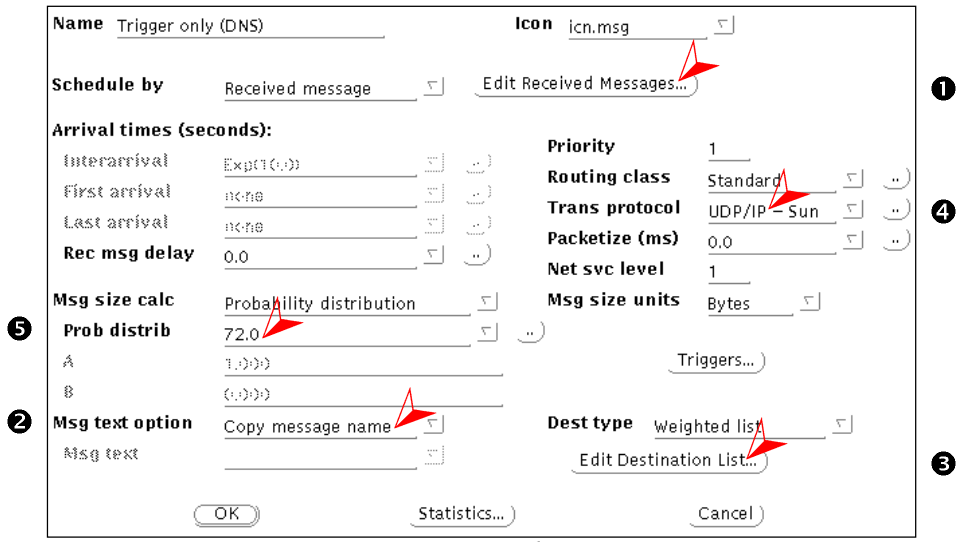


Figure 2.7.9. Trigger only DNS Message

Figure 2.7.10.: The *look for MT* message ❶ is sent from the DNS to the APS ❷ whenever the (forwarded) *handover request response* is received ❸. This is a 72 byte UDP message ❹❺.

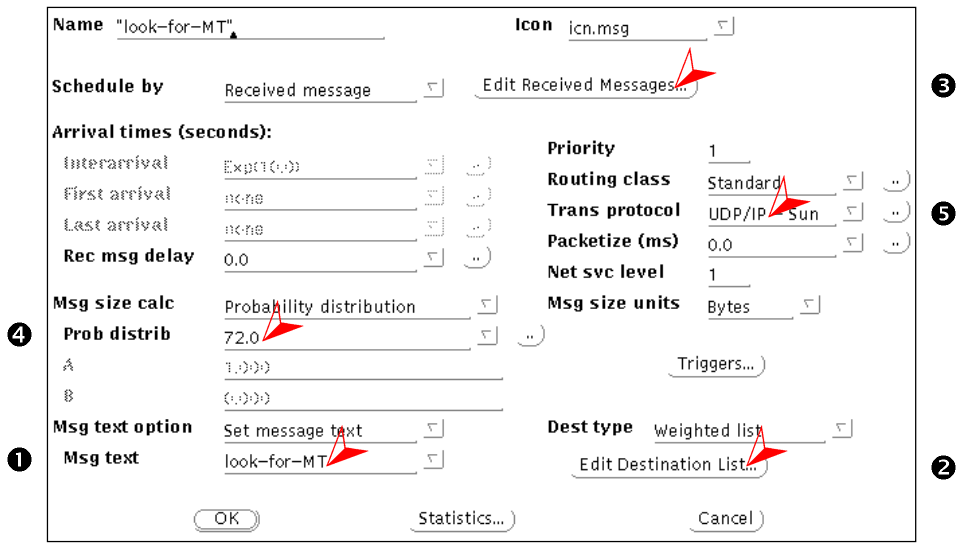


Figure 2.7.10. Look for MT Message

Figure 2.7.11.: The *message MSS need retransfer* has a long inter-arrival time ❶, constant size ❷ and uses UDP ❸. The destination is always the APS ❹.

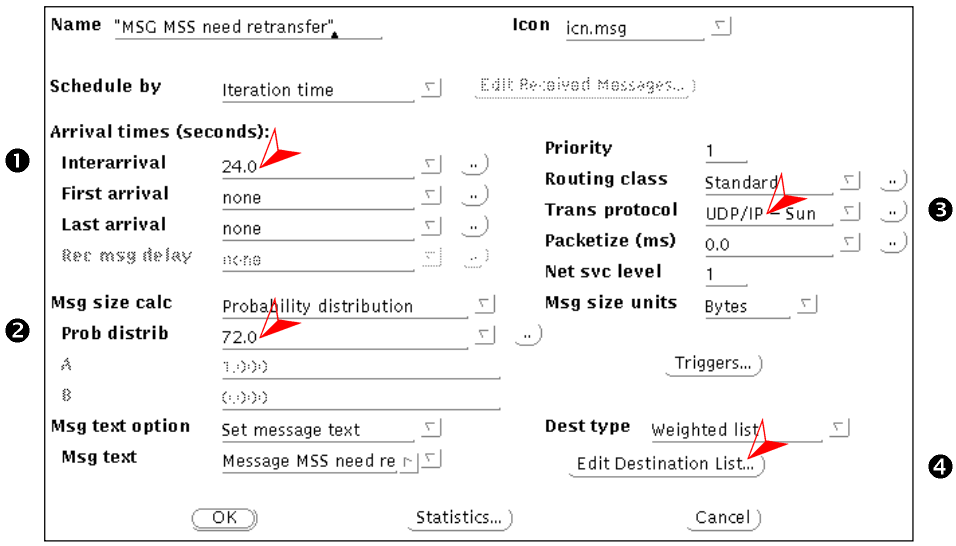


Figure 2.7.11. Message MSS Need Retransfer

Figure 2.7.12.: The *user data message* ❶ simulates the traffic to the MTs from the Internet. The message size ❷ and inter-arrival time ❸ is taken from [BJUR98] and scaled up for 10 MTs. A TCP/IP connection ❹ is assumed and the destination ❺ is the APS.

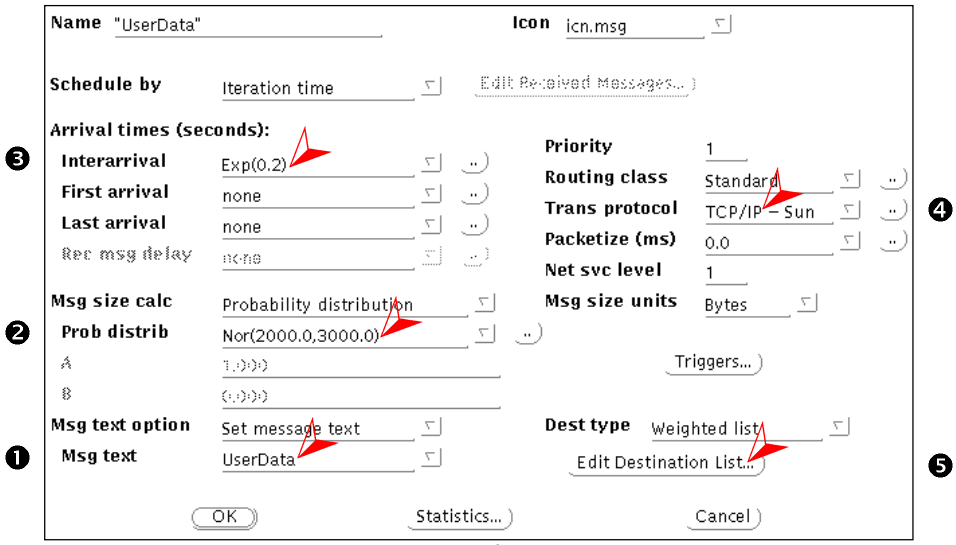


Figure 2.7.12. External User Data Message

The *Extra Traffic Source* sends packets of constant size and IAT to the *DataSink* node to simulate a background utilization of the network. In the recent simulation runs (Section 2.4.2.) the corresponding node is switched off by setting the nodes *current state* to *down*.

2.7.4. Links of the Model

The *links* or *networks* in the simulation use predefined defaults for well known types of networks, with the exception of the wireless link that demands its own profile. Most links are considered to be point-to-point T1 links ❶, as shown in Figure 2.7.13. below.

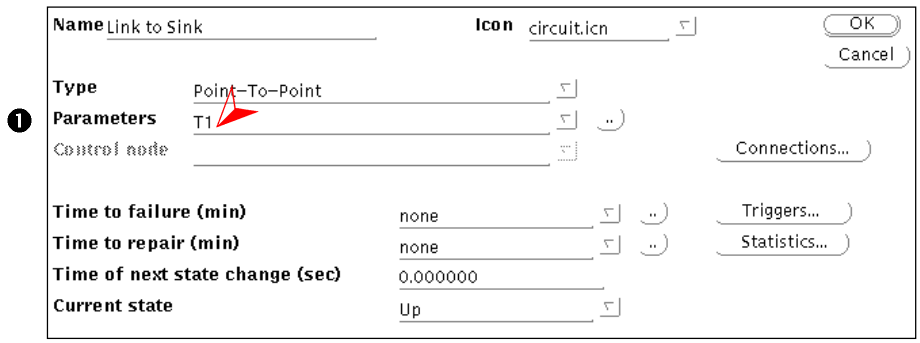


Figure 2.7.13. Default Point-to-Point Link

Figure 2.7.14. shows the link dialog of the *backbone network*, which is assumed to be a standard 10BASE-T ethernet ❶. This is the realistic assumption for 3 BSs or less.

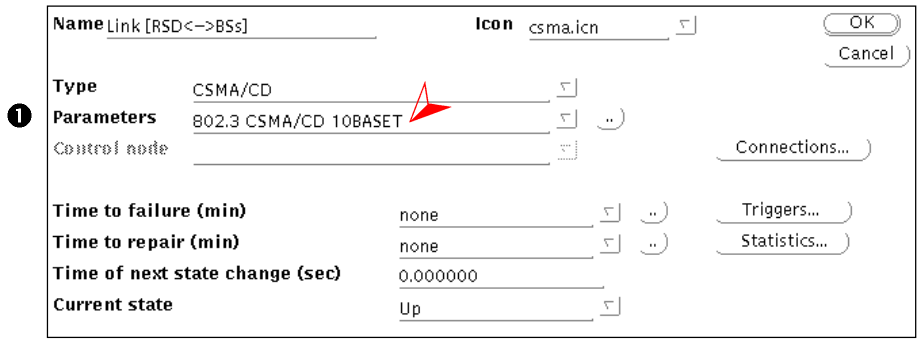


Figure 2.7.14. The BackBone Network

Figure 2.7.15. shows the link dialog and detailed parameters of the *wireless* link, which is based on the parameters of a 10BASE-T link, with the bandwidth changed to 1Mbps.

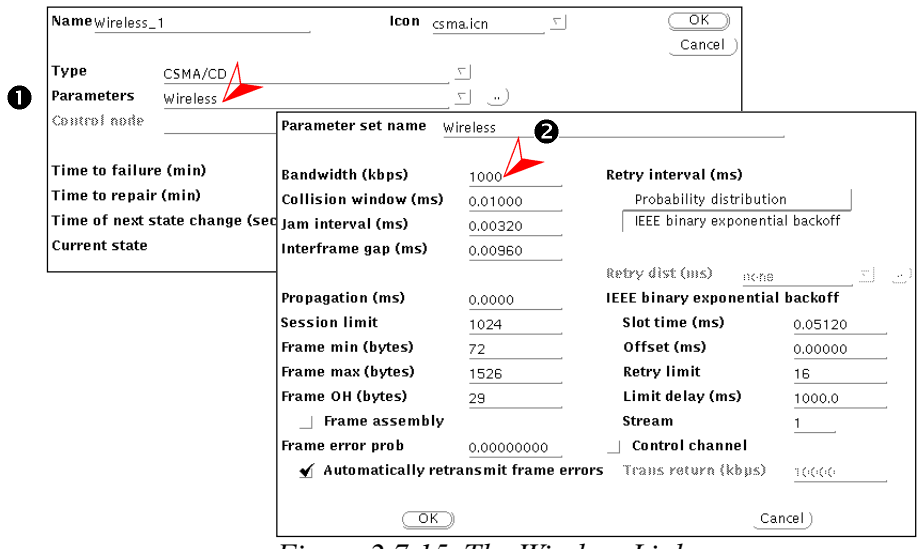
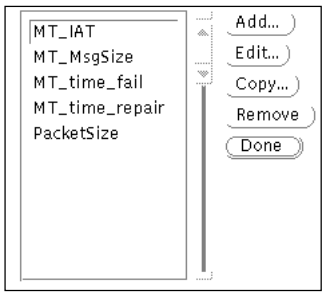


Figure 2.7.15. The Wireless Link

2.7.5. Global Variables



Some parameters are defined as global variables, accessible via the *Define/User Distributions* menu, since they are subject to frequent change and apply to several objects.

Chapter 12 of [COMN96] describes all statistical distributions COMNET III uses.

MT_IAT: Is defined as an Exp(2) distributed interarrival time in seconds and applies to the *UserData* messages of all MTs.

MT_MsgSize: Is defined as Nor(2000, 3000) distributed message size in bytes for the *UserData* messages of all MTs.

MT_time_fail: Is defined to be Nor(0.017, 0.005) distributed with unit minutes! This is the time the MT is up until it fails - which is the simulation of the MT's power down mode.

MT_time_repair: Is Nor(0.17, 0.05) distributed by default. This is the time in minutes that defines how long the MTs are asleep. Thus the sleep:awake ratio is 10:1.

PacketSize: This parameter defaults to 1400 bytes and specifies the size of the *MT REG* and *H/O REQ* messages sent by the MTs.

3. Chapter III – Partitioning the Functionality between BS and APS

As suggested in Section 1.5.4.1. the functionality of the base station is partitioned between the BS itself and the APS. The Radio-MAC part should be partitioned as Figure 1.5.5. indicates, and in particular, as suggested in Section 2.4.1., memory resources should be located at the APS to keep the BS's structure simple and its cost low. Part one of [MEDE23] addresses the issue of partitioning as well.

3.1. Data Flow between Internet, APS, BS, and MT

The previous assumptions on power management and partitioning the functionality among the nodes lead to the following two alternative modes for transporting a packet from and to the mobile terminals.

When the MT sends data to the Internet this is always accomplished directly without APS intervention and the intermediate BS acts purely as a bridge. Assuming TCP packets, this scenario is shown in Figure 3.1.1. However, in the *buffered mode* the ACKs flow via the APS.

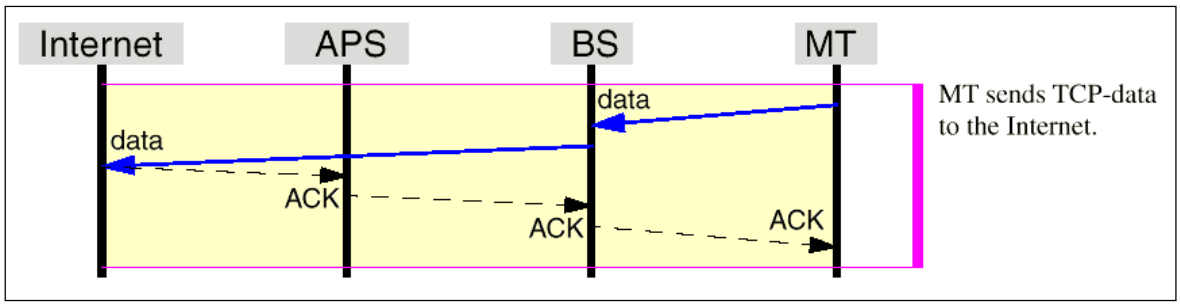


Figure 3.1.1. MT sends data

The reception buffering and forwarding of packets destined for a MT can be and is initially handled by the APS. The APS has knowledge about the sleep mode and cycle of all the MTs in it's domain and, when a particular MT is awake, it will forward the buffered packets to that BS, that is responsible for the MT.

See Figure 3.1.2. below for an illustration.

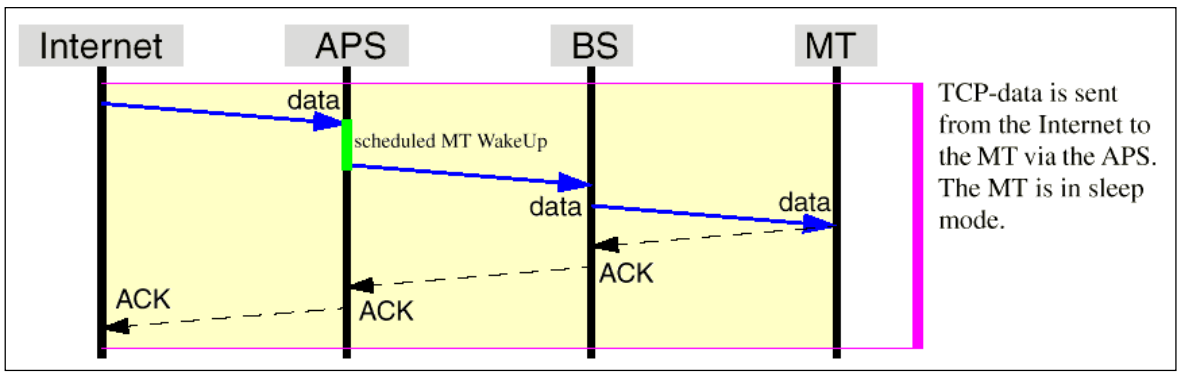


Figure 3.1.2. MT receives data

However, under suitable circumstances the BS can process the packets for one of it's MTs itself. We call this *stream mode*.

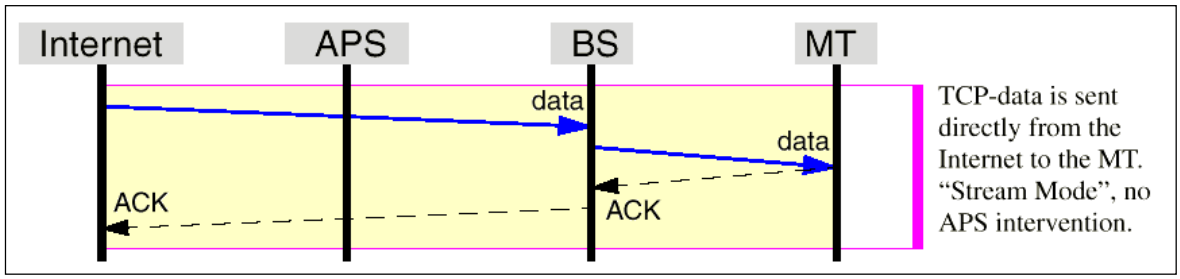


Figure 3.2.3. MT receives data in stream mode

If a BS seems to be able to handle and upon detection of a stream of data to the MT the APS can initiate entering the *stream mode*. The APS shall not enter this mode after a BS has repeatedly left it. If the BS experiences prolonged input-buffer overflows or unavailability of a MT it can fallback to *buffered mode* and thereby leave the *stream mode*.

The advantage of less delay and delay jitter of this mode of operation is obvious.

3.2. Messages exchanged between APS and BS

Resulting from the data flows shown in the last section the following messages are necessary for proper collaboration between APS and BS:

- **`MT_Mode (BS→APS)`**
  The message allows the APS to sync with the MT's sleep mode and cycle. The power management is still controlled at the BS, but with this message it keeps the APS informed.
- **`Enter_Stream_Mode (APS→BS)`**
  With this message the APS informs the BS to process packets for a given MT on its own.
- **`Leave_Stream_Mode (BS→APS)`**
  The BS informs the APS that it delegates the processing of packets for a given MT back to the APS.

The additional network load due to this control messages is negligible and the overall network load can only be less than the simulated case (see Section 2.3.2. and 2.4.2.) since a packet sent in *stream mode* saves 50% of the time needed to receive, store, and forward a packet which is processed in the *buffered mode*.

4. Chapter IV – Providing Data for other Models

4.1. First stage of testing the interface functionality:

A first approach to verify the proper functionality of the radio- and ethernet-MACs, as implemented in SDL, is to connect the two interfaces back-to-back. Consider an IP bridge as shown in Figure 4.1. We can simplify this to a frame bridge as shown in Figure 4.2. by simply using an identity transform for our IP bridging. Thus the address-mapping between 802.3 and 802.11 in this simplified case is trivially the identity mapping.

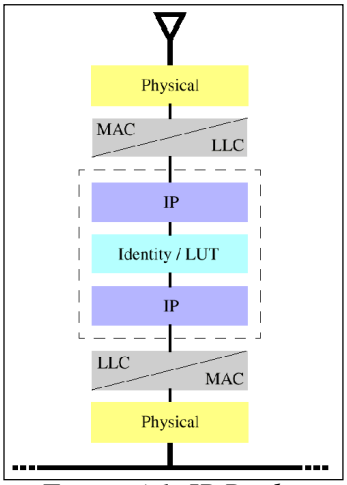


Figure 4.1. IP Bridge

Furthermore the two communication-directions could be tested separately and independently assuming the media is full-duplex. However, as the media is half-duplex we must add a test for the case when the media is busy.

4.1.1. Providing Stimulus-Data for the SDL-Model

In order to verify the correctness and measure the estimated performance[1] of the SDL-Model, selected "real-world" traffic is used as a stimulus to the simulation. The tcpdump program [URL006] seems to be the right tool to provide this input. For illustration and details of the various TCP/IP protocols as viewed with tcpdump see [STEV94].

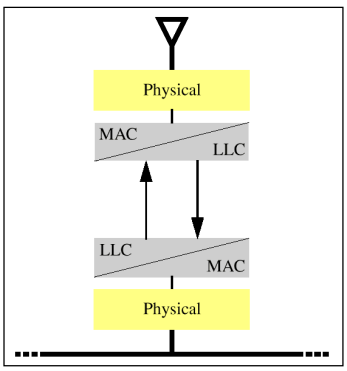


Figure 4.2. Frame Bridge

The following command dumps all traffic going to or coming from the stated address:

```
tcpdump -w ping.cap icmp or ip or tcp and src or dst client2.electrum.kth.se
```

The next two lines discriminate the sent and received traffic:

```
tcpdump -w ping_txd.cap -r ping.cap dst client2.electrum.kth.se
tcpdump -w ping_rxd.cap -r ping.cap src client2.electrum.kth.se
```

The resulting files looks like this (viewed with i.e., tcpdump -r ping_txd.cap):

```
10:57:14.365759 lab2.it.kth.se > client2.electrum.kth.se: icmp: echo request (DF)
10:57:15.360860 lab2.it.kth.se > client2.electrum.kth.se: icmp: echo request (DF)
10:57:16.360769 lab2.it.kth.se > client2.electrum.kth.se: icmp: echo request (DF)
10:57:17.360737 lab2.it.kth.se > client2.electrum.kth.se: icmp: echo request (DF)
```

The prepended timestamp makes it easy to 'measure' the simulated propagation delay in an environment like the one showed in Figure 4.3.

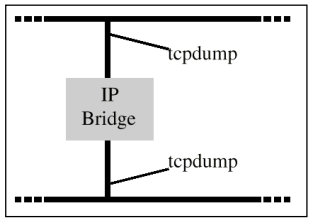


Figure 4.3. Network

[1] The current SDL tool provides no time- and therefore no performance-information.

5. Future Work

This thesis is written in the framework of the MEDIA project. The aim of this long term research project is to select or provide the technologies that fit best to accomplish the goal of efficiently supporting terminal mobility, and combining them to a system.

In the next phases of the MEDIA project (limited to the scope of this thesis) the following points at issue will be dealt with:

- The choice to use of tcpdump to acquire real network traffic is obvious. The question which of the tcpdumps should be used and fed to which tools still remains open.
- The functionality of the SDL models is still unverified and the question how to interface with them is not yet settled.
- Up to now the simulated traffic sources of our model were of statistical nature. “Real” traffic, possibly acquired as well with the tcpdump tool, shall be fed to the simulations.

6. Conclusions

At the end of my thesis-work I conclude that:

- The feasibility of the architecture has been shown in chapter 2.2.1. and specifically the link between MSS and RSD has been found to be properly dimensioned.
- Open questions on the message delay that can be expected are satisfactory answered in chapter 2.3.
- Chapter 2.4.2. recommends the use of an APS and shows its feasibility. The estimated memory usage of this node is moderate and thus furthermore supports this approach.
- By investigating the foreseeable future of network technology in chapter 2.5. it is shown that our architecture will scale well with covered area and technological enhancements over the time.
- An architectural simplification by merging the APS and MSS is found to possible in chapter 2.6.

Literature References

[BJUR98] “Simulering av datanät” by Per Bjurström, KTH/IT.
This report is available at URL `http://mail.tel.fmv.se/~per/thesis/`

[COMN96] “COMNET III User’s Manual” by CACI Products Company. Release 1.3, Nov. 1996.

[ETSI90] “ETSI/GSM Recommendations”, URL `http://www.etsi.org`

[GOOD97] “Wireless Personal Communication Systems” by David J. Goodman, Addison Wesley, ISBN 0-201-63470-8.

[HÄKA95] “Packet Radio Service for the GSM Network” by J. Hämäläinen and H. Kari in the publication of The Second International Workshop on Mobile Multimedia Communication, April 11-14, 1995.

[HERR86] “ISDN: the opportunity begins”, by T.J. Herr and T.J. Pleyrak,
IEEE Communications Magazine, 1986, 24, (p. 11).

[HUIT98] “IPv6 (second edition) The New Internet Protocol” by Christian Huitema, Prentice Hall, ISBN 0-13-850505-5.

[IEEE11] The Specification of IEEE 802.11

[JLGM97] “Towards a Future Ubiquitous Wireless Mobile Network Architecture”, Paper by Juntong Liu and Gerald Q. Maguire Jr. for the Media Project Meeting at Berlin, 23. Sept. 1997.

[JLIU98] “A Network Architecture for highly integrated Access Points for use by Multimedia Mobile Terminals”, licentiate thesis by Juntong Liu, ISRN KTH/IT/AVH--98/01--SE, March 1998.

[JOAN91] “IP-based Protocols for Mobile Internetworking” J. Ioannidis, D. Duchamp and G.Q. Maguire Jr. In Proc. SIGCOMM 91, ACM, Zurich, Sept. 1991, pp. 235-245.

[KATS97] “Voice Over IP: Policy and Regulatory Issues” by Mark Steven Katsouros,
May 10, 1997. Available at URL `http://dcs.umd.edu/~mark/631paper.html`

[MCKN98] “Internet Telephony and Open Communications Policy.” by Lee McKnight. Harvard University, 3-5 December 1997 and forthcoming in similarly titled volume from MIT Press, 1998.

[MD1297] “ESPRIT 21929 - MEDIA, Strategic Research in Network Systems for Multimedia Mobile Computing” Deliverable D1.2 by Liu and Maguire, Release date 6. October 1997.

[MEDE23] “ESPRIT Project MEDIA Deliverable D2.3” Draft Version 0.1 by Jussi Ryömä, Tampere University of Technology.

[MKBL98] “Internet Telephony: Costs, Pricing, and Policy.” by Lee W. McKnight and Brett Leida. Paper presented at the Twenty-fifth Annual Telecommunications Policy Research Conference, Alexandria, Virginia, September 1997 and forthcoming in TPRC volume comprised of selected papers, Lawrence Erlbaum Associates, 1997.
See also: URL `http://itel.mit.edu:/itel/pubs/itel.tprc97.pdf`

[MURO81] “GMSK modulation for digital mobile radio telephony” by K. Murota and K. Hirade, IEEE Trans., 1981, COM-29, (7), pp. 1044–1050.

Literature References

[NATV88] “Speech coding in the Pan European digital mobile radio system” by J.E. Natvig, Special Issue, Speech Communication, 1988, No. 1.

[NMKS95] “Technology Policy and Information Infrastructure.” by Lee McKnight and W. Russell Neuman. In *National Information Infrastructure Initiatives: Vision and Policy Design*, edited by Brian Kahin and Ernest Wilson. Cambridge, Mass.: MIT Press, 1995.

[NMKS97] “The Gordian Knot: Political Gridlock on the Information Highway.” by Russel W. Neuman, Lee McKnight, and Richard Jay Solomon. Cambridge, Mass: MIT Press, 1997.

[PERK98] “Mobile IP - Design Principles and Practices” by Charles E. Perkins, Addison-Wesley, ISBN 0-201-63469-4.

[SEIF98] “Gigabit Ethernet” by Rich Seifert, Addison-Wesley, ISBN 0-201-18553-9.

[STEV94] “TCP/IP Illustrated, Volume 1” by W. Richard Stevens, Addison-Wesley, ISBN 0-201-63346-9.

[URL001] “The GSM route to third-generation” by Ericsson Radio Systems AB, 1998. URL `http://www.ericsson.se/systems/gsm/gsm3g.html`

[URL002] “GSM 06.10 lossy speech compression”. URL `http://wwwwbs.cs.tu-berlin.de/~jutta/toast.html`

[URL003] “World Cellular Report (44) - World GSM 98” by EMC. URL `http://www.emc-database.com/reports.nsf/report+index/wcr44`

[URL004] “GSM Half-Rate Speech Codec: Bit Exactness versus DSP-Architecture”. URL `http://www.ind.rwth-aachen.de/veroeffentlichungen/tf_dspdeutschland95.html`

[URL005] “Allocating enough spectrum for cellular”, by Dr William Webb. URL `http://www.smithsys.co.uk/smith/public/tech/allocating.html`

[URL006] The tcpdump utility for UNIX is available at URL `ftp://ftp.ee.lbl.gov/tcpdump.tar.Z`

[WAKI87] “Coming to OSI: network resource management and global reachability” by S. Wakid et al, Data Communications, December 1987.

Appendix — Acronyms and Abbreviations in Alphabetic Order

1000BASE-X ..... Gigabit Ethernet technology
100BASE-T ..... 100 Mbps Ethernet technology, using UTP wire
10BASE-T ..... 10 Mbps Ethernet technology, using UTP wire
AAAA ..... DNS address record for IPv6 addresses
AB ..... (swedish) aktiebolaget, stock corporation
ACK ..... Acknowledge
ACTA ..... America's Carriers Telecommunication Association
ADM ..... Administration
AGCh ..... Access Grant Channel
AoC ..... Advice of Charge
AP ..... Access Point
APS ..... Access Point Server
APS+BS ..... Access Point Server merged with Base Station
ARP ..... Address Resolution Protocol
ATM ..... Asynchronous Transfer Mode
AuC ..... Authentication Center
BAIC ..... Barring of All Incoming Calls
BAOC ..... Barring of All Outgoing Calls
BCh ..... Broadcast Control Channel
BER ..... Bit Error Rate
BIC-Roam ..... Barring of Incoming Calls when Roaming
BOIC ..... Barring of Outgoing International Calls
BOIC-exHC ..... Barring of Outgoing International Calls except to Home Country
BS ..... Base Station
BSC ..... Base Station Controller
BSS ..... Base Station sub-System
BT ..... Modulation depth, bandwidth data-rate product
BTS ..... Base Transceiver Station
CCH ..... Committee on Harmonisation
CCITT ..... Consultative Committee on International Telephony and Telegraphy
CDMA ..... Code Division Multiple Access
CEPT ..... Conference of European Postal and Telecommunications administrations
CFB ..... Call Forward on Busy
CFNRc ..... Call Forward on not Reachable
CFNRy ..... Call Forward on no Reply
CFU ..... Call Forward Unconditional
CLIP ..... Calling Line Identification Presentation
CLIR ..... Calling Line Identification Restriction
CoLP ..... Connected Line Identification Presentation
CoLR ..... Connected Line Identification Restriction
CRC ..... Cyclic Redundancy Check
CSMA/CA ..... Cyrrier Sense Multiple Access / Colision Avoidance
CSMA/CD ..... Cyrrier Sense Multiple Access / Colision Detect
CUG ..... Closed User Group
CW ..... Call Waiting
D1.3 ..... Deliverable 1.3
DCh ..... Dedicated Control Channel

Appendix — Acronyms and Abbreviations in Alphabetic Order

DHCP ..... Dynamic Host Configuration Protocol
DNS ..... Domain Name Service
DPM ..... Differential Phase Modulation
DRAM ..... Dynamic Random Access Memory
DRX ..... Discontinuous Receive
DSP ..... Digital Signal Processor
DSSS ..... Direct Sequence Spread Spectrum
e.g. ..... (lat.) exempli gratia, for instance
EIR ..... Equipment Identity Register
etc ..... (lat.) et cetera, and so on
ETSI ..... European Telecommunications Standards Institute
FACCh ..... Fast Associated Control Channel
FCC ..... Federal Communications Commission
FDDI ..... Fiber Data Distribution Interface
FDMA ..... Frequency Division Multiple Access
FMAT ..... Flow and Multicast Address Binding Table
FPAA ..... Field Programmable Analog Array
FPGA ..... Field Programmable Gate Array
FSK ..... Frequency Shift Keying
GMSK ..... Gaussian Minimum Shift Keying
GPRS ..... General Packet Radio Services
GSM ..... Global System for Mobile communications
H/O ..... handover
HLR ..... Home Location Register
HRBT ..... Home Register Binding Table
HSCSD ..... High Speed Circuit Switched Data
i.e. ..... (lat.) id est, that is to say, which means ...
IANA ..... Internet Assigned Numbers Authority
IAT ..... inter-arrival time
ICMP ..... Internet Control Message Protocol
ID ..... identification
IEEE ..... Institute of Electrical And Electronics Engineers
IETF ..... International Engineering Task Force
IGMP ..... Internet Group Management Protocol
IMSI ..... International Mobile Subscriber Identity
IN ..... Intelligent Network
IP ..... Internet Protocol
IPv4 ..... Internet Protocol Version 4
IPv6 ..... Internet Protocol Version 6
IPX ..... Internetwork Packet Exchange, a networking protocol used by the Novell
IR ..... Infra Red
ISDN ..... Integrated Services Digital Network
ISO ..... International Organization for Standardization
ISP ..... Internet Service Provider
ITC ..... Internet Telephony Consortium
ITSP ..... Internet Telephony Service Provider
LAN ..... Local Area Network

Appendix — Acronyms and Abbreviations in Alphabetic Order

LEC ..... Local Exchange Carrier
LSB ..... Least Significant Bit
LSP ..... Link State Advertisement
LTP ..... Long-Term Predictor
MAC ..... Media Access Control
MBONE ..... Multicast Backbone
MIB ..... network Management Information Base
MoU ..... Memorandum of Understanding
MPTY ..... Multi Party
MS ..... Mobile Station
MSB ..... Most Significant Bit
MSC ..... Mobile services Switching Center
MSR ..... Mobile Support Router
MSS ..... Mobility Support Server
MT ..... Mobile Terminal
MTP ..... Message Transfer Part
MTU ..... Maximum Transmission Unit
NBMA ..... NonBroadcast Multiple Access
NetBEUI ..... NetBios Enhanced User Interface
NLA ..... Next Level Aggregator
NMC ..... Network Management Center
NMT ..... Nordic Mobile Telephone
NS ..... Network sub-System
NSAP ..... Network Service Access Point (address)
O&M ..... Operations and Maintenance
OMC ..... Operations and Maintenance Center
OS ..... Operating System
PC ..... Personal Computer
PCh ..... Paging Channel
PDA ..... Personal Digital Assistant
PICC ..... Presubscribed Interexchange Carrier Charge
PIN ..... Personal Identification Number
PSTN ..... Public Switched Telephone Network
PTT ..... National Telecommunications Provider (Post, Telephone and Telegraph)
RACh ..... Random Access Channel
REG ..... Registration
RELP ..... Residually Excited Linear Predictive
REQ ..... Request
RF ..... Radio Frequency
RISC ..... Reduced Instruction Code
RSD ..... Routing/Switching Device
RSSI ..... Received Signal Strength Indication
RSVP ..... The Resource ReSerVation Protocol
SACCh ..... Slow Associated Control Channels
SAP ..... (NetWare) Service Advertising Protocol
SCCP ..... Signalling Connection Control Part
SCP ..... Service Control Point

Appendix — Acronyms and Abbreviations in Alphabetic Order

SDL ..... Description Language used by Telelogic's SDT tool
SIM ..... Subscriber Identity Module
SLA ..... Site Local Aggregator
SLC ..... Subscriber Line Charge
SMS ..... Service Management System, Short Message Service
SNMP ..... Simple Network Management Protocol
SS7 ..... Signalling System No. 7
SSP ..... Service Switching Point
TC ..... Transaction Capability
TCP ..... Transmission Control Protocol
TCP/IP ..... Transmission Control Protocol / Internet Protocol
TDMA ..... Time Division Multiple Access
TLA ..... Top Level Aggregator
TMSI ..... Temporary Mobile Subscriber Identity
TTL ..... Time To Live
UDP ..... User Datagram Protocol
URL ..... Uniform Resource Locator
UTP ..... Unshielded Twisted Pair wire
VLR ..... Visitor Location Register
VLSI ..... Very Large Scale Integration
VRT ..... Visiting Register Table

For acronyms not listed herein or -further explanation
try URL `http://webopedia.internet.com/`

[1] This last page only contains the famous recursive footnote[1].