AEÜ, Vol. 47
(1993), No. 5/6

Modelling and Performance Analysis of Common Channel Signalling Networks

Marcos Bafutto, Paul J. Kühn, and Gert Willmann

Modelling and Performance Analysis of Common Channel Signalling Networks

Connection control and the realization of Intelligent Network functions in digital networks is supported by common channel signalling based on the CCITT Signalling System No. 7. In this paper, a generic modelling approach for the signalling network is developed which yields the loading of all network resources and the network performance. The models are obtained considering the signalling system protocol functions specified by the CCITT, as well as the information flows through these functions. The approach allows the derivation of virtual processor models, which in a further step are mapped onto the implementation according to individual architectures. Using these principles, a signalling network planning tool has been developed. The capabilities of the planning tool are demonstrated by an example, where the new service Universal Personal Telecommunication (UPT) is introduced additionally to ISDN and mobile communication services.

Modellierung und Leistungsanalyse von Zentralkanal-Signalisiernetzen

Die Verbindungssteuerung und die Einrichtung von Funktionen des "Intelligent Network" in digitalen Netzen wird mit Hilfe der Zentralkanalsignalisierung auf der Basis des CCITT-Signalisiersystems Nr. 7 realisiert. In diesem Artikel wird ein generischer Modellierungsansatz für das Signalisiernetz entwickelt, über den die Auslastung von Netzkomponenten und die Leistungsfähigkeit des Netzes ermittelt werden können. Die Modelle basieren direkt auf den vom CCITT standardisierten Protokollfunktionen und den Informationsflüssen durch diese Funktionen. Der Ansatz ermöglicht es, virtuelle Prozessormodelle anzugeben, die in einem weiteren Schritt auf spezifische Implementierungen abgebildet werden können. Darauf basierend wurde ein Werkzeug zur Planung von Signalisiernetzen entwickelt. Die Möglichkeiten dieses Werkzeugs werden anhand eines Beispiels aufgezeigt, bei dem der neue Dienst "Universal Personal Telecommunication" (UPT) zusätzlich zu ISDN- und Mobilkommunikationsdiensten eingeführt wird.

1. Introduction

Connection control and the realization of Intelligent Network (IN) [2]–[4] functions in digital networks is supported by common channel signalling based on the CCITT Signalling System No. 7. The protocol architecture of the signalling network was object of standardization by the CCITT [5], and a set of recommendations is available since the last decade. A good introduction to the subject can be found in [6]–[10].

In the signalling network domain, a telecommunication service can be viewed as a number of packets transmitted to and from the related nodes. The early applications supported by the signalling network were mostly circuit-related, e.g., voice services. A characteristic of these services is the concentration of the signalling activity mainly in the connection set-up and release phases.

The signalling network scenario became more complex with the introduction of new services, such as supplementary services, IN applications, mobile communication services, and Universal Personal Telecommunication (UPT). In these new services, there is also an exchange of signalling information which is not related to the set-up of connections over the transport network, e.g., database transactions or remote procedure invocations. Here, signalling activities may be present in any phase of a call or even when there is no call active, e.g., in the procedure of keeping track of a mobile subscriber location.

The introduction of these new services requires additional information to be transported over the signalling network increasing the signalling network load. This load increase may lead to a performance degradation of the signalling network, which affects not only the Quality of Service (QOS) parameters of the new service, but also the services already offered by the network.

In this article, a modelling approach which yields the loading of all network resources and the network performance is presented. These concepts have been implemented in a planning tool suitable to support the planning of new signalling networks according to given service, load, and grade of service figures, or to detect bottlenecks or possible deficiencies in case of resource outages.

Received July 14, 1993.

Eng. M. Sc. M. Bafutto, Prof. Dr.-Ing. P. J. Kühn, University of Stuttgart, Institute of Communications Switching and Data Techniques, Seidenstraße 36, D–70174 Stuttgart, Germany.
Dipl.-Ing. G. Willmann, Alcatel SEL AG, Public Switching Systems Division, Lorenzstraße 10, D–70435 Stuttgart, Germany.
M. Bafutto is on leave of Telegoiás, and supported by CAPES Brazil, grant 9532/88-15.

2. Signalling Network Planning Problems

The problems faced by the network planner in a medium term is the balance between required capacity, expected traffic, available budget, and target grade of service to be achieved. In a medium term, the nonstationary behavior of the network, e.g., overload control and traffic burstiness, are not primarily subject of the planning process, and the traffic variables may be considered as mean values of a busy-hour call attempt.

For an already existing network, additional problems arise from the introduction of new services. The user behavior with respect to new services may be different from that for the existing ones, and patterns of traffic variation as well as Grade of Service (GOS) concepts may also be different. In this case, a careful analysis must be carried out directed to the new service as well as to its impact on the currently supported services.

Within the signalling network domain, a telecommunication service is characterized by a number of signalling messages exchanged between the involved nodes. According to the service type, different network capabilities may be required. As a result, changes of the network load and, therefore, of the QOS parameters may be caused by the introduction of a new service, by variations in the traffic intensity of a particular service, or by temporary outages. The various network components are typically affected in the following way:

Links: The offered signalling link load per call may change drastically with the introduction of new service concepts.

Signalling Points (SPs) and Signalling Transfer Points (STPs): The increasing signalling traffic load requires more processing capacity. Since this additional load may not be homogeneously distributed over all processes within an SP or STP, only some processors may become overloaded.

Service Control Points (SCPs): Since the capacity of an SCP is determined by the number and types of transactions and by the offered functionality, the effect of the introduction of a new service can be estimated by the resource capacity required to process the related transactions.

The complexity of the signalling network environment and its importance to the network have motivated the development of various signalling network planning tools [1], [11]–[16]. These tools cover different aspects of the signalling network planning, e.g., network structure, dimensioning, costing, quality of service, etc., and demonstrate the preoccupation with the impact of the fast evolution of the signalling network technology and of services on the network.

3. Signalling Network Protocol Architecture

The functional structure of Signalling System No. 7 is divided into the Network Service Part (NSP) and various User Parts (UPs). According to the OSI Reference Model, the NSP corresponds to the first three OSI Layers and provides a reliable message transfer service. The upper OSI Layers are represented by the UPs.

The NSP consists of the Message Transfer Part (MTP) and the Signalling Connection Control Part (SCCP), while the UPs are the Telephone User Part (TUP), the Integrated Services Digital Network User Part (ISUP), the Data User Part (DUP), and the Transaction Capabilities (TC). The TC can be further subdivided into the Transaction Capabilities Application Part (TCAP) and the Intermediate Service Part (ISP), which is still empty. There is no agreement concerning the national implementations of the UPs; some Operating Companies consider the functions of the TUP and DUP as part of the ISUP [9], while others define a National User Part (NUP) [17] or, e.g., a Handover User Part (HUP) [18] for supporting mobile communication services.

The MTP provides a simple datagram service, and its addressing capabilities are limited to the identification of a certain UP in one specific node. The SCCP extends the MTP capabilities in order to identify a subsystem of a UP or to translate an address that does not contain MTP routing information (e.g., a Global Title).

The ISUP offers the signalling functions that are necessary to support voice and non-voice applications in an ISDN network. The TUP provides the call control functions for ordinary telephone calls. Since these functions, as well as the data transport capabilities of the DUP, are included in the ISUP, a tendency for substituting the TUP and DUP by the ISUP can be observed. Finally, the TC provide a set of capabilities in a transaction-based and connectionless environment that support applications which require remote procedure calls or database queries.

4. Modelling Framework

The analysis of a real network initially requires a modelling methodology able to represent the internal processes of the protocols. The consideration of all mechanism of a complex protocol may create mathematical difficulties making the resulting model intractable. In the literature, however, methodologies to determine analytically the throughput and delay behavior of communication architectures may be found [19]–[24]. In [19]–[21], a method based on multiple chain product-form queueing networks is proposed. Another approach based on decomposition and aggregation techniques is presented in [23], [24]. A comparison between some of these methodologies can be found in [25].

4.1 Generic Submodels

The modelling methodology presented in [23], [24] is used as baseline. Each submodel is derived directly from the CCITT functional specifications with consideration of internal mechanisms such as segmenting/forking of messages and scheduling strategies, thus, reflecting the internal behavior of the underlying blocks. Hence, this approach is relatively independent of specific implementations leading to a generic model of the signalling network composed of generic submodels. The generic submodels are finally transformed into realistic models where the actual assignment of functional processes to real processors are taken into account.

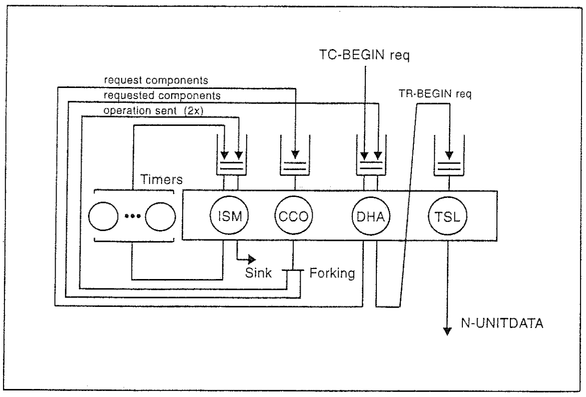


Fig. 1. Message chain for the Dialogue Begin message.

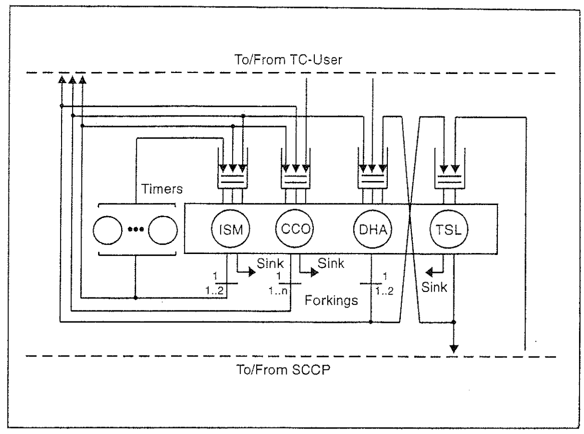


Fig. 2. Generic submodel for the TCAP functional block.

The principles of this methodology can be explained using the TCAP block as an example of derivation of a generic submodel. According to [5, Fig. A-2a/Q.774], the TCAP is composed of two subblocks: the Transaction Sub-layer (TSL) and the Component Sub-layer. The Component Sub-layer consists of the Dialogue Handling (DHA) and Component Handling. The Component Handling is further subdivided into the Component Coordinator (CCO) and the Invocation State Machine (ISM). From this, a processor model comprising four distinct processing phases (TSL, DHA, CCO, and ISM) and four message input queues is derived. Inside this submodel, there are different message routing paths, i.e., message chains. As an example, the message chain corresponding to an outgoing Dialogue Begin message containing two Invoke components is described below and depicted in Fig. 1.

- The primitive "TC-BEGIN req" received from the TC User is processed by DHA. Any Invoke components with the same Dialogue ID (in this case two components) are then requested from the CCO through a "request components" signal;
- The CCO processes the "request components" signal and generates three outputs signal (fork): two "operation sent" signals to the ISM (one for each Invoke component) and then one "requested components" signal to the DHA;
- Under reception of each of the two "operation sent" signals, the ISM starts an invocation timer. No output is generated (sink). The case of time-out can be modelled by a message branching with the branching probability given by the time-out probability;
- When the DHA receives the "requested components" signal, it composes a "TR-BEGIN req" primitive to the TSL;
- The TSL processes the "TR-BEGIN req" and requests the service of the Signalling Connection Control Part (SCCP) through an "N-UNITDATA" primitive.

The TCAP functional block is not restricted to the Dialogue Begin message, and comprises a set of messages consisting of a combination of all possible types for the transaction portion and component portion. The representation of all these message chains requires the extension of the model depicted in Fig. 1 through additional chains. The full model, which is depicted in Fig. 2, is obtained by considering the set of all possible message chains for the TCAP functional block.

The models for the functional blocks of the Levels 3 and 4 of the signalling network protocol architecture (MTP Level 3, SCCP, and ISUP) are obtained in the same way. The reader interested in more details about these models is referred to [23], [24].

MTP Level 1 is simply modelled as an infinite server with a service time representing the signalling link propagation delay. The modelling approach described before cannot be applied to the MTP Level 2 entities, because they are closely coupled via the error correction and flow control mechanisms on Level 2. Therefore, the approach included in the CCITT recommendations is adopted, where the corresponding queueing delay formulas are given explicitly.

4.2 Complete Model

In order to embed all these models in a realistic environment, it is necessary to extend them by adding traffic sources representing the traffic generated by the users, traffic sinks, response times of exchanges and users, database access delays, etc.

A better overview of a complete model for the Levels 3 and 4 can be obtained by representing the individual submodels in a reduced form, as depicted in Fig. 3. In this case, it is assumed that each functional block is implemented in a single processor. In MTP Level 3, the identified processes are the Message Discrimination (HMDC), Message Distribution (HMDT), and Message Routing (HMRT). The ISUP contains the Call Processing Control Incoming (CPCI), Call Processing Control Outgoing (CPCO), Message Distributing Control (MDSC), and Message Sending Control (MSDC). The processes of the SCCP are the Connection Oriented Control (SCOC), Connectionless Control (SCLC) and Routing Control (SCRC) and since they are used in both directions, transmitting and receiving, they can be split in one processing phase for each direction.

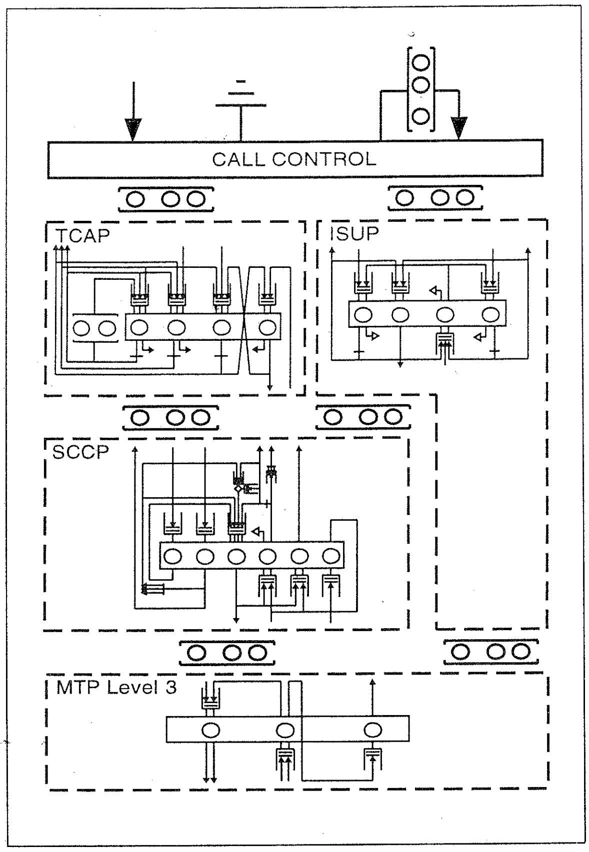


Fig. 3. Submodels for the functional blocks in an SP.

5. Analysis Outline

The complete model for the entire signalling network includes extended queueing network elements like, e.g., full-duplex flow-controlled links, priority processors, multiple-chain multiple-class traffic streams, segmenting and reassembling of messages, etc. The exact analysis of such a large and complex system is far beyond the current knowledge, and an approximation based on a combined decomposition and aggregation technique is used.

5.1 Decomposition and Aggregation

The principle of decomposition is to break up a complex system into its subsystems in order to achieve a reduction in the complexity of the whole system. This assumption is valid if the interactions between the subsystems are largely dominated by the local interactions inside each subsystem [26], [27]. This can be assumed to be sufficiently satisfied in this case. The network is decomposed into link sets, SPs, and STPs. Then, in the second decomposition step, the link sets are further decomposed into single signalling links, whereas the SPs and STPs are decomposed into their submodels, i.e., MTP Level 3, SCCP, ISUP, and TCAP.

The basic idea behind the signalling traffic aggregation is the observation that a particular subsystem is shared by a large number of connections, and, because of this, message streams belonging to individual connections need not be distinguished from the corresponding aggregate traffic streams. The aggregate arrival processes to each subsystem are approximated by suitable point processes, e.g., Poisson processes.

5.2 Analysis of Submodels

The application of these principles allows the analysis of each subsystem in isolation. For MTP Level 2, the formulas given in Recommendation Q.706 [5] can be used. The performance of models for the upper Levels depends on additional factors like, e.g., the priority and the distribution of the message processing times of each process.

In the models for the remaining functional blocks (MTP Level 3, SCCP, TCAP, and ISUP), a message departing from one processing phase can be fed back to another process in the same processor or even to the same process. In addition, one message can be forked or segmented upon feedback. This functionality is implemented in the switching system software, and a priority-based strategy is often used to schedule the process execution. This leads to the classification of the corresponding processor models into a rather general class of M/GI/1 priority queueing systems with feedback.

An exact analysis of the models of Levels 3 and 4 may become quite complex, but a mean value analysis, initially introduced in [28], is still possible in most cases. Previous work has been done in the analysis of priority queueing systems in which a message may feed back and change priority after having been served [29]–[32]. An algorithm considering also bulk arrivals, forking and branching of messages, and different preemption strategies is described in [30]. The application of this algorithm results in a system of linear equations for the mean sojourn times of all message chains in each process. The global performance results, e.g., message transfer times or call set-up delays, are obtained by composition of the submodel results.

6. A Planning Tool

Using the described methodology, a signalling network planning tool has been implemented. This planning tool takes into consideration the network configuration, the routing plan, the impact of the individual messages on the network components, the mix of scenarios, and the traffic matrix.

6.1 Basic Concepts

The network configuration and the routing plan are network-specific information, and include aspects related to the topology, equipment type, and routing strategy. The impact of the individual messages on the network components is obtained by a detailed description of the unitary message load characteristics.

The contribution of a message to the network load comprises the message length and the sequence of visited processes in the network elements on the path between its origin and destination. The information about the functional blocks visited by a message is obtained from the SDL (Functional Specification and Description Language) diagrams of the CCITT specifications, e.g., as shown in Fig. 1 for the TCAP Dialogue Begin message. A catalogue with the description for a large number of messages is avaliable, and the user just has to provide the message length and the processing time on each process of the underlying implementation.

The messages are used to compose the mix of scenarios representing the services supported by the network. It should be noted that even a single service may generate a set of subscenarios, e.g., a normal ISDN voice call consists of at least three subscenarios, i.e., a successful call, no answer, and destination subscriber busy. Each one of these subscenarios corresponds to a distinct message exchange over the signalling network and must be considered individually. The traffic matrix gives the amount of call attempts per second between two local exchanges. The execution of the tool can be divided into two main steps:

6.2 Message Flow Analysis

The analysis of the signalling network is started with a message flow analysis from given data about the network configuration, routing plan, mix of services types, and a traffic origination-destination matrix. This message flow analysis yields the message flow rates on each transmission or processing resource with respect to all message types. From message length and processing time distributions, the resource utilizations follow straightforwardly. In order to provide a better survey on the impact of the introduction of a new service, the load information is given with respect to each scenario at this stage of the analysis.

6.3 Performance Analysis

In a multivendor environment, the implementation of the functional blocks among the processors may not be the same for all SPs and STPs. For example, in a distributed SP architecture each functional block may be implemented in isolated processors, while in a centralized one, various functional blocks may share the same processor. The tool provides the user with the flexibility of defining her or his own architecture with arbitrary distributions of functional blocks among processors.

The analysis of the processor models requires the priority assignments of the processing phases within each processor. This priority assignment represents one of the parameters that determine the sojourn time for a particular message type. Depending on the priority assignment and traffic load, the differences between the sojourn times of distinct message types can be significant. Furthermore, the processor can be overloaded for some message types, but still be able to process higher priority messages; such a behavior is typical for priority systems.

With the remarks above, the user is required to provide – for each SP, SCP, and STP along the path – the number of processors and the distribution of the functional blocks between them. For each processor, the number of units and the priority assignment of the processing phases are also necessary.

These input data and the individual message arrival rates for each submodel obtained from the traffic flow analysis are then used to carry out a performance analysis for the submodels. In the cases, where there is more than one processing unit available, it is assumed that the incoming traffic is homogeneously distributed between the units. The mean sojourn times for the individual message chains are calculated using the algorithm described in [30].

The end-to-end transfer time of a particular message is computed by the summation of the individual transfer times, sojourn times, and other delays along the path through the network. As in the message flow analysis, this is again done by routines working with the network topology and routing strategy. With the end-to-end transfer time of a particular message sequence, it is possible to evaluate response delay parameters, such as connection set-up delays, data transfer delays, and database query delays.

7. A UPT Case Study

In this section, a simplified case study is provided to demonstrate the capabilities of the described planning tool. The introduction of the UPT service in a network supporting ISDN and PLMN is considered.

The UPT is a new service concept that improves subscriber mobility by deattaching the subscriber identity from the terminal identity. A UPT subscriber can register on specific access points across different networks (PSTN, ISDN, or PLMN) and can be reached by dialling her or his Personal Telecommunication Number (PTN). In an access point, the UPT subscriber can initiate calls, and the charge is provided on the basis of the PTN. According to the CCITT view [33], [34], the elements of the UPT functional architecture comprise a Registration Point (RP) for storage of dynamic and static data, a Call Routing Point (CRP) for UPT call handling, and an Access Point (AP) through which the subscriber utilizes the service.

7.1 Network Structure

The network topology depicted in Fig. 4 is considered. On the highest hierarchical level (level 1) there are three interconnected STPs. The next hierarchical level (level 2) contains 2 transit SPs of the ISDN subnetwork and 1 Gateway Mobile Switching Center (GMSC) for the PLMN. Under the transit SPs there are three SPs of the lowest level (level 3). Three Mobile Switching Centers (MSCs) are linked to the GMSC. Each link set consists of 4 links in the highest level and 2 links in the remainder of the network. The transmission capacity

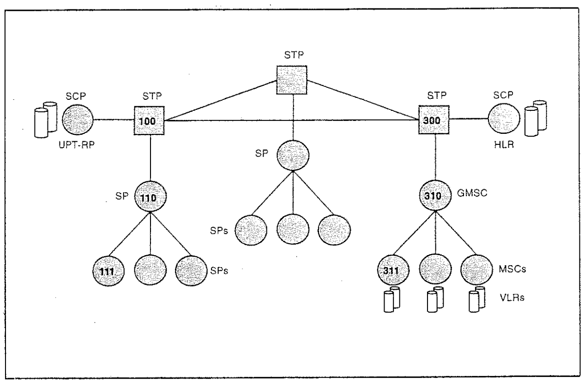


Fig. 4. Structure of the example network.

of each link is 64 kbit/s, and the propagation delay is 5 ms. The routing is assumed to be strictly hierarchical.

The information related to the UPT RP functionality is stored in an SCP connected to the STP identified by the code "100". For the PLMN, the Visitor Location Register (VLR) is integrated into the MSC, and the Home Location Register (HLR) is centralized in an SCP linked to the STP with code "300".

7.2 Service Mix and Network Data

The ISDN and PLMN voice services are in operation in the example network. The typical scenarios for these services can be classified into three subgroups: normal call, subscriber busy, and no answer. Each service comprises 70% successful calls, 20% subscriber busy, and 10% no answer. The PLMN total traffic corresponds to 10% of the ISDN total traffic.

The composition of the mobility-related signalling scenarios in the PLMN requires some assumptions to be made: at location update, the user is identified by her or his International Mobile Subscriber Identity (IMSI) instead of the Temporary Mobile Subscriber Identity (TMSI), which avoids an additional query to the old VLR; the VLR has the necessary information to manipulate an outgoing call; the HLR knows the Mobile Station Roaming Number (MSRN), etc. The consideration of all transactions produces a spectrum of messages generally with different message lengths. In order to simplify the example, the transaction-related messages were divided into two subgroups, i.e., short and long messages, according to the amount of information. A database query in one of the SCPs consumes 200 ms. The message types and their corresponding lengths are listed in Table 1.

The traffic between the ISDN and PLMN subnetworks is assumed to be unbalanced and characterized as follow: 5% of the ISDN traffic is directed to the PLMN, and 95% of the PLMN traffic is homogeneously distributed between the ISDN nodes.

The generated traffic corresponding to the support of the mobility procedures is a function of some network characteristics, e.g., cell size, ground topology, subscriber moving speed, etc. In this example, a simplified stationary approach is considered, with a PLMN subscriber generating a location update request for each call, and a handover being performed in 30% of the calls. Subsequent handovers are not taken into account.

Table 1. Messages considered in the example (A, B, and C refer, respectively, to ISDN, PLMN, and UPT; message lengths are in byte).

Type	Designation	Length		
		A	B	C
IAM	Initial address message	59	65	70
ACM	Address complete message	17	17	17
ANM	Answer message	15	15	15
REL	Release message	19	19	19
RLC	Release complete message	14	14	14
INV1	Invoke message (short)	-	20	25
RES1	Response message (short)	-	30	35
INV2	Invoke message (long)	-	40	45
RES2	Response message (long)	-	60	65

With the introduction of the UPT concept, it is also necessary to differentiate between the calls destined to from those originated from a UPT user currently registered in an ISDN or PLMN terminal. A percentage of the ISDN or PLMN traffic is substituted by the UPT service. The traffic between UPT users is considered to be negligible.

It is assumed that the UPT functions are merged into normal ISDN/PLMN call set-up, e.g., by including new UPT information in the related messages. The user is authenticated on registration, deregistration, incoming calls, and outgoing calls. The storage of the UPT authentication information is centralized in the UPT-SCP.

The comparison between the different possibilities of physical implementation of the SS7 functionality is beyond the scope of this work. In this example, an arbitrary architecture has been chosen, where the MTP and SCCP are implemented in the same processor. The ISUP and TCAP functions are assumed to be performed by isolated processors.

The priority strategy of the processors is non-preemptive. The priority assignment for the processes in a decreasing order of priority is:

ISUP processor: MSDC, CPCO, CPCI, and MDSC.

TCAP processor: TSL, DHA, CCO, and ISM.

MTP and SCCP processor: HMDC, HMDT, HMRT, SCLC-(R), SCLC-(T), SCRC-(R), and SCRC-(T).

The database query is performed using the TC. The processing times in the TCAP block of an SP are assumed to be 2 ms for the DHA, and 1.5 ms, 1 ms, and 0.5 ms for the CCO, TSL, and ISM, respectively. All processes of the SCCP have a processing time of 1 ms. For the ISUP, the processing time is 1 ms for the CPCI and CPCO, and 0.5 ms for the MSDC and MDSC. In the MTP, the processing time for a message in the HMDT and HMRT is 1 ms, and 0.5 ms in the HMDC.

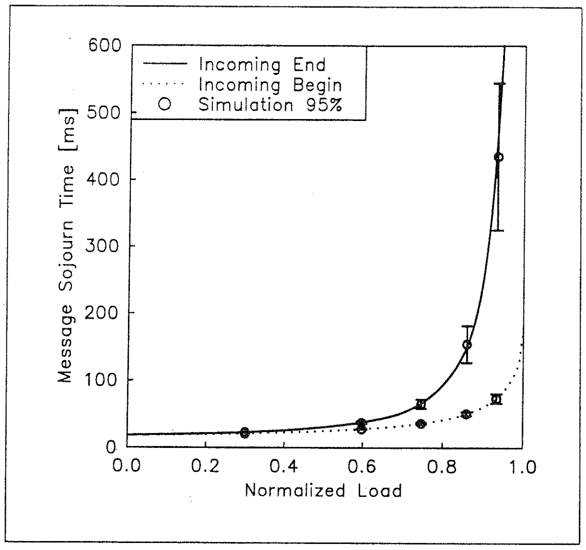


Fig. 5. Analysis and simulation results for an MSC.

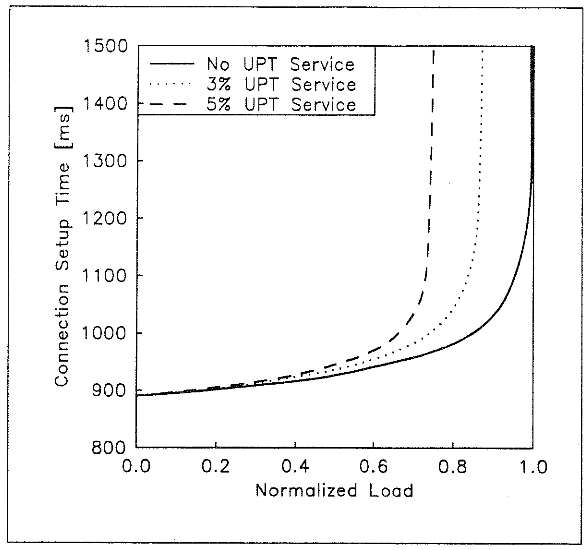


Fig. 6. Influence of the UPT concept on the ISDN–PLMN call set-up delay.

7.3 Accuracy

Before evaluating the network parameters, the accuracy of the adopted modelling approach has to be checked through a simulation study. A good agreement between analysis and simulation results suggests that the assumptions of the analysis do not introduce significant errors. Simulations were performed with 10 "batches", each simulating 90 seconds of an MSC. The warm-up phase comprised 60 seconds. The input data were obtained from the output data of the message flow analysis. This information was taken as input data for the signalling point analysis, i.e., the arrival rate of all messages destined to, originated from, and passing through the underlying node. The analysis and simulation results for the sojourn time in an MSC of an incoming TCAP Begin message with an Invoke Component and an incoming TCAP End message with a Result Component is shown in Fig. 5. The results are depicted with the call attempt loading normalized with respect to the maximum call attempt loading of the underlying MSC.

The difference between the sojourn times of the messages results from the fact that the incoming TCAP Begin message is not processed by the ISM process. The assignment of the lowest priority to the ISM implies that a message in the ISM must wait until the queues of all the remaining processes are empty, and the probability that this is true decreases with increasing load. This example illustrates a case where a processor is almost overloaded for a particular message type, while it continues to operate normally for other message types.

The total CPU time for the simulation run of a single load case was about 75 minutes, contrasting with the 50 seconds consumed for the analysis, both on a MicroVAX 3600. The computer resources required by the simulation highlights the limited applicability of the simulation techniques to large signalling networks.

7.4 Results

Firstly, the impact of the introduction of the UPT concept on both networks is studied. The UPT concept is introduced in a limited basis, and the parameter of interest is the connection set-up time between ISDN and PLMN. The results for a substitution of 3% and 5% of the voice service of the ISDN and PLMN are depicted in Fig. 6. The call attempt loading is normalized with respect to the maximum call attempt loading of the network carrying only ISDN and PLMN voice services.

The introduction of the UPT concept and the support of the mobility functions like, e.g., call routing, registration, authentication, etc., represent an additional load for the signalling network in the form of transactions to retrieve the necessary information. This load is not homogeneously distributed on the network and varies with the amount of introduced UPT services. In the cases of no UPT and 3% UPT service, the bottlenecks are located in the ISUP block of the transit ISDN-SP and in the ISUP block of the GMSC, respectively. For an introduction of 5% UPT traffic, however, the overload situation is observed in the TCAP block of the UPT-SCP. The QOS parameters of the existing ISDN and PLMN services are also affected by the UPT introduction.

The complexity of these new scenarios and the corresponding effect on the call routing, the number of database queries, the signalling network load, and the set-up times suggest the use of alternative architectures. These alternative architectures should consider factors like, e.g., database to database interactions, integration of databases, or allocation of IN functionality into the MSC.

The individual message transfer time between two nodes allows the computation of the delay involved in the exchange of any message sequence. With this information, there are various parameters of interest

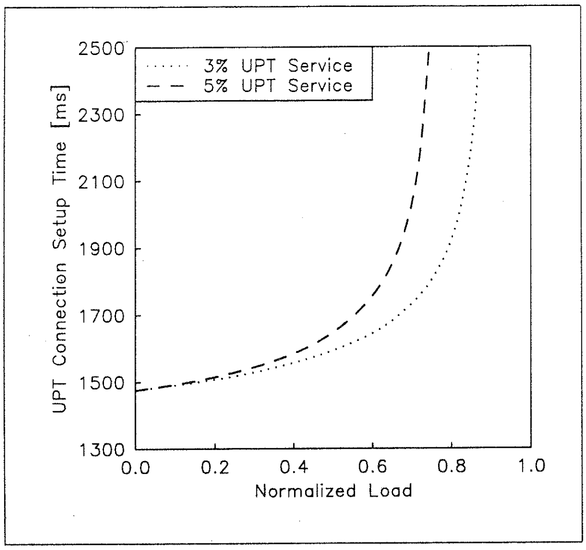


Fig. 7. Connection set-up delay from a UPT user registered in ISDN to a PLMN user.

that can be evaluated. The results corresponding to the connection set-up time for a call originated from a UPT user registered in the ISDN to a PLMN user is shown in Fig. 7.

The shape of the curves in Fig. 7 differs from those obtained for the connection set-up between ISDN and PLMN. This difference is explained by the larger number of elements involved in the UPT scenario.

8. Conclusion

The results show that quality of service parameters of future networks will be strongly influenced by the performance of the signalling network. The introduction of new services which require database interactions, such as UPT services and mobile communication services, will impact the signalling network performance, e.g., the processing load of signalling points and the end-to-end message transfer delays.

The presented modelling approach and the consideration of physical implementation aspects, such as the distribution of the processes among processors and priority assignment, cover important characteristics present in a multivendor environment. This modelling framework has been implemented in a tool suitable to support the planning of new signalling networks according to given service, load, and grade of service figures, or to detect bottlenecks or possible deficiencies in case of resource outages.

References

[1] Bafutto, M.; Kühn, P. J.: Capacity and performance analysis for signalling networks supporting UPT. Proc. 8th ITC Specialist Seminar on Universal Personal Telecommunication, Santa Margherita Ligure (Genova), Italy, 1992. 201–213.

[2] Duran, J. M.; Visser, J.: International standards for intelligent networks. IEEE Commun. Mag. **30** (1992), 34–42.

[3] Richards, P. S.: Rapid service delivery and customization in a developing network infrastructure. Performance Evaluation **25** (1993), 1031–1039.

[4] Robrock, R. B.: The intelligent network – Changing the face of telecommunications. Proc. of the IEEE **79** (1991), 7–20.

[5] CCITT: Blue Book, Recommendations Q.700–Q.795, Specifications of Signalling System No. 7. Vol. VI.7–VI.9. Geneva: Int. Telecommun. Union, 1988.

[6] British Telecom: CCITT signalling system no. 7. British Telecommun. Engng. **7** (1988), 1–79.

[7] Jabbari, B.: Common channel signalling system number 7 for ISDN and intelligent networks. Proc. of the IEEE **79** (1991), 155–169.

[8] Manterfield, R.: Common-Channel Signalling. London: Peter Peregrinus Ltd., 1991.

[9] Modarressi, A. R.; Skoog, R. A.: Signaling system no. 7: A tutorial. IEEE Commun. Mag. **28** (1990), 19–35.

[10] Modarressi, A. R.; Skoog, R. A.: An overview of signaling system no. 7. Proc. of the IEEE **80** (1992), 590–606.

[11] Antonios, M. N.; Perez, C. E.: Resource planning in common channel signaling (CCS) networks. Proc. 5th Int. Network Planning Symp., Kobe, Japan, 1992. 253–258.

[12] Bartolomé, A.; Brea, T.; García, J. A.: A computer tool for planning and performance analysis of common channel signaling #7 networks. Proc. 5th Int. Network Planning Symp., Kobe, Japan, 1992. 259–264.

[13] Bafutto, M.; Kühn, P. J.; Willmann, G.; Zepf, J.: A capacity and performance planning tool for signalling networks based on CCITT signalling system no. 7. – In: Intelligent Networks – The Path to Global Networking. Bayliss, P. W. (ed.). Washington: IOS Press, 1992, 368–379.

[14] Gómez, F. O.; Marrodán, M. P.; Gil, J. S.: Methodology for planning the common channel signalling network. – In: Network Planning in the 1990's. Lada, L. (ed.). Amsterdam: Elsevier, 1989, 293–300.

[15] Sancho, I. M.; Gallego, R. C.: Signalling network planning: A pragmatic approach in the Spanish Telecommunication Network. – In: Network Planning in the 1990's. Lada, L. (ed.). Amsterdam: Elsevier, 1989, 307–312.

[16] Tcha, D. W.; Kim, H. J.; Chang, S. G.; Lee, Y. H.: An interactive common channel signalling network planning system (SINEPS). – In: Teletraffic and Datatraffic Socio-Economic Aspects (ITC-13). Jensen, A.; Iversen, V. B. (eds.). Amsterdam: Elsevier, 1992, 445–450.

[17] Mitchell, D. C.; Collar, B. E.: CCITT signalling system no. 7: National user part. British Telecommun. Engng. **7** (1988), 19–31.

[18] Heidermark, A.; Halvorsen, T.: Signalling system number 7 in the nordic countries. Proc. 12th Int. Switching Symp., Stockholm, Sweden, 1990. 43–48.

[19] Conway, A. E.: Performance modeling of multi-layered OSI communication architectures. Proc. IEEE Int. Conf. Commun. (ICC), Boston, USA, 1989. 651–657.

[20] Conway, A. E.: Queueing network modeling of signaling system no. 7. Proc. IEEE Global Telecommun. Conf. (GLOBECOM), San Diego, USA, 1990. 552–558.

[21] Kritzinger, P. S.: A performance model of the OSI communication architecture. IEEE Trans. Commun. COM- **34** (1986), 554–563.

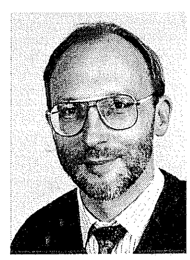
[22] Reiser, M.: Communication-system models embedded in the OSI-Reference model, a survey. – In: Computer Networking and Performance Evaluation. Hasegawa, T.; Takagi, H.; Takahashi, Y. (eds.). Amsterdam: North-Holland, 1986.

[23] Willmann, G.: Modelling and performance evaluation of

multi-layered signalling networks based on the CCITT no. 7 specification. – In: Teletraffic Science for New Cost-Efective Systems, Networks and Services, Part 1 (ITC-12). Bonatti, M. (ed.). Amsterdam: Elsevier, 1989, 930–940.
[24] Willmann, G.; Kühn, P. J.: Performance modeling of signaling system no. 7. IEEE Commun. Mag. **28** (1990), 44–56.
[25] Conway, A. E.: A perspective on the analytical performance evaluation of multilayered communication protocol architectures. IEEE J. Sel. Areas Commun. JSAC- **9** (1991), 4–14.
[26] Courtois, P. J.: Decomposability: Queueing and Computer System Applications. New York: Academic Press, 1977.
[27] Lavenberg, S. S. (ed.): Computer Performance Modeling Handbook. New York: Academic Press, 1983.
[28] Cobham, A.: Priority assignment in waiting line problems. J. Operations Res. Soc. of America **2** (1954), 70–76.
[29] Paterok, M.; Fischer, O.: Feedback queues with preemption-distance priorities. ACM SIGMETRICS Performance Evaluation Review **17** (1989), 136–145.
[30] Paterok, M.: Warteschlangensysteme mit Rückkopplung und Prioritäten. Dissertation. Universität Erlangen Nürnberg, Institut für Mathematische Maschinen und Datenverarbeitung, 1990.
[31] Simon, B.: Priority queues with feedback. J. of the ACM **31** (1984), 134–149.
[32] Villén-Altamirano, M.; Fontana, B.: Models to evaluate response times in single-processor systems and their application to a multiprocessor system. – In: Teletraffic Science for New Cost-Efective Systems, Networks and Services, Part 1 (ITC-12). Bonatti, M. (ed.). Amsterdam: Elsevier, 1989, 402–411.
[33] CCITT Study Group XI: Report on the meeting held in Geneva from 8-12 april 1991. Tech. Rept. COM XI-R 60, Geneva: Int. Telecommun. Union, 1991.
[34] CCITT Study Group XI: Report on the meeting held in Geneva from 21-25 september 1992. Tech. Rept. COM XI-R 262-E, Geneva: Int. Telecommun. Union, 1992.

Marcos Bafutto was born in 1963 in Rubiataba, Brazil. He received the B.E. degree in Electrical Engineering from the Federal University of Goiás in 1984, and the M.Sc. degree also in Electrical Engineering from the Federal University of Uberlândia in 1989. He joined Telecommunications of Goiás S.A. (Telegoiás) in 1984. Since 1989, he is on leave of Telegoiás working at the Institute of Communications Switching and Data Techniques of the University of Stuttgart towards a Dr.-Ing. degree. His interests are the modelling and performance evaluation of signalling networks, mobile communications, and Intelligent Networks.

Paul J. Kühn was born in 1940 in Grüssau, Germany. He received the Dipl.-Ing., Dr.-Ing., and Dr.-Ing. habil degrees from the University of Stuttgart in 1967, 1972, and 1981, respectively. He held the position of Research Head for Computer Performance Modelling at the University of Stuttgart from 1973 to 1977. In 1977, he joined Bell Laboratories in Holmdel, NJ, USA, where he worked in the field of computer communications. In 1978, he was appointed full professor for Communications Switching and Transmission at the University of Siegen, Germany. Since 1982, he is holding the Chair for Communications Switching and Data Techniques of the University of Stuttgart. His areas of interests include communication switching, computer and communications systems performance evaluation, and computer networks. Professor Kühn is a member of IEEE, ACM, ITG (German Communications Society), and GI (German Informatics Society). In 1977, 1979, and 1983, he became a member of the Communications Switching Committee of the NTG, the International Advisory Council of the International Teletraffic Congress, and Working Group 7.3 of the IFIP, respectively. He was appointed Vice President of the ITC in 1985, and Governor of the ICCC in 1987. In 1989, he was elected Fellow of the IEEE. He was also Program Chairman of the International Conference on Computer Communication (ICCC) in 1986, and various national conferences on computer communication and performance modelling. In 1990, Professor Kühn was appointed Professeur Associé at the Télécom Paris/ENST. In 1991, he was elected Chairman of the International Advisory Council of the International Teletraffic Congress (ITC).



Gert Willmann was born in 1954 in Münnerstadt, Germany. He received the Dipl.-Ing. degree in Electrical Engineering from the University of Siegen in 1983. He then joined Standard Elektrik Lorenz AG (SEL) in Stuttgart, where he was involved in the design of communication networks and in software engineering for digital switching systems. From 1985 to 1989, he was a member of Scientific Staff at the Institute of Communications Switching and Data Techniques of the University of Stuttgart, where he worked mainly in the modelling and performance analysis of common channel signalling networks and in queueing network analysis. During 1990 and 1991, he worked for a research project at the University of Stuttgart and as a consultant in the area of common channel signalling. In 1992, he joined Alcatel SEL AG, Stuttgart, as a supervisor in the Public Switching Systems Division. His current activities are in the area of traffic and performance modelling, analysis, and optimization of communication systems, and of communication network planning. G. Willmann is a member of the IEEE.