Signalling Networks for ISDN, IN and Mobile Networks - Modelling, Analysis, and Overload Control

P.J. Kühn, M. Schopp
University of Stuttgart
Institute of Communications Switching and Data Technics
Seidenstraße 36
D-70174 Stuttgart

Abstract

Signalling networks are the spine of the control in telecommunication networks. Besides the original purpose of connection control, signalling networks form the basis for Intelligent Network applications, mobility management in mobile networks and for personal communication services. Finally, signalling networks are the basis for operations administration and maintenance applications. According to these new features, the load on signalling links and signalling control functions increases rapidly and has to be taken into consideration in planning signalling resources for capacity and grade of service provisioning.

In this paper a rigorous approach is outlined for the systematic modelling of signalling networks, traffic flow analysis, capacity planning, and performance analysis under both stationary regime as well as overload situations.

1 Introduction

Common channel signalling is a concept for the out-of-band exchange of control signals by which separate data link channels are used between stored programme controlled switching nodes. Control signals, such as for connection establishment or release are carried by packetized messages which are transported through the signalling network in a connectionless mode. This form of control allows a fast and safe transfer of signals directly between the processors of the switching nodes. A layered protocol architecture provides the enhancement of the basic signalling message transfer capabilities to support various application contexts such as Telephone, ISDN or transaction oriented User Parts.

Originally, signalling networks have been introduced mainly for connection control. The extension of telecommunication networks to intelligent, mobile, and personal communication services makes use of these universal signal transport capabilities. On the other hand, these types of applications add traffic loads to the signalling network which interfere with the connection-related signalling traffic on signalling links or within signalling control processing elements affecting traffic loading and real-time performance. This has motivated our studies for a universal approach to the analysis of signalling networks about ten years ago.

Signalling networks have been subject of intensive performance studies as a basis to understand their working and for planning. A fairly good state of the art has been presented in the last IEEE JSAC on Common Channel Signalling Networks: Performance, Engineering, Protocols and Capacity Management, see [15]. In the following, an outline of the universal modelling and analysis approach is given. the paper is structured as follows: In Chapter 2 a short description of the architecture of the Signalling System No.7 (SS7, [7]) is given with its basic modules (parts) and an example signalling scenario. Chapter 3 addresses aspects of new functions such as for IN, Mobile, or UPT. The main Chapter 4 outlines the modelling and analysis approach. the basic concept of it has been published before, see, e.g., references [1], [2], and [3], consisting of a functional model, a traffic model, traffic flow analysis and decomposition. Chapter 4.1 addresses in particular the Flow Analysis approach for IN and Personal Communications, see also [4], [5], and [6]. In the following sub-chapters, features of the resulting component models and the performance evaluation are summarized. Chapter 5 addresses issues of overload and flow control in signalling networks, particularly under the view of IN and new services impacts. The final Chapter 6 reports on an illustrative case study example, which evaluates the introduction of two IN services into an example network.

2 The Signalling System No. 7 (SS7)

2.1 SS7 Protocol Architecture

Fig. 1 shows the structure of the SS7 protocol architecture and its relationship to the OSI layering. The initial specification of SS7 was oriented towards the needs of circuit-related telephony control. It was specified in four functional levels: The Message Transfer Part (MTP) comprising levels 1-3 and all other functional blocks as level 4 user parts. The SS7 evolution has been influenced by the need to encompass a broader role and it became more aligned with the OSI reference model. The result of this evolution is, that functional levels and OSI layers co-exist in SS7. For example, the Signalling Connection Control Part (SCCP) is a level 4 user part which provides, together with the MTP, OSI network layer services to other level 4 user parts like the ISDN User Part (ISUP) and to the Transaction Capabilities (TC). In the following, an introduction to the different functional blocks will be given.

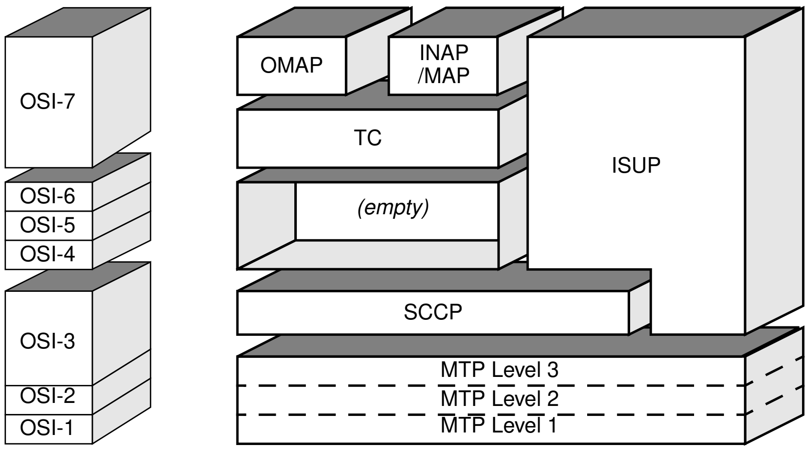


Figure 1: SS7 protocol architecture and its relationship to the OSI layering

2.2 The Message Transfer Part (MTP)

The MTP provides reliable transfer and delivery of signalling information across an SS7 network and has the ability to react to system and network failures with the aim to ensure that reliable transfer is maintained. MTP addressing is limited to signalling points within a single signalling network, e.g. within the signalling network of a network operator. Internetwork signalling can only be performed via MTP users like the ISUP or the SCCP. The MTP provides a sequenced connectionless service and consists of three functional levels: the signalling data link functions (level 1), the signalling link functions (level 2), and the signalling network functions (level 3).

MTP level 1 fully complies with the OSI definition of the physical layer. The recommended bit-rate for the international standard is 64 kbit/s. MTP level 2 corresponds to the OSI data link layer. Error correction is usually performed using a positive/negative-acknowledgement go-back-N retransmission error correction method (Basic Error Correction Method). When congestion is detected at the receiving end of a signalling link, a flow control procedure is invoked, which periodically notifies the transmitting end of a busy condition and which withholds acknowledgements. Congestion onset and abatement detection is an implementation dependent function. In case of a level 3 failure, a processor outage condition indication is sent by level 2 to the far end level 3.

MTP level 3 signalling network functions are divided into two basic categories: signalling message handling functions and signalling network management functions. The purpose of the signalling message handling functions is to ensure the delivery of signalling messages to a specific user part at a specific signalling point as indicated by a sending user part at another signalling point. .As illustrated in Fig. 2**,** three functions are defined as signalling message handling functions: the

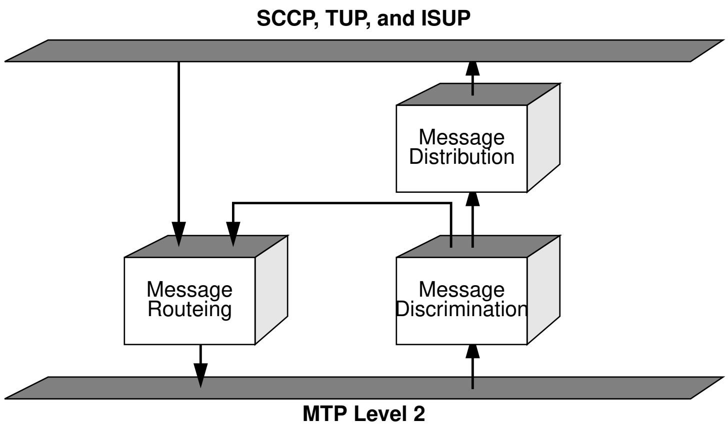


Figure 2: MTP message handling functions

message discrimination, the message routeing, and the message distribution function. Message routeing is based on the Destination Point Code (DPC) and the Signalling Link Selection (SLS), which are part of the routeing label of signalling messages. SLS is used for load sharing. The routeing label consists of DPC, OPC (Originating Point Code), and SLS. Messages with the same routeing label usually take the same route through the signalling network, which - together with level 2 error correction methods - enables a *sequenced* connectionless service

The signalling network management functions provide reconfiguration of the signalling network in case of congestion or blockage. The goal is, that, when a failure or congestion occurs, reconfiguration is carried out, so that messages are not lost, duplicated, or put out of sequence.

2.3 The Signalling Connection Control Part (SCCP)

The SCCP enhances the services of the MTP to provide OSI network layer services for circuit-related and especially for non-circuit-related signalling. It provides four classes of services: two for connectionless and two for connection-oriented services. Class 0 is pure connectionless. Class 1 is sequenced connectionless, based on the MTP service. Class 0 and class 1 are used by the Transaction Capabilities to support applications like Intelligent Network (IN) services and mobile communications (GSM). Class 2 is connection-oriented. Sequencing is based on the MTP sequenced connectionless service, and segmenting and reassembly is performed if necessary. Class 2 is used by the ISUP for end-to-end signalling. In class 3, the features of class 2 are enhanced by flow control and detection of message loss or missequencing.

SCCP addressing is based on DPC and Subsystem Number (SSN). The SSN identifies a SCCP user (application process) at a signalling point. A very important enhancement to these addressing capabilities is the possibility to address SCCP users by using global titles (GT). Global titles are addresses, such as dialled digits, that will not be recognized in the SS7 network. For routeing, a translation capability in the SCCP is required (Global Title Translation) which translates a GT into a DPC and possibly also a new SSN and a new GT. By using global titles, worldwide routeing of signalling messages becomes possible. The Home Location Register (HLR) of a mobile subscriber, for example, can be addressed by using the International Mobile Subscriber Identity (IMSI).

No congestion control procedures in response to MTP congestion are defined for the SCCP. SCCP takes no other action than informing all its subsystems (SCCP user) about the congestion. In case of signalling point or subsystem inaccessibility or failure, SCCP management might change global title translation to backup addresses.

2.4 The user parts

The Telephone User Part (TUP) and the ISDN User Part (ISUP) provide the signalling functions to support basic bearer services, as well as supplementary services. They use the MTP for reliable in-sequence transport of signalling messages between exchanges. Fig. 3 shows a typical call setup and release procedure between an originating and a destination exchange via a transit exchange using ISUP messages.

ISUP also provides signalling functions supporting supplementary services which might need end-to-end signalling between originating and destination exchange. End-to-end signalling can either be realized by using the pass-along method, where messages are passed between the ISUPs involved in call control, or by using the services of the SCCP.

2.5 The Transaction Capabilities (TC)

TC provides the means for non-circuit-related communication between application entities. It provides the means to exchange operations and replies via a dialogue. Typical TC users are application entities of the Mobile Application Part (MAP), the Intelligent Network Application Part (INAP), and the Operations Maintenance and Administration Part (OMAP).

3 The changing Role and Characteristics of Signalling

TUP is defined primarily to control the establishment and release of speech paths. TUP signalling is circuit-related. Signalling relations exist between exchanges which are directly interconnected by speech circuits. Associated signalling, where signalling is performed via a direct signalling

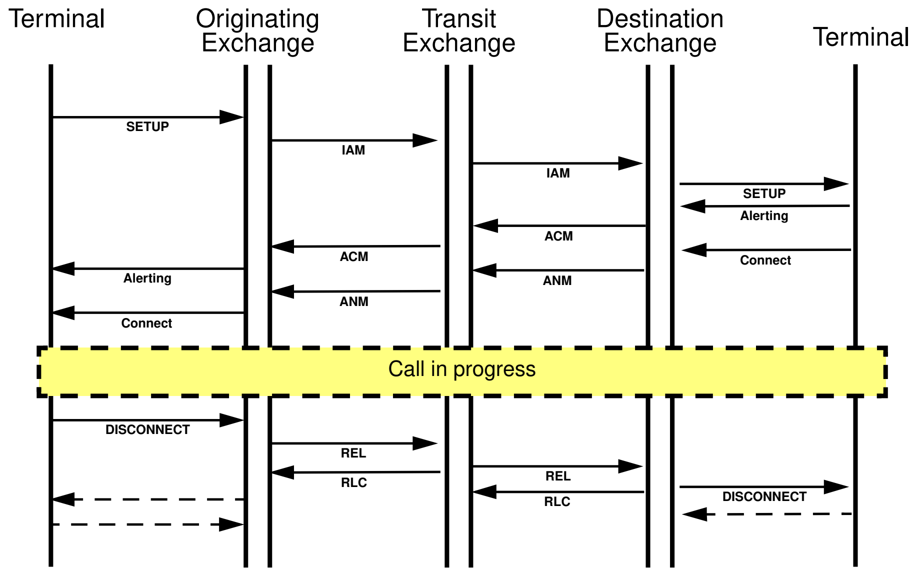


Figure 3: ISDN call setup and release signalling scenario

link between two signalling points, is very common here. But even in the case of quasi-associated signalling, where messages are routed via other signalling points (signalling transfer points), signalling traffic characteristics are easy to predict.

With the introduction of ISUP, not only the amount of signalling is increased (due to more parameters and additional supplementary services), but also the characteristic of the signalling traffic is slightly changed. End-to-end signalling via the SCCP not necessarily follows the traffic path (speech path). Global title addressing in some supplementary services even needs Global Title Translation (GTT), which might be provided by dedicated SCCP relay points, which could be placed anywhere in the network.

Services like the 800/FREEPHONE service need number translation. An 800-call is routed to an exchange which is able to query an 800-database. Call processing is suspended and the 800-database is queried via non-circuit-related signalling (TC and SCCP) and replies with the translated number in the same way. Addressing might be based on global titles. The call then is routed according to the translated number. Since the exchange of query and response messages adds to the post-selection delay, it has to be done on a real-time basis and the signalling network has to have a good performance. The characteristics of this non-circuit-related signalling differ significantly from the characteristics of TUP or ISUP signalling. Nevertheless, the according signalling scenarios are rather simple and predictable.

Signalling in PLMNs like GSM on the other hand is much more complex. Fig. 4 shows a GSM signalling scenario where a mobile terminal (MT) performs a location update which involves its previous Visitor Location Register (pVLR), its new Visitor Location Register (VLR), its Home Location Register (HLR) and a Mobile Switching Centre (MSC). A large number of different signalling scenarios can be identified. They make intensive use of non-circuit-related signalling and are even often non-call-related. The frequency of their occurrence and between which physical entities they take place depends on many factors like user behaviour, network structure, and protocol parameters. When the characteristics of the signalling traffic have to be predicted, all these factors have to be taken into account and additional user and protocol modelling is needed. Also international roaming and Global Title Translation has to be taken into account. Since users may roam between different networks, signalling to distant HLRs might be required. HLRs are usually addressed via global titles, like e.g. the IMSI. Global title routeing is the basis for international roaming.

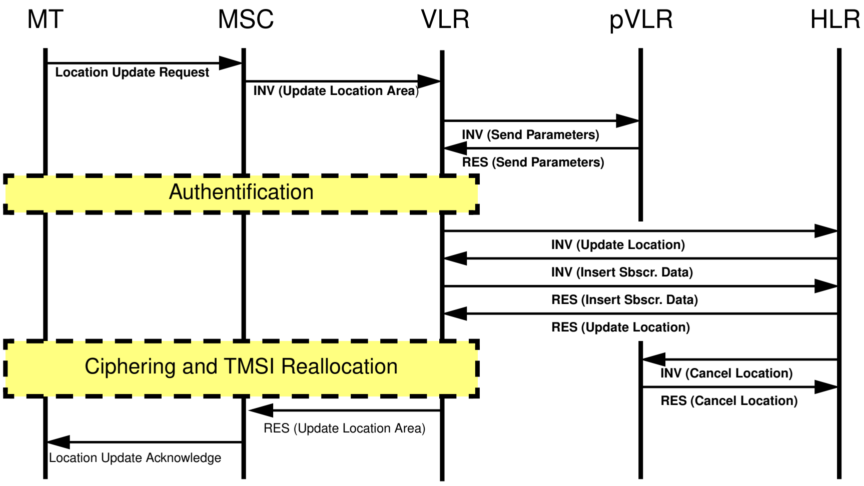


Figure 4: GSM location update signalling scenario

The basic idea of the Intelligent Network (IN) concept is the separation of call control and service control. Service control is usually provided on a centralized basis and signalling between call control and service control is non-circuit-related. The 800/FREEPHONE service is an early realization of this concept. GSM can also be regarded as an intelligent network. But while in GSM the specified interfaces between call control and service control are service-specific, the ITU-T recommendations on the Intelligent Network [12] define service independent interfaces in order to enable the fast and flexible introduction of many different new services.

Signalling in an IN environment very much depends on the offered services, their realization, the way they are used by subscribers, and the location of service logic and service data within the network. Each service might result in a number of signalling scenarios. The flexibility of the concept, the large number of expected new services, and the dependability on user behaviour makes it difficult to predict the actual signalling traffic and its characteristics.

The ITU-T IN Capability Set 1 (IN CS-1, [12]) defines a very limited number of interfaces between functional entities in an IN. This simplifies the derivation of signalling scenarios from service specification. Other capability sets which will be defined in the future might have many more interfaces and will make it more difficult to derive the according signalling scenarios.

At the moment a working group within ETSI GSM standardisation[1], called CAMEL (Customized Applications for Mobile network Enhanced Logic), defines a concept to provide IN-type services to mobile subscribers within GSM networks. This concept is similar to the ITU-T IN concept with special care being taken to enable service provisioning over several GSM networks. Signalling is comparable to IN CS-1 signalling.

[1] Under control of SMG3 WPC and SMG1 WPA

With the introduction of personal numbers and Universal Personal Telecommunications (UPT) services, signalling will increase dramatically. Every call will involve non-circuit-related signalling. When tens or hundreds of millions of people shall be able to communicate anytime and anywhere, a significant amount of signalling will be needed for location management. The location of a user in relation to service data and data about the user's or other user's location becomes important. Location management will probably be based on distributed databases (e.g. in third generation mobile communication systems) and has to be able to cope with an multi-operator environment. The more service logic and service data will be distributed, the more difficult it will be to predict signalling traffic in detail. The introduction of TINA (Telecommunications Information Networking Architecture) concepts will be another step in this direction.

4 Capacity Planning and Performance Modelling

During the design and development phase of a signalling network, or when new services are to be introduced into an existing network, capacity and network performance figures are needed for planning purposes. Overall signalling network performance aspects such as, for example, post-selection delays or end-to-end signalling information transfer times are important quality of service criteria. A method to obtain these figures is presented here.

In a first step, a global traffic flow analysis has to be performed to determine loading of signalling links and of physical entities which process signalling network functions or application processes.

In the case of TUP, ISUP, FREEPHONE, and even GSM signalling, flow analysis can be carried out on the basis of representative signalling scenarios, in conjunction with more or less complex traffic matrices. In the case of IN signalling, the flexibility of the IN concept and the large number of new services make the extension of this approach necessary.

4.1 Flow Analysis for IN and Personal Communications

We present a hierarchical modelling approach where the load generated by new IN services on the signalling system and on the IN physical entities is derived from the service description on the global functional plane of the IN conceptual model. This is done in several steps including the derivation of service scenarios from the service description, the linearization and the refinement of service logic, and the mapping onto the different planes of the IN conceptual model, resulting in a signalling message flow and functional entity actions in physical entities of the signalling network and the IN.

A framework for the design and description of the IN architecture is given by the IN conceptual model (Fig. 5). This model is divided in four planes addressing service aspects, global functionality, distributed functionality and physical aspects of an IN.

The service plane represents an exclusively service-oriented view and no aspects regarding the implementation of the services are handled. The services of the service plane are mapped onto the service logic in the global functional plane. A service logic is a chained sequence of standardized, monolithic, and reusable network capabilities called Service Independent Building Blocks (SIBs). The SIBs are service independent and describe a single complete activity.

The functionality required to support the realization of the SIBs are contained in the functional entities (FE) of the distributed functional plane. Important functional entities of IN CS- 1 are Service Switching Function (SSF), Service Control Function (SCF), Specialized Resource Function (SRF), and Service Data Function (SDF).

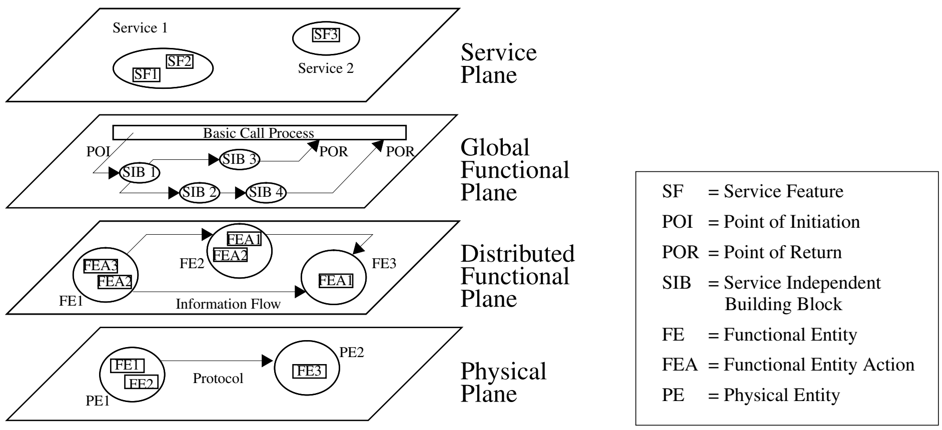


Figure 5: IN conceptual model

These functional entities are mapped onto the physical entities of the physical plane. Typical physical entities of the IN Capability Set 1 are Service Switching Points (SSP), Service Control Points (SCPs), Intelligent Peripherals (IPs), and Service Data Points (SDPs).

The basic components of IN service definitions are the SIBs. A SIB can result in distinct information flows in the physical plane. The information which SIB is executed is usually not sufficient to perform a flow analysis. A SIB can result in very different information flows depending on several factors. The User Interaction SIB, for example, impacts the network in different forms depending on factors like, e.g., number of announcements to be played, amount of information to be collected, the location of the IP, etc. Therefore, a more detailed view of the SIBs must be taken into consideration.

In [4],[5] and [6] all the possible combinations for the input and the possible results of a SIB are identified and isolated as a new element denominated sub-SIB. The objective of the sub-SIB is to cater for the set of combinations which influence global network performance. A new service is then defined as a chain of these sub-SIBs. Generally, the service logic depends on decision events, and consequently the chain of sub-SIBs results in a complex structure.

Sub-SIBs are made available in a catalogue as the new basic elements for the service logic definition in a planning environment. A new service is defined as a chain of these sub-SIBs. Generally, the service logic depends on decision events and consequently the chain of sub-SIBs approaches a tree structure. For a flow analysis in a tool environment, this tree structure is not feasible and all possible paths are represented linearly in scenarios of the service logic. The resulting information flow between FEs for each individual sub-SIB can be exactly determined from the ITU-T recommendations. The FEs are, in this point only functional elements that are further mapped on the physical entities (FE). The information flow in the distributed functional plane is mapped onto an exchange of signalling messages in the physical plane. This mapping takes the actual physical network architecture into account.

The flow analysis includes the following steps:

- Definition of service logic by using sub-SIBs
- Identification of typical service scenarios and linearization
- Mapping onto the distributed functional plane (sub-SIB catalogue)
- Mapping onto the physical plane.

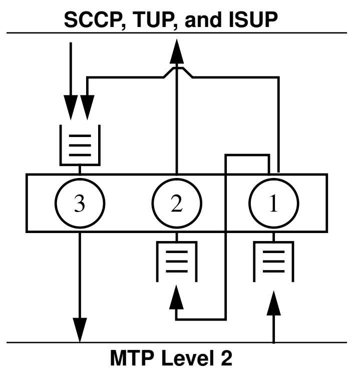
This approach yields loading of physical entities of the IN as well as of the signalling network. The limited number of interfaces in the ITU-T IN CS-1 recommendations makes this approach feasible. In an environment, where more interfaces exist, like e.g. between SCF and SCF or between SDF and SDF, the number of sub-SIBs and scenarios might explode, and the mapping of information flow onto the physical plane becomes much more complex. In a personal communications environment with distributed databases and sophisticated mobility management and data distribution techniques, it is difficult to determine the resulting information flow. Adequate user and protocol modelling is required, which has to be studied in conjunction with the above mentioned mobility management and data distribution techniques.

4.2 Performance Modelling

The flow analysis yields important capacity figures. In order to be able to obtain performance figures, a performance modelling and evaluation methodology is needed, which is applicable to the signalling network as well as to the application processes. The methodology presented here follows [20] and is based on the development of generic submodels for all higher-level functional blocks of SS7. Each submodel is directly derived from the corresponding functional specifications. Block-specific management processes are not taken into consideration. For convenience, and without loss of generality, a generic approach is taken, where each functional block may be assumed to have its own processor, and each function inside a functional block is represented by a corresponding service phase of the processor queueing model. Other system structures can be accomplished by simply changing this generic assignment of functions to processors. Similarly, one function can also be distributed among several processors.

Generic submodels for MTP levels 1 and 2 are not introduced in this article. A collection of formulas extracted from several publications on the performance analysis of signalling links can be found in recommendation Q.706 [7]. These formulas will later be used for analysis of level 2 delays.

The submodel of MTP level 3 only considers the signalling handling functions. The processor model is depicted in Fig. 6 and has three distinct processing phases and three message input queues - each of these for one of the three signalling handling functions.



1:Message Discrimination
2:Message Distribution
3:Message Routeing

Figure 6: MTP level 3 submodel

Since SCCP signalling is dominated by class 0 and class 1 services, a submodel is presented here which only covers functions related to the connectionless services. They are: Routeing Control Receiving, Routeing Control Transmitting, Connectionless Control Receiving, and Connectionless Control Transmitting (Fig. 7).

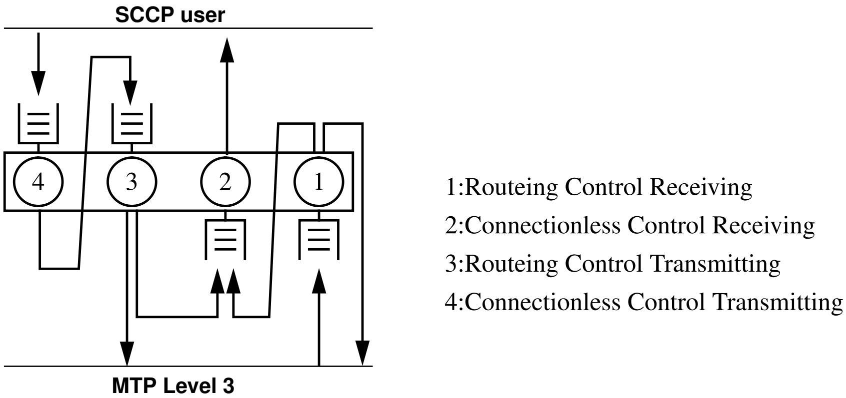


Figure 7: SCCP submodel for connectionless services

The submodels for ISUP and TC are slightly more complex and can be found in [20]. They include segmenting and forking of messages and timers. The processing phases in the ISUP submodel are: Message Distribution Control, Message Sending Control, Call Processing Control Incoming, and Call Processing Control Outgoing. The processing phases of TC are: Transaction Sublayer, Dialogue Handling, Component Coordinator and Invocation State Machine.

Application Processes like Call Control, INAP, or MAP are not in the scope of this article, since the focus is on signalling. As generic submodels simple $\Sigma M_i/G_i/1$ systems are used.

In [1], [2] and [3] the derived submodels are designated as ‘virtual‘ in the sense that they can be implemented in different ways. They have to be mapped onto physical implementations for each signalling point, which allows performance evaluation in a multi-vendor environment.

4.3 Performance Evaluation by Decomposition

In the preceding section, we have described a generic modelling methodology. Starting from this modelling framework, an approximate performance evaluation methodology based on decomposition techniques is presented in [20]. Under certain conditions, the concept of decomposition allows to break up a complex system into distinct subsystems which then can be analysed - at least approximately - in isolation. Basically, a signalling network can be considered to belong to the class of open networks with requests coming from several external traffic sources. However, the large number of individual communication relations that coexist in a signalling network makes it impossible to take each single communication relation into account. The observation that each node (processor) is shared by many individual signalling relations, where each relation does not contribute much to the total loading, allows to approximate the aggregated arrival process by a Poisson process.

The analysis of a processor is based on a rather general class of $\Sigma M_i/G_i/1$ priority queueing models with feedback, where feedbacks might result in multiple feedback messages (segmenting, forking). The analysis of such queueing models is based on Cobham‘s original publication [9], which describes the solution of feedback-free nonpreemptive priority queueing models by a mean value analysis technique (method of moments). This method has later been extended by Herzog

[11] and Krämer [14], to cover preemptive and preemption-distance priority schemes as well as simple feedback. Further generalizations ([19]) have led to a publication by Paterok and Fischer [17], [18]. With this, most of the internal mechanisms in the presented generic submodels and some possible implementations can be taken into account, and an analysis using this technique yields mean values of sojourn and waiting times of messages for each processing phase. From these results, together with the MTP level 2 delays, mean end-to-end transfer times of interest can be obtained.

4.4 Open Issues

Some issues must be considered for further research. Different options how to decompose the system exist. Each option must be validated by comparison with simulation or measurement. The characterization of end-to-end delays by mean values might not be sufficient. Modelling and analysis methods which yield higher moments and/or distributions are needed.

5 Overload Control

Overload in signalling networks may be originated by various causes such as misrating signalling traffic, introduction of new services, focused overload by subscriber behaviour (special events), or by outage of signalling or communication network resources. Each of these causes may be detected by a special pattern of indicators such as message buffer occupations, link loads, processor loads, message transfer delays, or delayed response signals. Accordingly, a set of different overload control actions has been suggested to defeat an overload situation; examples are

- message flow control by credit or explicit start/stop message mechanisms
- selected dropping of excessive signalling messages
- priorization of important signalling traffic
- routeing of signalling messages
- delaying of deferrable work

which have to be activated after overload detection. An extensive review of overload control in signalling networks has been reported in a recent PhD thesis [22].

Overload control actions can be further categorized with respect to the range of control, such as link level, processor, nodal or network wide control. Overload control has already been taken up as a part of the standardization within the CCS No.7 system:

- error and flow control on level 2 of the MTP
- load sharing, link changing and flow control on level 3 of the MTP
- stepwise load reduction and time-controlled resumption of signalling load on the ISDN User Part level
- automatic call gaping in transaction oriented application parts.

These mechanisms have been systematically modelled and analysed by simulation as well as analytically [22]. Other studies on similar problems have also been reported, see e.g. [10], [13], [16].

As an example, the basic mechanisms of the dynamic behaviour of an overload control scheme can be modelled by a multi-load-level single server queue, see Fig. 8. The server stands for either an overload link downstream towards the Service Control Point (SCP), where IN related traffic concentrates or for the SCP itself. Distant traffic sources S_i can be controlled, i.e., switched on/off or throttled, dependent on the current load level of the queue. Note, the load levels for the onset of

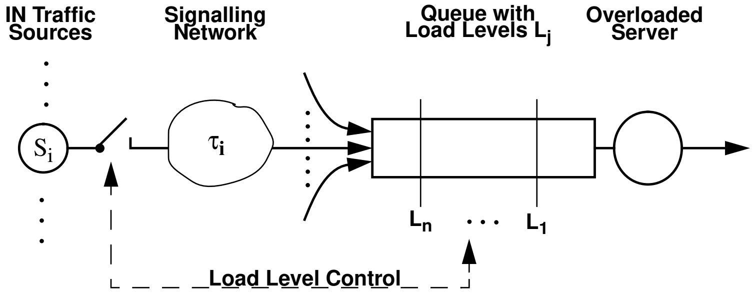


Figure 8: Basic Multi-Load-Level Control Model

control for load reduction and for the reset for resuming the full load may be different (hysteresis). This model has been analysed analytically as a modified finite buffer, multi-source M/G/1 queue [22].

6 A Simple Case Study

In order to provide an overview on the application of the modelling concept, a simple case study is provided. A network topology consisting of an SCP, an IP, and two SSPs interconnected by an STP is considered. A real network would be much more complex, with a larger number of SSPs and different topology arrangements, but nevertheless, the analysis can be carried out in the same way.

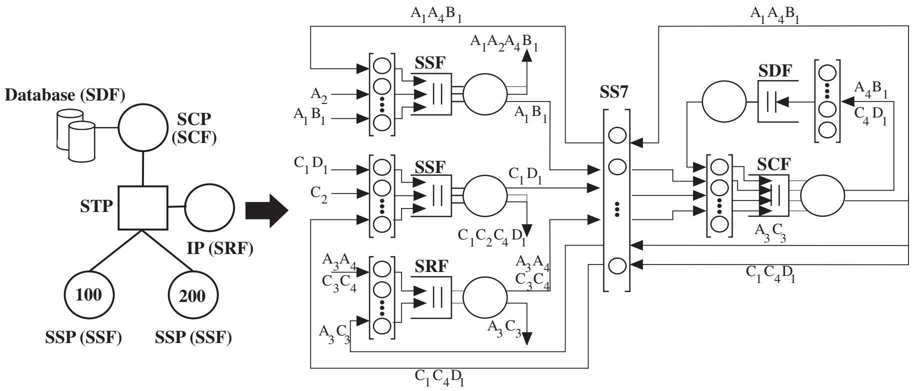


Figure 9: Topology and derived model for the case study

The IN services implemented in the network are the Freephone and the Credit Card Service. The service script of the Freephone service requires an address translation to determine the call routeing and may be designed using a sub-SIB of the Translate SIB. The Credit Card service is performed sub-SIBs of the User Interaction SIB and the Screen SIB. The topology and the corresponding model for the network supporting the underlying services are depicted in Fig. 9. The chains identified by (A) and (B) correspond, respectively, to the Credit Card and Freephone service originated at the SSP with code 100. The same services generated at the SSP with code

200 are represented by the chains (C) and (D) respectively. The resulting model in this case is a multiple-chain open queueing network. The number of chains is a function of the complexity of the related services and may be large for networks with a great variety of services.The realization of the IN services requires the exchange of messages over the signalling network for the setup and release of connections. These signalling scenarios are not represented in the IN model, but they must be considered in the analysis of the signalling network. The credit card service, for example, requires additionally the setup of connection between the originating SSP and the IP. These scenarios completes, from the point of view of the signalling network, the description of the information exchange. In addition, the message processing times in the servers of the IN model must be provided. The IN planning process aims to accommodate a given service demand with a specific grade of service. However, the characteristics of the IN environment require the provision of the necessary infrastructure without any forecast of service mix or traffic levels. Therefore, the effect of several traffic profiles on the network must be carefully analysed. In this case study two traffic profiles are considered. The traffic profile 1 consists of 20% credit card calls and 80% Freephone calls. In a second case, the same traffic values are assumed for both services: 50% each. The originating IN traffic is the same for both SSPs.

The grade of service of interest for the Freephone service is the retrieval time of routeing information in the SCP (operations B1 and D1). In the credit card service, the time between the end of the dialling and the reception of the recorded message requiring the user identification is considered. This time interval for the SSP with code 100 is given by the summation of the times required for operation A1, operation A3, and the connection setup time between the SSP and the IP. The analysis is interrupted as soon as any IN element is overloaded. All results are presented as a function of the IN call rate.

The IN message flow analysis yields the mean load of the IN components, i.e., the impact of the IN services on the various elements. This information allows the dimensioning of the IN elements according to the given service mix and traffic figure. Another output of this phase is the IN information flow through the signalling network. The IN information flow through the signalling network is represented by TCAP messages.

The messages are mixed with the signalling traffic of the non-IN services. The signalling network related parameters are assumed to be the same as in [2], with the non-IN signalling traffic represented by the ISDN scenarios. It is assumed that the non-IN traffic between the SSPs are represented by ISDN calls. The ISDN call rate in each direction is 4.0 Calls/s. The analysis results for the retrieval time of routeing information in the SCP considering the assumed traffic profiles are depicted in Fig. 10. Fig. 11 shows the results for the time between the end of the dialling and the reception of the recorded message requiring the user identification of the credit card service.

7 Conclusion

This paper aimed at the outline of a universal approach to model signalling networks and their various applications as a basis for flow analysis, performance evaluation, signalling network capacity planning and overload control. The analysis method is based on decomposition. We are aware of the possible inaccuracies incurred by this method; the results, however, have been subject to intensive simulations where these assumptions were not necessary. Fortunately, the superposition of many single signalling activities of real network environments support a randomization effect which is responsible for the relatively accurate results obtained by the decomposition method. On the other hand, decomposition seems the only feasible approach to attack such complex models and allows easily extension according to new service types and network modifications. This aspect is of particular interest when analysis is part of a planning tool.

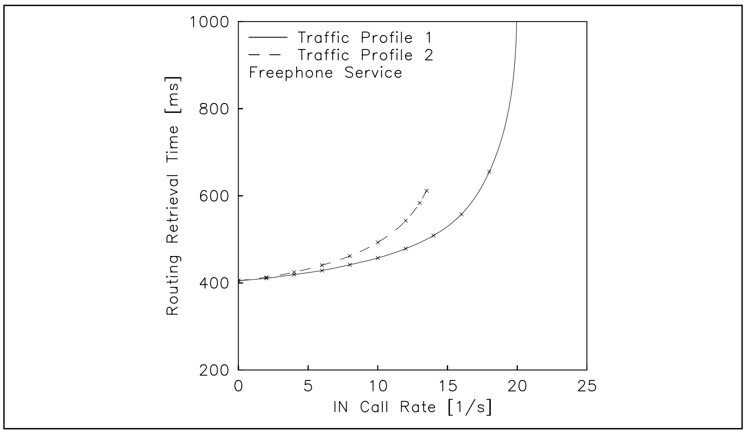


Figure 10: Routeing retrieval time in the Freephone service

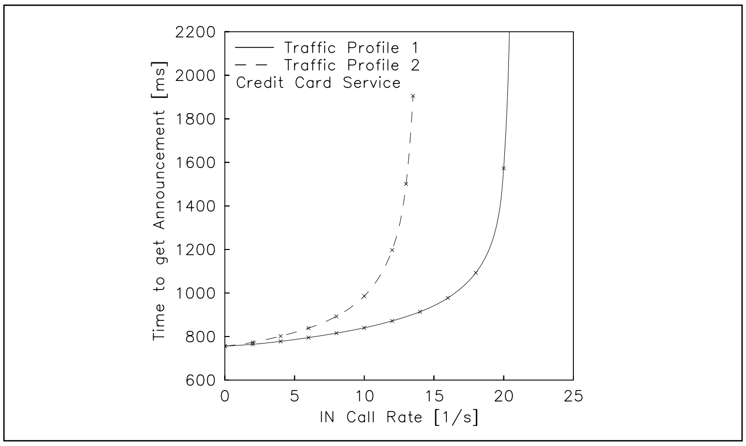


Figure 11: Time between end of dialling and reception of recorded message in the Credit Card Service

References

[1] M. Bafutto, P.J. Kühn, G. Willmann, J. Zepf, "A Capacity and Performance Planning Tool for Signalling Networks based on CCITT Signalling System No. 7", in Intelligent Networks -- The Path to Global Networking, P.W. Bayliss (Editor), IOS Press, Washington, pp. 368-379, 1992.

[2] M. Bafutto, P.J. Kühn, "Capacity and Performance Analysis for Signalling Networks Supporting UPT", Proceedings of the 8th ITC Specialist Seminar on Universal Personal Telecommunication, Santa Margherita Ligure (Genova), Italy, pp. 201-213, October 1992.

[3] M. Bafutto, P.J. Kühn, G. Willmann, "Capacity and Performance Analysis of Signalling Networks in Multivendor Environments", IEEE Journal on Selected Areas in Communications, vol. 12, no. 3, pp. 490-500, April 1994.

[4] M. Bafutto and M. Schopp, "Planning the Capacity and Performance of Intelligent Networks based on the ITU-TS Capability Set 1", Proc. Regional ITC-Seminar South Africa, Pretoria, 4-8 September 1995, pp. 282-296

[5] M. Bafutto and M. Schopp, “Network Performance and Capacity Figures of Intelligent Network based on the ITU-TS IN Capability Set 1”, International Workshop on Advanced Intelligent Networks (AIN'96), Passau, March 1996

[6] M. Bafutto and M. Schopp, “Performance Evaluation of IN-based Mobility Management”, IEEE Intelligent Network Workshop (IN'96), Melbourne, April 1996

[7] CCITT, Blue Book, Volume IV, Fascicles VI.7-VI.9, “Specifications of Signalling System No. 7”, Recommendations Q.700-Q.795, International Telecommunication Union, Geneva, 1989.

[8] CCITT, Blue Book, Volume IV, Fascicles VI.10-VI.11, “Digital Subscriber Signalling System No. 1 (DSS 1)”, Recommendations Q.920-Q.940, International Telecommunication Union, Geneva, 1989

[9] A. Cobham, “Priority Assignment in Waiting Line Problems”, Op. Res., vol. 2, no. 1, pp. 70-76, Feb. 1954.

[10] G. Hébuterne, L. Romoeuf, R. Kung, “Load Regulation Schemes for the Intelligent Network”, Proc. XII Int. Switching Symposium, Stockholm, Vol. V, pp. 159-164, May, 1990

[11] U. Herzog, “Priority Models for Communication Processors Including System Overhead”, Proc. of the 8^{th} ITC, Melbourne, Australia, vol. 2, sess. 62, paper 3, Nov. 1976

[12] ITU-T Recommendations Q.1200-Q.1219, International Telecommunication Union, Geneva, 1993

[13] U. Körner, C. Nyberg, B. Wallström, “The Impact of New Services and New Control Architectures on Overload Control”, Proc. ITC-14, Antibes Juan-les-Pins, France, pp. 275-283, June, 1994

[14] W. Krämer, “Processor Priority Model with Different User Tasks and Operation System Phases”, Computer Performance, K.M. Chandy and M. Reiser (Editors), Amsterdam, North Holland, pp. 305-325, 1977

[15] P.J. Kühn, C.D. Pack, R.A. Skoog, “Common Channel Signaling Networks: Past, Present, Future”, IEEE Journal on Selected Areas in Communications, vol. 12, no. 3, pp. 383-393, April 1994.

[16] H. Nyberg, B. Olin, “On Load Control of an SCP in the Intelligent Network”, Proc. Australian Telecommun. Networks & Applications Conf., Melbourne, pp. 751-756, December, 1994

[17] M. Paterok, O. Fischer, “Feedback Queues with Preemption-Distance Priorities”, ACM SIGMETRICS Performance Evaluation Review, vol. 17, no. 1, pp. 136-145, May 1989.

[18] M. Paterok, “Warteschlangensysteme mit Rückkopplung und Prioritäten”, Arbeitsberichte des Instituts für Mathematische Maschinen und Datenverarbeitung, Band 23, no. 12, Erlangen, Germany, October 1990.

[19] B. Simon, “Priority Queues with Feedback”, Journal of the ACM, vol. 31, no. 1, pp. 134-149, January 1984.

[20] G. Willmann, P.J. Kühn, “Performance Modelling of Signalling System No. 7”, IEEE Communications Magazine, vol. 28, no. 7, pp. 44-56, July 1990.

[21] J. Zepf, G. Rufa, “Congestion and Flow Control in Signaling System No. 7 - Impacts of Intelligent Networks and New Services”, IEEE Journal on Selected Areas in Communications, vol. 12, no. 3, pp. 501-509, April 1994.

[22] J. Zepf, Modellierung und Bewertung von Überlastabwehrmechanismen in Signalisiernetzen - Modelling and Performance Evaluation of Overload Control Mechanisms in Signalling Networks, 61. Bericht über verkehrstheoretische Arbeiten - 61th Report on Studies in Congestion Theory, Institut für Nachrichtenvermittlung und Datenverarbeitung, Prof. Dr.-Ing. habil. P.J. Kühn, Stuttgart, Germany, 1995