# An Overview of Signaling System No. 7

ABDI R. MODARRESSI, MEMBER, IEEE, AND RONALD A. SKOOG, MEMBER, IEEE

*Invited Paper*

*In modern telecommunication networks, signaling constitutes the distinct control infrastructure that enables provision of ALL other services. The component of signaling systems that controls provision of services between the user and the network is the access signaling component, and the component that controls provision of services within the network, or between networks, is the network signaling component. There are international standards for both access signaling and network signaling protocols. From a network structure viewpoint, access signaling structures generally provide point-to-point connectivity between the user and a network node, while network signaling structures provide network-wide communication capability (directly or indirectly) between the nodes of the public network(s). Since the network signaling system acts as a traffic collector/distributor for many access signaling tributaries, its functions are more complex, its structure more involved, and its performance more stringent. This paper provides an overview of modern network signaling systems based on the Signaling System No. 7 international standard.*

## I. INTRODUCTION

In the context of modern telecommunications, signaling can be defined as the *system* that enables stored program control exchanges, network databases, and other 'intelligent' nodes of the network to exchange a) messages related to call setup, supervision, and tear-down (call/connection control); b) information needed for distributed application processing (inter-process query/response, or user-to-user data); and c) network management information. As such, signaling constitutes the control infrastructure of the modern telecommunication network.

Modern signaling systems are essentially data communication systems using layered protocols. What distinguishes them from other data communication systems are basically two things: their real time performance and their reliability requirements. No matter how complex the set of network interactions are for setting up a call, the call setup time should still not exceed a couple of seconds. This imposes quite a stringent end-to-end delay requirement on the signaling system. On the other hand, because of the absolute reliance of the telecommunication network on its signaling system, requirements for signaling network reliability (message integrity, end-to-end availability, network robustness, recovery from failure, etc.) are extremely demanding. For example, current objectives require the down-time between any arbitrary pair of communicating nodes in the signaling network not to exceed 10 min/year. This is at least two orders of magnitude smaller than the corresponding requirement in a general-purpose data network. Requirements on real-time performance and reliability of signaling systems are likely to become even more stringent with advances in technology and new application needs.

Over the last century or so, signaling has evolved with the technology of telephony, although the pace of this evolution has never been faster than in the last two decades, a period characterized by the marriage of computer and switching technologies. The advent of the Integrated Services Digital Network (ISDN) has further accelerated the pace of development and deployment of signaling systems to support an ever increasing set of "intelligent network" services on a worldwide basis. When viewed as an end-to-end capability, signaling in ISDN has two distinct components: signaling between the user and the network (access signaling), and signaling within the network (network signaling). The current set of protocol standards for *access signaling* is known as the Digital Subscriber Signaling System No. 1 (DSS1). The current set of protocol standards for *network signaling* is known as the Signaling System No. 7 (SS7).

This paper provides an overview of Signaling System No. 7. It is a somewhat abridged and updated version of a tutorial on SS7 that was published in 1990 [1]. Following this introduction, the salient features of SS7's Network Services Part (NSP) are described in Section II. Functionally, NSP corresponds to the first three layers of the Open System Interconnection (OSI) Reference Model. This section also provides a discussion of signaling network structures that, in conjunction with the NSP, provide ISDN nodes with a highly reliable and efficient means of exchanging signaling messages. Once this reliable signaling message transport capability is realized, each network node has to be equipped with capabilities for processing of the transported messages in support of a useful function like setting up of a call (connection). In an increasingly large number of cases, call setup has to be preceded by invocation of some distributed

Manuscript received October 23, 1991; revised December 18, 1991.

A. R. Modarressi is with AT&T Bell Laboratories, Columbus, OH 43213.

R. A. Skoog is with AT&T Bell Laboratories, Holmdel, NJ 07733.

IEEE Log Number 9108075.

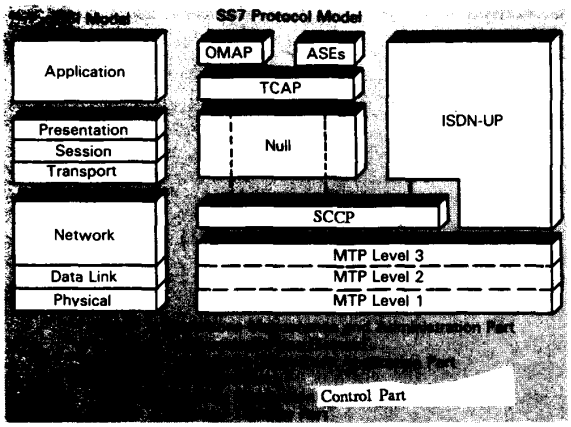


**Fig. 1.** SS7 protocol architecture.

application processes, the outcome of which determines the nature as well as the attributes of the subsequent call or connection control process. These nodal capabilities of call control and remote process invocation and management are part of the Signaling System No. 7 User Parts, which are described in Section III. In Section IV, we dwell on the very stringent performance requirements of signaling systems. These requirements reflect the critical nature of signaling functions and their real time exigencies. Finally, in Section V we sketch a broad outline of the likely evolution of network signaling in the remaining years of this century.

## II. Signaling System No. 7 Network Services Part (NSP)

In this section, we describe the Signaling System No. 7 protocols that correspond to the first three layers (Physical, Data Link, and Network) of the OSI Reference Model. This component of the Signaling System No.7 protocol is called the Network Services Part (NSP), and it consists of the Message Transfer Part (MTP) and the Signaling Connection Control Part (SCCP). Figure 1 shows how these relate to each other and to the other components of the protocol. MTP consists of levels 1–3 of the Signaling System No. 7 protocol, which are called the Signaling Data Link, the Signaling Link, and the Signaling Network functions, respectively. SCCP is an MTP user, and therefore is in level 4 of Signaling System No. 7 protocol stack. MTP provides a connectionless message transfer system that enables signaling information to be transferred across the network to its desired destination. Functions are included in MTP that allow system failures to occur in the network without adversely affecting the transfer of signaling information. SCCP provides additional functions to MTP for both connectionless and connection-oriented network services.

MTP was developed before SCCP and it was tailored to the real time needs of telephony applications. Thus a connectionless (datagram) capability was called for which avoids the administration and overhead of virtual circuit networks (one of the disadvantages of CCS6). Later, it became clear that there were other applications that would need additional network services (full OSI Network service capabilities) like an expanded addressing capability and connection-oriented message transfer. SCCP was developed to satisfy this need. The resulting structure, and specifically the splitting of the OSI Network functions into MTP level 3 and SCCP, has certain advantages in the sense that the higher overhead SCCP services can be used only when needed, allowing the more efficient MTP to serve the needs of those applications that can use a connectionless message transfer with limited addressing capability.

Sections II-A and II-B provide an overview of MTP and SCCP, respectively. Section II-C describes the signaling network structures that can be used to implement the Network Services Part.

### A. The Message Transfer Part (MTP)

The overall purpose of MTP is to provide a reliable transfer and delivery of signaling information across the signaling network, and to react and take necessary actions in response to system and network failures to ensure that reliable transfer is maintained. Figure 2 illustrates the functions of MTP levels, and their relationship to one another and to the MTP users. These three levels are now described.

*1) Signaling Data Link Functions (Level 1):* A *Signaling Data Link* is a bidirectional transmission path for signaling, consisting of two data channels operating together in opposite directions at the same data rate. It fully complies with the OSI's definition of the physical layer (layer 1). Transmission channels can be either digital or analog, terrestrial or satellite.

For digital signaling data links, the recommended bit rate for the ANSI standard is 56 kb/s, and for the CCITT International Standard it is 64 kb/s. Lower bit rates may be used, but the message delay requirements of the User Parts must be taken into consideration. The minimum bit rate allowed for telephone call control applications is 4.8 kb/s. In the future, bit rates higher than 64 kb/s may be required (e.g., 1.544 Mb/s in North America and 2.048 Mb/s elsewhere), but further study is needed before these rates can be standardized.

*2) Signaling Link Functions (Level 2)*: The Signaling Link functions correspond to the OSI's data link layer (layer 2). Together with a signaling data link, the signaling link functions provide a *signaling link* for the reliable transfer of signaling messages between two directly connected signaling points. Signaling messages are transferred over the signaling link in variable length messages called *signal units*. There are three types of signal units, differentiated by the length indicator field contained in each, and their formats are shown in Fig. 3. The Signaling Information Field (SIF) in a Message Signal Unit (MSU) must have a length less than or equal to 272 octets. This limitation is imposed to control the delay a message can impose on other messages due to its emission time (which is limited by the maximum standardized link speed of 64 kb/s).

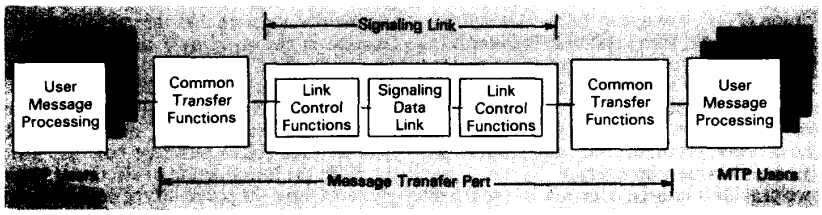


**Fig. 2.** MTP functional diagram.

The SS7 link functions show a strong similarity to typical data network bit-oriented link protocols (e.g., HDLC, SDLC, LAP-B), but there are some important differences. These differences arise from the performance needs of signaling (e.g., lost messages, excessive delays, out-of-sequence messages) that require the network to respond quickly to system or component failure events. The standard flag (01111110) is used to open and close signal units, and the standard CCITT 16-bit CRC checksum is used for error detection. However, when there is no message traffic, Fill-In Signal Units (FISU's) are sent rather than flags, as is done in other data link protocols. The reason for this is to allow for a consistent error monitoring method (described below) so that faulty links can be quickly detected and removed from service even when traffic is low.

*a) Error correction:* Two forms of error correction are specified in the signaling link procedures. They are the *Basic Method* and the *Preventive Cyclic Retransmission (PCR) Method.* In both methods only errored MSU's and Link Status Signal Units (LSSU's) are corrected, while errors in FISU's are detected but not corrected. Both methods are also designed to avoid out-of-sequence and duplicated messages when error correction takes place. The PCR method is used when the propagation delay is large (e.g., with satellite transmission).

The Basic Method of error correction is a non-compelled positive/negative acknowledgment retransmission error correction system. It uses the "go-back-N" technique of retransmission used in many other protocols. If a negative acknowledgment is received, the transmitting terminal stops sending new MSU's, rolls back to the MSU received in error, and retransmits everything from that point before resuming transmission of new MSU's. Positive acknowledgments are used to indicate correct reception of MSU's, and as an indication that the positively acknowledged buffered MSU's can be discarded at the transmitting end. For sequence control, each signal unit is assigned forward and backward sequence numbers and forward and backward indicator bits (see Fig. 3). The sequence numbers are seven bits long, which means at most 127 messages can be transmitted without receiving a positive acknowledgment.

The PCR method is a non-compelled positive acknowledgment cyclic retransmission, forward error correction system. A copy of a transmitted MSU is retained at the transmitting terminal until a positive acknowledgment for that MSU is received. When there are no new MSU's to be sent, all MSU's not positively acknowledged are retransmitted cyclically. When the number of unacknowledged MSU's (either the number of messages or the number of octets) exceeds certain thresholds, it is an indication that error correction is not getting done by cyclic retransmission. This would occur, for example, if the traffic level was high, which causes the retransmission rate to be low. In this situation a *forced retransmission* procedure is invoked. In this procedure new MSU transmission is stopped and all unacknowledged MSU's are retransmitted. This forced retransmission continues until the unacknowledged message and octet counts are below specified threshold values. These threshold values must be chosen carefully, for if they are set too low, and the link utilization is large enough, the link will become unstable (i.e., once a forced retransmission starts, the link continues to cycle in and out of forced retransmission [2]).

*b) Error monitoring:* Two types of signaling link error rate monitoring are provided. A *signal unit error rate monitor* is used while a signaling link is in service, and it provides the criteria for taking a signaling link out of service due to an excessively high error rate. An *alignment error rate monitor* is used while a signaling link is in the proving state of the initial alignment procedure, and it provides the criteria for rejecting a signaling link for service during the initial alignment due to too high an error rate.

The signal unit error rate monitor is based on a signal unit (including FISU) error count, incremented and decremented using the "leaky bucket" algorithm. For each errored signal unit the count is increased by one, and for each 256 signal units received (errored or not), a positive count is decremented by one (a zero count is left at zero). When the count reaches 64, an excessive error rate indication is sent to level 3, and the signaling link is put in the out of service state. When loss of alignment occurs (a loss of alignment occurs when more than six consecutive 1s are received or the maximum length of a signal unit is exceeded), the error rate monitor changes to an octet counting mode. In this mode it increments the counter for every 16 octets received. Octet counting is stopped when the first correctly-checking signal unit is detected.

The alignment error rate monitor is a linear counter that is operated during alignment proving periods. The counter is started at zero at the start of a proving period, and the count is incremented by one for each signal unit received in error (or for each 16 octets received if in the octet counting mode). A proving period is aborted if the threshold for the

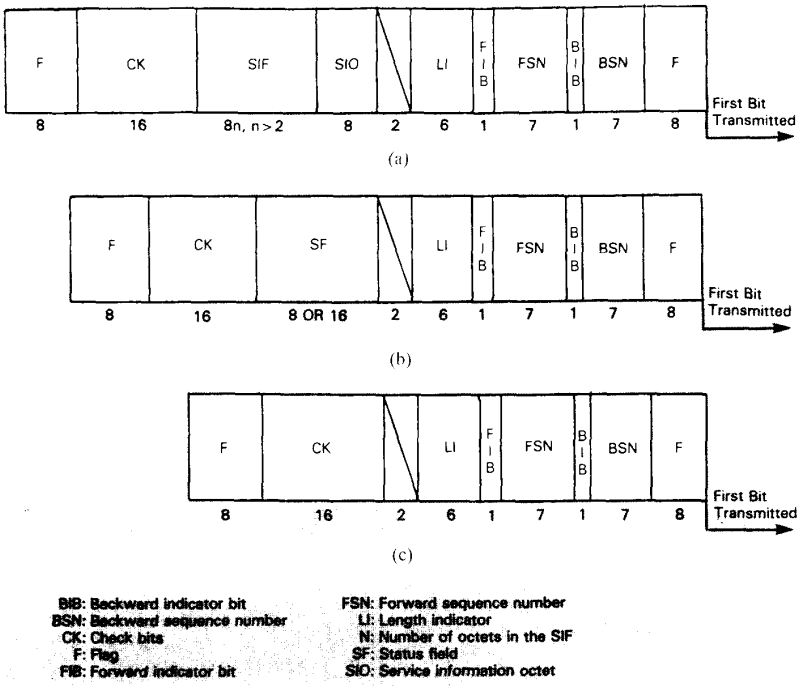


**Fig. 3.** Signal unit formats.

alignment error rate monitor count is exceeded before the proving period timer expires.

*c) Flow control:* The flow control procedure is initiated when congestion is detected at the receiving end of the signaling link. The congested receiving end notifies the transmitting end of its congestion with a link status signal unit (LSSU) indicating busy, and withholds acknowledgment of all incoming signal units. This action stops the transmitting end from failing the link due to a time-out on acknowledgment. However, if the congestion condition lasts too long (3–6 s), the transmitting end will fail the link.

A processor outage condition indication is sent by level 2, called signaling indication processor outage (SIPO), whenever an explicit indication is sent to level 2 from level 3 or when level 2 recognizes a failure of level 3. This indicates to the far end that signaling messages cannot be transferred to level 3 or above. The far-end level 2 responds by sending fill-in signal units and informing its level 3 of the SIPO condition. The far-end level 3 will reroute traffic in accordance with the signaling network management procedures described as follows.

*3) Signaling Network Functions (Level 3):* The signaling network functions correspond to the lower half of the OSI's Network layer, and they provide the functions and procedures for the transfer of messages between signaling points, which are the nodes of the signaling network. The signaling network functions can be divided into two basic categories: *signaling message handling* and *signaling network management*. The breakdown of these functions and their interrelationship is illustrated in Fig. 4.

*a) Signaling message handling:* Signaling message handling consists of message routing, discrimination, and distribution functions. These functions are performed at each signaling point in a signaling network, and they are based on the part of the message called the *routing label*, and the Service Information Octet (SIO) shown in Fig. 3. The routing label is illustrated in Fig. 5 and consists of the Destination Point Code (DPC), the Origination Point Code (OPC), and the Signaling Link Selection (SLS) field. In the international standard the DPC and OPC are 14 bits each, while the SLS field is 4 bits long. For ANSI, the OPC and DPC are each 24 bits (to accommodate larger networks), while the SLS field has 5 bits, and there are 3 spare bits in the routing label. The routing label is placed at the beginning of the Signaling Information Field, and it is the common part of the label that is defined for each MTP user.

When a message comes from a level 3 user, or originates at level 3, the choice of the particular signaling link on which it is to be sent is made by the message routing function. When a message is received from level 2, the discrimination function is activated, and it determines if it is addressed to another signaling point or to itself based on the DPC in the message. If the received message is addressed to another signaling point, and the receiving signaling point has the transfer capability, i.e., the Signal Transfer Point (STP) function, the message is sent to the message routing function. If the received message is addressed to the

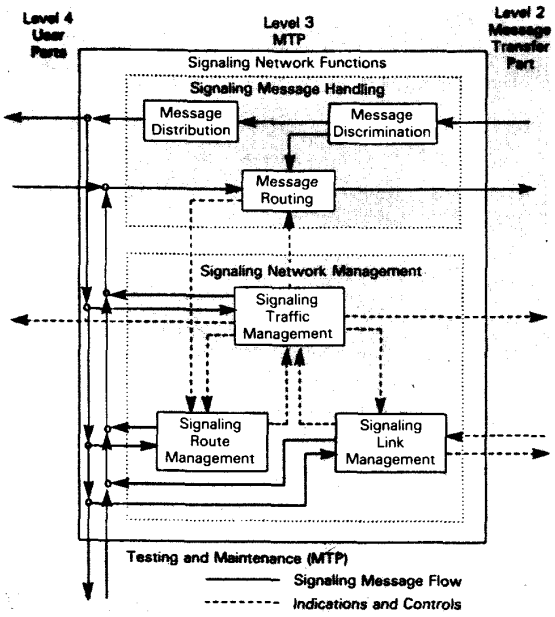


**Fig. 4.** Signaling network functions.

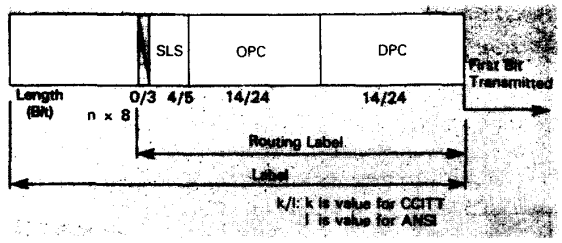


**Fig. 5.** Routing label structure.

receiving signaling point, the message distribution function is activated, and it delivers the message to the appropriate MTP user or MTP level 3 function based on the service indicator, a sub-field of the SIO field. Message routing is based on the DPC and the SLS in almost all cases. In some circumstances the SIO, or parts of it (the service indicator and network indicator), may need to be used.

Generally, more than one signaling link can be used to route a message to a particular DPC. The selection of the particular link to use is made using the SLS field. This is called load sharing. A set of links between two signaling points is called a *link set*, and load sharing can be done over links in the same link set or over links not belonging to the same link set. A load sharing collection of one or more link sets is called a *combined link set*.

The objective of load sharing is to keep the load as evenly balanced as possible on the signaling links within a combined link set. For messages that should be kept in sequence, the same SLS code is used so that such messages take the same path. For example, for trunk signaling with ISUP (see Section IV-A) the same SLS code is used for all messages related to a particular trunk. In order to ensure proper load balance using SLS fields, it is critical that the SLS codes are assigned such that the load is shared evenly across all the SLS codes. Even then, the SLS load sharing method does not provide a fully balanced loading of signaling links in all cases. For example, if there are six signaling links in a combined link set, the 16 SLS codes would be assigned so that four signaling links would each carry three SLS codes and two of the signaling links would each carry only two SLS codes.

*b) Signaling network management:* The purpose of the signaling network management functions is to provide reconfiguration of the signaling network in the case of signaling link or signaling point failures, and to control traffic in the case of congestion or blockage. The objective is that, when a failure occurs, the reconfigurations be carried out so messages are not lost, duplicated, or put out of sequence, and that message delays do not become excessive. As shown in Fig. 4, signaling network management consists of three functions: signaling traffic management, signaling route management, and signaling link management. Whenever a change in the status of a signaling link, signaling route or signaling point occurs, these three functions are activated as summarized below.

The *signaling traffic management* procedures are used to divert signaling traffic, without causing message loss, missequencing, or duplication, from unavailable signaling links or routes to one or more alternative signaling links or routes, and to reduce traffic in the case of congestion. When a signaling link becomes unavailable, a *changeover* procedure is used to divert signaling traffic to one or more alternative signaling links, as well as to retrieve for retransmission messages that have not been positively acknowledged. When a signaling link becomes available, a *changeback* procedure is used to reestablish signaling traffic on the signaling link made available. When signaling routes (succession of links from the origination to the destination signaling point) become unavailable or available, *forced rerouting* and *controlled rerouting* procedures are used, respectively, to divert the traffic to alternative routes or to the route made available. Controlled rerouting is also used to divert traffic to an alternate (more efficient) route when the original route becomes restricted (i.e., less efficient because of additional transfer points in the path). When a signaling point becomes available after having been down for some time, the *signaling point restart* procedure is used to update the network routing status and control when signaling traffic is diverted to (or through) the point made available.

The *signaling route management* procedures are used to distribute information about the signaling network status in order to block or unblock signaling routes. The following procedures are defined to take care of different situations. The *transfer-controlled* procedure is performed at a signaling transfer point in the case of signaling link congestion. In this procedure, for every message received having a congestion priority less than the congestion level of the signaling link, a control message is sent to the

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

**LAW FIRMS**
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

**FINANCIAL INSTITUTIONS**
Litigation and bankruptcy checks for companies and debtors.

**E-DISCOVERY AND LEGAL VENDORS**
Sync your system to PACER to automate legal marketing.