S38.3115 Signaling Protocols – Lecture Notes

Lecture 5 – Common Channel Signaling System Nr 7 (CCS7 or SS7)

Introduction

The specification of SS7 began during the 1970´s when it started to become clear that the future switching systems would be program controlled, that their switching fabrics would also be implemented in silicon and closer to the end of the decade that even subscriber access in those switching systems would be digital.

AT&T, RBOCs[1], BT, other large European and Japanese operators played key roles in specifying SS7. They sought to remove the limitations of analogue signaling systems that were in use in electromechanical switching systems in their networks. The limitation included: a limited set of signals, difficulty of signaling after call setup, difficulty of signaling independent of the voice circuit, high level of use of network capacity because of slow signaling and dedicated signaling resources

Specification of SS7 took place in parallel and even earlier than the work on OSI in the International Standardization Organization. The same principle of hierarchical layers as in OSI was used in both. However, the purpose and the design goals of the two protocol sets are different and as a result there are

---

[1] Regional Bell Operating Companies.

significant differences in the allocation of functions on different *layers in OSI* and *levels in SS7*.[2]

There are two variants of SS7. One for ETSI (European Telecommunications Standards Institute) and the other for ANSI markets. The latter is mainly in use in the Americas, South Korea and Japan. This presentation is mainly based on ETSI specifications and makes only marginal references to ANSI.

Design considerations: Performance, Capacity and Reliability

The solution to the limitations of analogue signaling was a fully digital, message based system that could carry all kinds of information needed for network services. Moreover, to make signaling independent of voice circuits it was obvious that an out-of-band, and more specifically a *common channel signaling* system, was needed. Common channel refers to the idea that one timeslot allocated for signaling will carry all signaling needed for many voice slots and even for many PCM –lines. Let us verify this idea concerning the *scalability* of the system.

Let us assume that 1000 bits are needed on average for signaling for a single call[3].

From our previous analysis for a similar 1000 bit flow of DSS1 and taking into account that the signaling channels in the trunk network are 4 times faster than in BRI, we expect that the *signaling transfer delay in SS7* is not a problem: transferring the whole signaling flow for a call will take about (1000/64=) 15…16 ms.

Let us further assume that an average call takes 3 minutes. It is easy to calculate that a PCM line can carry no more than 30 call minutes in a minute and each call minute on average will create 333 signaling bits. For a PCM full of calls this would give a total of 10 kbit/minute = 167 bit/s.

It follows from this simple calculus that one 64 kbit/s timeslot can carry call signaling for many PCM lines. If we take as the dimensioning criteria that signaling channels should be filled not more than 20% of the time, the number of PCM-lines we can serve with a single signaling timeslot would be:

Nrof PCM lines = 0,2 * 64 000/167 = 76 ≈ 2300 timeslots.

On the other hand, let us recall the discussion on the structure and capacity of an exchange. We gave an example of an exchange with 8000 PCM lines and calculated that about 7000 of them might be used for connections to other

---

[2] Layer and level are kerros and taso in Finnish respectively.
[3] For DSS1 we showed a signaling flow trace for a call that was about 1000 bits in length. SS7 signaling is binary like DSS1 and uses a similar coding style as DSS1. So, the assumption is rather realistic.

exchanges. What if there are e.g. 1000 PCM lines connecting two exchanges. These PCMs would carry the maximum of:

$1000 \times 30 = 30\ 000$ timeslots or simultaneous telephone calls.

With our earlier assumptions, one signaling channel can support approximately 76 PCM lines and ca. 14 signaling channels would be required to support the call traffic on 1000 PCM lines with 30 000 timeslots.

The above calculus is not meant to be exact but rather show the orders of magnitude of capacity requirements for signaling and the carriage of voice.

Another major consideration in SS7 design was *reliability of signaling*. During the 1970's operators and users were used to noisy transmission. On the other hand from the business perspective a lost or a misrouted call may mean lost revenue for the operator or an unsatisfied subscriber.

Figure 5.1 shows the telephone network service reliability as a serial reliability diagram depicting that all the boxes must work for the service to work. For such a system the weakest link determines the reliability of the whole system. By eliminating the impact of signaling with a limited cost it became possible to leverage the expensive long-term investment into the voice path transmission plant and the operator's other infrastructure.
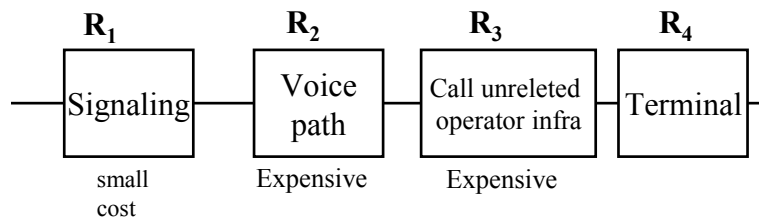

Figure 5.1 Network service reliability diagram

It follows from the reliability requirements that it would be unacceptable to loose signaling connectivity between two exchanges that between them carry for example 1000 simultaneous calls because of a failure of a single signaling channel (even 100 simultaneous calls might trigger to say the same). Also sometimes a connection between two exchanges uses more than 80 PCM lines. From these two considerations, we have a requirement that signaling channels need to be able to replicate each other and also share load. Replication means that it should be possible to detect signaling channel failures quickly and to automatically move the load off the failed signaling channel to another channel that has been preconfigured to handle the load. Sharing the load means that at least it must be possible to split the PCMs between two exchanges onto different signaling channels. This split can be a configuration matter because adding and reducing the number of PCMs between two exchanges is not a

frequent event and always requires management actions from the operator. However, it would be best if any of the signaling channels between two neighboring exchanges could carry signaling for any of the voice channels.

The designers of SS7 had in mind a rich set of services and that the services would be produced using not only exchanges but other network nodes as well. All the nodes would be connected using SS7 as the basic signaling system.

SS7 concepts

SS7 defines a signaling network that is made of *signaling points* (network nodes) and *signaling links*. In practice most times a signaling link is a PCM-timeslot[4]. Among the signaling points from a call perspective are the *originating point* (OP) and the *destination point* (DP). Obviously, addresses of the OP and DP need to appear in signaling messages. These addresses are called *signaling point codes*. OPC stands for Originating Point Code and DPC for Destination Point Code.

SS7 also specifies that intermediate signaling points that do not process the call itself may appear on the way from the OP to the DP. These intermediate points are called Signaling Transfer Points or STPs and work either on MTP (message transfer part) or SCCP levels. SCCP stands for Signaling Connection Control Part.

STPs can be seen as signaling message routers. They are useful e.g. in a large country for the purpose of concentrating or aggregating signaling from a large number of local or transit exchanges to a small set of Intelligent Network (IN) Nodes. We will come back to IN later on this course. Such transfer points are also useful for the purpose of implementing mobility management and global roaming.

In ETSI networks signaling point codes are 14 bits long. In ANSI networks the point codes are 24 bits long. In an ETSI signaling network there may be some 16 000 unique point codes. Since the codes are allocated to Exchanges and other network nodes but not to users, this is quite enough. However, this means that an SS7 signaling network is not global in itself. Instead, the "global signaling network" is formed from independent islands that each of which is an SS7 signaling network. The islands are bridged by higher-level functions that process telephony signaling. The "bridging" takes place on administrative boundary nodes.

A *signaling link* is a pre-configured connection between two (adjacent) signaling points. A *signaling link set* is a set of signaling links such that all

---

[4] In some early networks modem connections at slow speeds were used as signaling links while most of voice transmission took place over analogue transmission. Lately, replacement of PCM transmission with broadband systems is taking place. In such a case a signaling link in SS7 may map to an IP or Ethernet "connection".

links have the same signaling end points. A sequence of signaling link sets between two end points forms a *signaling route*. A signaling route differs from a link set because of the possible intermediate STPs. The set of all signaling routes connecting two signaling points is a *signaling route set*. These concepts are depicted in Figure 5.2.
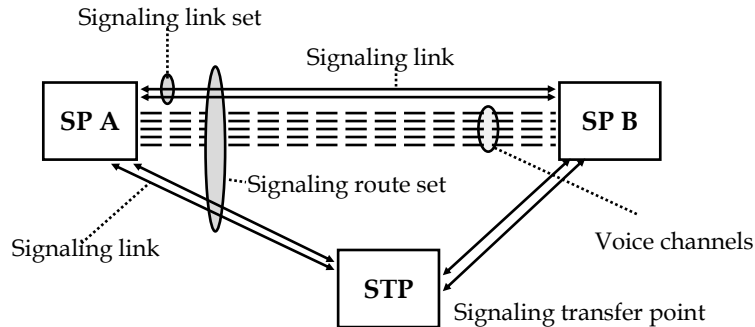

Figure 5.2: SS7 signaling network concepts

Signaling link sets, signaling routes and route sets are pre-configured. An SS7 node will provide configuration commands for the operator for the creation of the needed configuration information. Exchanges do not provide automatic means to re-configure any of the above, always operator involvement is required.

Recalling the capacity requirements, we see that the concepts of signaling link set and signaling route set help to meet the requirement of high signaling capacity between two large exchanges through load sharing. Because the dimensioning rule is that each signaling link is supposed to carry less than 0.2 Erl during a busy hour, we can always take one signaling channel out of use in a signaling link set and the rest will still carry all the signaling traffic due to redundancy. (NB: a well-implemented SS7 will process signaling messages on wire speed, i.e. even if the full capacity of 64kbit/s is used.)

In Channel Associated Signaling timeslot 16 was dedicated to signaling. The same agreement could be used also for SS7. However, all timeslots in a PCM system except timeslot zero are identical. In a digital telephone network it is up to the operators to configure the use of those timeslots as they please. In Finland, the agreement was that timeslot 1 was reserved for signaling purposes in SS7 networks.

SS7 Protocol Architecture

Figure 5.3 depicts the levels and the user and application parts in the SS7 architecture.

On the left the Figure shows comparison to OSI layers. MTP stands for Message Transfer Part, SCCP for Signaling Connection Control Part, TC for

Transaction Capabilities, INAP for IN Application Part, CAP for CAMEL Application Part that is used in GSM for a similar purpose as INAP in wire line networks. MAP stands for Mobile Application Part, BSSAP for Base Station Subsystem Application Part, ISUP for ISDN User Part, TUP for Telephony User Part and MUP and HUP are legacy systems that we do not need to worry about. There is also OMAP that is Operations and Maintenance Application Part.
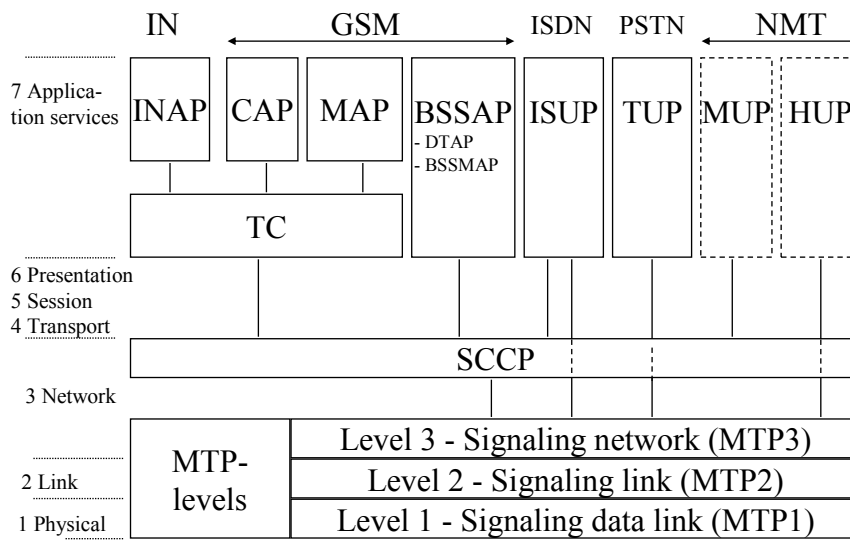

Figure 5.3: SS7 protocol architecture.

MTP covers two and a half OSI layers. Compared to OSI what is missing is global reachability for messaging. This capability is added with the help of SCCP. MTP + SCCP then together correspond to three lower layers in OSI.

The primitive interfaces between levels are left for vendors to design. Therefore, instead of layers we talk about levels. One reason that justifies looser layering in SS7 than in OSI is that performance requirements were seen very important for SS7. Strict layering was sacrificed for the sake of high performance and high reliability. Let's recall that this design decision was made sometime late 1970's or early 1980's and with the level of computing power available at that time was probably reasonable.

The task of MTP is to carry signaling messages between exchanges within an administrative domain using point code addressing. These MTP messages will carry in their payload Application Part or User Part messages. MTP messages can be routed by SS7 nodes within a single addressing domain over many hops, i.e. MTP is enough to implement the concept of "signaling route" in Figure 5.2.

MTP service is connectionless, connections (i.e. signaling links, link sets etc) between signaling points are pre-configured. In other words, MTP message routing is based on static routes. This is natural, because business relationships are rather stable between operators and telephone traffic is rather predictable. MTP uses DPCs and OPCs for addressing.

SCCP uses a richer set of information items for addressing than MTP. Global addressability among telephone exchanges and other network nodes is achieved by using routable telephone numbers in SCCP message addressing. SCCP supports both connectionless and connection oriented messaging services.

User Parts contain the telephone call signaling functions. Application Parts such as MAP and INAP add non-call related signaling capabilities. Application Parts use the Transaction Capabilities (TC) services and TC uses SCCP. TUP never uses SCCP but rather sits directly on top of MTP. ISUP may, in principle use SCCP but in practice it does not need it for regular telephony services and thus also in practice relies only on MTP for messaging.

Hop by hop and end to end in telephone networks

Usually telephone calls are routed and signaling is processed hop-by hop. This means that each exchange on the path from the originating local exchange till the terminating local exchange will make a routeing decision and for that it also needs to process the call signaling not only on the MTP level but also on the ISUP and call control levels. The routeing decision implies some physical resource reservations (such as timeslots) at each of the nodes.

Call routeing is based on dialed digits that in case of SS7 are carried in a *Called Party Address* information element in ISUP. A similar information element possibly named differently exists in other signaling systems as well. Moreover, it may be that dialed digits are split between several signaling messages. For the route decision, an exchange looks up the digits one by one in an analysis tree progressing deeper into the tree on each digit. Analysis trees are configuration information that is based on the allocation of telephone numbers to exchanges and the network topology. The result of analysis is a *destination* that is the set of possible routes. Next, the exchange looks at the reservation state of the current route among the alternatives by scanning the current *circuit group* in the route. Scanning progresses through the possible alternatives in the exchange database using some local algorithm until a free circuit (timeslot) is found.

Exchange based routeing assumes that dialed digits form routable telephone numbers. These are numbers that are allocated to local exchanges block by block (block size can be for example 1000 telephone numbers). This being the

case, each exchange on the call path except the terminating exchange need only look at the leading digits and the tail of the telephone number is needed only in the terminating exchange to select among its extensions. The reason for this limitation is scalability to a wide area network with many exchanges. If we were to allow free numbering among subscribers in for example a 50M subscriber network with say 500 exchanges, each exchange would have to be aware of the location of each subscriber: this is not feasible because there would be too many errors in different nodes related to some of the subscribers and because it would take too long to make changes in the numbering.

If Number Portability for example in a city of 1 000 000 subscribers is supported, telephone exchange routeing is not sufficient. Porting one single number would require configuration changes into many exchanges creating a data management headache and would make the configuration error-prone. Therefore, NP is most times implemented using IN. For example, it may be agreed that each originating exchange will trigger an IN query to a server with a database that will map a *telephone directory number* that was dialed by a user to a *routeing number* that is aligned with the network topology and can be handled by the exchanges on the call path.

From this discussion we gather that telephone routeing is *static*. Compare it to Internet routing that uses routing protocols to create forwarding tables – Internet routing is dynamic. By dynamic IP routing we refer to the fact that routing learns about link and node failures, removals and additions on the fly.[5]

There are no routing protocols for telephone networks. Telephone number portability would correspond to IP address portability in the Internet. *It is not supported* because the wide area routing in the Internet would not scale to support *host routes* that would emerge due to ported IP addresses.

Some information elements in signaling may be uninteresting to the intermediate nodes. Such information may be carried *end to end* through the intermediate nodes from the originating LE till the terminating LE. For example in R2 signaling the tail of the dialed digits are carried end to end.

SS7 supports an *associated mode of signaling* and a *non-associated mode of signaling*. Associated mode means that along the circuit group that connects two neighboring exchanges there is also a signaling link set connecting the same two exchanges. Each exchange will process call signaling using ISUP hop by hop. In non-associated mode, signaling messages are sent through a chain of nodes to a remote destination. Intermediate nodes may be either MTP level STPs or also engage SCCP function for message routing. The non-

---

[5] Some Telephone networks support dynamic routing. In the context of telephone networks this means that routing tries to adapt to network load rather than link or node failures (which is the case in IP networks). Also in this case there are not dynamic routing protocols.

associated mode is not particularly useful for call related signaling because the endpoints need to have a direct connecting circuit group for voice anyway.

The non-associated mode is, however, very useful for example for location updates in mobile networks. Think of an example where an Elisa GSM/3G phone from Finland is visiting Madrid in Spain. The phone will regularly make location updates and when it moves from an area handled by a particular Mobile Switching Center/Server to an area handled by another MSC/MSS[6], the update needs to be routed to Helsinki Finland through probably several SS7 networks. Endpoints of the signaling are one of the MSCs/MSS in Madrid and the Home Location Register in Helsinki. On MAP level this location update signaling is end to end from the visited MSC/MSS to the HLR. Intermediate message routing can be handled on MTP and SCCP levels. On SS7 network borders SCCP processing will be needed.

MTP – Message Transfer Part

MTP Reliability requirements

MTP assumes a low level of reliability from the underlying transmission infrastructure. The assumptions are:

- Long term bit error rate $\leq 10^{-6}$
- Medium term bit error rate $\leq 10^{-4}$

Let us assume that an erroneous bit will make a signaling message useless but that using a checksum it is rather easy to detect the error. On the medium term bit error rate, if a message has 100 bits, 1% of messages would be erroneous and would need to be resent.

The target requirements for MTP are:

- unavailability of signaling route set ≤ 10 min/annum
- share of undetected faulty signaling messages: $\leq 10^{-10}$
- loss probability of signaling messages $\leq 10^{-7}$
- probability of reordering or replication of signaling messages $\leq 10^{-10}$.

Assuming that a message with a single bit error is lost, these requirements mean that MTP must be implemented such that reliability of message transfer is at least 3 to 5 orders of magnitude better than what is assumed of the underlying infrastructure. This is reached using different time and space redundancy mechanisms. Time redundancy means that time is sacrificed to

---

[6] Late 1990's MSC was split to MSS and a separate data plane node (media gateway, switching fabric etc.) for the purpose of being able to scale the control and the data planes separately depending on the characteristics of the traffic such as signaling flow length and call length.

resending when needed. Space redundancy comes in many different forms such as: parity bits, checksums, replicated channels, replicated computers etc.

MTP levels

Figure 5.4 illustrates MTP functions.

MTP level 1 implements a transmission channel usually in a timeslot and switching functions. The latter are implemented so that normally, the signaling channel from a neighboring exchange is connected to the exchange switching fabric along with all other PCM timeslots. The switching fabric is used to map the trunk channel to an internal PCM leading to a signaling (and possibly call control) computer for SS7.

Level 1 is concerned with transmitting a bit stream from one control computer in a network node to a control computer in another neighboring network node.
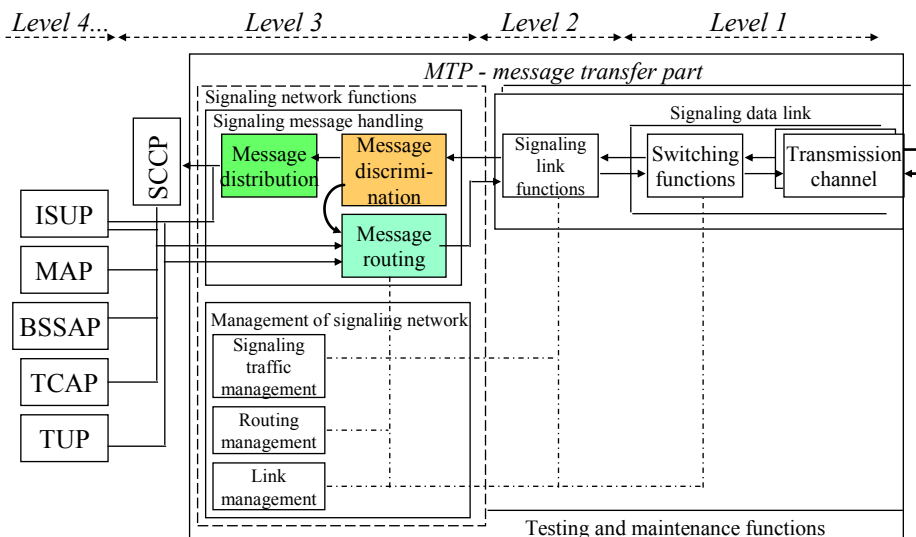
Figure 5.4 MTP in action

MTP level 2 is also called the LAPB protocol. It implements signaling link functions. These include for example processing frame alignment flags allowing the cutting of the level 1 signaling channel continuous bit stream into frames. By processing checksums, locally significant message numbers and their acknowledgements level 2 provides message error and link failure detection. By storing all sent messages until reception of an acknowledgement in addition of fast failure detection, level 2 provides signaling channel recovery and retransmission of message units. Message ordering is based on locally significant message numbers and the acknowledgement numbers. Message unit acknowledgements follow the windowing principles similar to what is found in DSS1, LAPD.

LAPB uses the same 01111110 flags for frame alignment as DSS1 and HDLC. Transparent transfer capability for all bit combinations in a byte is achieved by bit stuffing like in LAPD and HDLC. For error detection the receiver needs to calculate the checksum over the frame and compare it to the checksum found in the frame.

LAPB may be implemented on a preprocessor to the main signaling processor.

MTP level 3 is responsible for load sharing among signaling links in a route set, MTP *message routing* and *distribution* to user and application parts.

MTP management functions play an important role for reliability enhancement. The procedures support *link failovers* and *switchovers* on management command, link restoration and re-routing.

On an **incoming message** MTP 3 first executes *message discrimination*. This means that the node decides whether the message is addressed to this node or needs to be sent to *message routing*. If it is sent to message routing, for this message the node will act as an STP. Message routing will decide which outgoing signaling channel is suitable for the message and send the message onto level 2 for the execution of link functions. If message discrimination decides that the message is addressed to this node, it sends it to *message delivery* functions. Message delivery will decide which user part or application part needs to be used and will send the message to it.

A user part or an application part, when it wishes to **send an SS7 message** to an SS7 destination, actually hands the message to SS7 MTP message routing that will decide which signaling link to use etc. Message routing in MTP is based on signaling point codes and uses *static routing tables* created by management commands by the operator. This fits with the idea that telephone traffic is quite predictable and the business relationships between operators are rather stable. The routing tables need to be modified only when transmission capacity changes are made to the network or new transport connections are either added or old ones removed.

MTP message structures

MTP supports three message types. A message that carries call signaling in its payload is called a Message Signaling Unit or MSU. LSSUs or Link State Signal Units are exchanged between two neighbors in the SS7 network on MTP level for MTP internal purposes. Fill-in Signal Units or FISUs are sent always when there is nothing else to send on a signaling channel.

Figure 5.5 shows the signaling unit structure.

8	16	8n, n >2		8	2	6	1	7	1	7	8
F	CK	SIF		SIO		LI	FIB	FSN	BIB	BSN	F

Routing Label

transmission

F - Flag (delimiter -01111110)
BSN - Backward sequence number
BIB - Backward indicator bit
FSN - Forward sequence number
FIB - Forward indicator bit

LI - Length indicator
SIO - Service information octet
SIF - Service Information field
= payload
CK - Check bits

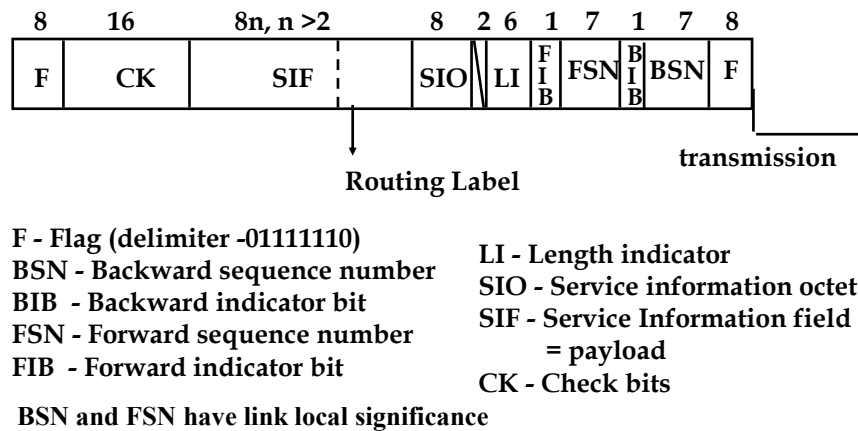**BSN and FSN have link local significance**

Figure 5.5: MTP message signaling unit structure (ETSI).

The difference between MSUs, LSSUs and FISUs is based on the LI – length indicator field. Message unit sequence numbering uses the FSN and BSN fields. Service Information Octet or SIO is split in the Service Indicator (SI) giving a hint of which user or application part is required for message processing. The other half of SIO is the sub-service field (SSF) allowing encoding a label for the particular SS7 network that is used. SS7 support 4 labels: 2 for different national SS7 networks and two for different international networks.

Initially, normal HDLC circuits did not implement FISU processing on silicon. Therefore, when LAPB had nothing else to send, the preprocessor implementing LAPB was the busiest because each FISU, the shortest possible frame, created an interrupt. Later this problem was solved by implementing FISU processing on silicon. Because the signaling channel is always transferring something, error detection is fast. The receiver needs to maintain a few timeouts and keep calculating checksums over the received frames.

The routing label has fields for the Destination Point Code (DPC) and the Originating Point Code (OPC). The latter is needed for acknowledgements. The routing label also has a 4 bit field for SLS – signaling link selection. Based on this field, the signaling link between the two nodes is selected. By keeping the same value of SLS for each message in a flow, all messages of the flow will use the same signaling link on the end-to-end connection. This helps to preserve message ordering end-to-end.

Because the SLS field has only 4 bits, a signaling link set can have the maximum of 16 signaling channels. We earlier estimated that a signaling link set with 14 signaling channels could handle the call signaling load from 1000 PCM lines.

MTP Reliability solutions

The reliability enhancement is achieved by several methods. Among them are checksums in all SS7 messages for detecting bit errors and locally significant message numbers in both directions in all messages. Message numbers allow detecting loss, duplication and re-ordering of messages on connections between network nodes. By locally significant we mean that only immediate neighbors in the signaling network understand the numbers. Each hop uses its own numbers. MTP requires that signaling messages on a signaling channel can be sent in both directions. Each timeslot transfers bits in one direction but for each timeslot there is a timeslot in the opposite direction. (For example, this means that MTP messages never collide on the channel). On both directions MTP transfers a continuous flow of numbered LAPB frames. In each flow, frame numbers indicate the order of the frames (FSN) and in addition carry another number (BSN) to indicate up-to which number frames have been received from the other endpoint of the link.

In addition, several higher-level methods are used for ensuring reliable delivery of messages. One of them is *replication* of signaling channels, automatic fail-over procedures for signaling flows if error level on a signaling channel grows too high and by *buffering all sent messages* until they are acknowledged by the next node on the path. Messages that are unavoidably lost during a failure prior to fail-over are resent from the sending buffer using an alternative signaling channel resulting is a significant increase is the reliability of message transfer with the cost of additional memory for buffering and some additional messaging delay. Due to continuous flow of messages, error detection is fast. Therefore, the added delay due to resending is small.

A fail-over takes place between signaling links in a signaling link set. The set is preconfigured to handle the fail-over.

For the fail-over procedures to work, under normal busy hour call traffic SS7 signaling links are dimensioned for 0,2 Erlang signaling traffic (20% of the 64 kbit/s timeslot capacity is in use). After the failover, some signaling link may need to carry a signaling traffic on 0,4 Erlang. Good solid implementations of SS7 in public network exchanges are able to run on wire speed. This means that even if the signaling link is sending signaling messages back to back, the exchange will not complain or loose any messages because of the load.

Also, some of the MTP functions in exchanges are implemented in replicated computers like we discussed earlier when we described exchange hardware architectures.

Currently, fiber transmission plant is significantly better than what SS7 assumes. Therefore, SS7 signaling is very reliable.

One additional feature helps to detect signaling channel failures quickly. When there is nothing to send, SS7 sends Fill-In Signaling Units (FISU), short messages that are nevertheless framed as messages. Having not seen a FISU or any other message for only a short period of time, a node can make the decision that the channel has been lost. Thus the period during which channel looses messages is quite short and channel fail-over can be launched fast.

Message ordering is preserved by MTP using several methods. First, messages are numbered on a signaling link between two neighbors (BSN, see the MSU structure). This helps to detect message loss and duplication. Actually, re-ordering could potentially take place during a fail-over or because of parallel processing of messages in several parallel intermediate nodes on a signaling route between two route end-points. The first reason for re-ordering can be tackled by locally significant message numbering but for the second the signaling link selection (SLS) is used. SLS code on level 3 is significant to all nodes on the path between the end points of a signaling route set. This helps to preserve message ordering even in case of signaling channel fail-overs. By letting a higher level such as ISUP to assign the SLS, the application can ensure that the order of messages is preserved when it matters such as for example for a telephone call whereas the order of messages between two different telephone calls does not matter.

Signaling Connection Control Part

A use case: location update

Let us take a look at the use case of location updates for GSM and 3G networks. Let us assume a phone that has a SIM (or USIM) card from the Finnish operator, Elisa, is visiting Madrid in Spain. The phone is switched on and it connects to a base station under a particular visited Mobile Switching Center. The user needs to be authenticated and the phone's location updated into the Home Location Register, a server, located in Helsinki. Authentication of the SIM card, i.e. the user and the location update are functions in the MAP protocol in the visited MSC and HLR. Based on a number retrieved from the SIM card, the visited MSC makes a message routing decision seeing that this is a foreign SIM card. Therefore, the first intermediate point on the messaging path is a gateway MSC. The GMSC is able to further route the message to an international network. In terms of SS7, the GMSC is connected to at least two different SS7 networks. One is the national SS7 signaling network and the other is an international SS7 network. In terms of point codes these two networks form two separate address spaces and thus MTP is unable to make the hop between the two SS7 networks. The first international switch in the international network also uses two SS7 networks: one for interconnection to the national network and another for the international network itself.

A similar situation for switching between MTP networks occurs on all SS7 network boundaries on the path from Spain to Finland. Finally, the messages arrive at the HLR in Helsinki.

In this example the function that is needed in the intermediate nodes on SS7 network boundaries does not need to concern itself with the details of whether the message is for authentication or location update or for some other purpose. After all the business relationship is between the visited network and the home network and the service required of the intermediate networks is that of message transport. All that is required is to correctly route the messages between two SS7 networks. This can be accomplished by executing SCCP in the boundary nodes.

SCCP service classes

SCCP provides both connectionless and connection oriented message transport services. The connectionless service comes in two variants: basic and order preserving. The connection oriented service means that a virtual connection is established between the communicating parties. The CO service also comes in two flavors: basic service and flow controlled CO service.

The basic connectionless SCCP service is quite similar to IP. The difference is that routing is based on routable telephone numbers rather than IP addresses and that only static routes based on configuration information are supported in SCCP while IP routers typically support a number of dynamic routing protocols in addition to static routes. Fault tolerance of connections in SCCP is achieved by replication of signaling channels on MTP level and replication of exchange functions. The benefits of dynamic routing protocols include a lesser need of configuration and the fact that the network can always recover if some level of connectivity remains in place. In terms of reliability of message delivery, the SS7 network wins over IP networks by a significant margin. This is both due to protocol functionality and strict engineering rules for the SS7 network.

It is the task of the upper level protocol to indicate which of the service classes it wishes to use.

Global Title

The address information for global routing is carried in the Global Title information elements for called and calling party in SCCP. The structure of GT is presented in Figure 5.6.

First octet in GT contains indicators about the presence of different optional parts in the rest of the GT. A GT can carry signaling point codes, subsystem numbers (SSN) and routable numbers from the telephone numbering space defined in E.164 and other numbering specifications. SSN complements what

MTP can do using the service indicator in the SIO field (turned out that 4 bits were not sufficient to label all possible users of MTP/SCCP).

Calling and called party in SCCP

8	7	6 5 4	3 2	1
National use	RI - Routing indicator	GT Indicator	SSN ind	PCode ind
Signaling Point Code				
Sub System Number (SSN)				
Global title translation type				
Numbering plan Coding method				
Address type				
Address information				

SSN (cmp. Port Numbers in TCP/IP)
1 - SCCP management
2 - TUP
3 - ISUP
4 - OMAP - Operation and
            Maintenance AP
5 - MAP - Mobile AP
6 - MAP/HLR
7 - MAP/VLR
8 - MAP/MSC
9 - MAP/EIR - Equipment Id reg
10 - MAP/AuC
11 - ISUP/SS ISUP supplementary
              services
12...247 , 249...252 reserved
248 - MUP (NMT Mobile UP
253 - OMC - Operation and
            Maintenance Center
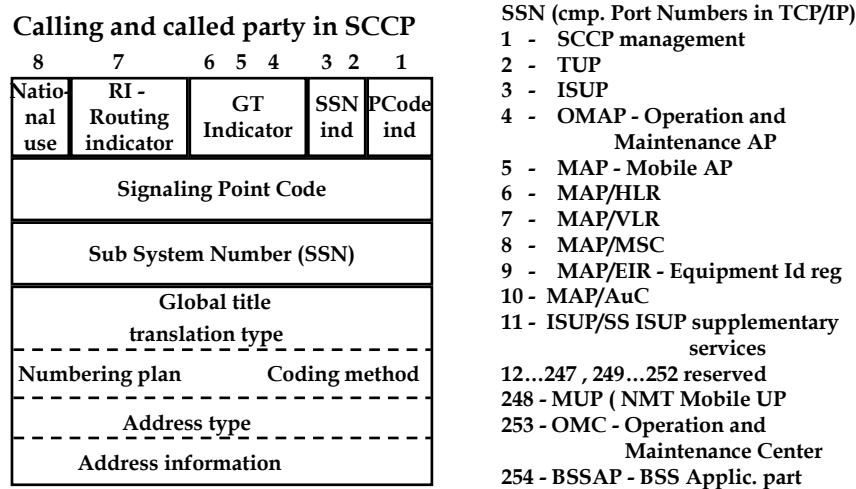254 - BSSAP - BSS Applic. part

Figure 5.6 Structure of Global Title in SCCP

Cellular systems such as GSM use two different routable numbering systems and also a combination of the two. One is the regular *directory numbering* (numbers than can appear in a telephone directory) used in Mobile Station ISDN numbers (MSISDN number) and the other is the International Mobile Subscriber Identity numbering specified in E.212. The hybrid is specified in E.214 and combines the familiar country codes and operator codes of E.164 with subscriber identities from IMSI. By using IMSI and mapping that to a temporary IMSI, cellular networks conceal the identity of their users and preserve their privacy. Privacy means that an effort is made by the network to hide the real identity of the users from 3rd parties listening on the air interface.

Having both MSISDN numbers and IMSI eases the implementation of Number Portability between operators in cellular networks. IMSIs are not public, so no harm to users is done if they change when the user switches from an operator to another. After the change, message routing can still be implemented efficiently using static routing tables in cellular network nodes.

More use cases of SCCP

SCCP implements global messaging over SS7 networks. The communication is out of band and completely independent of voice circuits. Due to this, it became possible to implement the signaling required for cellular networks. Imagine that for location updates one would need to reserve a voice circuit because of architectural properties of signaling systems!

IN user parts, both for wire line networks, i.e. INAP and mobile networks, i.e. CAP make use SCCP messaging services. SCCP allows aggregating IN

requests from a geographically large network to a node that contains service logic or a database or both. SCCP even allows aggregating such traffic globally from visited networks to a service node that resides in the home network of the subscriber. This is needed in CAP.

We mentioned that in principle ISUP may use SCCP. Two services where this is needed would be (1) a lookup service to find out the busy or free state of a destination subscriber without reserving a voice circuit and (2) CCBS supplementary service. The latter is the abbreviation for Call Completion to a Busy Subscriber (Call Back When Free). It works in the following way: A calls to B who is busy, A activates CCBS and the terminating exchange stores this state against B. When B becomes free, B's exchange looks up the state and sends a message to A's local exchange. A's LE alerts A and when A answers, initiates a call setup towards B automatically. If B is still free, the call is successfully set up. Networks rarely support this service and people use it even more rarely, so the link between ISUP and SCCP is not often used even if it was implemented at all.

Summary of SS7

SS7 significantly reduced call setup delays making the use of network resources more efficient. SS7 signaling is native to digital processors and program-controlled exchanges. SS7 signaling is out of band and uses a common channel for a large number of voice circuits, so it can continue irrespective what is happening with the voice circuit. SS7 signaling can even be sent and received without reserving a voice circuit at all. SS7 signaling is very reliable. And finally, SS7 made it possible to create mobile networks and allows implementing a rich set of services across a wide area network.

The reliability in SS7 was achieved with hard to implement requirements. Therefore, for a long time SS7 implementations (particularly good ones) were rare. Particularly challenging are distributed implementations of SS7. When SS7 was designed, strict layering was sacrificed for the sake of high performance and high reliability. This decision was fine when thinking of the vertically integrated products such as exchanges or Central Offices. If we think of signaling more generally as an infrastructure for implementing network services and hope to create a competitive market, probably this was not the best choice. However, loose layering with vendor specific interfaces between levels was a sound decision at the time when it was taken.

SS7 was designed in an era when most big countries had a government monopoly operator taking care of telephony services. The focus of thinking was on traditional telephony services and possibly a few additional tricks to add some limited amount of value to the service. Each operator wished to control the use of all the resources it had in its network, so the thinking was that calls are processed hop-by-hop and that each operator handles its own

network autonomously. By selecting short 14 bit addresses for MTP, operators were looking for isolating their SS7 signaling networks from each other. At least for signaling network stability this was a sound decision. We will see on later lectures that the idea of hiding the network topology from other players is still popular among the operators.

SS7 carries the heritage of a closed market with a limited number of players. It was not designed to be easily extended to operators that are independent of network operators that would for example implement additional value added services for telephone users. For such cases addressing scales poorly on MTP level and security needs to be carefully considered so that two such players do not for example interfere with each other's services. SS7 is not extended to PABXs, instead DSS PRI is used (actually this is politics. Few technical obstacles can be identified why SS7 could not be used to connect PBXs to a public network). Sometimes SS7 is used to connect Voice Mail systems to exchanges.

Today, global statistics tells that there are about 7 Billion mobile subscriptions. Telephone calls to and from most of these subscribers are still circuit switched. Between network nodes, these calls are signaled using CCSS7. CCSS7 has also been ported to the IP-environment, we will come back to this at the end of the course. Due to the very large installed base, it will be very hard to get rid of CCSS7 in the years to come. In terms of designing new signaling protocols, due a rich set of included protocol mechanisms, CCSS7 acts as a model and a point of reference.