Reliable Signalling Transport in Next Generation Networks

E. Vázquez, M. Álvarez-Campana, A. Hernández, and J. Vinyes

ETS Ing. Telecomunicación, Universidad Politécnica de Madrid
Av. Complutense s/n, 28040, Madrid, Spain
{enrique,mac,albertoh,vinyes}@dit.upm.es

Abstract. Next-generation network architectures aim at the convergence of services (e.g. telephony, television, Web access, online games, and new services) over a common IP core. Additionally, the integration of wired and wireless access technologies will offer users ubiquitous access to all those services. Telecommunications networks rely on signalling protocols to perform key service control functions such as locating and authenticating users, establishing communication sessions, reserving resources, charging, etc. Signalling is crucial for network operators and so requires a transport service with very high availability and low delay. Traditionally, operators have deployed dedicated and highly redundant signalling networks following ITU-T standards. More recently, and motivated by the current evolution towards IP, the IETF has defined SIGTRAN, a new protocol suite intended to transport signalling protocols over IP. This paper addresses the problem of configuring the protocol mechanisms existing at different layers of the SIGTRAN architecture in a coordinated way, with the goal of creating a redundant network topology able to meet the stringent availability levels required for signalling in carrier-grade networks. The proposed solution is described and evaluated numerically.

Keywords: Signalling, SIGTRAN, IP, network availability.

1 Introduction

The fixed telephone network and mobile networks such as GSM, GPRS, and UMTS employ the Signalling System no. 7 (SS7) standardized for international use by ITU-T [1]. SS7 is a packet-switched network designed specifically to transport call control messages (e.g. call setup, call release) among telephone exchanges with low loss, low delay, and very high availability. An SS7 network exists in parallel to the telephone network to which it serves. A few trunk circuits are dedicated to carrying signalling messages, and the remaining circuits carry user calls.

Over the years, the SS7 architecture was successfully adapted to support Intelligent Network (IN) services, broadband networks, and 2G/3G mobile networks [2]. In this evolution process, new types of nodes have been connected to the SS7 networks in addition to telephone exchanges, for instance computers that provide IN services, databases that store information about mobile subscribers, and radio base station controllers. Also, new signalling protocols have been added in the upper layers of the SS7

architecture to communicate with the novel elements. These protocols implement transaction control, mobility management, security procedures, and others that greatly extend the basic call control functionality of the original SS7. On the contrary, the lower layers of the SS7 architecture, which are responsible for the reliable transfer of signalling messages, have remained essentially unchanged.

Recently, the IETF defined the architecture for signalling transport over IP networks known as SIGTRAN [3]. SIGTRAN includes several adaptation protocols and a common transport which normally is the new Stream Control Transmission Protocol (SCTP) [4]. See Fig. 1. In applications where high availability is not required, TCP may replace SCTP. (Also, SCTP is increasingly used for other purposes, different from signalling transport.) Depending on the adaptation chosen, SIGTRAN can transport different signalling protocols.
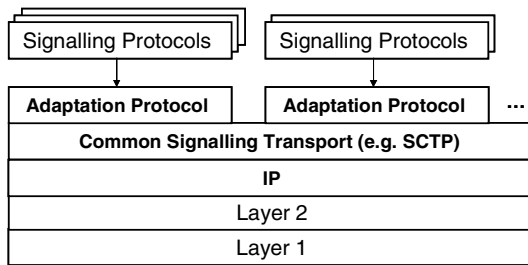
Signalling Protocols	Signalling Protocols	
Adaptation Protocol	**Adaptation Protocol**	...
Common Signalling Transport (e.g. SCTP)		
IP		
Layer 2		
Layer 1		

Fig. 1. Signalling Transport (SIGTRAN) over IP

A typical usage of SIGTRAN is in Signalling Gateways that allow new voice over IP networks to interwork with traditional telephone networks using circuit switching and SS7. The gateway receives the signalling messages coming from the SS7 network and forwards them on the IP side with SIGTRAN. In the opposite direction, the messages received from the IP side are forwarded by the gateway to the SS7 network.

SIGTRAN can also be used by fixed and mobile network operators as one step in the evolution to a common IP-based transport infrastructure, not only for user traffic but also for signalling. Mobility control messages increase the signalling traffic in mobile networks compared to fixed ones. Moreover, in GSM the enormous success of the Short Message Service (SMS) is another important factor that raises the signalling network load, because short messages are transported over SS7. In this situation, SIGTRAN can replace parts of the SS7 network in order to offload traffic to IP. This can be done at layer 2 or at layer 3. In the first case, reliable SCTP connections over IP (called *associations*) replace selected links in an SS7 network. The adaptation protocol used over the SCTP association provides the service interface expected by the SS7 network layer, which remains unchanged. The SS7 routing and management procedures are kept to provide network monitoring and recovery from failures.

This paper focuses on the second case, in which the three lower layers of the SS7 network, called the Message Transfer Part (MTP), are completely replaced by an IP network with SIGTRAN. The end-to-end SS7 protocols located above the MTP do not change. Inside the network, signalling traffic is carried over IP, while SIGTRAN/MTP gateways provide connectivity to external SS7 networks. This approach has important advantages for the operator. Network nodes such as telephone switches, user databases, and short

message service centers send signalling and SMS traffic over IP interfaces. The MTP protocols are required in the gateways only. If necessary, the transport capacity available for signalling or SMS can be increased in a more flexible and economical way than in a traditional SS7 network using dedicated circuits. Additionally, for a mobile network operator the adoption of SIGTRAN fits well in the evolution path to All-IP 3G/4G networks.

However, the replacement of the Message Transfer Part in a real SS7 network is not trivial. The SIGTRAN protocols and the underlying IP network have to be carefully configured in order to maintain the quality of service, security and, most important, the high availability required by signalling. This is the problem addressed here.

Both SCTP and the adaptation protocols include mechanisms designed to handle redundant configurations and improve availability, most notably SCTP multi-homing, which protects against network failures, and support for active/backup server processes at the adaptation layer, which protects against node failures. While multi-homing and other SCTP features have been widely studied, there are comparatively few references in the literature about configuration of the adaptation protocol.

The next section gives necessary background on SS7 networks, focusing on the functions that are more relevant for this work. The contributions of the paper are presented in sections 3 and 4. Section 3 analyzes the transport and adaptation functions that increase the SIGTRAN network availability, and then defines a configuration of SIGTRAN gateways with multi-homed associations and redundant servers which can be applied to migrate from SS7 to SIGTRAN. Section 4 evaluates the load in the proposed SIGTRAN network. To conclude, section 5 reviews existing related work, and section 6 summarizes the contributions of the paper.

2 Overview of Signalling System no. 7

2.1 Architecture

As mentioned in the introduction, Signalling System no. 7 is an ITU-T standard [1] that defines a packet-switched network and its corresponding protocols for transporting signalling information among the different elements (telephone exchanges, computers, data bases, etc.) of a telephone network. The end nodes making use of the SS7 network services are called Signalling Points (SP), and the intermediate switching nodes within the SS7 network are called Signalling Transfer Points (STP). See Fig. 2. The connections among them are realized by means of Signalling Links, which are usually based on 64 kbit/s channels of E1 or T1 lines.

To cope with node or link failures, redundant network topologies are employed, so that there are several alternatives for routing messages towards any particular destination point. Adjacent SS7 points are usually connected by a Link Set formed by two or more signalling links. Furthermore, end points are attached to two different transfer points, and these are interconnected in a full-meshed topology. Typically, the operator of a large SS7 network will divide it into several regions, each of them having a mated pair of transfer points that serve the signalling points in the region, and will interconnect the different regions with a full mesh of signalling links.

Fig. 2 shows the SS7 protocol architecture. The Message Transfer Part (MTP) provides a basic datagram-like network service for the exchange of messages between signalling points. Further details of MTP are given in section 2.2. The Signalling Connection Control Part (SCCP) provides additional service capabilities such as connection-oriented transfer, segmentation and reassembly, and extended addressing.

The upper SS7 components can be classified into two categories: User Parts and Application Parts. User Parts are traditional signalling protocols oriented to establish, maintain, and terminate calls. The Integrated Services User Part (ISUP) is the standard for establishing and releasing calls in telephone networks, both fixed and mobile.
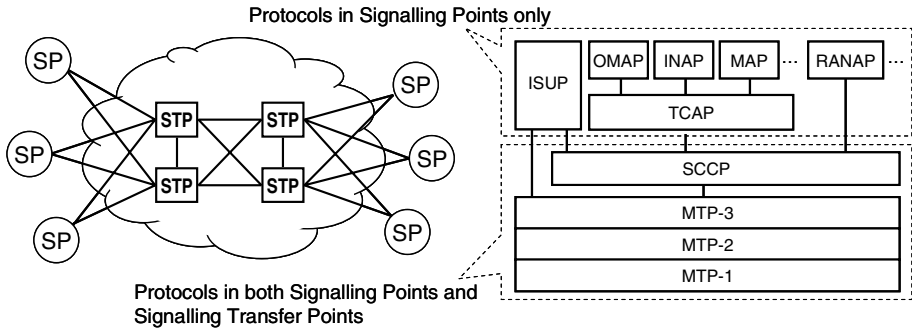

Fig. 2. SS7 Network and Protocol Architecture

Application Parts perform an increasing number of functions which divert from basic circuit/call control, for example the Intelligent Network Application Part (INAP), the Mobile Application Part (MAP), and the Operation and Maintenance Application Part (OMAP). These elements follow a client-server model, based on remote operations supported by the Transaction Capabilities Application Part (TCAP). Radio Access Network Application Part (RANAP) is one of the new additions to the SS7 protocol family, standardized for UMTS networks. User Parts and Application Parts do not change when the Message Transfer Part is replaced by SIGTRAN.

2.2   Message Transfer Part

The SS7 Message Transfer Part provides a connectionless, point-to-point transfer service used by protocols located in the upper layers of the SS7 architecture. Each signalling message contains two addresses called Origin Point Code and Destination Point Code. The MTP may send the message via a direct link from origin to destination or, most often, via one ore more transfer points that route the message towards its destination. The SS7 data-link layer protocol (MTP2) handles error control and retransmissions in each hop. When the signalling message reaches its destination, the Octet Indicator Service field indicates the upper-layer protocol to which the message has to be delivered. (This is similar to the function of port numbers in TCP and UDP.)

As mentioned above, signalling links and routes are normally provisioned redundantly, and the MTP performs load sharing among the available links and routes. As a result, two signalling messages addressed to the same destination may follow different routes and be

delivered out of order (recall that MTP provides a connectionless service). For example, in Fig. 3 the possible routes from A to B in absence of failures are A-STP1-STP2-B, A-STP1-STP4-B, A-STP3-STP2-B, and A-STP3-STP4-B. The links between transfer points of the same pair, i.e. STP1-STP3 and STP2-STP4, are used only in case of failures.

The MTP supports a mechanism that allows the sender to force the same route for a sequence of messages addressed to a given destination point, so they are delivered in order. It makes use of a field called Signalling Link Selector (SLS) that is included in every message. If several messages are sent from A to B with the same selector value, the MTP must choose the same route for all of them. For example, all ISUP messages related to the same call are sent with the same link selector, so that they are delivered in sequence and the signalling dialog for that call is not altered.

The network layer protocol (MTP3) includes management procedures that monitor the status of the SS7 network and divert traffic from a failed route to an alternative one if necessary [5]. Fig. 3 shows different cases of link or node failures and the alternative routes that can be followed in each case to reach node B. In Fig. 3a) the link STP1-STP2 has failed. Therefore, STP1 suspends the load sharing between STP2 and STP4, and temporarily routes all messages in transit for destinations B and C via STP4 regardless of their link selectors. If STP1-STP4 also failed, STP1 would divert traffic to STP3 as a last option. Fig. 3b) shows a similar case. If STP2 goes down, both STP1 and STP3 divert traffic to STP4. Finally, in Fig. 3c) the link STP2-B has failed, but STP2 can still deliver messages to B via STP4.
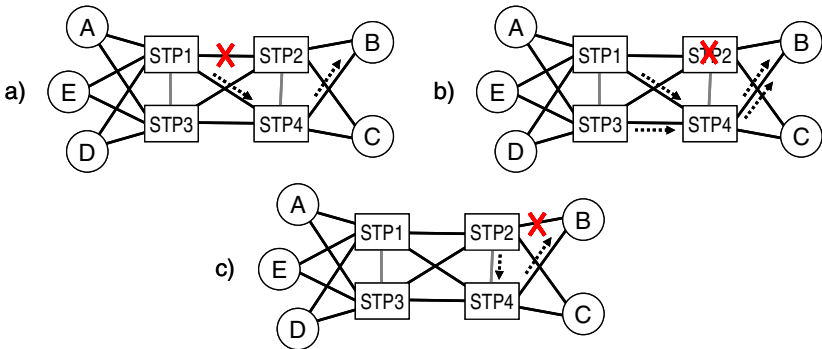

Fig. 3. Examples of Possible Failures and Alternative Routes

Network redundancy and the MTP routing functions outlined above offer a reliable transfer of messages between any two signalling points, meeting strict requirements, for example: message loss probability smaller than 10^{-7}, probability of message delivered out-of-sequence (including message duplication) smaller than 10^{-10}, probability of message delivered with undetected errors smaller than 10^{-10}, and signalling route set unavailability smaller than 10 minutes per year [6].

In addition, the MTP offers low message transfer delays (except for messages transmitted over satellite links, due to the long propagation delay). The delay in each intermediate transfer point is minimised by overdimensioning the capacity of the outgoing links. The performance and availability of SS7 networks based on mated STP pairs have been extensively studied in the literature [7].

3   Reliability in SIGTRAN Networks

The IETF SIGTRAN working group addressed the transport of signalling protocols (ISUP, TCAP, etc.) over IP, taking into account the functional and performance requirements of traditional signalling networks. For this purpose, the group has defined a general architecture [3] (see Fig. 1 and Fig. 4), the Stream Control Transmission Protocol (SCTP), and several adaptation protocols. The MTP3-User Adaptation Layer (M3UA) is the most important for this paper. SIGTRAN uses IP and existing IETF quality of service and security standards. As mentioned in the introduction, SIGTRAN is useful for new operators with IP networks that need to interwork with SS7 networks through signalling gateways, as well as for incumbents that are migrating their legacy circuit-switched infrastructure to IP. The functions of SCTP and M3UA that are more relevant for this paper are analyzed in sections 3.1 and 3.2 respectively.

3.1   Stream Control Transmission Protocol (SCTP)

SCTP [4,8] is a reliable, connection-oriented protocol that offers acknowledged, error-free, non-duplicated transfer of user messages. A connection established between two SCTP endpoints, called an *association*, may contain one or several *streams* in each direction. Sequence numbers are defined per stream, so SCTP guarantees ordered delivery within each stream only. In this way, message losses and retransmissions in one stream do not delay other messages going in the same direction but in a different stream. The stream identifier is 16 bits long, so there is ample room to define as many streams as necessary. Some SCTP messages may be marked with the U (unordered) bit set to 1. They are delivered to the upper layer as soon as they are received. The concept of stream in SCTP can be related to that of Signalling Link Selector. As stated in section 2.2, SS7 messages transmitted with different selector values are not required to maintain their relative order. Therefore, these messages may be transported over different SCTP streams.

SCTP supports multi-homing. Each association endpoint can have several IP addresses, and the protocol includes procedures for actively monitoring the reachability status of each peer address by means of periodic "heartbeat" messages. Each endpoint keeps a transmission error count for each of the addresses of its peer. An address is marked inactive if its error count exceeds a predefined threshold. When a heartbeat acknowledgement is received from a peer address, it is considered active again.

The destination and source IP addresses that an endpoint normally puts into outbound packets are called the primary path. Alternative active addresses, if available, are used for retransmissions or when the primary path is interrupted due to a network failure. If the underlying IP network topology is configured in such a way that packets sent to different peer addresses follow disjoint paths in the network without common points of failure, SCTP is able to maintain the association and provides effective recovery from network failures. If, in spite of multi-homing, the peer endpoint becomes unreachable, e.g. due to multiple network failures or to a failure of the endpoint itself, the SCTP association is lost. In this case, the redundancy mechanisms of the upper layer (see next section) come into play.

3.2 MTP3-User Adaptation Layer (M3UA)

In the SS7 architecture, ISUP and SCCP utilize the services provided by the MTP3 layer. The MTP3-User Adaptation Layer (M3UA) [9] has been designed to transport these protocols over IP without change. The 3rd Generation Partnership Project (3GPP) specifies the use of M3UA over SCTP as an option for signalling transport in the UMTS Terrestrial Access Network (UTRAN) interfaces Iub, Iur, and Iu.

M3UA can be used between a Signalling Gateway (SG) and an IP node, for example a Media Gateway Controller (MGC), an IP Home Location Register (HLR), etc. See Fig. 4. The M3UA gateway terminates all MTP protocols on the SS7 side.
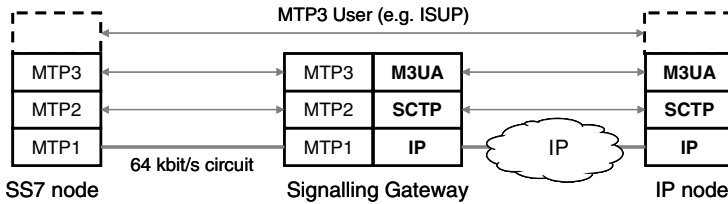

Fig. 4. Signalling Gateway with M3UA

From the point of view of the SS7 network, a link towards an M3UA gateway can route messages addressed to one or several signalling points, which terminate the higher-layer SS7 protocols (ISUP, SCCP, TCAP, MAP, etc.) Consequently, an M3UA gateway behaves as a Signalling Transfer Point (STP) in the SS7 terminology.

M3UA allows a completely distributed implementation of the protocol entities that process the SS7 messages on the IP side. The signalling messages that arrive from the SS7 network may be distributed by the gateway to physically different hosts located anywhere in the IP network, in order to achieve load sharing or redundancy. M3UA defines several logical entities intended to provide a flexible and dynamic allocation of signalling messages to protocol entities able to handle them. A Routing Key (RK) defines a subset of signalling messages based on conditions set upon the values of selected message fields. In a general example, RK1 may correspond to all messages addressed to destination X, and RK2 to ISUP messages addressed to destination Y.

An Application Server Process (ASP) is a process instance capable of handling messages for one or several RKs. At a given point in time, a process may be in different states for each of its RKs. For example, ASP A may be active for RK1 and RK2 (i.e. A is currently processing messages of those RKs), but inactive for RK3 (i.e. A does not process messages of RK3 now, but can do it later if necessary.) Each ASP is connected to the gateway via an SCTP association which transports all signalling messages handled by that process.

More than one ASP may be able to process messages for the same RK. For example, let us assume that processes A and B (possibly located in different hosts) can handle messages identified by RK1. One of them may be active and the other inactive as a backup. Alternatively, both processes may be active at the same time in load sharing (some messages go to A, and the others to B) or broadcast mode (each message goes to both A and B). The set of all Application Server Processes associated to a given routing key is called the Application Server (AS) for that key. See Fig. 5.
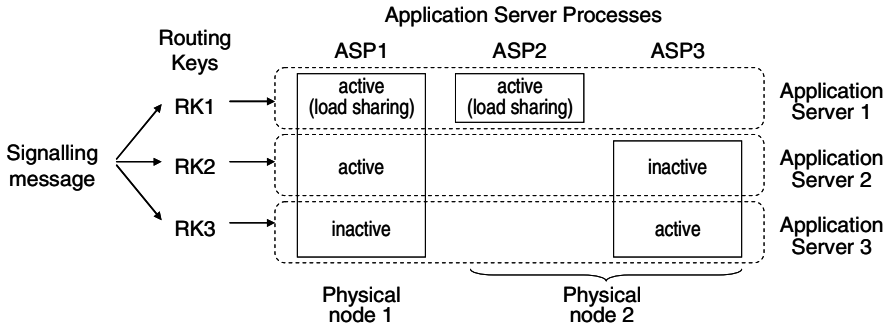
Fig. 5. Routing Keys, Application Server Processes, and Application Servers

The signalling gateway maintains a list of currently defined routing keys and monitors the state of all configured remote processes. For each signalling message coming from the SS7 network, the gateway matches the message to one key, determines the destination process (or processes in case of broadcast), and forwards it over SCTP.

In summary, by using M3UA with a redundant configuration of active and backup ASPs located in different machines, the SIGTRAN network can be effectively protected against host failures. A redundant IP network topology, together with the heartbeat and multi-homing procedures of SCTP associations presented in section 3.2, add protection against network failures. With an adequate configuration of M3UA and SCTP protocols, SIGTRAN networks can match the high reliability levels of traditional SS7 networks. The next section illustrates this with a detailed example.

3.3 Proposed SIGTRAN Network Configuration

Based on the analysis of relevant SIGTRAN mechanisms made above, this section defines a SIGTRAN network design equivalent to the basic mesh or quad of Signalling Transfer Points that typically forms the building block of SS7 networks. (This basic mesh, explained in section 2, is reproduced in Fig. 6. For clarity of the explanation, no links between transfer points of the same pair, i.e. STP1-STP3 and STP2-STP4, have been included in the example.) This provides a general solution in which large SIGTRAN networks are built by systematically repeating the configuration of interconnected signalling gateway pairs described here.

Let us recall that Signalling Transfer Points (STPs) are deployed in pairs, and each STP pair gives access to a subset of Signalling Points (SPs). In absence of failures, the traffic of each area is shared by the two STPs serving that area. For example, in Fig. 6 STP1 and STP3 serve area L, while STP2 and STP4 serve area R. STP1 will normally split the traffic addressed to destinations located in area R among STP2 and STP4 depending on the Signalling Link Selector (SLS) values. Let us assume that messages with even selector values are sent via STP2, and messages with odd values are sent via STP4. If the link STP1-STP2 fails, or if STP2 is down, STP1 will route all traffic for area R through STP4. Conversely, if the failure affects STP4, STP1 will route all traffic through STP2. The other three transfer points behave in a similar way.
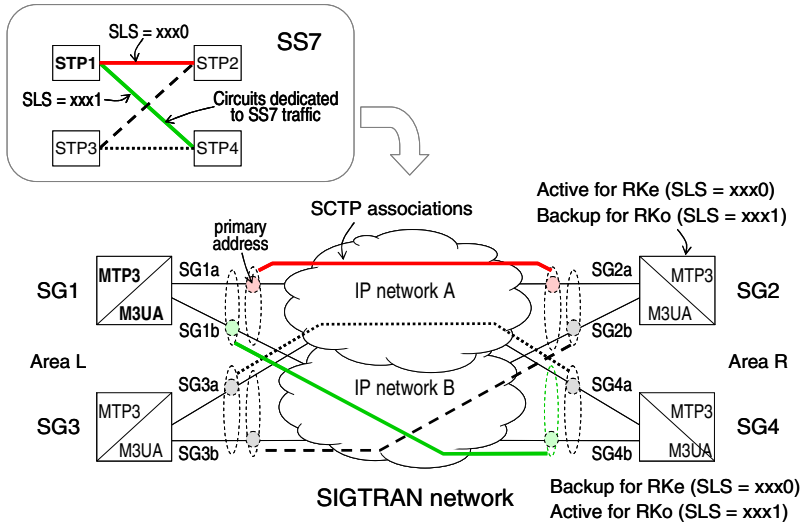
Fig. 6. Basic SS7 Mesh Network and Equivalent SIGTRAN Network

This behaviour can be reproduced in a SIGTRAN network with M3UA as follows. Each transfer point is upgraded to a Signalling Gateway (SG) that implements the MTP protocol stack on the links that come from the end points in its service area, and the M3UA/SCTP/IP stack on the links that go to gateways serving other areas. Each signalling gateway is assumed to have separate interfaces to two IP networks A and B. Let SG1a and SG1b denote the IP addresses of the two network interfaces in gateway SG1. (In the following we take SG1 as example. The other three gateways are configured in a similar way.)

The signalling link SG1–SG2 is implemented as a multi-homed SCTP association between the address list (SG1a*, SG1b) on one side and (SG2a*, SG2b) on the other side. A second SCTP association is established between (SG1a, SG1b*) and (SG4a, SG4b*), corresponding to the link SG1–SG4. The stars indicate the addresses selected for the primary path in each association. If SCTP detects that the primary path is unavailable, e.g. due to a network failure, it automatically switches traffic to the alternative path without breaking the association. Therefore, from the point of view of M3UA the network failure is transparent and the connectivity between the signalling gateways is maintained.

As indicated above, gateway SG1 must send messages with even link selector values to STP2, and messages with odd values to STP4. To force this behaviour, we define two routing keys: RKe matches traffic going to SG2/SG4 with even selector values and RKo matches traffic with odd values. RKe is associated with an Application Server formed by two Application Server Processes: SG2 (active) and SG4 (backup). RKo is associated with another server formed by the same two processes as before, but now the initial active and backup roles are reversed: SG4 is the active process and SG1 is the backup. In this way, if SG2 fails (or the SCTP layer is unable to maintain the association SG1–SG2) SG1 will send all traffic to SG4 as expected. Obviously, if the failure affects SG4, traffic goes to SG2.

Finally, in SS7 networks the MTP is not required to maintain the relative order of messages with different Signalling Link Selector values. Therefore, in the proposed SIGTRAN network different selector values may be mapped to different streams within each SCTP association between gateways.

Based on the preceding description, it is straightforward to deduce the configuration of SCTP associations and active/backup processes for the remaining gateways SG2, SG3, and SG4. It is also possible to generalize this example to include links between STPs of the same pair or for networks with more than two STP pairs.

The proposed configuration can be applied, for example, by circuit-switched network operators that have deployed large SS7 networks (fixed-line incumbents, mobile operators) and want to gradually migrate to IP and SIGTRAN. In this case, an initial step may be moving the inter-STP traffic to IP. In addition to the redundancy and failure recovery capabilities provided by M3UA and SCTP, the existing signalling links can be temporarily kept as additional alternative routes until the reliability of the new SIGTRAN network has been validated. In a second step, the access links between Signalling Points and STPs may be migrated to SIGTRAN as well.

4   Evaluation of Signalling Load in the SIGTRAN Network

The SIGTRAN traffic will increase the load in the IP network of the operator. An approximate estimation of the SIGTRAN traffic volume can be obtained from measurements taken in the existing SS7 network, adjusting the values to take into account the different protocol overheads of MTP and SIGTRAN. While MTP2 and MTP3 add a total of 11 octets per message, the combined headers of M3UA, SCTP, and IP are considerably longer. M3UA data messages include a common header (8 octets), an additional header (16 octets), and two optional fields (8 octets each), giving a total of 40 octets. An SCTP message with common header (12 octets), DATA block and Selective Acknowledgement block (16 octets each) adds 44 octets more. Adding an IP v4 header (20 octets), the resulting overhead is 104 octets per message, plus the overhead of any protocol layers below IP. This value may be further increased if IPsec security procedures are used [10]. Considering that many of the messages generated by ISUP and other SS7 protocols are short, e.g. a few tens of octets, it is clear that the SIGTRAN overhead has a significant impact on the total signalling traffic load.

The estimation of the SIGTRAN traffic volume is useful if this traffic is going to receive a preferential treatment in the IP network, for example by means of Differentiated Services (diffserv), MPLS Traffic Engineering (MPLS-TE), or a combination of both [11]. These mechanisms help to offer a better quality of service (e.g. low delay) to signalling flows transported over the IP network. Additionally, MPLS offers fast reroute capabilities below IP [12] that reduce the number of network failures that have to be recovered by SCTP or M3UA.

Besides reliability, quality of service and security are two fundamental requirements for every signalling network, including IP-based ones. Although IPsec, diffserv, and MPLS-TE are not considered further in this paper, they are standards that complement SIGTRAN in order to meet such requirements. Therefore, the necessity of using them should be assessed as part of the overall SIGTRAN network design.

Instead of relying on measurements taken in a real SS7 network, or as a complement to them, the signalling traffic may be modelled analytically or by simulation. The effort required to develop a detailed signalling model is not small, because it is necessary first to identify the services and procedures of interest, and then to analyze the possible SS7 message exchanges generated by each of them. However, the advantage of such model is that it relates signalling load to service usage. Consequently, it serves to study the signalling network not only in the current situation, but also in alternative scenarios that may be of interest for the operator and for which there are no measurements. For example, the model can be used to estimate the impact on the signalling network caused by changes in current services or user profiles, introduction of new services, etc.

To illustrate this, we consider here the signalling messages used in GSM mobile networks to support voice calls and the Short Message Service (SMS), including the signalling procedures triggered by user mobility, for example change of Visitor Location Register. For each service, we have analyzed the possible cases that produce different signalling sequences. For example, in voice calls it is necessary to distinguish calls coming from an external network, calls going to an external network, intra-network calls between mobile users in the same region, calls between mobiles in different regions, etc. For each service and case, we identify the network nodes involved: access and transit Mobile Switching Center (MSC), Home Location Register (HLR), Visitor Location Register (VLR), Short Message Service Gateway (SMSG), Serving GPRS Support Node (SGSN), Signalling Transfer Point (STP), etc., as well as the signalling messages transmitted and received by each of them.

For example, Table 1 shows the messages exchanged during a call between two mobile users located in the same region (so their respective switching centers are connected to the same pair of STPs). Similar tables for other services are not shown here due to lack of space, but they can be found in [13].

Table 1. Signalling Messages During a Voice Call

Network node		SS7 Integrated Services User Part (ISUP) Messages	SS7 Mobile Application Part (MAP) Messages
Calling MSC	Send	IAM, REL	SRI
	Receive	ACM, ANM, RLC	SRI_Ack
Called MSC/VLR	Send	ACM, ANM, RLC	Provide_MSRN_Ack
	Receive	IAM, REL	Provide_MSRN
Transit MSC	Send	IAM, ACM, ANM, REL, RLC	
	Receive	IAM, ACM, ANM, REL, RLC	
HLR	Send		SRI_Ack, Provide_MSRN
	Receive		SRI, Provide_MSRN_Ack

The analysis includes several parameters that may be configured in order to adapt it to the scenarios of interest, namely network topology, number of users, traffic per user and service in the busy hour, message lengths, and protocol overheads. As result, we obtain the signalling load in each node interface, and the load in the links between STPs. For example, let us assume a GSM network with 10 million subscribers uniformly distributed over 50 MSCs and 10 HLRs. The network has 8 transit MSCs,

8 SGSNs, and 4 SMSGs. The SS7 network consists of three pairs of STPs interconnected by a full mesh of signalling links. The parameter values that define the traffic per subscriber have been derived from real data of the Spanish mobile market, taken from the annual report published by the Spanish national telecommunications regulator 'Comisión del Mercado de las Telecomunicaciones' (CMT) [14]. Based on this report, we have selected values for short messages sent and received per subscriber, voice calls with subscribers of the same mobile network, calls with subscribers of other mobile networks, and calls with subscribers of fixed telephone networks.

If SIGTRAN is deployed in all interfaces (i.e. inter-STP links and access links of MSCs, HLRs, etc.), the estimated signalling load values computed by the model are shown in Fig. 7. The highest load, 1482 kbit/s, is found in the transit MSCs. MSC/VLRs, HLRs, and SMSGs give smaller values, between 146 and 355 kbit/s. For each node, the graph indicates the fractions of the total load generated by the different services considered.

Note that only MSC/VLRs and HLRs are involved in all services. Transit MSCs are not concerned by SMS or location updates, and SMSGs deal only with the Short Message Service. The small signalling load computed for SGSNs, 4.3 kbit/s, corresponds to routing area updates where subscribers move from their current node to a new one.

The values given in Fig. 7 correspond to the flow of messages generated by each network node. The signalling load in the opposite direction, i.e. messages received by each node, may be higher or lower because the procedures are not symmetric. Although not shown in the graph, we estimated the load values in both directions and the differences found are small: between 0.6% and 5.3%.
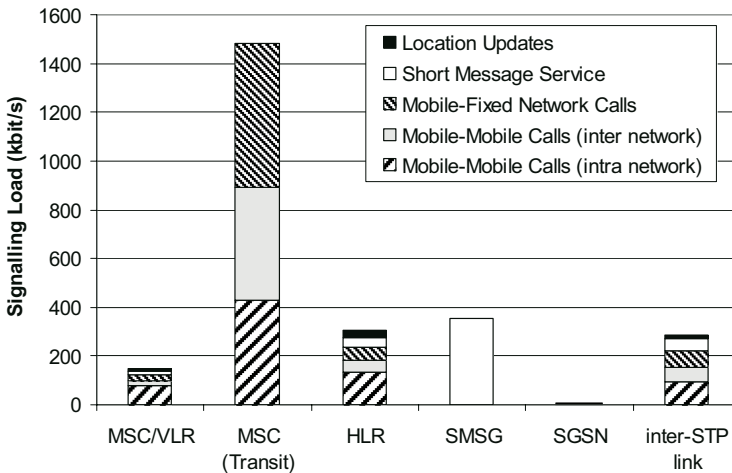

Fig. 7. Signalling Load when All Nodes Use SIGTRAN

The type of data commented above helps to quantify the new signalling traffic flows that must be transported over IP if the traditional SS7 network is migrated to SIGTRAN, in order to ensure that these flows, which are crucial for the operator, receive an adequate quality of service.

5 Related Work

As mentioned previously, multi-homing and other SCTP features have been extensively studied, but there are comparatively few references in the literature about the M3UA protocol. The operation of SCTP in failover scenarios has been investigated for example in [15,16,17]. These papers evaluate SCTP-controlled failovers under different parameter settings (e.g. timeout values, round-trip delay, traffic load, etc.) The correctness of the failover algorithm has been analysed in [18], where the authors have detected problems in some circumstances and proposed solutions. Other works study applications of SCTP beyond its original purpose of signalling transport, such as HTTP transport over SCTP [19] and mobility management at the transport layer based on multi-homing plus extensions for dynamic address reconfiguration [20].

The configuration of process redundancy mechanisms in M3UA is touched upon in [21] and [22], but we have not found any papers dealing with this issue in detail. In [23], the authors point out some limitations of redundancy procedures in M3UA and other SIGTRAN adaptation protocols, and present solutions based on the new Reliable Server Pooling architecture being defined by the IETF. Rserpool aims at creating a unified redundancy solution supporting high availability and scalability of applications through the use of pools of servers. At the time of writing this paper, only the requirements for reliable server pooling have been published as an informational RFC, while the architecture and the protocols are still in draft state.

6 Conclusions

This paper focuses on the practical application of the IETF SIGTRAN standards in order to facilitate the migration towards IP of traditional SS7 networks used by telecommunications operators. Special attention is paid to the high network availability and quality of service required by signalling traffic when it is moved from SS7 to IP.

The paper makes two contributions. Firstly, it describes a configuration of SIGTRAN gateways with SCTP multi-homed associations and M3UA redundant servers that matches the signalling route redundancy offered by the basic arrangement of meshed STP pairs typically employed in SS7 networks. This configuration provides a general building block that can be used to deploy large SIGTRAN networks. Secondly, it presents a detailed analysis of signalling message exchanges that serves to estimate the SIGTRAN network load in different network configuration and service scenarios. As a practical case, we describe the application of the proposed SIGTRAN configuration to a GSM mobile network, and evaluate the resulting signalling load, taking as input real data of subscriber behaviour in Spanish mobile networks.

Acknowledgements

References

1. TU-T Recommendation Q.700: Introduction to Signalling System No. 7 (1993)
2. Kuhn, P.J., Pack, C.D., Skoog, R.A.: Common channel signalling networks: past, present, future. IEEE Journal on Selected Areas in Communications, vol. 12, no. 3 (1994)
3. Ong, L., Rytina, I., García, M., Schwarzbauer, H., Coene, L., Lin, H., Juhasz, I., Holdrege, M., Sharp, C.: Framework Architecture for Signalling Transport. RFC 2719 (1999)
4. Stewart, R.: Stream Control Transmission Protocol. RFC 4960 (2007)
5. ITU-T Recommendation Q.704: Signalling network functions and messages (1996)
6. ITU-T Rec. Q.706: Message Transfer Part signalling performance (1993)
7. Chung, M.Y., You, J.U., Sung, D.K., Choi, B.D.: Performability Analysis of Common-Channel Signalling Networks Based on Signalling System #7. IEEE Transactions on Reliability, vol. 48, no. 3 (1999)
8. Fu, S., Atiquzzaman, M.: SCTP: state of the art in research, products, and technical challenges. IEEE 18th Annual Workshop on Computer Communications (CCW), Dana Point, California, USA (2003)
9. Morneault, K. (ed.), Pastor-Balbas, J. (ed.): Signalling System 7 (SS7) Message Transfer Part 3 (MTP3) - User Adaptation Layer (M3UA). RFC 4666 (2006)
10. Loughney, J., Tüxen, M., Pastor-Balbas, J.: Security Considerations for Signalling Transport (SIGTRAN) Protocols. RFC 3788 (2004)
11. Fineberg, V.: QoS Support in MPLS Networks. MPLS/FR Alliance (2003)
12. Pan, P. (ed.), Swallow, G. (ed.), Atlas, A. (ed.): Fast Reroute Extensions to RSVP-TE for LSP Tunnels. RFC 4090 (2005)
13. Guénon, G., Vázquez, E., Álvarez-Campana, M.: SS7 Signalling Transport over IP. Universidad Politécnica de Madrid (2006)
14. Comisión del Mercado de las Telecomunicaciones (CMT), http://www.cmt.es
15. Jungmaier, A., Rathgeb, E., Tüxen, M.: On the use of SCTP in failover scenarios. 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI), Orlando, Florida, USA (2002)
16. Grinnemo, K-J., Brunstrom, A.: Impact of Traffic Load on SCTP Failovers in SIGTRAN. 4th International Conference on Networking (ICN), Reunion Island (2005)
17. Jungmaier, A., Rathgeb, E.: On SCTP multi-homing performance. Telecommunication Systems, vol. 31, no. 2-3 (2006)
18. Noonan, J., Murphy, J., Murphy, S., Perry, P.: Stall and Path Monitoring Issues in SCTP. IEEE INFOCOM 2006, Barcelona, Spain (2006)
19. Natarajan, P., Iyengar, J.R., Amer, P.D., Stewart, R.: SCTP: an innovative transport layer protocol for the web. 15th International Conference on World Wide Web, Edinburgh, Scotland (2006)
20. Koh, S.J., Chang, M.J., Lee, M.: mSCTP for soft handover in transport layer. IEEE Communications Letters, vol. 8, no. 3 (2004)
21. Gradischnig, K.D., Tüxen, M.: Signalling transport over IP-based networks using IETF standards. Third International Workshop on the Design of Reliable Communication Networks, DRCN'01, Budapest, Hungary (2001)
22. Lee, H., Lee, B.: A redundancy method of AS traffic in signalling gateway using load sharing scheme. 6th International Conference on Advanced Communication Technology, Phoenix Park, Korea (2004)
23. Dreibholz, T., Rathgeb, E.P.: RSerPool - Providing Highly Available Services using Unreliable Servers. IEEE EuroMicro 2005, Porto, Portugal (2005)