S38.3115 Signaling Protocols – Lecture Notes

Lecture 6 – ISDN User Part in SS7

Introduction

ISUP is a national and international telephony network signaling protocol for call setup, supervision and release between exchanges in wire line and mobile networks. It supports a rather wide range of supplementary services for ISDN networks. Support of supplementary services achieved a sufficient level in 1992 and ISUP deployment in Finland started in 1994. ISUP mainly runs on top of MTP of SS7.

The first telephony signaling protocol for SS7 was the Telephony User Part or TUP. It was designed before DSS1. It achieved fast call setup because of the use of SS7 signaling infrastructure but supported only a limited set of supplementary services. National variants of TUP exist for many countries due to the need to take into account interworking requirements with different national analogue signaling requirements. For example in the UK, the user part corresponding to TUP is called the National User Part or NUP, in Finland it is just called national TUP. The message coding style in TUP slightly differs from DSS1 in favor of using a lesser amount of bits. TUP was difficult to extend while preserving backward compatibility.

Operators and exchange vendors were looking for a further step towards more global markets based on a wide support of supplementary services and automation of compatibility establishment features. Therefore, instead of trying to extend TUP, a decision was made to specify a new user part better aligned in terms of services with DSS1 and also taking into account better support for backwards compatibility.

Compatibility of two versions or variants of a protocol means that two instances conforming to the different versions of the specification can talk to each other and can agree on the set of services support by both versions. This eases the deployment of the new versions because deployment can progress node by node and two ends of a signaling link can run using different versions for a time. This is important because of the high availability requirements placed on exchanges and the high cost of software upgrades. A major software upgrade in an exchange may take a whole working day for a team of engineers and is often executed during a night to reduce disturbance to live traffic. Besides the cost of the software itself, one needs to consider also the operational costs incurred.

ISUP deployment was rather easy because the underlying infra of MTP was already in place and had been in use with TUP.

Most recent developments in ISUP after the year 2000 include the addition of charging messages suitable for European countries although for example in the Finnish version charging messages were present even earlier. They are useful when operators change their tariffs frequently and use special campaign prices. With the help of charging messages other operators do not need to change their tariffing configuration when another operator makes a tariff change and still subscribers of all operators can benefit from the campaign prices.

Another development has been the specification of how to carry ISUP over IP -networks. This is called the SIGTRAN architecture and we will discuss it on the last lecture.

Services support in ISUP

TUP did not support the transfer of compatibility information nor the transfer of User to User Information available in DSS1. The latter also supports a procedure of service mobility or like it is called in ISDN: suspending a call in one extension and resuming it in another. For smooth operation over a wide area, this requires transferring the Suspend and Resume messages of DSS1 across the trunk signaling network and further to the other party in the call. Such a procedure was not supported in TUP.

Operators and vendors widely believed that support for a broad set of standard supplementary services was necessary for getting the ISDN phone market off the ground which did not seem to happen although implementable DSS1 specifications were produced in 1988. A set of supplementary services specifications called Euro-ISDN was produced in 1992. TUP naturally did not support it but ISUP did. We will describe some of the Euro-ISDN services in a later section.

In addition to bearer services implementable over 64 kbit/s timeslots such as speech, audio at 3.1 kHz, audio at 7 kHz and unrestricted 64 kbit/s (data) channel, ISUP also supports multi-slot services (2 × 64 kbit/s, 384 kbit/s, 1536 kbit/s and 1920 kbit/s all in the unrestricted mode). These services were deemed necessary for video conferencing and data connections. In practice their provision over PCM infra is however cumbersome because transmission resources in practice need to be dedicated to either 1 × 64 kbit/s services or to a multi-slot service at a certain speed. When service penetration is low, this is costly. The penetration stayed low for example because telephone operators did not know better than charge N times the cost of one slot for N time-slots. Clearly, this was unattractive to users when Internet with flat rate charging came along very quickly on the heels of ISUP deployment.

In terms of services support DSS1 and ISUP form the uniform services offering of *Narrow band ISDN*. If a service is supported in DSS1 and needs some signaling over wide area, that support is either present or can be added in ISUP.

Identification in ISUP

Subscribers (A – calling party, B called party and C if needed in a supplementary service) are naturally identified with telephone numbers. ISUP assumes that these numbers are routable in exchanges. If Number Portability is supported in the network, directory numbers need to be translated into routing numbers prior to ISUP signaling.

An exchange processes many simultaneous calls. Identification among simultaneous calls between two exchanges is based on Circuit Identification Code (CIC) and optionally on Call References. A CIC is a 12 bit field just after the routing label that we showed when discussing MTP. The value of CIC is *an integer* and each exchange needs to maintain a configuration table that maps particular PCM lines and timeslots to CICs and vice versa. The combination of routing information for an ISUP message is depicted in Figure 6.1.

The use of mandatory CIC for call identification ties the ISUP state machine for the call to the time-slot that is used for the call. This means that there can be no call signaling unless a voice circuit is reserved.

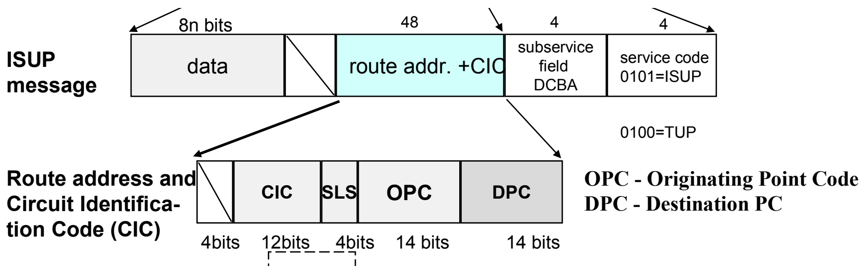

Figure 6.1 Message routing information in ISUP and MTP

One can claim that this is an architectural limitation. On the other hand ISUP offers the use of arbitrary call references similar to what is used in DSS1. These messages are for national use only. The use of CIC for multi-slot services is possible because only consecutive time-slots are used for them due to data transfer integrity reasons. For call identification the smallest CIC for the time-slots is used.

We will come back to this architectural issue when discussing IN.

ISUP message encoding

ISUP messages are binary encoded similar to DSS1 (but not the same, so once again a protocol specific message parser is needed). A generic way to define binary encoding would be to use ASN1 from the OSI protocol stack. Neither ISUP nor DSS1 use ASN1. They both have their own coding style. This means that message parsing is integrated in the signaling protocol implementation. Each message has a *type field* that can be found in a fixed place in the message. After the type field follow the *mandatory information elements* for the particular message type. Information element may have a fixed or variable length. Last in the message follow *optional information elements*.

An information element is identified by its *type field*. A variable length information element will of course have a *length field*. Each element may carry one or several *information items*. Each information item may be stored in a program variable by the message parser.

Message type gives a pretty good idea of the state change that follows from the reception of that message.

Due to binary encoding, ISUP messages are quite compact in size and nicely fit into 64kbit/s signaling channels.

Successful call setup

A successful call setup in ISUP (see Figure 6.2.) starts with the Initial Address Message (IAM). IAM is a message type. In case of overlap sending, the caller side exchange may send the reminder of digits in Subsequent Address Messages (SAM) towards the called party. If en-block sending is used, all digits are sent in the IAM. Voice path is built hop-by-hop on IAM in the backwards direction. In the forward direction the voice path must be broken at least in the originating exchange so that no bidirectional communication can take place prior to initiating charging. In practice SAM is not needed in mobile originated calls because the user of a handheld phone first has to select the number and only after that push the call button. If the call originates from an analogue PSTN phone and the dialing plan assumes variable length phone numbers, SAMs are useful. They speed up the call setup. SAMs are of limited use, if the PSTN dialing plan is such that for each area code, the number of digits in a subscriber number is fixed. In that case, the originating exchange may as well be programmed to receive the required number of digits and send the IAM only when all digits have been received. Under Number Portability, the dialed digits are translated in an IN query issued by the originating exchange to an IN node. Consequently, in case of NP, SAMs in ISUP are not needed because ISUP signaling starts only after the IN query that has resulted in a fully routable number of the called subscriber.

When the terminating exchange has received enough digits for identifying the callee extension, it will send the Address Complete Message in the backward direction. After the ACM, the network will not receive any more digits.

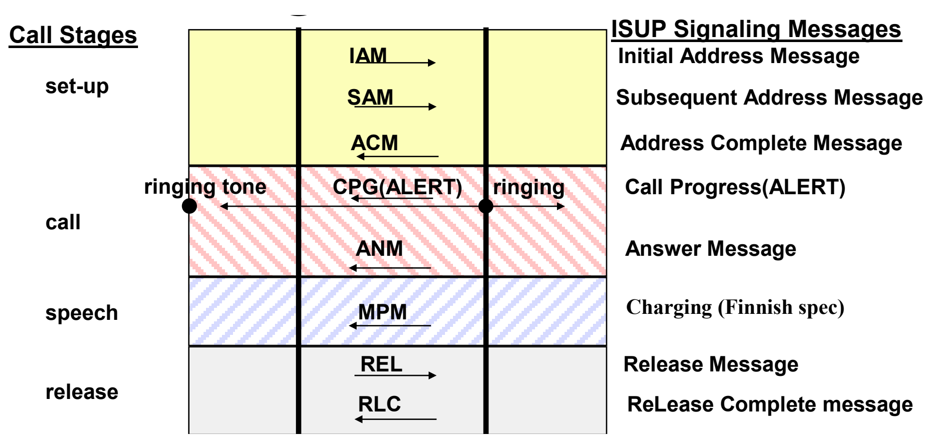

Figure 6.2. Successful ISUP call setup.

When the terminating exchange finds out that the callee is free and sends the ringing signal (a voltage or current that will ring the phone's bell and/or possibly some signaling message: Setup….alerting sequence in DSS1) to the callee, it will also report the new state backwards using ISUP Call Progress Message (CPG) with an Alerting indicator. This message is accompanied by the ringing tone in the backwards direction on the voice path.

When the callee picks up the phone, an Answer message (ANM) is sent in the backwards direction. ANM initiates call charging in the *charging points* usually including the originating exchange. Charging points are ISUP nodes that generate or collect charging information. On reception of ANM in all exchanges on the way, bi-directional through-connection state is established for the voice path.

For release, ISUP uses two messages: Release Message (REL) starts the release procedure, in terms of call control in all exchanges, after REL, incoming and outgoing sides are independent. On REL, the through-connection is torn down and charging is stopped. The Release Complete Message acknowledges that the next exchange has indeed received the REL and that all resources that were reserved for the call can be released.

For the purpose of establishing the call tariff dynamically on a call by call basis, ISUP supports *charging messages*. Figure 6.2 shows the charging messages in the Finnish ISUP. Later in this lecture we will show the messages in the ETSI ISUP. The idea is that each operator may set its tariffs independently and that users should reap the benefits of lower campaign or regular tariffs immediately without other operators having to alter their own charging by a management command. If no charging messages are exchanged on a call by call basis, the result is that the originating exchange has to calculate the tariff based on its own configuration information that can not take into account the tariff changes that may have been made by the transit or terminating operator. This would be detrimental to competition between operators.

PCM management operations in ISUP

It is the responsibility of an exchange to supervise the technical condition of the PCMs that it is connected to. An exchange monitors the time-slot zero and maintains frame synchronization on the PCM. If frame synchronization is lost or if an alarm is received on the PCM, the exchange should take the PCM line out of use and not offer any new calls to that PCM. The exchange may also need to release existing calls upon the alarm. An exchange may need to release calls from a particular PCM on operator command. The exchange may need to tell the neighbor (on the other end to the PCM) using the out-of-band SS7 signaling to block new calls to a particular (possibly failed) PCM.

Behind all those requirements lays the requirement that users do not like *silent calls* or to be mislead during the call setup or the call itself. In addition, it must be possible to make rearrangements in the PCM plant in live exchanges.

To meet the needs of managing PCMs ISUP has several messages. Blocking (BLO) tells the neighbor not to place calls onto the CIC. The neighbor acknowledges the command by send Blocking Acknowledgement (BLA). The blocking state ends on reception of the Unblocking Message (UBL). Unblocking is acknowledged by Unblocking Acknowledgement (UBL).

Other messages for management purposes are present in ISUP as well.

Forward compatibility rules and procedures

Upgrading live exchanges, while keeping downtime to minimum, is costly. Upgrading a signaling system in such a way that the upgrade would need to be synchronized in neighbor exchanges is cumbersome, error-prone and even more costly.

To help with upgrading procedures, from 1992 onwards ISUP has conformed to forward compatibility rules and procedures. The idea is that two versions of ISUP should be able to talk to each other and agree on the minimal level of functionality supported by both.

To achieve the goal, a message in a new version may carry different indicators referring to new information elements that were not present in the older version. These indicators may tell the receiving ISUP to

- Pass on the information element without processing it
- Release the call if the element is unknown
- Notify the sender if the element is unknown
- Discard the message if the information element is unknown
- etc

In addition, the types of changes that are possible are limited. The following should not be changed in ISUP:

- Protocol procedures, messages, information elements, coding, except to correct an error in the protocol.
- Semantics of existing info elements.
- Formatting and coding rules.
- Adding new parameters into mandatory part of Messages, Optional part can be extended.
- Order of information items in an Information Element of variable length, new items can be added to the end of the IE.

It is also agreed that

- Information item value =“all 0” == non-significant value.

Examples of Supplementary Services Support in ISUP

Calling Line Identification Presentation and Presentation Restriction (CLIP/CLIR)

CLIP requires that ISUP transfers the directory number of the Caller (Calling Party Number) to the terminating exchange. This number is carried in the IAM message along with the indicators whether this number actually can be presented. If the A-party’s number is a secret number or the user has forbidden the presentation of her number, the callee will not see the caller’s number without special priviledges (these may be given for example to the Police).

CLIR allows restricting the presentation of caller’s number and overriding the restrictions given appropriate rights by the law.

These services have become commonplace with mobile services. They are also supported in ISDN networks.

Connected Line Identification Presentation (COLP) and Connected Line Identification Presentation Restriction (COLR)

The idea of the service is to present to the caller the number she or he actually is connected to.

This pair of services is a mirror image of CLIP/CLIR. COLP/COLR are needed because due to number translations, call forwarding, call transfer and possibly other services, it may be that a caller is connected to a quite different number than what she/he dialed.

The restriction service once again is required due to privacy reasons. In addition to the number itself, the service also gives an indication whether the number is assured by the network or provided by the user.

Due to the combination of CLIP/CLIR and COLP/COLR ISDN networks support the idea of *non-repudiation of communication*. This is useful for conducting business using the network.

User to user signaling (and billing)

Both DSS1 and ISUP support the supplementary service of *user to user signaling*. UUS can also be viewed as a bearer service. The idea is to let user’s send a small amount of data through the signaling channel during call setup or during the call itself.

The introduction of UUS was painful because the billing systems did not initially support the service and the operators had a hard time making up their minds how they wanted to treat the service in terms of billing. Traditional operator thinking was that the users should not be allowed to communicate at all unless billing is started. Also, the billing was expected to reflect network resource usage by the user.

When ISDN introduction started, billing was based on counters of call minutes for local, transit and international calls for each subscriber. In many countries local calls in PSTN were traditionally free (e.g. US, Russia). Billing based on Call Detail Records with the accuracy of seconds was first widely introduced in mobile networks.

Terminal Portability or Suspend/Resume

We mentioned this service in DSS1 as one of the motivations of introducing ISUP to replace TUP. Indeed, a user can port the call from one terminal to another. The Suspend message is sent to initiate the service and the Resume message sent by the new terminal will restore the service. These messages help to synchronize the applications in the user devices.

Call Forwarding

Call Forwarding No-Reply (CNFR)
The idea is that when the callee does not answer within a timeout (15…30s), the call is forwarded by the terminating exchange into a new number (C-number). A subscriber has procedures (dialing sequences) for activating and deactivating the service and modifying the C-number.

The service is used for example to forward the incoming call to a voice mail in case of no-answer.

Call Forwarding Unconditional
Immediately, when a new call arrives at a number, the call is forwarded to a new number (C-number). A subscriber has procedures for activating and deactivating the service and modifying the C-number.

This service is useful during longer periods of absence of the called party or when the called party is out of coverage or otherwise non-reachable by the network.

Call Forwarding on Busy
If an arriving call hits a busy condition of the callee, the call is forwarded to a new number (C-number). Once again, a good use case is forwarding an incoming call to a voice mail when the callee is busy.

NB: Voice mail services provided by the network were widely first deployed in mobile networks. The most common use cases of different call forwarding services are connected to Voice Mail. Probably this development is connected to the fact that mobile service is personal while calling to a fixed number is often about calling a place in which any number of people might answer the phone.

Call Transfer

Call transfer is a service where the callee answers the call and then after the call has been established (and having probably talked to the caller) transfers the call to a new number (C-number).

This service was initially used by call attendants who transferred incoming calls to people in a company but the service was later provided to any subscribers.

Categorization of Supplementary Services

Supplementary services can be classified into A-subscriber services (short number service, CLIR etc) and B-subscriber services (CFNR etc) based on where the service data resides and which of the users have control over the service.

One should take note that not all supplementary services require support from trunk signaling. Some services may be implemented in the terminal or alternatively in the network (e.g. short numbers). Many services need support from the access signaling (e.g. DSS1) but can be completely implemented in the local exchange.

Featurism

It is a typical development in communications technology (and more widely in ICT) that systems tend to grow in complexity over time. New and new requirements are assigned and implemented in systems that have achieved some level of penetration or become popular. When a system is sold and the buyer has to answer the question: "is it better to have this added feature or not to have it", the answer almost invariably is "yes, of course it is better have it". This applies even when the person answering the question does not know anything about the feature or has made any analysis of the usefulness of the feature in terms of business. This person, although a decision maker, typically has no idea about the cost of implementing these features either.

The result is that systems become gradually more and more complex. We call this phenomenon "featurism".

A good example of featurism is the long lists of supplementary services that have been specified and implemented in public network switching systems and PABXs. There are hundreds of them. Only a few are really useful. In our presentation we tried to pick out examples from the most common and useful ones.

End to end view or Interworking of signaling systems

So far we have left it open how does an ISUP implementation decide to send a particular message. In a network, access and trunk signaling are interlinked with each other through the process of call control.

The implementation of call control is left for the vendors. However, ITU-T, ETSI and other specifications describe the semantics of linking two signaling systems together.

Interworking of signaling systems is mapping of events and information between two signaling systems.

We also talk about signaling interworking when two exchanges use the same signaling system to carry calls but the exchanges are implemented by different vendors. This last concept comes very close to the notion of compatibility we talked about earlier. Interworking of signaling systems and compatibility are quite different concepts.

The challenge of interworking specification lies in the fact that if we have N signaling systems, there are N^2 signaling interworking cases. For ensuring correct operation of services, all these cases need to be specified and tested.

Another challenge related to interworking is that different signaling systems support different information elements and different messages. Therefore, it often happens that some information is lost in the interworking node.

Legacy method of interworking specification

For analogue signaling system interworking (e.g. analogue subscriber signaling and R2) ITU-T (or actually its predecessor CCITT) used a method based on *telephone events*. These were abstractions of actual events taking place in exchanges and the signaling systems concerned. Through abstraction some level of generalization was achieved so that some work could be reused for several signaling systems. Telephony events were of three different types:

- FITE – Forward Interworking Telephone Event (e.g. a digit)
- BITE – Backward Interworking Telephone Event (e.g. an ack on receiving a digit)
- SPITE – Switching Processing Interface Telephone Event (e.g. a bidirectional through-connection)

In the telephone event method, the interworking specification will give the l*ogical procedures for incoming signaling*, the *interworking logic* and the *logical procedures for outgoing signaling*. All these are specified in SDL and the SDLs use the lists of forward and backward telephone events in sent and received signals and switching processing events in tasks. The incoming and outgoing logical procedures pick out enough detail of the concerned signaling systems to nail down the correct semantics of signaling interworking. The interworking procedures show how the incoming and outgoing signaling are interlinked and how the switching system itself changes state on reception of incoming signals.

When layered signaling systems, more particularly DSS1, were introduced, the method based on telephone events was seen as too detailed. Operators wanted to leave more of the implementation to vendors. A new method for specifying interworking of signaling systems was adopted. We call this the layer based method.

Layer method of interworking specification

In layer based interworking specifications the model is that an incoming signaling system is connected to an outgoing signaling system through *call control* (see Fig. 6.3.). Incoming signals trigger *primitives* that go from the incoming signaling system to call control. These are *indications* if the incoming message was a "new request" or *confirmations* if the incoming message is seen as an acknowledgement to an earlier message. Call control will issue *Request primitives* towards the outgoing signaling system related to earlier events on the incoming side. Based on an event that is seen as an acknowledgement in the backward direction, call control will issue a *response primitive* towards the incoming signaling system.

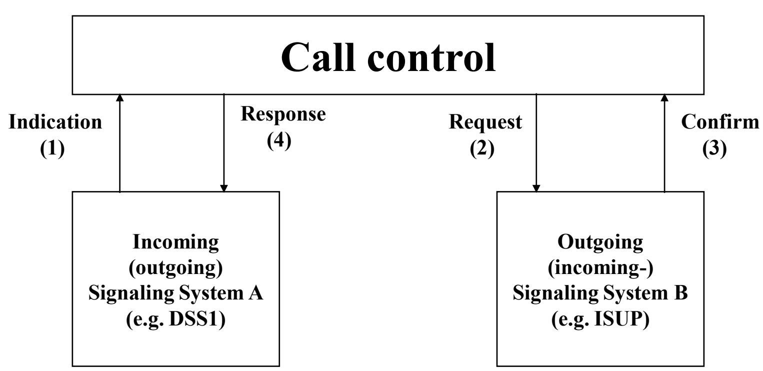

Figure 6.3. Interworking specification using the layer model

Using the layer method, interworking of for example DSS1 and ISUP has been specified using signaling *message flow charts* between DSS1, Call Control and ISUP for a significant number of signaling scenarios. In addition information mapping tables are provided. The tables give details on how information from one system is mapped into information elements in the other.

Example: ISUP and DSS1

Figure 6.4 shows the interworking of DSS1 and ISUP in case of successful call setup on a course level.

The flow chart in Figure 6.4 shows the mapping of messages in DSS1 and ISUP.

Such flow charts are needed for B-subscriber busy, no-answer, call forwarding etc.

In a *detailed interworking flow chart* the interworking specification shows the *actions* taken by call control such as: reservation of outgoing timeslot, forward through-connection, backward through-connection, bi-directional through-connection and tear-down of through-connection.

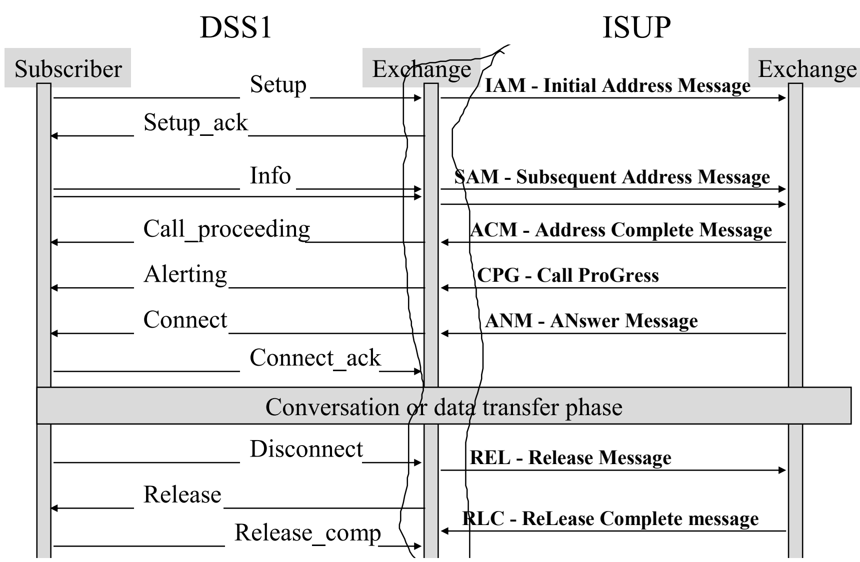

Figure 6.4: Interworking of DSS1 and ISUP for Successful call setup

Charging support in ISUP

One of the latest additions in ISUP was the ETSI specification for charging support. This is depicted in Figure 6.5.

When a Charging Determination point wishes to apply a tariff to a call, it will send on a call-by-call basis an Advice of Charge message (CHG) in the backward direction. The charging registration and generation point, usually the originating exchange, will collect the CHG messages and compile the total tariff based on them. It will also acknowledge CHG by sending CHGA, a Charging Acknowledgement message. There may be several Charging determination points involved in a single call (probably one in each network).

Finally, when all charging messages have been acknowledged, ACM will complete call setup and charging determination.

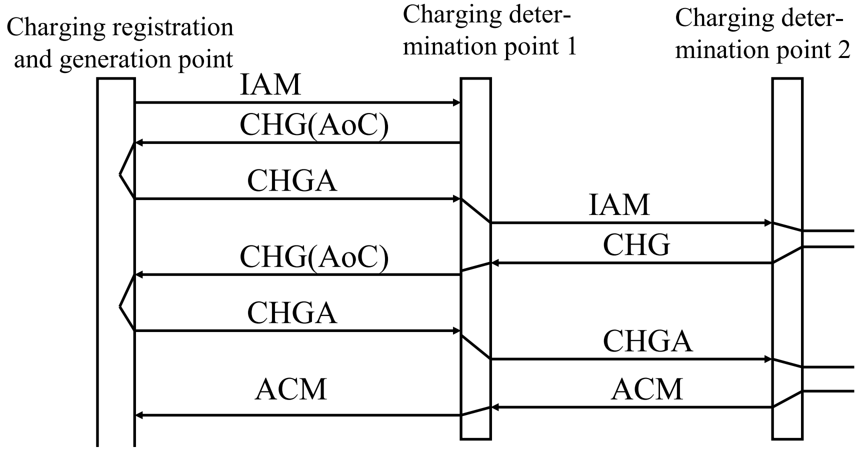

Figure 6.5: Charging support in ISUP.

Discussion why DSS1 and ISUP are two different protocols

Most recently, people who specify signaling for IP networks have argued that the same signaling protocol should be used across the whole network and that there are no good reasons for making a distinction between access and network-to-network signaling. This may be true for broadband networks but let us consider this notion in its historical context.

The reasons that favored the specification of two different protocols for access and trunk signaling purposes in ISDN can be listed:

- ISDN/PSTN network stakeholder model contains subscribers and many network operators. Charging on a network-to-network interface differs from charging on subscriber to network interface. On the network-to-network interface, we may ignore who made the calls. It may suffice to count call minutes or seconds in both directions.
- The architecture assumes 64kbit/s timeslot infra. When used for signaling, 64 kbit/s can carry signaling for more than 1000 voice time slots. It is wasteful to use 64kbit/s for a single subscriber or even allocate one timeslot on each PCM connection since each signaling channel requires its own signaling terminals in exchanges. Users do not like silent calls. Therefore, exchanges supervise the condition of PCMs. Non-functional PCMs must be taken out of use, tested and repaired, calls may need to be released on failed connections and never offered to a failed connection. Also, it must be easy to add new PCMs into the system and easily take them into use. All this requires managing states of PCMs by signaling (BLO etc). On a subscriber line, the signaling (DSS1) is carried on the same physical connection as user

data/voice. So, in this case monitoring of the connection is simpler than monitoring of PCMs.

- If one signaling channel is responsible for >1000 calls it must be more reliable than a channel that is responsible just for < 30 simultaneous calls.
- Some supplementary services are implemented in a local exchange needing support from subscriber signaling but not from an exchange to exchange signaling (e.g. call forwarding, call transfer, speed dialing, …)
- Identity assurance is needed at the subscriber interface but not on a network-to-network interface.
- When ISDN was designed optimizing performance, resource usage and complexity were important!

Through experience we have learned a lot about protocol specification. It is true that in hindsight it would probably be possible to meet all the above requirements with a design of a base protocol plus extensions architecture that has become popular in protocol design lately. However, when ISDN was designed, implementation was done by telecom vendors into their vertically integrated exchanges and other products. This basic assumption stressed the importance of reliability and efficiency requirements but left the problems of *flexibility of design* for vendors to solve.