S38.3115 Signaling Protocols – Lecture Notes

Lecture 8 – GSM and IN Architecture – a common component: TCAP

Introduction

The purpose of this lecture is to discuss the design choices and commonalities in two major developments in networking technology, namely the GSM system and Intelligent Networks. IN was initially designed in the US while GSM was specified by ETSI in Europe. Work on the architectures started in the 1980's and continued for many years. Many of the GSM specifications were inherited by 3G and 4G mobile networks, so the work is still continuing.

By looking at the commonalities we conclude that a common protocol component was needed for both IN and GSM. This is the Transactions Capabilities Application Part in the SS7 system. These two examples, GSM and IN also demonstrate that call signaling is the infrastructure on top of which services are implemented in communication networks.

The motivation of IN was that operators were solely dependent on switching systems vendors in their services. Also, the penetration of services varies greatly. The basic call is used by everyone while many supplementary services are used only by a few users. Irrespective of this, an exchange centric services implementation model requires that software for all services is available at least in all local exchanges. This is expensive and leads to long lead times of deploying new services.

Also, even if an operator used switching systems from two or more vendors, it was difficult to make available the union of services available in the switching systems by different vendors. Rather what easily happened was that the operator could make available to the users only the intersection of the set of supplementary services that was available in the switching systems from the different vendors. The benefits of a particular switching system by a vendor might be related to, for example, support of some business customer related signaling and services. Nevertheless, as the operators saw that deregulation will increase competition, they wished to find new ways of differentiating on the market.

IN specification started in the US in the 1980's and later moved to ITU-T and to ETSI when IN was applied to extend the basic capabilities of GSM. The idea of IN was to create a platform for implementing services in telephony networks such that software and computer companies could be attracted to become telecoms equipment vendors and compete with the traditional switching system vendors. Also the idea was to support easy and fast creation of new services, operator programmability of services and implementation of new services in such a way that software changes in exchanges could be avoided or at least concentrated in a single or a few network nodes. To make this happen IN concentrates service logic in a function that is called *Service Control Function (SCF)*. Usually this function is mapped to a network node implemented on a general-purpose computer system (e.g. a cluster).

For specification of IN, the standardization organizations used ***a functional method of specification***. In this method function and information flows between functions are specified. It is up to the operators to decide how the functions are mapped to equipment such as network nodes or "points of different type". For example the Service Control Function, when mapped to a node, typically becomes a *Service Control Point (SCP)*.

GSM MOU (Memorandum of Understanding) was signed in 1987. The co-signers made a commitment to develop and deploy a digital mobile service that would support *roaming of users* across networks operated by different operators in different countries. The first GSM network in the world went live in 1991 in Finland. The ideas of IN were known to GSM designers and IN designers knew about the preceding first generation, analogue mobile systems. Instead of a Service Control Point, GSM has a Home Location Register or HLR that is the centralized node for keeping track of the whereabouts of the GSM subscribers.

HLR can be seen as the SCP for mobility management in a "conceptual IN architecture". GSM specification follows this idea on a conceptual level but does not follow it in detailed implementation. For example different protocols are used: IN Application Part in IN and Mobile Application Part (MAP) in GSM. The specification of GSM was driven by the liberalized mobile services

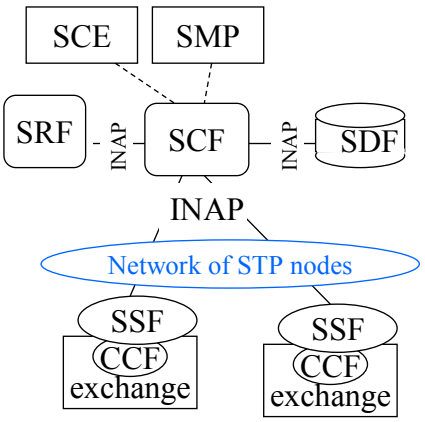
market and network operators and vendors (such as Nokia) that were looking for new business opportunities. On the other hand, many large incumbent (wire line) operators either did not have a license for GSM or did not believe that the service would pick up and reach a wide penetration[1]. At the same time they were driving the specification work on IN.

At some point there was a half hearted attempt by wire line operators to interfere with the business of mobile operators by suggesting that INAP should be developed further to support mobility.

IN Architecture

IN is a way of implementing services independent of the underlying exchanges in the network. Examples of services are *free phone*, *premium rate calling*, a *virtual private network* service for a corporation etc. Figure 8.1 illustrates the architecture.

In order to deploy IN in a PSTN or ISDN network, the exchanges or at least some of the exchanges have to be upgraded to support IN. The call control system in the exchanges is upgraded to support *service switching functions* (SSF) for IN services triggering. The triggering is a way of matching calls to pre-set criteria and asking guidance from a centralized server when the criteria match. The IN specifications also describe how the SSF functions relate to call control, so they mention the call control function that also resides in the exchanges.



INAP = IN Application Part
= main protocol
CCF – Call Control Function
SSF - Service Switching Function
maintains call state with CCF
SCF - Service Control Function
implements service logic
SRF - Special Resource Function
processes in-band signals
SDF - Service Data Function
is a database
SCE - Service Creation Environment
for creating new service logic
SMP - Service Management Point
implements mgt functions

Figure 8.1: IN Architecture

The idea of *IN service triggering* is that the exchange stores a number of conditions or filters that need to be checked on call or signaling message arrival. When there is a match to the conditions, the trigger fires and the SSF

[1] Actually, not even the developers of GSM believed during the 1980's or early 1990's that mobile service would ever overtake wire line penetration.

function will form a request to the Call Control Function (SCF) identified for that trigger using the INAP protocol. The SSF puts the call state on hold and waits for further instructions from the SCF.

SCF contains service logic programs. For service processing it may make use of service data that is stored in the *Service Data Function* (SDF). This is a database.

It is possible that for correct operation the service requires announcements to be played to the user and possibly even further data collection from the user. Often in-band DTMF signaling is used because all phones (including modern mobile phones) support it. The service may for example ask the user to select among different alternatives and push a button on the phone indicating the selection. To implement the playing of announcements and data collection from in-band signaling, a *Service Resource Function* (terms like intelligent terminal etc are also used) is required. It has an interface to voice circuits and it has an INAP interface for receiving instructions from the SCF and reporting its responses to the service logic.

In addition, IN has a *Service Creation Environment* for the purpose of programming logic for new services. IN also may have a *Service Management Function* for the purpose of managing the services that are created by SCE or work within the IN framework.

The advantage of IN compared to exchange based services is that services data can be maintained in a centralized place instead of the need to configure the data into numerous exchanges that are scattered around a country.

Take for example the simple service of *free phone*. This service is very popular for example in the US. The idea is that a company that wishes to attract calls from customers provides advice to customers without cost to the potential customers. So, not the caller, but rather the callee has to pay for the call. Without IN this service would require either configuring the free phone number in all local exchanges or signaling support from the terminating exchange of the callee to the originating exchange. At least the originating exchange would need to understand the signaling. The transit exchanges would need to pass the additional signaling transparently. This is quite cumbersome if similar changes are required for each new service.

When free phone is implemented using IN, let us assume that the LE provides the SSF function. In the SSF the call is triggered based on the free phone number to the SCF. This may be based just on the free phone “area code”. The SCF may process the call differently during office hours and at night. The calls may be processed in several call centers while the SCF may be responsible for distributing the call to a particular call center. The SCF will instruct the LE not

to charge the caller. Instead, it may create a CDR itself assigning the cost to the callee.

What is typical of IN is that during service processing the SSF needs to maintain call state. To make this possible a *Basic Call State Model* (BCSM) is specified in IN. The purpose of BCSM is to de-couple the state of the call from the state of the services logic. The weakness of this idea lies in the fact that a telephone call is also a service. It is difficult to define a call model once and for all possible future services.

Actually, we may ask: why do we not send ISUP messages directly to the SCF? Why is a different protocol needed for services? First, let us recall that in ISUP the call session is identified based on the Circuit Identification Code. I.e. a voice circuit would need to be allocated between the SSF and the SCF in order to use ISUP! This obviously does not make any sense. Also managing the circuits that are required for the call is a burden that needs to be taken care of even if signaling itself would be circuit independent.

It would be possible to map ISUP or subscriber signaling to DSS1 with its call references architecture and run the DSS1 to a an SCF but then one would need to carry it across the SS7 MTP/SCCP network which would not be based on standards. In such a case for the purpose of controlling SRF and accessing the database different protocols would need to be used. To make this work, some modifications to DSS1 would be needed. NB: This has not been done. But we will later see that in IP telephony, instead of a separate protocol, one protocol, namely SIP, is used for access and trunk signaling and for accessing intelligent services logic in service nodes.

If you will take a look the INAP specification you will immediately notice that the protocol does not have a modular structure. It is just a collection of messages for all kinds of purposes that people have invented on the way. So, although IN allows introducing new supplementary services and even operator programming of new services flexibly, IN does not apply the good idea of modularity in protocol design.

GSM Architecture

Figure 8.2 depicts the original GSM architecture (so called version 2) that supported circuit switched services only.

Mobile Stations (MS) (handheld phones) have a radio connection with a cell. One or a set of cells is supported by a Base Transceiver Station (BTS). A BTS may have an antenna high above the ground on top of a tower in order to increase the coverage of the cells. Cells cover the area where GSM users are reachable. Cells may be organized into several layers. Large cells are useful for fast moving Mobile Stations and small cells are needed to increase network

capacity – the number of MSs that can be served simultaneously in a particular spot. A call may start in one cell, the MS may traverse through a number of cells while the call is on and the MS may be located in a cell under a different BTS or even a different MSC when the call ends. The action of changing a cell during a call is called a *handover*. Several BTSs are controlled by a Base Station Controller (BSC). There may be many BSCs under the control of one Mobile Switching Center (MSC). An MSC controlling BSCs is in the same position in GSM as a Local Exchange in PSTN or ISDN. The difference is that an MSC does not "own" its subscribers. Rather, all the MSs it is controlling are visitors. A Visitor Location Register (VLR) for storing information about the users and MSs visiting this MSC is separately specified but always resides in the MSC. Recall that also a wire line Local Exchange contains a subscriber database. A VLR contains service and subscription information but also location information and information for mobility management.

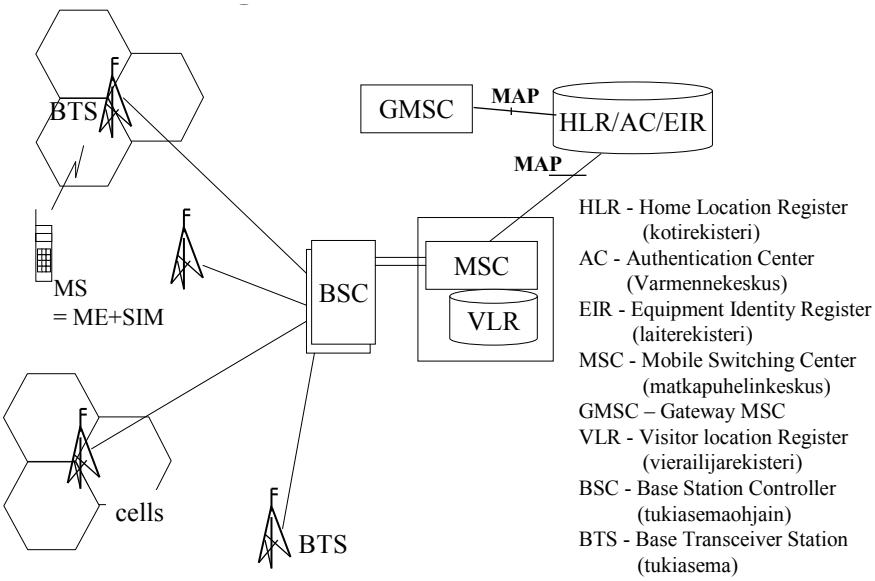


Figure 8.2: Original GSM Architecture

Cells normally forming a continuous geographical area are organized into a *Location Area*. A location area is a logical concept and gives the accuracy with which the location of MSs is maintained in the HLR.

Each MS has a MSISDN number. It is a logical or a directory number. An MSISDN number maps to a particular HLR based on its leading digits. Thus MSISDN is routable to the HLR that is supposed to know where the MS is located. MSISDN is not routable to the subscriber's current location.

The HLR also houses the Authentication Center (AC) for storing authentication information about the users. The HLR has the Equipment Identity Register (EIR). Each MS has an Equipment Identity on silicon and thus stolen equipment can be identified, traced and removed from the network.

One should note that a meaningful ***commercial mobile service is not feasible without authentication***. If the users were not authenticated, anyone could fake to be anyone else. As a result, not even flat rate charging for the service would be enforceable. Businesswise the consequences are significant: mobile operators cannot survive unless they know their users. Knowledge of users is power on the market. This is one of the reasons of profitability of mobile services in the long run.

NB: Compare this to fixed telephony. In fixed circuit switched telephony the authentication of the caller is based on the physical local loop from which the call originates. Since, physical access to the local loop is rather well protected, authentication is sufficiently secure for charging based on usage of the network service (usually time based charging in telephone networks). To generalize back to authentication: *usage based billing and mobile service are the only 2 good reasons to provide authentication of subscribers by the operators.* The logic is that authentication does not come for free – investment to the authentication infrastructure and the operational expenses must be paid for, i.e. there must be a reasonable business case for authentication.

MSs are reachable because they regularly make location updates to the VLR. The VLR will authenticate the user and establish that the user can be billed for the use of the network resources by contacting the HLR. When the MS first establishes a connection with the network or changes the Location Area, VLR will update the location of the MS in the user's HLR.

For call setup the GSM system uses ISUP between exchanges and BSSAP as the access signaling between BSS and the Circuit Core (in practice the visited MSC). Due to mobility, additional signaling functionality is required in the Core network. This is mainly provided by the Mobile Application Part (MAP). Making a call from an MS (Mobile Originated call) is quite similar to the wire line case. Terminating a call to a Mobile (Mobile Terminated call) is different.

When an exchange sees that it has received a mobile number for the callee, it will route the call to a gateway MSC. The GMSC will send a request using the SS7 signaling network and the Mobile Application Part signaling to the HLR. The HLR will return the *Mobile Station Routing Number* (MSRN) to the GMSC. The MSRN like its name tells is routable in all switching systems i.e. based on leading digits only. It has earlier been dynamically assigned by the VLR for the call or for the duration of the visit of the subscriber and given to the HLR to be used for terminating calls.

When a call arrives to a visited MSC, the MSC finds out the location of the MS in the VLR with the accuracy of several cells. The MSC will *page* the MS in all those cells. Paging means that the MSC sends a call signal to the mobile using signaling channels over the air in several cells simultaneously. The MS will respond using a signaling channel and the cell that it sees best.

There is one more significant difference in call establishment in GSM as compared to wire line networks. Radio resources are seen as very expensive and should be preserved as much as possible. Therefore, when an MS makes a call, no radio resources for user plane data or for voice are allocated at the originating cell until it is known that the callee is not busy, not out of coverage and that the callee is also willing to take the call. So, the reservation of timeslots on the air interface of the originating side takes place later in the call establishment process.

From the very beginning GSM supported *roaming*. This means that subscribers of one operator can move into the area of another and make the use of the other operator's network. Usually, only *international roaming* is supported for business reasons. I.e. the two operators that allow their subscribers to use each other's networks cover different countries. International roaming has been a major benefit of GSM to the users and has helped to consolidate the positions of the GSM operators. Most times, in certain matters, they act as a family to pursue their common interests. Roaming increases the value of all GSM networks (Metcalfe's law).

National roaming would be technically the same as international roaming but it is not typically supported for the reasons of creating a competitive market. One competitive advantage an operator has is the quality of its coverage of the country. If one allows competitor's customers into one's network, the advantage may be lost leading to disinterest to invest into coverage. For this reason, national regulators and operators have most times been against national roaming.

Note however, that in many established markets there are different types of mobile virtual operators (MVO) that do not have their own radio access network and may be not even exchanges. In this case the business relationship between the network operator and the MVO is asymmetrical. In case of national roaming, we would be talking about a symmetric arrangement between two or more network operators.

GSM version 2+

In late 1990's a major upgrade to GSM was introduced. Mobile operators were predicting that packet switched services would add a significant new revenue stream to their business. Packet switched services bring a new subsystem to GSM. This is called the Packet Core network. The Packet Core network resides in parallel with the Circuit Switched Core network made of MSCs and the HLR. The Packet Core also makes use of the services of HLR, so this element belongs to both subsystems. (See Figure 8.3)

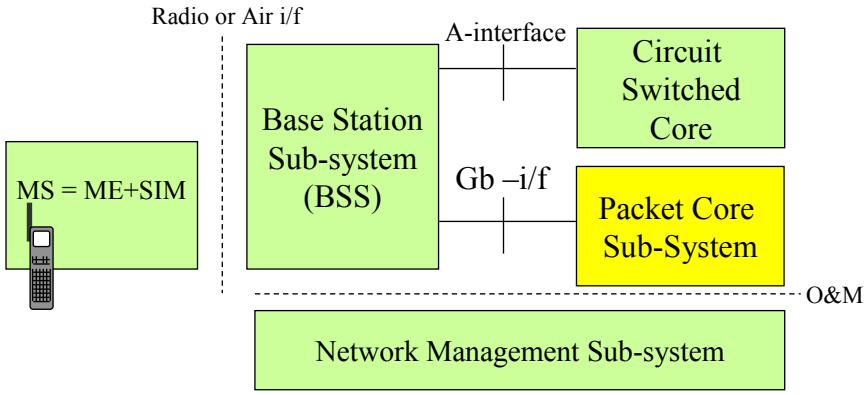


Figure 8.3: GSM subsystems.

The task of the Packet Core network is to manage the mobility of packet terminals and provide *tunneling* of user packets from Base Stations Controllers to an external packet data network such as the Internet or a company Intranet. A tunneled packet is created or consumed by the mobile and carried between the BSC/RNC and the gateway to the Internet as in an envelope across the packet core. Information on the envelope takes care of the mobility.

A new element that can be seen as a special purpose packet router called SGSN (serving GPRS support node) is in a similar role as the visited MSC/VLR for packet terminals. Another special purpose router, called GGSN (gateway GPRS support node) provides access to external packet data networks like the Internet. So, GGSN is the gateway to the Internet. Role of GGSN is somewhat similar to the role of GMSC with the difference that GGSN deals with packet traffic only and GMSC with circuit switched traffic only. We will discuss the similarities and differences of these elements later.

In addition to the Base Station Subsystem, the Circuit Switched Core and the Packet Core Subsystems, GSM specifies the Network Management Subsystem that uses OSI protocols to manage the other three subsystems. In practice, many network equipment vendors have implemented their own management interfaces for their network elements and the OSI protocol based management enjoys only limited support.

Business boundaries in GSM

In the GSM system the A-interface has historically been the most important *multivendor interface*. A-interface is between the Base Station Subsystem and the Circuit Core. In terms of signaling this interface is based on SS7 and uses the Base Station Subsystem Application Part (BSSAP). A-interface has been a business boundary for many mobile operators: they have bought the BSS from one or two vendors and the Circuit Core possibly from yet another vendor.

Also the interface between the HLR and the rest of the network elements is sometimes a multivendor interface in practice. This happens for example when

a Mobile Virtual Operator has its own HLR for managing its subscribers and the underlying GSM Core is provided by another vendor (and owned by a different company). For a mobile network operator owning its network and trying to compete with a broad set of services it is, however, probably not advisable to procure the HLR from one vendor and the Core networks from another. This may lead to being able to support just an intersection of services that are supported by the different vendors rather than the union. So, the result may be the opposite of what the goal was.

Also the Gb interface between the BSS and the Packet Core has been a solid multivendor interface in practice. Operators make the Packet Core buying decisions separately from the rest of the subsystems.

CAMEL Application Part (CAP)

CAMEL comes from the words Customized Application for Mobile network Enhanced Logic. (Figure 8.4 shows that this oxymoron was invented to reflect the two humps in the figure). CAP is originally a subset of ETSI core INAP customized for GSM networks. CAP is the protocol used in CAMEL.

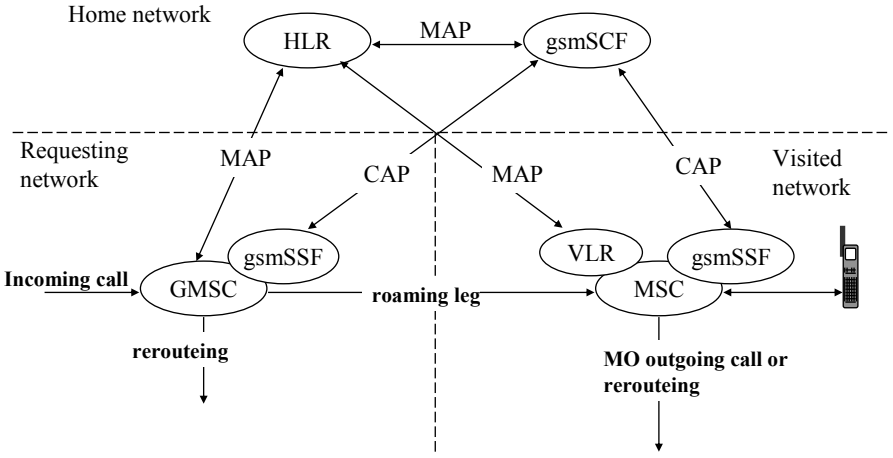


Figure 8.4: CAMEL architecture.

The original GSM model actually means that services are provided to roaming subscribers collaboratively by the visited and the home network. This makes it difficult for operators to differentiate and create services for example for mobile office concepts. The home network operator turns out to be dependent on the rest of the GSM operators for supporting services.

To overcome this architectural limitation, CAMEL was introduced. The MSCs are enhanced with SSF features including a Basic Call State Model suitable for mobile calls and a triggering capability. During a call setup, an MSC can request for instructions from a gsmSCF residing in the home network. This

allows the home network to provide additional service logic to call handling even for cases where its subscriber is roaming in a foreign network.

CAMEL was introduces in several releases (capability sets): version 1 had just 7 operations, version 2 already had 22 operations and was already useful for introducing a substantial set of supplementary services. Later versions 3 and 4 were released[2]. Many GSM networks have deployed CAMEL and the CAP protocol.

Summary of IN and GSM

IN in wire line networks provided added value to the operators and the architecture can be considered to have been rather successful. However, in my opinion this was true only in some countries such as the US. In many other networks, the interest to supplementary services that could be implemented with number translations and a few other tricks was quite limited. The cost of implementing IN for the vendors was high. It required upgrades to several systems and products. BSCM and triggering make call processing heavier for all calls. This increases performance requirements on switching systems. Differentiation of service by the operators using IN is possible if the operator is quick to deploy IN. However, very shortly afterwards the rest of the operators will follow suit and the advantage is lost. Creating new services using SCE may be fast. From the operators perspective this, however, is only part of the story. A large operator may have more than 100 software based systems supporting its business processes. IN is just one of them. When new services are deployed, upgrades are often needed in many other business critical software systems (related to billing, customer service etc). Also the organization must be taught to help customers with the new services. These changes may dilute the advantage from IN service creation.

The GSM creators really thought that creating this marvelous technology is going to be fun but that probably only some 10% of customers will ever want to own and use a mobile phone. This prediction turned out to be false and the rest is history. In hindsight GSM was may be the biggest hit in the history of communications services and stands very high even on the list of ICT success stories (on par with may be the PC) of all times.

There is a commonality of technical nature in IN and GSM. They both rely on SS7 for signaling. Both introduce their own Application Parts. Those application parts are quite complex, i.e. contain many modules and deal with several aspects of services.

[2] http://www.3gpp.org/TB/CN/CN2/camel-contents.htm contains an overview of CAMEL phases 1 to 4.

Examples of such aspects are: authentication of the subscriber during location updates and handovers, checking and validating the equipment identity, downloading services data from HLR to a VLR, removing such data, requesting call routing instructions, managing voice circuits already reserved for the call for example in order to process in-band voice or DTMF signals and announcements etc.

For modularity reasons it is good to keep those aspects as separate as possible. If each of the modules would communicate over SS7 directly using SCCP, this would lead to many roundtrips during call establishment. This would increase delay. Long delay of call setup is undesirable. Therefore, both IN and GSM can benefit from an additional common component between the Application Part and SCCP. This common component is called Transaction Capabilities Application Part or TCAP.

TCAP provides services that are friendly to Application Parts and allows separating remote communication events from application states. Each part of service logic in SCP or HLR can do their own thing and once all sub-modules have created their data that needs to be communicated to the remote MSC or SSP, the actual messaging can be initiated minimizing roundtrips. For example, during a handover, HLR and MSC/VLR may be concerned with authentication of the user, validating the Mobile Equipment and storing subscriber data into the VLR.

Transaction Capabilities Application Part

TCAP provides generic services supporting the execution of distributed transactions. Parties in the transactions can be exchanges, service nodes, data bases etc.

TCAP offers a way to implement services that are independent of network resources.

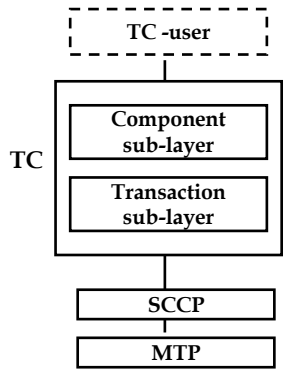


Figure 8.5: TCAP structure

TCAP is used by MAP, INAP and CAP. MAP uses it to support roaming and mobility management, IN more generally for implementing remote transactions. Usually the term *transaction is defined as a set of operations that are tied together and thus either all take place or none of them happens.*

TCAP makes use of SCCP messaging services over the SS7 infra.

The internal structure of TCAP is presented in Figure 8.5.

TCAP has two sub-layers. The upper sub-layer is called the component sub-layer. Its function is to support the data units of the TC users and to support the requests and responses in a context. Context is provided by the concept of a dialogue.

The transaction sub-layer is concerned with the communication related to transactions with the remote system either maintaining a "virtual connection" or dialogue or in a connectionless manner.

TCAP has a lot of commonality with two OSI protocol stack application service elements on OSI layer 7, namely ROSE – Remote Operation Service Element and ACSE – Association Control Service Element. SS7, IN and GSM designers were, however, more concerned with optimizing network performance than conforming to some standards created elsewhere. Therefore, TCAP is its own protocol in the SS7 family.

Identification in TCAP

Network elements that process TCAP are engaged in many TCAP communication flows that are independent of each other. One node processing TCAP may be talking to several network elements simultaneously.

TCAP makes use of all addressing modes supported by SCCP to point the remote system. The result is that TCAP communicating parties can reside anywhere in the global telephony network.

A TCAP flow may be a dialogue. Dialogues are identified by a pair of Originating Transaction Identifier (OTID) and the Destination Transaction Identifier (DTID). This means that when an element starts a TCAP flow, it will assign its OTID that is unique locally. The remote system that is the target of the transaction will assign DTID that is unique locally. The pair of (OTID, DTID) then uniquely identifies the dialogue or transaction for the two involved network elements.

One TCAP message exchange may carry data related to several operations that are of interest to different parts of the Application. Such operations are identified by Invoke_Id in TCAP. As a result, TCAP can report the results of the operations to different modules of the Application. A user can have many simultaneous operations. Operations can be chained by TCAP. This can be for example used to download a subscriber data file taking several SS7 messages into a VLR.

A TCAP dialogue

Figure 8.6 gives an example of a TCAP dialogue.

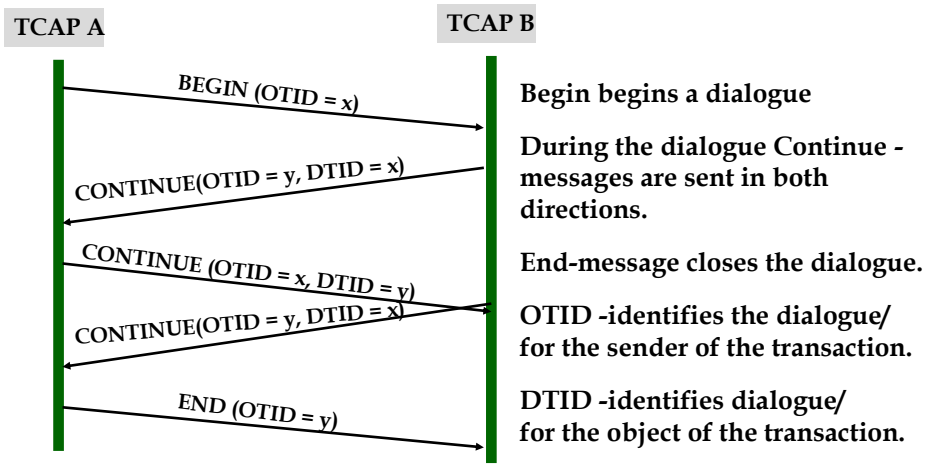


Figure 8.6: A TCAP dialogue

A (structured) Dialogue starts with a BEGIN message. During the dialogue the elements can exchange CONTINUE messages as they please. Normally, the dialogue ends with the END message.

There are four types of Operations in TCAP:

- Class 1 - Both success and failure are reported.
- Class 2 - Only failures are reported.
- Class 3 - Only success is reported.
- Class 4 - Nothing is reported

Note that reporting failures only speeds up recovery. A positive acknowledgement that something has actually taken place is however more useful. Having received a positive acknowledgement, the application can move on.

The result of an operation in the remote system can be success or failure or a reject. Prior to sending the result there can be any number of chained operations between the two network nodes.

Non-structured dialogue

Figure 8.7 shows a non-structured dialogue for Class 4 operations leaving sequencing of operations to the Application.

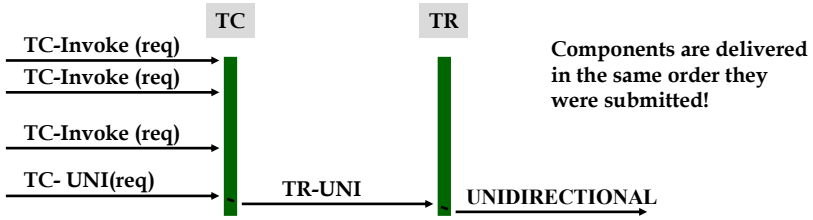


Figure 8.7: A non-structured dialogue.

A non-structured dialogue transfers one or many operations or components. They are all carried in a single UNIDIRECTIONAL TCAP message. They all have the same dialogue id.

Structured Dialogue

A structured dialogue was already presented in Figure 8.6. The same with the primitives is presented in Figure 8.8.

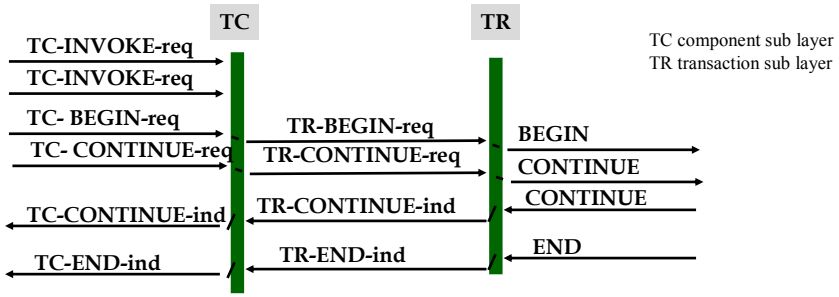


Figure 8.8: A structured dialogue

BEGIN causes a transaction identifier to be reserved. The remote system can either continue the transaction or close it. Closing can be by pre-arrangement or normally with END. Abnormally, the transaction can be aborted.

Component sub-layer

The component sub-layer with the primitives it provides to applications is depicted in Figure 8.9.

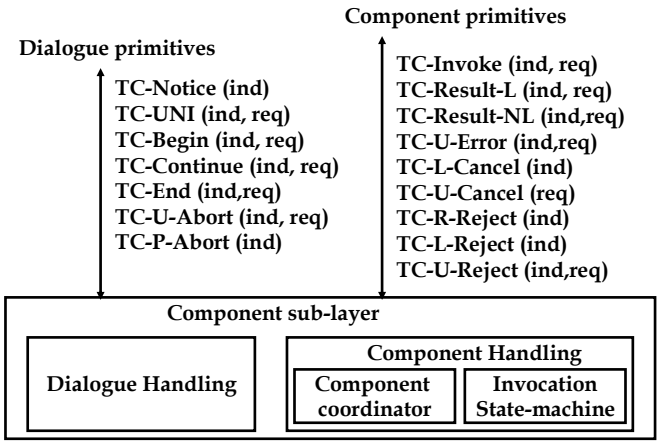


Figure 8.9: Component sub-layer

The component sub-layer can be divided into Dialogue handling for processing Dialogue primitives and Component handling for processing the operations primitives. The "_L" in primitive names refers to "Last" and "NL" to non-last.

Transaction sub-layer

Figure 8.10 depicts the transaction sub-layer with the primitives and messages with the peer system.

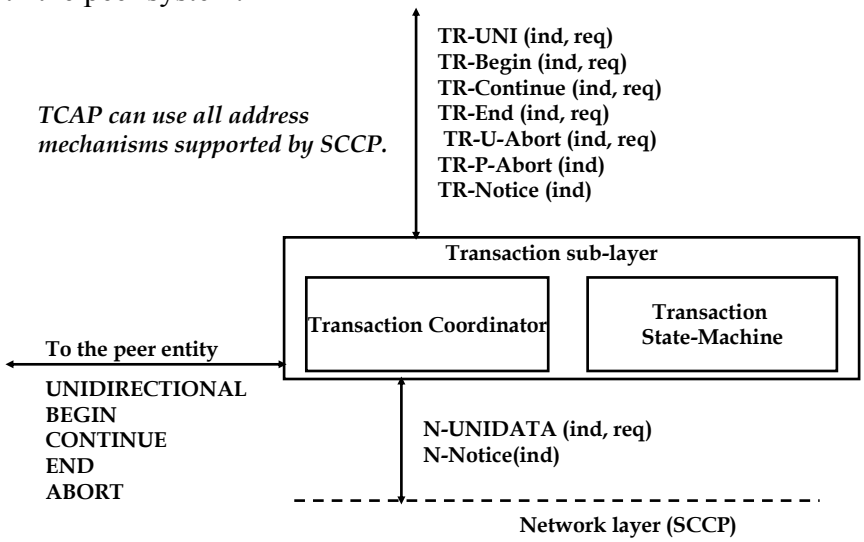


Figure 8.10: Transaction sub-layer.

The Transaction sub-layer handles the interfacing with the network layer and takes care of communication states with the remote system.