Chapter 8

UMTS Application Support

The CEC sees UMTS as the broadband mobile access system to complement fixed B-ISDN access into people's homes, schools and offices [UTF]. In confined pico-cellular environments UMTS will be able to fulfil this role with ample bandwidth (2Mbit/s) to fully support multimedia applications. However, UMTS is also supposed to provide communication at any time, in any place, supporting a total European market penetration of 50% of the population. It is recognized that priorities will differ in different environments, with some customers demanding high quality services while others might accept lower quality while demanding ubiquitous mobility. As was shown in section 2.5.6 and illustrated in figure 9, the reduction in bandwidth required to achieve ubiquitous mobility is large and the applications that UMTS supports will need to adapt to these diverse requirements.

This thesis has demonstrated why satellite access to UMTS will be valuable and how the UMTS Network Architecture will support satellites and their spectrally efficient channel assignment schemes. However, satellite access to UMTS will be of little use unless applications can adapt to it, a topic that has received less attention than it requires. The satellite component does not have to support all applications in full and application designers will want to specify what aspects of communication are most important for their applications. This chapter provides a survey of how UMTS might support its applications, as an introduction to work still required on UMTS and FPLMTS. It is instructive to compare this with [NAKAJIMA], a survey of standardization in this area in 1994.

8.1. Voice Telephony in Today's Networks

To date, the driving force for telecommunications has been voice telephony. Terminals supporting voice telephony are ubiquitous, well understood by customers and simple to use. Networks carrying voice traffic have been carefully designed around the specific requirements of the application - a low delay, no echo and a 300Hz to 3.1kHz frequency response to match that of human speech. In different parts of networks, voice telephony is carried in different ways. Analogue is still used in most copper access networks, carrying the raw base band analogue signal. In transit networks a 64kbit/s PCM (pulse code modulation) [G.711] version of the base band signal is carried, which allows a faithful representation of the original information. Over core network radio links and cordless telephone channels, a 32kbit/s ADPCM (adaptive differential PCM) version is often used to conserve bandwidth, with some degradation of the original information. In radio access networks (such as second generation cellular networks), predictive voice coding is used to reduce the bandwidth to between 6.4kbit/s and 16kbit/s. To reduce the information carried, the voice coder and decoder have a model of the human voice permanently programmed in so that the information transferred on the communications link are the control signals required to make the model "speak" instead of the audio information required to recreate the original audio signal.

At each switch in the network consideration is given to voice telephony being handled using different coding on different links. In such cases the switch must be able to transcode between the two coding schemes with minimum loss of quality. This is usually achieved by conversion to 64kbit/s PCM, the standard G.711 interface for 3.1kHz voice telephony codecs, then conversion to the outgoing coding standard. Voice coding techniques are all lossy in different ways, so each conversion degrades the voice quality.

In OSI terms, transcoding is a presentation layer function which means that even a basic switching function in the PSTN may need to work at the presentation layer rather than at the network layer, which is as far as switching functionality should go in an OSI protocol stack. Figure 53 shows how it is handled[1].

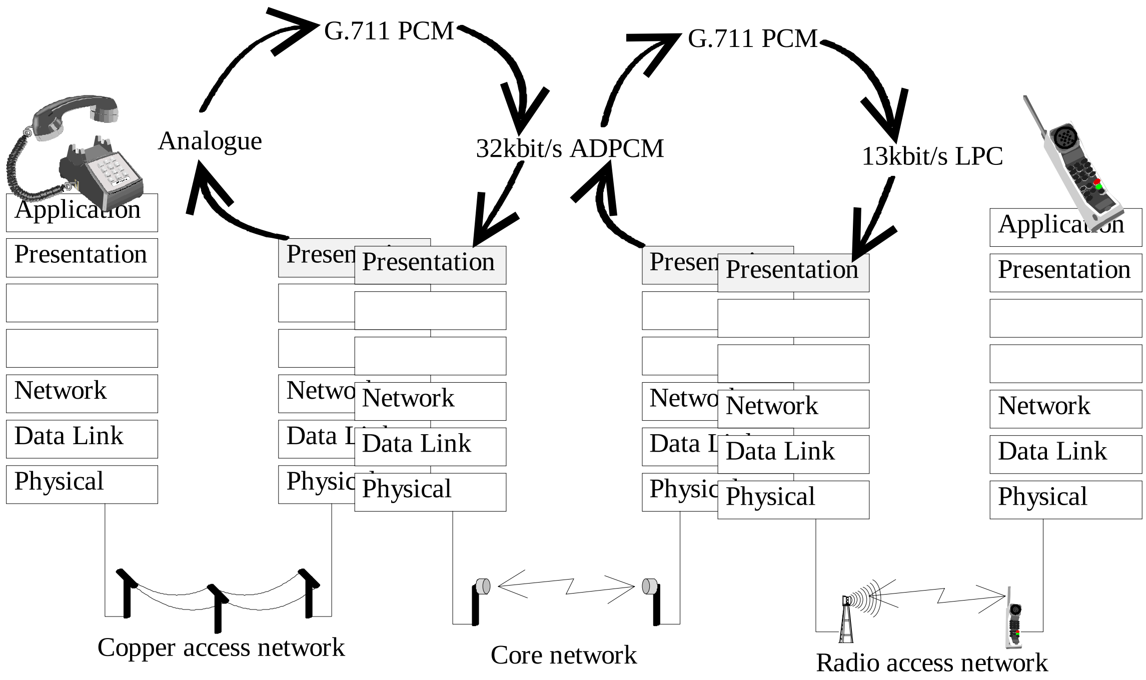

Figure 53   Voice telephony crossing networks on the OSI protocol stack

[1]The OSI model is not very suitable to illustrate the voice telephony application. There are no session or transport layer functions to speak of, so the presentation layer (voice transcoders) and network layer (switching functions) are next to each other in the protocol stack anyway.

8.2. Layered Communications Protocols

The OSI reference model layers protocols so that different applications can share the same transport protocols and networks. In a layered system, the interface between one protocol and that in the layer above and below is standardized so that a range of applications can work with a range of physical communications links. Layered communications protocols are favoured today in the computer networking industry because their encapsulation of functions into layers with defined interfaces makes system development simpler and allows layers to be re-used in different situations.

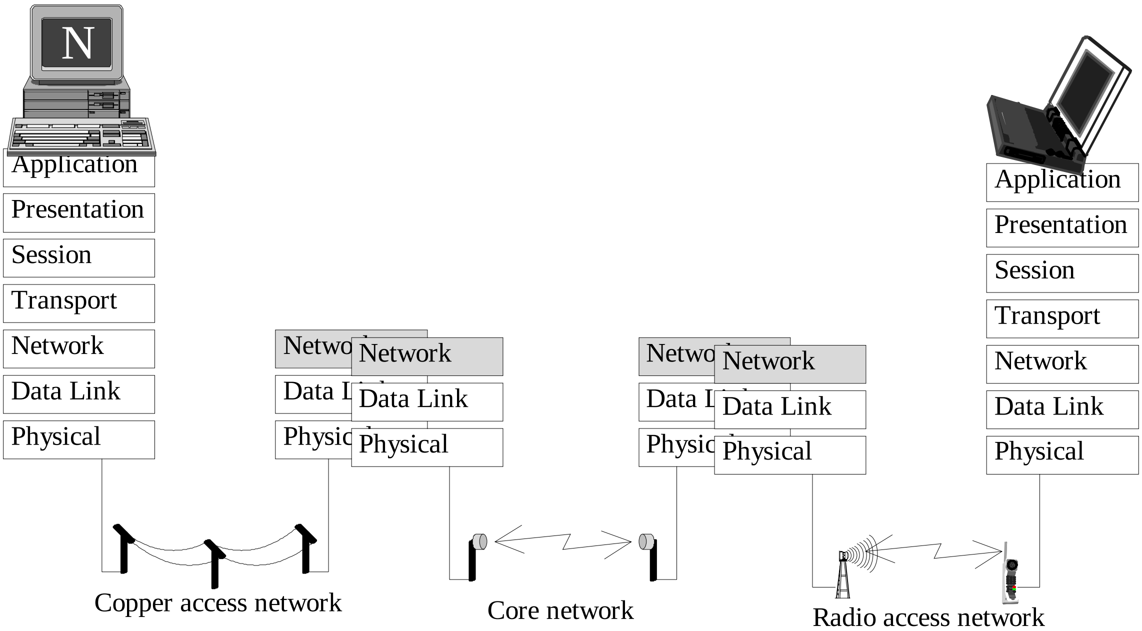

Figure 54 WWW browsing over networks using TCP/IP

The most widespread example of a layered protocol stack is TCP/IP (transmission control protocol and Internet protocol) which has made application development for the Internet as simple as possible even though the underlying communications links form the most diverse and complex communications network on Earth today. Without the clear, well-defined distinction between network and application, WWW (world-wide web) browsing would have taken much longer to develop than the 12 months it took to dominate the Internet. Without TCP/IP, WWW would also have needed the direct support of the communications network switches and routers. Figure 54 shows how WWW browsing is an end-to-end application with no such application support in network switches.

8.3. Second Generation Facsimile Support

After voice telephony, facsimile is the next most popular application on public networks today. It is instructive to see how this has been supported in networks such as GSM and Inmarsat networks.

Facsimile was designed for the terrestrial voice telephony network using modem tones to convey the digital information. Thus the output of a fax modem is an analogue voice-band signal containing information to be carried transparently from end-to-end by analogue and 64kbit/s PCM networks. In setting up the fax call, the fax machines at each end of the link negotiate by testing short samples of modem data at various rates over the link to determine the highest data rate that the link can support with an adequately low bit error rate . It is effectively an end-to-end protocol detecting the capabilities of the communications network(s) in between to carry modem tones. It works well, for example detecting that a 32kbit/s ADPCM link in the network would not be able to support 28.8kbit/s or even 14.4kbit/s modem tones but can support 9.6kbit/s.

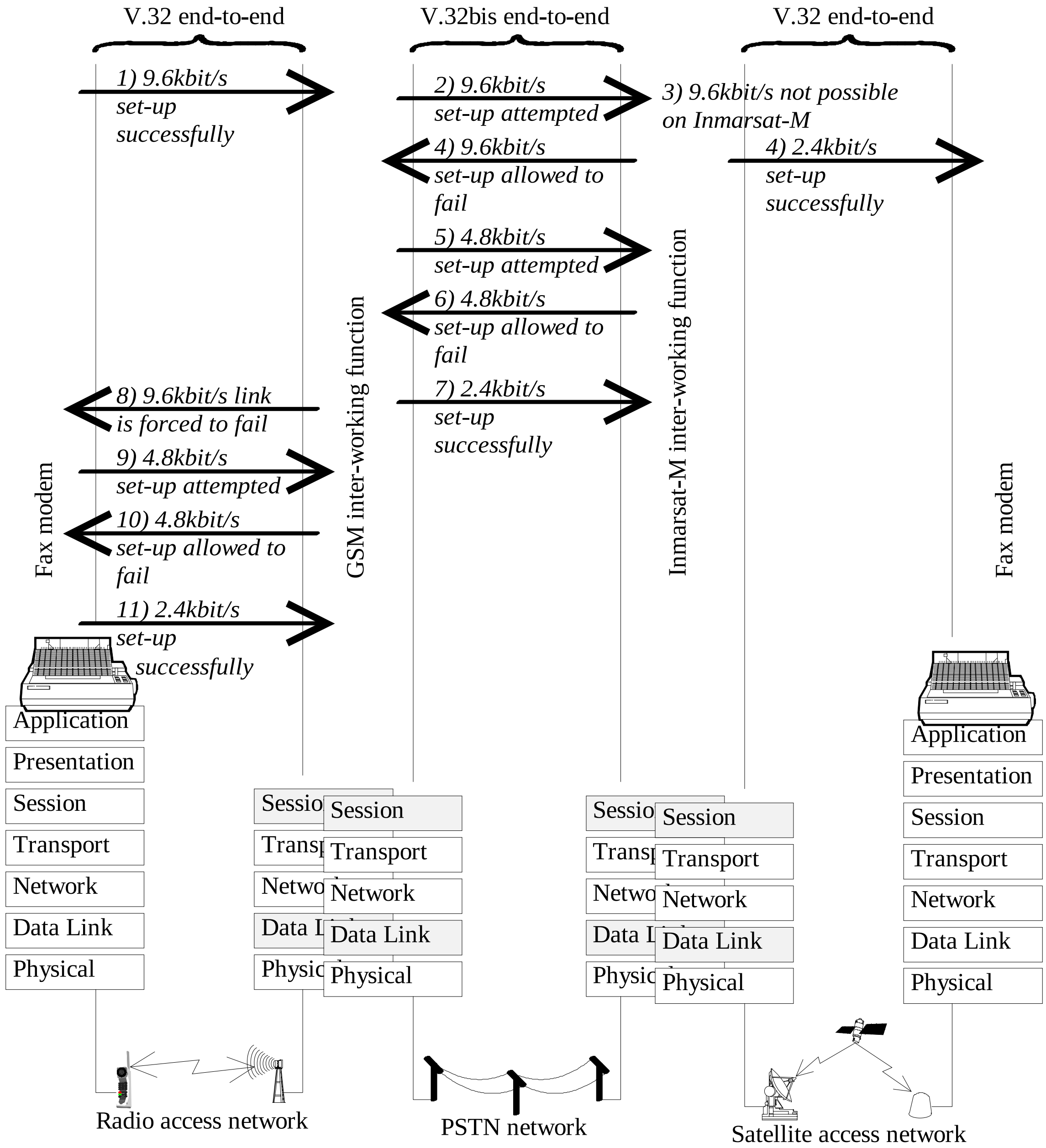

Figure 55 Facsimile crossing networks on the OSI protocol stack

In mobile networks bandwidth is a valuable and limited resource. Using 32kbit/s to carry sampled modem tones that only convey 9.6kbit/s of information is an unacceptable waste of resources. The solution is to demodulate the modem tones on entry to the mobile link and convey the demodulated data directly on the digital network using error correction designed for data. This breaks up the end-to-end protocol that was the original design for modem data as and makes it look more like the voice telephony model in figure 53 with transcoding at switches between dissimilar networks. Figure 55 shows the modem standards used as the level of interworking for fax and all other modem data applications[2]. These V-series ITU modem specifications are used unchanged as hop-by-hop protocols in this model even though they were designed as end-to-end protocols. This potentially complicates call set-up where two mobile networks are involved. For example, consider a fax call from a GSM terminal to the core PSTN then on to an Inmarsat-M terminal. The maximum data rates available may be 6.4kbit/s over Inmarsat M, V.34 (28.8kbit/s) modem tones on 64kbit/s PCM through the core network and 9.6kbit/s over GSM. The GSM fax inter-working function would negotiate 9.6kbit/s for the GSM terminal's fax application over the GSM link and the core network. The Inmarsat-M inter-working function finds that 9.6kbit/s does not work over the Inmarsat-M link and negotiates 6.4kbit/s with the Inmarsat-M terminal's fax application instead. Now there is a problem because the GSM link and core network link both use end-to-end protocols as hop-by-hop protocols even though they are completely unaware of each other. To allow the Inmarsat-M link's protocol to communicate back to the other two they must fall back further to 6.4kbit/s. The only way to do this is to stall the call and force a re-negotiation of the data rate from the Inmarsat-M end. Whilst the process (usually) works, it requires many seconds of signalling to do.

Looking forward to third generation mobile systems, the configuration of these links will be able to change on handover between different FPLMTS. Clearly a more advanced way to rapidly re-establish the best possible link from the resources available is required.

8.4. ISDN

PSTN applications like facsimile rely on V-series modem standards to inter-work with each other because they were the only well-defined interfaces in the industry. Even H.324, the new digital PSTN video-phone, uses V.34 modem tones to enable it to work with different networks without direct support for the application in inter-working functions. Arguably, the interworking of the V-series modem recommendations is a function of the session layer in the OSI protocol stack, even though the V-series recommendations themselves refer to the data link and physical layers. To date, mobile network operators have concentrated on implementing this interworking function reliably to support non-voice telephony applications. This has been difficult because the ITU-T recommendations do not standardize modem control. The Hayes command set is often implemented but frequently incompletely or with variations.

---

[2]Again, the OSI model is not ideal to illustrate the facsimile application. The ITU-T V series modem standards are data link layer specifications but the interworking functions that support the dialogue between co-operating fax processes on different systems are better described as the session layer. V.42 error control, if it is used, would operate hop-by-hop, arguably in the transport layer.

ISDN (integrated services digital network) was designed as a digital bearer network from the start, with signalling enhancements over the PSTN to make better use of the capabilities. Application requirements are signalled during call set-up, including bandwidth negotiation and ensuring that the right type of terminal answers the call (preventing the fax machine answering a voice call, for example). This was necessary to encourage the design of new applications to use the 64kbit/s data links to the full, such as H.320 video-conferencing (which is incompatible with H.324). B-ISDN (broadband ISDN), on which UMTS is based, also includes this feature. Unfortunately, the ISDN, like the PSTN before it, was designed for fixed terrestrial networks where bandwidth is not scarce and therefore lacks some features that would be useful for mobile networks. GSM has been built on ISDN principles, its signalling identifying the application to be carried.

Terrestrial ISDNs assume very low bit error rates from the communications channel, as do fax and video phone applications on the PSTN. Mobile channels suffer from bursts of errors making bit error rates much higher than the terrestrial channels unless error correction is used, raising bandwidth requirements. Experience with voice coding in digital mobile networks shows that there is a trade-off between voice coding efficiency and susceptibility to degradation of service in the presence of bit errors. When the redundancy in raw information is reduced the effect of errors in the information that is left is much more critical. In the GSM voice codec, for example, FEC is applied at different rates to different parts of the coded voice information depending on the importance of the information to the quality of the voice at the decoder. The voice information is arranged so that it can withstand the impairments expected when it passes through the communications channel.

Service	**Bit Error Rate**
National 64kbit/s ISDN	1.3×10^{-8} (95% error free minutes)
International 64kbit/s ISDN	2.2×10^{-8} (92% error free minutes)
GSM 9.6kbit/s transparent	10^{-3}
GSM 4.8kbit/s transparent	10^{-3}
Inmarsat HSD 64kbit/s transparent	10^{-6}
Inmarsat B 16kbit/s transparent	10^{-3}
Inmarsat B 9.6kbit/s transparent	10^{-5}
Inmarsat M 2.4kbit/s transparent	10^{-5}
GSM 9.6kbit/s non-transparent	10^{-8}
GSM 4.8kbit/s non-transparent	10^{-8}

Table 7 Bit error rates of various ISDN services. Error free minutes to bit error rate conversion assumes non-bursty noise

8.5. B-ISDN

B-ISDN will use an ATM (asynchronous transfer mode) interface which in the future will encourage splitting an application's data into a number of virtual channels, each of which has its own quality of service requirements. This will be important for UMTS because the very important information can travel in a VC with high priority and the less important information travels in a VC that will be conveyed if there is capacity spare but will be discarded if the network is congested. This information will allow mobile networks to error correct the different parts of an application's data in different ways and intelligently carry the most important data with an appropriate level of error protection as the application's designers would intend. Then if a 6.4kbit/s mobile satellite network is required to access a terminal for a multimedia application, the network can selectively drop some parts of the multiplex whilst conveying the most important 6.4kbit/s, so that the customer can still use the application, albeit with reduced fidelity.

For example, a PC user may be editing a shared document whilst talking using audio and video to a person at the other end of a communications link who is contributing to the editing of the document. The PC may be capable of generating 2Mbit/s of data to provide a very high quality video conference and document exchange over a high-speed LAN, for example. To work over a basic rate ISDN link, however, the application needs to scale the quality down to a total of 128kbit/s and it is likely that different elements of the data multiplex should ideally be scaled down differently. A B-ISDN application would multiplex the data in say 8 different VCs, as shown in the example in table 8. The right-hand column shows what data a 9.6kbit/s satellite link might convey, with a non-transparent acknowledgement protocol such as TCP running end-to-end for the critical document and control data.

VC	Content	Priority	Peak bit rate	Acceptable BER	Implementation over satellite link
1	Control channel	Very high	256bit/s	10^{-8}	Non-transparent
2	Document data	Very high	2.4kbit/s	10^{-8}	Non-transparent
3	3.1kHz basic telephony	Very high	4.8kbit/s	10^{-6}	3/4 rate FEC
4	3.1kHz telephony enhancement	High	16kbit/s	10^{-6}	Not carried
5	MPEG4 basic video	Medium	16kbit/s	10^{-8}	Not carried
6	7kHz audio enhancement (G.722 SB-ADPCM)	Low	48kbit/s	10^{-8}	Not carried
7	MPEG2 basic video	Low	720kbit/s	10^{-8}	Not carried
8	MPEG2 video enhancement	Very low	1Mbit/s	10^{-8}	Not carried

Table 8 Virtual channels in B-ISDN multiplex of an example shared document and video conference application

8.5.1. An API for B-ISDN

ATM now has Q.2931 and UNI 3.1 and UNI 4.0 signalling protocols defined to set up these VCs on demand at layer 2 of the OSI model. Elements of what these support are evident from the terrestrial and mobile systems described so far and from ISDN's CAPI (common application programming interface). Clearly, to enable rapid deployment of new applications such as Netscape, the application developer needs a stable communications platform that allows visibility and adaptation of the numerous different networks through which the communications channel is routed. IP is under revision at the moment and IP version 6 promises to support varying bandwidths and priority guarantees for VCs at OSI layer 3.

8.6. Mobile Multimedia

The UMTS task-force is now driving ETSI's development of UMTS. This initiative will ensure that UMTS develops as a broadband mobile solution to support higher data rate applications than GSM and DECT can evolve to. Second generation systems may benefit from a degree of compatibility with the network architecture of UMTS just as mobile networks benefit from interconnection with the PSTN and ISDN. Satellite networks will not be able to offer very broadband services to a hand-held terminal but will benefit from compatibility with UMTS insofar as applications can adapt to the communications channels available in the environment that they are used in. This is why scaleable applications will be so important to UMTS.

8.6.1. T.120

To guarantee inter-operation between similar multimedia applications, ITU-T has designed T.120 as the data multiplex protocol for multimedia conferencing. T.120 should enable data conferencing to work over a heterogeneous network in the future. At the moment, H.320 is used on the ISDN, H.234 on the PSTN and ProShare or CU-SeeMe is used on LANs to video-conference, for example. All of these are incompatible with each other. T.120 offers multi-point connections over heterogeneous networks, too. Unfortunately, a look at the T.123 transport stack profiles defined so far shows all the fixed networks are supported but there is no mention of mobile. For example, the bursty nature of errors on radio links is not considered in the design of the T.120 multiplex, resulting in a tendency to lose synchronization when errors occur.

8.7. Future Work

Fortunately, B-ISDN and T.120 are still being developed and will be completed in about the same time scale as UMTS. This currently presents an opportunity for developers of UMTS to work with the developers of B-ISDN and T.120 to create a mobile-friendly multimedia network standard instead of retrospectively adopting a sub-optimum fixed standard later. The development of IP version 6 might try to encapsulate mobility in the Network Layer of the OSI protocol stack [BHAGWAT], which might prove to be a restrictive generalization. The UMTS Network Architecture provides network services to allow handover of live calls between different networks as customers move about. Because of the diverse range of FPLMTS, applications will need to be able to adapt rapidly to new channel characteristics at handover to make the handover as transparent as possible to the customer. This aspect of FPLMTS will take considerable time to resolve and be implemented in correspondent networks, so work towards this must start now. CU-SeeMe already intelligently adapts its presentation of the application depending on the data throughput it can maintain through the communication link [CI]. At its worst, CU-SeeMe looks like a slide-show with reasonable quality narration but can scale up to full-motion video as communication resources increase. This automatic adaptation will be the key to FPLMTS' applications' success.