Private ATM Networks.

David J Greaves - Olivetti Research Ltd, UK.
Derek McAuley - University of Cambridge, UK.

From proceedings of IFIP WG 6.4 workshop ‘Protocols for high speed networks’ held at the Swedish Institute of Computer Science, Summer 1992. (C) North Holand 1992.

Abstract.

This paper advocates the use of local area networks which use 48 byte ATM cells. Hosts connected to the network are fitted with ATM interfaces and run a new protocol stack, up to the network level, which avoids multiplexing and efficiently handles the out-of-band signalling used by ATM.

The private network may be of WAN, MAN or LAN dimensions and contain several different network technologies, provided each is able to perform the basic function of carrying ATM cells from one point to another. The private network may be connected to the B-ISDN at one or more points.

1 Introduction.

Our research is directed at the provision of a distributed multimedia application environment. To attain this we are concerned with both communication architectures that can provide the desired services and how such services are supported within end-systems.

In this paper we discuss why we consider that an end-to-end Asynchronous Transfer Mode (ATM) network provides a useful basis for the communications architecture for such systems and hence leads to the desire for private ATM networks for local working and, via B-ISDN, ATM wide area interconnection.

We believe that besides cost reduction aims, functional requirements will lead to different solutions for local and wide area ATM networks and that these solutions will change with time and technology. To address this we have developed the Multi-Service Network Architecture (MSNA). We present a brief description of

MSNA, with both reasons for its internetworking approach and examples of its use.

While we advocate the use of ATM networks on an end-to-end basis, we also realise their usefulness in the interconnection of packet switched networks; the paper also includes a brief description of our experience with the use of MSNA.

2 Motivation for using ATM in the private local area.

The arguments which convinced the CCITT to recommend ATM as the solution for Broadband-ISDN also operate in the local area for private multiservice networks. Research work to establish this point has been carried out at the University of Cambridge Computer Laboratory and at Olivetti Research Ltd. Increasingly computer manufacturers are promoting the use of ATM techniques in privately owned and operated digital networks. A specific example is the 'Emerald' ATM switch recently announced by the BBN company [1].

By 'ATM techniques' we mean the use of a fixed-length cell as the primary means of information transfer, where the periodicity of cells is not known by the receiver in advance, but is indicated by a circuit identifier in the cell header. (This definition is taken from the CCITT recommendations.)

The attractive properties of ATM itself include:

- Support of a mixture of traffic types, including fixed, variable rate, and bursty traffic.
- Low jitter owing to short cell size and reduced switching delay due to cut through of multi-cell blocks.[1]

The fact that B-ISDN has adopted ATM as its transfer mode add two further, consequential advantages:

- Increased availability of special purpose VLSI devices.
- Opportunity for ease of interoperability between private and public networks.

[1] *Cut-through* switches are a class where the start of a message may already have left the switch on the appropriate output port, before the end of the message has been received at the input.

These advantages can only be fully realised if the private networks use the same size cell *payload* as the public networks, namely 48 bytes. However, the header size and *format* is not particularly important, since in ATM, headers can be manipulated by each switching entity, while the payloads are passed unaltered from one point to another.[2]

2.1 Comments on cell payload size.

It is not universally agreed that 48 bytes is the optimum payload size for a general purpose network. On the other hand, if one agrees that a fixed-size packet (or cell-based) network is preferable to one which supports variable length packets, then it is clear that, at least, all components of the network infrastructure should support the same cell size. This implies that private ATM networks should employ the same 48 byte payload size as the CCITT's B-ISDN.

Before returning to issues of fixed size versus variable size in Section 2.2, we comment on technical issues of cell size.

It has often been complained that a 48 byte payload is ridiculously small for computer data applications. The principle objections are that even the smallest messages sent by the current generation of distributed applications are much larger (e.g. a keystroke message using the X protocol over TCP/IP is 72 bytes in size) and that no computer would like to take interrupts for every 48 bytes received. These statements are generally correct, and the answer that this paper offers is that sensible computer network interfaces to ATM networks ameliorate these objections.

A payload of the order of 48 bytes was chosen by the CCITT so that the packetisation delay of 64 kilobit voice, when filling cells with 44 to 48 samples, was acceptably low: i.e. 6 milliseconds. Computer network users, who do not see speech as the predominant network traffic and who expect only low average utilisations of their networks, have argued that a larger cell size could easily be used, provided the cell is only partly filled when used for speech. If compact disc quality stereo sound becomes the norm, then the larger cell is filled already with the same packetisation delay. (E.g. a CD stream at 1.4 MBit/second would require a cell of 1050 bytes for 6 ms duration.)

These application and host interface arguments may indicate that a larger cell size is appropriate. Some multiple of a video-RAM shift register length has been suggested as suitable for easy implementation [2]. However, no single cell

[2] Our definition of MSDL in Section 3.1 will imply the control of other header bits, such as cell loss priority and payload type; we treat the coding of in-band indications as data-link specific so that their setting and interpretation is performed by MSDL, on a per-VCI basis, and according to the QoS of that VCI.

size will suit all applications; most existing data-oriented applications require variable length messages and will continue to do so. Clearly adaptation of these variable length messages to a sequence of fixed sized cells is required. ATM implies this approach, and once the need for such adaptation is recognised, whether implemented in hardware or software, the argument of what size cell to use becomes masked from the application, in all respects except delay and jitter performance.

2.2 Switching characteristics of short, fixed-length cells.

The basic principles of ATM switching is that fixed length cells are good for switching with low 99-percentile of jitter, while 'short' cells reduce the jitter and delay (and hence buffering) nearly in proportion to their length[3]. It is well known that in a variable length message switching system the 99-percentile of jitter is proportional to the packet size found in the tail of the packet size distribution [3]. Therefore if all messages are to be kept short, it is sensible to make them all the same size. This eases hardware design and in particular the buffer management. The reduction of delay with cell size is seen from a simple dimensional analysis, since both the duration of a cell (for a given transmission rate) and the queue sizes in switches are proportional to the cell length.

Short cells also simplify priority mechanisms, since there is no need to preempt transmission of a long, low priority message, in order to send higher priority messages; there are no such long messages. Finally, short cells imply that the speed penalty from not providing cut-through switches is minimised, and indeed, cut-through designs for ATM switches are rare, if not non-existent.

When we measure delay and jitter in seconds, rather than bits, cell duration, rather than length, is clearly the critical factor. In particular, cell size is only important in terms of multiplexing performance when the lowest rate link(s) of an ATM system are considered. The lowest rate links are likely to be between 100 and 150 Mbit/second. This is the limit of inexpensive multi-mode LED fibre technology and 10 K series ECL. For example, the basic SONET-based access to B-ISDN operates at 155 Mbit/second. The Advanced Micro Devices TAXI/FOXI chip set uses this technology at rates including 100 Mbit/second.

At rates of 100Mbit/second, the 48 byte payload cell duration is about 4 microseconds. When ten or so streams are statistically multiplexed and utilisation of the link is relatively high, the mean waiting delay for a cell will be about 20 microseconds and the 99-percentile between five and ten times greater, giving at worst, 200 microseconds. Although these example figures are quite specific,

[3] The 'nearly' is due to the bulk size increasing as messages from a fixed distribution are fragmented into a greater number of cells as the cell size is decreased.

values in practice depend on many factors, particularly on arrival discipline and any priority mechanisms. The importance of the example lies in that the 200 microsecond 99-percentile would be multiplied by ten to 2 milliseconds if a 480 byte cell was used. Clearly, 2 milliseconds of jitter from a single switch or link may have a significant effect on certain real-time applications, especially voice. With more than one multiplexing point in the system, the situation becomes worse, since jitter increases as switches are connected in tandem (although at a power which is less than linear with the number of stages). The conclusion of this is that if a network of cheap 100 to 150 Mbit links is to be used, increasing the cell size significantly beyond 48 bytes, or 4 microseconds is unattractive.

3 Protocol Architecture for private ATM networks.

In this section we present a protocol architecture for an *ATM internet.* We assume each end host or other network user is fitted with an ATM interface and runs the protocol stack. Issues related to the support and integration of existing network architectures with new ATM and B-ISDN networks are discussed in Section 3.2.

Our protocol architecture takes an internetworking approach in an attempt to address the problem of heterogeneous cell-based networks. While this could be said to be an historical problem at Cambridge due to a surfeit of different types of ATM networks, standardization activities have already defined two different implementations of the ATM service in B-ISDN and DQDB[4]; it would seem that as with packet networks, this heterogeneity will only increase with time.

One reason for this heterogeneity will be related to cost reduction for local area networks. Already several computer manufacturers have suggested the use of TAXI transmission systems instead of SONET due to the cost saving. Similar cost reduction exercises are bound to happen in switching systems; e.g. the B-ISDN virtual path service may not provide the management benefits in a LAN to make it worthwhile supporting.

Another reason for heterogeneity is a change in the required functionality. When we constrain ourselves to consider all ATM networks as composed of switches interconnected by point-to-point links, an argument can be made for following B-ISDN standards; however this is a naive approach. Consider an ATM based low power radio network for hand held computers: firstly, the requirements for radio media access require significantly different ATM headers from B-ISDN,

[4] Note here we refer to the DQDB access protocol not the higher level functions of 802.6 or SMDS.

while the requirement to deal with multiple simultaneous receptions of cells by different base stations and their consequent resolution requires that there be information in the header to identify duplicated and out of sequence cells.

Both these arguments come together when considering rings and dual bus solutions for ATM switching; these solutions can lead to cost reduction while their implementation requires a different cell header from B-ISDN to accommodate media access.

3.1 MSNA.

The protocol architecture presented is the Multi-Service Network Architecture [4]. An operational MSNA internet exists in Cambridge UK and spans the sites of Olivetti Research Ltd and the University Computer Laboratory. The physical network components consist of Fairisle ATM switches [5], slotted rings [6] and [7], and, in the near future, radio based ATM systems.

ATM networks allow users to specify differing qualities of communication service. MSNA is designed so that its implementation can be integrated closely with the end system processor scheduler to allow the provision of quality of service to the actual application. The desire is to be able to identify and schedule the relevant active entity in the end system with the minimum of protocol processing. One particular mechanism used in MSNA is to minimise, if not eradicate, the layered multiplexing often found in communications systems.

The multi-service network layer (MSNL) is an ATM internetworking service which is based on the idea of a lightweight virtual circuit providing raw ATM access, that is an end-to-end stream of ATM cells. We choose a connection oriented approach at the MSNL level as we see the virtual circuit as the obvious unit to which to attach a quality of service; for applications not requiring traffic guarantees we use a pipelined circuit set up mechanism to achieve a more rapid start up.

We use the term ‘lightweight’ to refer to the fact that, at the ATM level, no flow or error control is *required* on a hop by hop basis[5], and that the resources allocated to the connection are neither to be thought of as valuable or permanent. One aspect of this lightweight nature is that MSNL connections may be unilaterally de-allocated by any of the MSNL entities involved, in particular due to garbage collection to recover VCI space from idle or dead connections. Applications using MSNL directly are responsible for re-establishing the cir-

[5] Of course, flow and error control may be implemented to increase system performance – the CFR [6] implements both – however, MSNL does not rely on all hops being totally reliable.

cuit, although for many 'datagram' oriented applications, a layer on top of MSNL provides re-establishment *on demand* without explicit interaction with the higher-layer software; e.g. such a service is used in the implementation of IP over MSNL.

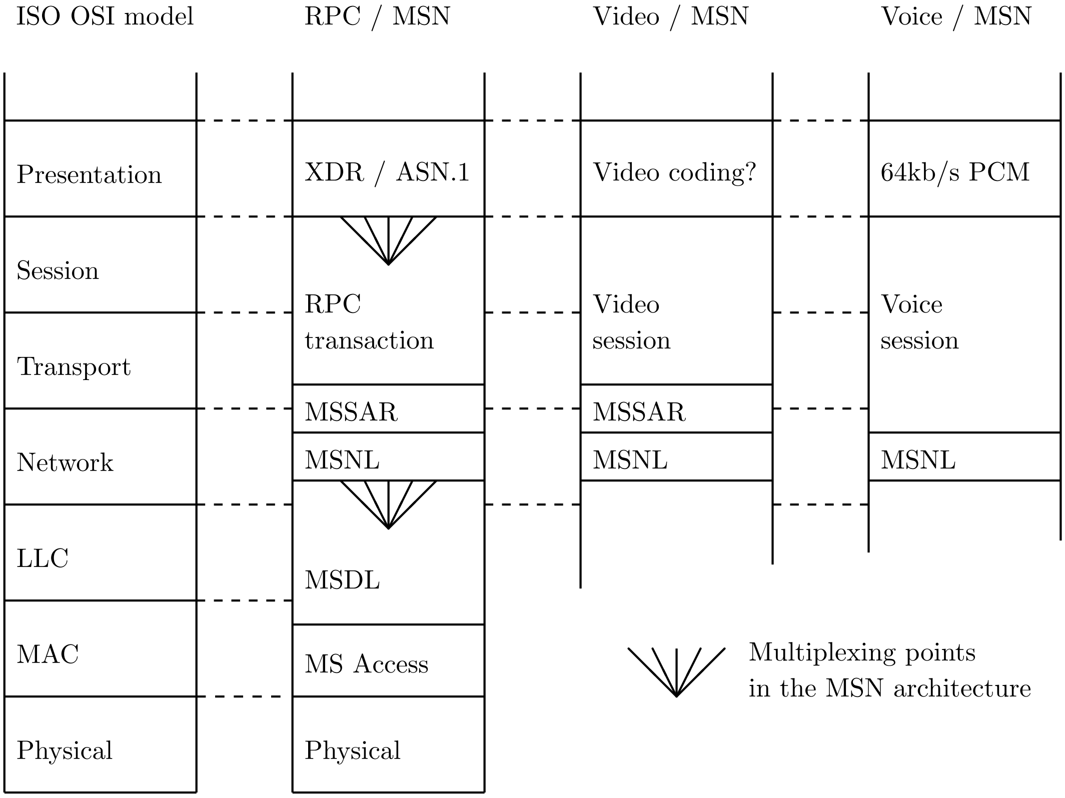

Figure 1: Relationship of functions within MSNA to the OSI Reference Model.[6]

To support MSNL, each different ATM network type is required to provide a generic virtual circuit service interface, called MSDL. An MSNL circuit (called a *liaison*) is formed from a concatenation of MSDL circuits (called *associations*).

There are three important aspects of MSNL:

[6] It is not possible to give an exact match between OSI and MSNA layers; the diagram indicates where the functions of the different MSNA and OSI layers have greatest overlap.

- it defines the MSNL address (MSAP),
- it defines an out-of-band liaison set-up procedure,
- it does not multiplex its liaisons over the MSDL associations.

Since MSNL liaisons are not multiplexed over MSDL associations, MSNL does not require in-band protocol headers in the service units. This means that MSNL introduces no processing overhead in the data path, and provides the same basic data interface as an MSDL association. The function of forwarding cells between heterogeneous networks requires that the VCI mapping function be capable of dealing with mapping between the two different header formats involved; this is straight forward in both hardware and software.

An MSNL liaison is established between two MSNL SAPs (MSAPs). These are unique and are allocated from a 64 bit global address space[7]. A host computer may have many MSAPs (loosely corresponding to the conventional idea of multiple ports), but on the other hand, there may be many computers sharing a single MSAP, such as individual controllers on the ports of a fast-packet switch. In general, for ease of routing decisions when a connection is set up, it is beneficial if the structure of the 64 bit numbers is actually hierarchical. In [4], the division into separate, 32 bit, *identifier* and *port* port fields is suggested. This optimises the typical case where multiple MSNL clients are situated at a single location (host). However, the individual client streams do not become multiplexed, owing to the separate liaisons for each client.

Setting up an MSNL liaison involves establishing a concatenation of MSDL association hops, with the MSNL address providing the addressing mechanism. During liaison establishment, routing is performed at each MSNL entity to select the appropriate MSDL instance for the next hop. The MSDL definition for each network type then maps the MSNL circuit establishment mechanisms to those appropriate for its underlying network type; this can involve both address mapping and, in some cases, protocol mapping. Once the liaison is established the routing of cells from one MSDL instance to another is normally performed by hardware; where the two instances are identical, this is the normal VCI mapping function seen in switches, where they are different, the hardware also translates the other fields of the header.

[7] The currently adopted approach at ORL and the University of Cambridge Computer Laboratory is to base 32 of the 64 bit address on IP addresses. This simply provides a convenient unique identifier space.

3.2 Interconnecting MSNA sites using B-ISDN.

The MSNA concept of providing an end-to-end service for cell payloads enables any adaptation layer to be used over MSNA. In theory, the same is true in B-ISDN, since the ‘Empty’ adaptation layer is available as a specified service, along with the four prescribed adaptation protocols. However it is unclear to the authors whether B-ISDN providers will allow users access to the network without one of the non-empty CCITT adaptation layers being imposed. Another doubt is whether VCI/VPI space will be available as cheaply and freely as in MSNA. With unfortunate pricing policies, customers may find it attractive to use an AAL which provides a circuit multiplexing function in order to save on active VCIs or VPIs.

Assuming these problems do not arise, or arise in a form which is easily overcome, local private MSNA domains can be interconnected over B-ISDN without impact on the MSNA protocol architecture. Since MSNA was designed to cope with heterogeneous networks with various signalling architectures and cell header formats, alternative control mechanisms for the virtual circuits used on the public network offer no new management problems; B-ISDN simply presents itself as an instance of MSDL.

4 Adaptation layer protocols

Our aim of producing an ATM internet, such that ATM is delivered into the real end system, means that our consideration of adaptation is concerned primarily with efficient implementation in end systems rather than as a means of interconnection of current packet or STM switched traffic types. However, as previously mentioned, by providing end-to-end service for cell payloads, MSNL can support any adaptation layers, in particular those being proposed by CCITT.

Within MSNA, and in keeping with the desire to minimise multiplexing, we have concentrated on adaptation layers which do not perform cell level interleaving of different segmented blocks over an single virtual circuit. If cell level interleaving is required, the VCI which is designed for the purpose is used, i.e. we use multiple MSNL streams. In B-ISDN terms this is equivalent to using a virtual path service where the end-to-end VCI bits distinguish cell interleaved blocks, rather than using a single VCI and sorting cells into blocks by another layer of multiplexing defined by the MID. In the MSNA architecture this block level service is known as MSSAR.

We have been using a particular adaptation protocol for a number of years, while investigating issues in adaptation [8]. In investigating the minimal information required to implement sufficiently secure yet efficient and simple to implement adaptation layers we have arrived, together with other researchers, at a proposal which requries that the ATM cell header now include a logical end-to-end user data bit[8].

The Bit can be used to implement a range of different adaptation layers. For example, for data services, the bit indicates last cell of a fragmented block; the last cell includes a 32 bit CRC over all the cells forming the block and a length field. The CRC detects the usual bit and burst errors, as well as most cell reordering[9], while the length is required to detect certain cell drop outs where the content and position of the cell contrive to provide no contribution to the CRC. This mechanism, now known as AAL-5 [9], is undergoing standardization activities in ANSI and CCITT.

We use this adaptation layer without further multiplexing in higher level protocols. For example a single VCI is used between a two threads to implement a reliable transport, while for RPC, a client acquires a VCI for the duration of a call to a server. This manner of working means that the VCI associated with received blocks can be used to identify the final recipient of the data (e.g. some thread). At this point the standard resource allocation mechanisms of the operating system associated with buffering and scheduling can be invoked without further protocol processing due to more layers of multiplexing.

4.1 Interconnection of existing networks.

Whilst aiming for a network infrastructure in both the local, metropolitan and wide areas based on interconnected ATM networks to provide integrated services to end-systems, we also appreciate the role ATM has to play as a integrated transport mechanism for current packet and synchronous traffic.

The MSSAR service provides many of the same facilities as Frame Relay or B-ISDN AAL-5, that is, a connection oriented data service without guarantees. Hence we have used this service to provide interconnection of packet networks between Olivetti and the Computer Laboratory. Using MSSAR implemented in the Wanda operating systems we have implemented bridging of 802.3 networks; this is in everyday use bridging XNS traffic between the Olivetti and Computer Laboratory Xerox systems using MSNA over the Cambridge Backbone Ring.

We have also implemented MSNA in a number of 4.3 BSD derived systems, and as well as providing both MSNL and MSSAR to user processes as a new protocol

[8] The December 1991 CCITT draft I.361 extends the payload type to three bits and defines this end-to-end user data bit.

[9] To detect whether two cells i apart have been swapped we need to select a polynomial such that the CRC of $x^{i*384}+1$ is non-zero.

family, MSSAR can be configured to present itself as an IP network interface, hence providing IP connectivity over our ATM internet. Circuits are established on demand to the next IP hop as indicated by IP routing; these circuits are cached to be used for future IP packets routed to the same hop and are deleted when that are observed to be idle for several minutes. Currently, about 150 machines at Olivetti and the University are able to communicate between each other using IP/MSNL; widescale use for everyday IP service awaits either our ATM network to become as reliable as Ethernet or we acquire some host based dynamic IP routing software able to perform automatic fail-over.

The most stringent synchronous requirement so far is within the Pandora system [10], where the audio component of the system uses standard 64 kbit/second sampling. This is transported over the network in 2 millisecond units consisting of 16 samples with a source time stamp in each ATM cell. The time stamps are used for resynchronisation at the receiver rather than any mechanism based on recovery of a 8kHz timing signal from a transmission system [11].

5 Conclusion.

We consider that the merits of ATM for wide area public service also apply in the local area and for private networks. Starting from this, the aim of our work can be described as ‘ATM everywhere’. That is we are primarily concerned with the delivery of integrated services based on ATM to end-systems, even where such end-systems are hand held mobiles. This leads to our internetworking approach and in to a different set of considerations from B-ISDN in the adaptation protocols we support. We consider that the B-ISDN supporting an empty adaptation layer plays a key role in this approach by the provision of wide area coverage for ATM.

References

[1] ‘BBN unveils broadband strategy with Emerald switch.’ ‘Communications Networks’ October 1991.

[2] ‘Micron Mos Data Book’ from Micron Technology Inc, 2805 East Columbia Road, Boise, Idaho 83706.

[3] ‘A Fast Packet Switch for Integrated Services.’ P Newman. IEEE JSAC 6(9) December 1988.

[4] ‘Protocol Design For High Speed Networks.’ DR McAuley. University of

Cambridge technical report 186. December 1989.

[5] ‘Fairisle: An ATM Network for the Local Area.’ Ian Leslie and Derek McAuley. Proceedings of SIGCOMM ’91, Zurich, September 1991.

[6] ‘The Cambridge Fast Ring Networking System.’ A Hopper and RM Needham. IEEE transactions on computers, Vol 37 no 10. October 1988.

[7] ‘The Cambridge Backbone Ring.’ David J. Greaves, Andy Hopper and Dimitris Lioupis. Proceedings of IEEE Infocom 90, San Francisco 1990.

[8] ‘Cambridge HSLAN protocol review.’ DJ Greaves, ID Wilson. Proceedings of IFIP WG6 International Workshop on ‘Protocols for high-speed networks’ edited by H Rudin and R Williamson, held at IBM Ruschlikon 1989. Elsevier 1989.

[9] ‘AAL-5 – A New High Speed Data Transfer AAL.’ IBM et al, ANSI Committee T1 Contribution T1S1.5/91-449, November 1991, Dallas Texas

[10] ‘Pandora - An experimental distributed system for multimedia applications.’ Andy Hopper. ACM Operating Systems Review, April 1990.

[11] ‘Network Compatible ATM for Local Network Applications.’ Phase 1, version 1.0, Apple et al, April 1992.