ATM Local Area Networks*

Peter Newman

N.E.T. Adaptive Division
800 Saginaw Drive
Redwood City, CA 94063
(newman@adaptive.com)

January 1994

Much interest is currently being shown in the application of Asynchronous Transfer Mode (ATM) switching technology to local area networking. ATM offers much greater capacity than current shared medium LANs and is capable of supporting multimedia traffic. This paper discusses LAN emulation and the design of ATM LANs. LAN emulation offers a best-effort, connectionless, packet transfer service at the MAC sublayer, implemented on top of a connection-oriented ATM network.

1 Introduction

In the last few years much interest has been expressed in Asynchronous Transfer Mode (ATM) technology due to its flexibility and support of multimedia traffic. Initially the interest came from the carriers and the manufacturers of wide area networking equipment. However, in the last year or so, increasing interest has been shown in the application of ATM technology to the local and campus area networking environment. ATM offers much greater capacity than existing shared medium LANs. It is scalable in that the capacity of an ATM system is not fundamentally limited by the technology itself. It is designed to support multimedia traffic and is capable of offering seamless integration with wide area ATM networks both public and private.

Much of the current discussion justifies the introduction of ATM technology into the local area on the basis of its ability to handle multimedia traffic. However, the move toward ATM is equally likely to be driven by ATM's more mundane benefits of increased bandwidth, and much greater manageability, for the continually increasing volume of regular data communications traffic.

2 LAN Service Requirements

If ATM technology is to be successfully introduced into the customer premises, it must offer a LAN-like service for data traffic compatible with the existing installed base of data communications protocols, applications, and equipment. A LAN offers a connectionless, best-effort service for the transfer of variable size data packets. The service is best-effort in the sense that lost or corrupted packets are not retransmitted. LANs offer point-to-point, multicast, and broadcast transfer and many current protocols rely on the broadcast capability. Users are not required to establish a connection before submitting data for transmission. Nor are they required to define the traffic characteristics of their data in advance of transmission. Users simply submit traffic to the LAN whenever they wish, as fast as possible, and the LAN dynamically shares the available bandwidth between all active users.

The majority of the installed base of LAN equipment conforms to the IEEE 802 family of protocols, fig. 1. In this architecture the data link layer

---

*To appear in the IEEE Communications Magazine, March 1994.

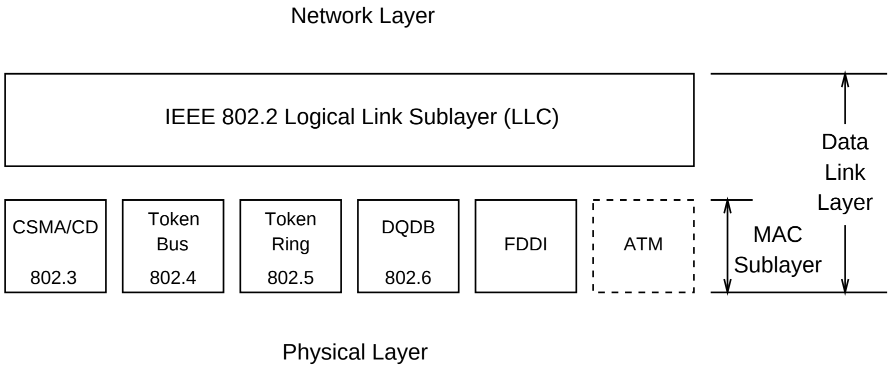

Figure 1: The IEEE 802 family of LAN protocols.

is split into the logical link control (LLC) sublayer and the medium access control (MAC) sublayer. The LLC sublayer offers a common interface to the network layer. Each of the different MAC protocols is specific to a particular design of LAN: CSMA/CD, Token Ring, Token Bus, etc. All stations on an IEEE 802 LAN are addressed using a globally unique 48 bit individual address with a flat address space. Group addresses may be defined for multicast groups and a well known broadcast address is also defined.

LANs are frequently interconnected with bridges and routers to form larger networks. Bridges operate at the MAC sublayer, and are popular because they require very little manual configuration and are transparent to the user. Bridges interconnect multiple LAN segments yet give the appearance to the user of a single LAN. Routers operate at the network layer but support only a finite set of network layer protocols (not all protocols in commercial use are routable). They offer greater control, better management facilities, and may be used to construct much larger networks than bridges. Two forms of transparent bridging have been defined: local and remote [9], fig. 2. A local bridge connects LANs that are directly attached to its ports. A remote bridge connects LANs across a non-IEEE 802 interconnecting medium, typically a wide area network such as X.25, frame relay, or T1 private lines. A remote bridge encapsulates each packet from the source LAN within a protocol specific to the interconnecting medium. The original packet is removed from the encapsulation at the remote bridge of the destination LAN. Remote bridging does not permit communication between a station on a LAN and a station connected directly to the interconnecting medium unless that station also runs the encapsulation bridging protocol.

It is possible to interface ATM directly to the transport layer or the network layer of the OSI model. This offers efficiency by avoiding the unnecessary complexity of the data link layer. However, there are many network layer protocols, and each one would have to be interfaced to ATM separately. To offer general compatibility with the installed base of networks and protocols, regardless of the network layer and upper layer protocol stack, and to support transparent MAC bridging, an interface at the MAC sublayer is required. This will permit the huge legacy of existing LAN applications to migrate to the ATM environment without major upheaval.

Thus a MAC sublayer should be developed for ATM LANs that offers the same connectionless MAC service as the IEEE 802 and FDDI MAC sublayers. Also we require the ability to perform MAC layer bridging between stations attached directly to an ATM LAN and stations connected to an IEEE 802 LAN. Such bridges should offer transparent local bridging. The ATM network is itself a LAN, and not simply a network across which LANs

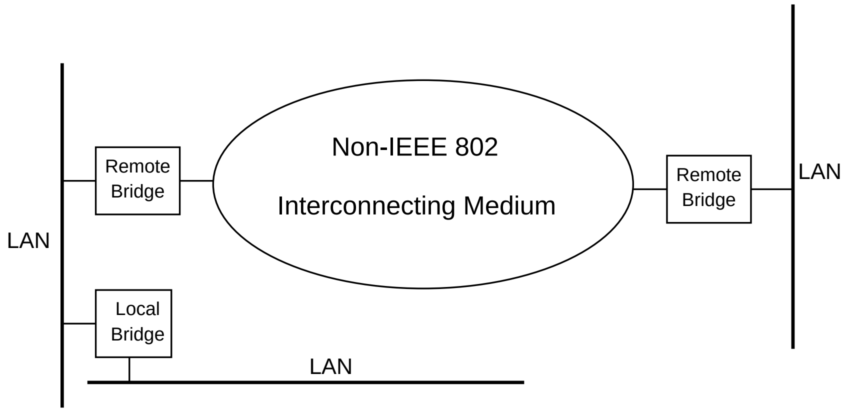

Figure 2: Local and remote bridging.

may be interconnected, so remote bridging would be inadequate. (Remote bridging would require a station on the ATM network to know whether the destination was also on the ATM network, or on an IEEE 802 network connected via a bridge, in order to select the required encapsulation protocol.)

3 Connectionless Service Implementation

3.1 Connectionless Server

ATM switches are connection-oriented and do not naturally support a connectionless service. Since we have decided that an ATM LAN must offer a connectionless service at the MAC sublayer, it is reasonable to consider implementing that service with a connectionless server (a CLSF in CCITT-speak) [10, 41]. This is the proposed implementation for the switched multimegabit data service (SMDS) [9] and also the B-ISDN connectionless data service (specified in the CCITT[1] I.364 recommendation).

In its simplest form a connectionless server is a packet switch attached to an ATM switch. All virtual channels carrying traffic that requires a connectionless switching service is directed by the ATM switch to the connectionless server. The connectionless servers are connected together with virtual paths through the ATM switches to form a 'virtual overlay network', fig. 3. This is basically the same architectural solution as narrowband ISDN — integrated access to separate switching facilities.

Implementing the connectionless server as a packet switch separately from the ATM switch has a number of disadvantages. Primarily, it substantially restricts the bandwidth available for switching connectionless traffic. Thus the ATM promise of high bandwidth and scalable capacity is lost. Also it negates the possibility of a single integrated switching mechanism for multimedia services. It is an approach based upon the existing solution to data networking. This approach is popular with the public carriers because it confines the statistical multiplexing to the connectionless server and avoids the requirement to support statistical traffic directly in the ATM network. It may be a useful public service for sending the occasional datagram, but in the local area, statistical switching directly at the ATM layer will offer far greater performance.

3.2 On-the-Fly Connectionless Implementation

An alternative implementation of the connectionless server integrates it into the port cards of the

---

[1] Now reorganized as the International Telecommunications Union, Telecommunications Standardization Sector (ITU-TS).

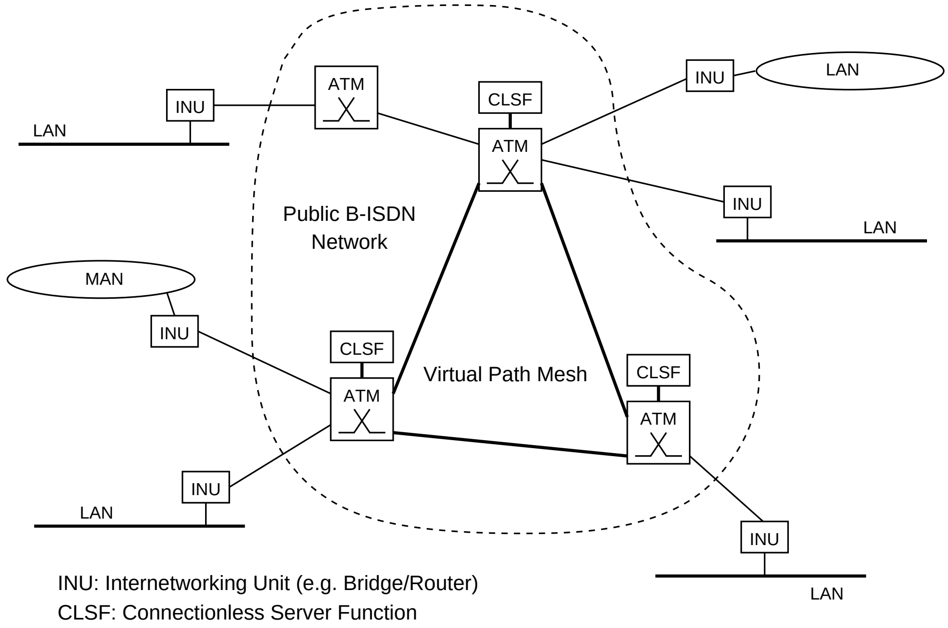

Figure 3: B-ISDN connectionless data service implementation.

ATM switch in a distributed manner [14]. Each port card detects virtual channels carrying connectionless traffic and inspects them for cells that contain the beginning of a packet (BOM cells). A routing operation is performed upon the destination address in the payload of the BOM cell. The forwarding table is then updated to transmit this cell, and the following cells belonging to the same packet, on the next hop toward its destination. The entry is removed from the forwarding table when the end of the packet (EOM cell) is detected. This method permits connectionless switching at the packet level without requiring that cells be reassembled into packets prior to switching. However, it does require the use of AAL 3/4 whereas the data community has shown much greater interest in the use of AAL 5 for data communication. This technique may be used to implement a centralized connectionless server [45, 25]. To achieve a distributed implementation requires fast route resolution followed by fast allocation of a free multiplexing identifier (MID), by the required output port, for every BOM cell [14]. This implies additional hardware, and therefore cost, but offers little increased functionality beyond direct ATM. Why add the hardware to perform fast routing and connection setup on a per packet basis when the same result may be achieved by performing routing and connection setup on a per call basis using permanent or switched virtual connections?

3.3 Connectionless Service over Permanent Virtual Connections

The simplest approach to implementing a connectionless service on top of a connection oriented network is to use a mesh of semi-permanent connections. Each end station has a virtual channel to every other end station in the network. This may be acceptable for a small number of nodes but the maintenance of the mesh of connections rapidly becomes unacceptable as the number of end-points grows.

A more practical approach is to interconnect customer premises ATM switches with a semi-permanent mesh of virtual paths, fig. 4. It is a similar approach to the 'virtual overlay network' of con-

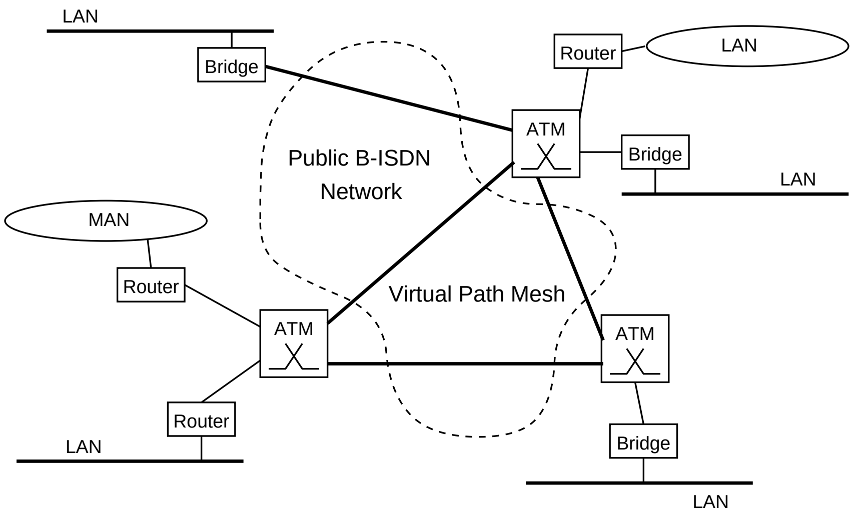

Figure 4: Private wide area network connectionless data service implementation.

nectionless servers except that we have removed the connectionless servers. The statistical multiplexing that the connectionless server provided is now implemented within the customer premises ATM switches.

This approach is well suited to the interconnection of ATM LANs across the public network to form a private wide area ATM network [40, 7]. The connectionless service is supported via switched virtual channels. Across the wide area the virtual channels are carried within the semi-permanent virtual paths so that the public network is not involved in the signalling, does not have to offer statistical multiplexing, and is not involved in reserving resources on a per virtual channel basis. Bandwidth for each virtual path is reserved across the public network and the customer premises ATM switch statistically multiplexes traffic for the wide area onto the virtual paths. If the public network only offers constant bit-rate virtual paths, reserving the bandwidth statically for each virtual path can result in an under utilization of bandwidth due to the very bursty nature of data traffic. Two techniques have been proposed for dynamically varying the bandwidth of each virtual path according to the load: bandwidth renegotiation [32, 20, 36]; and bandwidth advertising [11]. A more efficient solution may be achieved if the public network offers variable bit-rate virtual paths with statistical multiplexing [23].

3.4 Connectionless Service over Switched Virtual Connections

To implement a connectionless service using switched virtual connections we must establish a connection for each data conversation. It is inefficient to establish and release a connection in order to transmit each individual packet. It places too high a load on the signalling service and imposes a connection setup delay before the transfer of every packet. However, the majority of data conversations will consist of the exchange of multiple packets. So if a connection is established for the duration of the conversation the overhead of connection setup is minimized. In the local area we are not required to reserve network resources for idle virtual connections. So a virtual connection that has no traffic to send need consume no network resources other than entries in the connection tables. Thus a large number of virtual connections may be cached at each station on the assumption that future conversations to those destinations are likely to arise.

Even in a pure connectionless implementation such as IP, each station must maintain a limited cache of mappings between network addresses and physical addresses. Furthermore, each mapping is obtained from an address resolution protocol which is an operation comparable to connection setup.

The above argument suggests that for an ATM network offering a local area data service, the distinction between connectionless and connection-oriented implementation is being eroded. This is due to a fundamental principle of ATM switching: the separation of the connection from the network resources (bandwidth and buffers) consumed by the connection. An ATM connection is a lightweight connection — it consumes no bandwidth when inactive and a very large number of inactive virtual connections may be supported by the network. Whereas, in previous connection-oriented packet networks such as X.25, resources (packet buffers) were dedicated to each virtual circuit by the network in order to offer reliable service. ATM gives us the opportunity to separate the virtual connection from any guarantee of service across that connection. Thus a best-effort connectionless service may be emulated using ATM virtual connections if the resources of the network are not reserved by individual connections but are dynamically shared between all active connections.

4 LAN Emulation

An IEEE 802 LAN offers a connectionless MAC service that supports arbitration among stations for access to a shared physical transmission medium (e.g. the coax cable or the hub backplane). In contrast, ATM offers a connection-oriented communication service based upon switched point-to-point physical media. To offer a connectionless MAC service implemented on top of ATM we must design a protocol layer above the ATM adaptation layer (AAL) that emulates the connectionless service of a LAN. We will call this the ATM MAC sublayer, fig. 5. The ATM MAC sublayer emulates the LAN service by creating the appearance of a virtual shared medium out of an actual switched point-to-point network.

In an existing IEEE 802 LAN segment, all communication (unicast, multicast, and broadcast) is broadcast to all stations on the shared physical medium. Each station filters out the packets it wants to receive. The properties of a physical LAN segment may be emulated in an ATM network by connecting a group of end stations on the ATM network to an ATM multicast virtual connection. The ATM multicast virtual connection emulates the broadcast physical medium of the IEEE 802 LAN. It becomes the broadcast channel of the ATM LAN segment. Any station may broadcast to all others on the ATM LAN segment by transmitting on the shared multicast ATM virtual connection.

In current IEEE 802 LANs the membership of an individual LAN segment is defined by physical connection to the physical shared medium. Membership of an ATM LAN segment is defined by logical connection to the multicast ATM virtual connection that emulates the broadcast channel for that ATM LAN segment. So membership of an ATM LAN segment is defined logically (stored in some management database) rather than physically, as it is for an IEEE 802 LAN. This offers terminal mobility and greatly increased flexibility in network management and has led to the use of the term 'virtual LAN' to describe an ATM LAN segment.

It would be possible to emulate a LAN segment by transmitting all of the traffic for the segment on its broadcast channel. However, most LAN traffic is unicast and it is much more efficient to support unicast communication using point-to-point ATM virtual connections. This approach offers greater security since the unicast traffic appears only at the two communicating stations and is not broadcast to all stations on the LAN segment. It enables an ATM LAN segment to offer much higher aggregate bandwidth than if all traffic were transmitted on the same broadcast channel. Also, the use of individual virtual connections for unicast traffic permits much greater control of the quality of service (throughput, delay, probability of cell loss, etc.).

To establish a point-to-point ATM virtual connec-

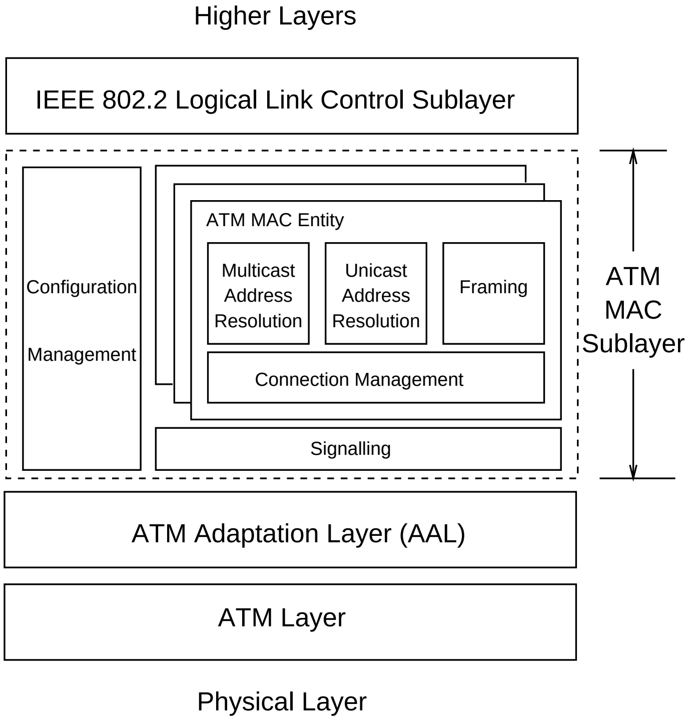

Figure 5: The ATM MAC sublayer structure for an end station.

tion for each instance of unicast communication, the current location of the destination end station must be discovered. This location must be expressed in the form of a destination address that the ATM signalling service can understand. This operation is called address resolution. The ATM signalling service must then be invoked to establish a point-to-point ATM virtual connection to the destination with the appropriate quality of service. Within the end station, this operation should be implemented in the software of the ATM MAC sublayer so as to offer a transparent service to the LLC sublayer, fig. 5.

4.1 Addressing

To emulate the service of an IEEE 802 LAN we must support addressing using the 48 bit MAC address. This address has a flat address space and identifies a network interface in the end station whether it connects to Ethernet, Token Ring, or FDDI etc. Each ATM MAC entity must therefore also be assigned a 48 bit MAC address, from the same address space, to identify it. The MAC address is assigned by the manufacturer of the network interface and is guaranteed to be globally unique. Thus it can be used to identify an end station (or a particular network interface on an end station). It allows the end station to be relatively mobile (they get disconnected and reappear at a different location fairly often) since the MAC address contains no hint of the location of the end station.

The direct use of a MAC address to communicate with an end station is acceptable in a single LAN segment or across a limited number of LAN segments interconnected via bridges. However, large bridged networks become very difficult to manage and introduce excessive broadcast traffic attempting to locate end stations. The address space of a large network is usually organized hierarchically

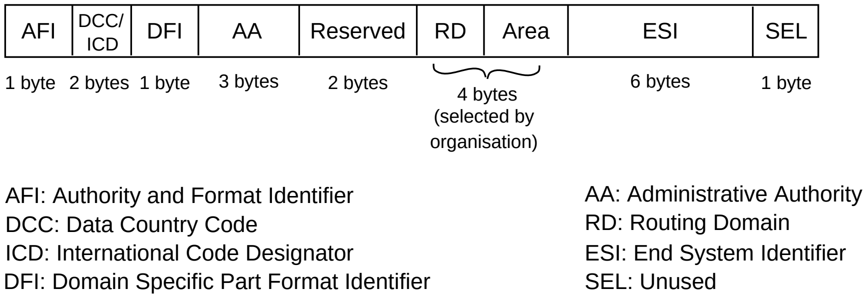

Figure 6: ATM address format specified by the ATM Forum.

(e.g. the telephone network: area code; central office number; customer's line number). This makes it much easier to locate any particular point on the network but such a structure greatly restricts the mobility of the addressed objects. This restriction is perfectly reasonable if the addressed object is relatively static, such as a particular phone jack in the office wall.

An ATM network will use a hierarchical address format to identify each ATM port on the network. The address resolution operation is used to bind the relatively mobile, end station MAC address, to the fixed physical address of the ATM port it is currently connected to. When an end station is attached to an ATM switch port a registration protocol will exchange MAC address and ATM address between the ATM network and the end station. Both the MAC address and the ATM address will also need to support group addresses for multicast connections.

The signalling protocol currently being developed for use in broadband-ISDN: Q.93B [15], permits an ATM address to be divided into two parts: an address and a sub-address. The usage recommended by the ATM Forum [19] is that the address describe the point of attachment to the public network (if connected to the public network); and that the sub-address identify a particular end station within a private network. The ATM Forum specification permits two address formats to be used to specify an ATM address. One is the hierarchical ISDN telephone numbering scheme E.164, and the other is a 20 byte address defined by the ATM Forum and modelled after the address format of an OSI Network Service Access Point, fig. 6. This address structure contains an initial string of seven bytes, allocated by national and international authorities, to identify a particular organization (e.g. an ATM service provider, a private ATM network, or an ATM vendor, etc.). This is followed by four bytes, the Routing Domain and Area, that the organization can allocate itself in some hierarchical manner of its own choosing. Next comes a field named the End System Identifier which contains a valid IEEE 802 MAC address. The SEL field is not used. An alternative format allows the eight bytes after the AFI field to contain an E.164 address. This option permits both public address and private sub-addresses to be combined into a single ATM address.

4.2 Address Resolution

Address resolution may be implemented with either a broadcast mechanism similar to IP ARP or by a distributed database mechanism. In both mechanisms the source sends an address resolution request containing the destination MAC address and its own MAC and ATM addresses.

In a broadcast mechanism, for unicast address resolution, the source broadcasts the request to all stations on the local ATM LAN segment and to all ATM LAN segments connected via bridges. All stations check the requested MAC address and the station that owns the requested MAC address replies with its current ATM address. The reply may be sent on the broadcast channel for the ATM LAN segment, or alternatively the destination may set up an ATM connection to the source and deliver the reply. For multicast address resolution an algorithm may be defined to convert from a group MAC address to a group ATM address. (The group MAC address is hashed onto a set of contiguous group ATM addresses pre-allocated to each ATM LAN segment.) Alternatively a simple server mechanism may be implemented for multicast addresses.

In a database address resolution mechanism, requests are received by an address server in the network. The server maintains a table containing MAC to ATM address mappings which is updated as part of the registration protocol every time an end station joins or leaves the network. Both unicast and multicast address resolution may be offered by the server. The server will need to be implemented as a distributed database to guard against hardware failure, so it is likely to require a more complex implementation than a broadcast implementation.

A complication arises when the destination is not directly attached to the ATM network but is attached to an IEEE 802 LAN connected to the ATM network via a bridge. In the broadcast implementation, the bridge may reply to the address resolution request with its own ATM address, as a proxy for the destination, if it contains the destination MAC address in its forwarding table. In the database approach the address server on the ATM network must contain entries not only for the directly attached stations but also for all stations attached to IEEE 802 LANs accessible via bridges. To achieve this each bridge must continually update the address server with the contents of its forwarding table. In large bridged networks this would result in a very large address table and require a substantial amount of traffic to keep it up to date.

The two approaches to address resolution may be combined transparently to the user. End stations should assume a broadcast mechanism is in use. If ATM multicast connections are implemented using a multicast server (see section 5.6) the multicast server can also act as an address server. It can intercept address resolution requests submitted for broadcast to an ATM LAN segment and respond with the requested address from its database. The address resolution database may be compiled from the exchange of addresses in the registration protocol and also by the use of a learning algorithm similar to those currently in use by transparent bridges. If no entry is found in the database for the requested address the address server should use the broadcast address resolution mechanism.

4.3 Virtual LANs

While it would be possible to configure the entire network as a single ATM LAN segment, there are advantages in partitioning it into multiple ATM LAN segments [4]. ATM LAN segments may be organized along administrative boundaries and increase the security across these boundaries. Partitioning increases the manageability of the network and limits the amount and extent of broadcast traffic. Also, partitioning into multiple segments facilitates interworking between ATM LANs and the existing base of installed LANs, bridges, routers, and protocols. The broadcast channel of each ATM LAN segment is limited to the members of that segment. Broadcast traffic on an ATM LAN segment will not escape the boundary of that segment unless segments are connected together via a bridge.

An example of a network with three ATM LAN segments is given in fig. 7. An ATM LAN segment is not confined to end stations connected to a single ATM switch but may accept members from any point of attachment in the ATM network. Assignment of end stations to ATM LAN segments may be determined by the network manager using a graphical network management tool. Thereafter the network will recognize each end station by its MAC address and may obtain the ATM LAN segments to which it belongs from a management database. Unknown MAC addresses may be assigned to a default ATM LAN segment. Alternatively, membership of ATM LAN segments may be determined by the physical port to which the end station is

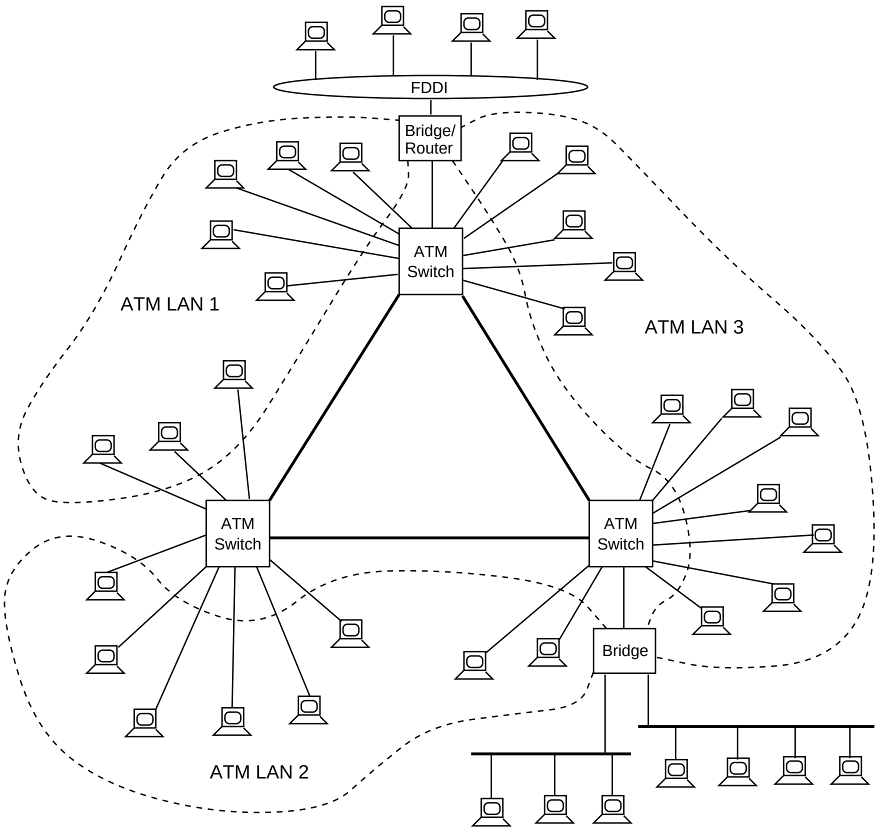

Figure 7: An example network showing three ATM LAN segments.

connected, with idle ports assigned a default LAN membership.

When the end station is plugged into an ATM switch port, the registration protocol informs the ATM network of the end station's MAC address. The network looks up the MAC address in its virtual LAN database to determine to which ATM LAN segment (or segments) the end station belongs. The network then informs the ATM MAC sublayer in the end station of the various parameters of each of its ATM LAN segments. One of these parameters is the ATM group address of the broadcast channel for each ATM LAN segment of which it is a member. The end station may then invoke the signalling mechanism to establish a connection to the ATM multicast virtual connection that emulates the broadcast channel for each of the ATM LAN segments to which it belongs. So each end station may be attached to any physical ATM port and still remain connected to its assigned virtual ATM LAN segment. Thus moves and changes may be made within the corporate campus without requiring a change of network layer address or any other action by the network manager.

4.4 Internetworking

By offering a LAN emulation service at the MAC sublayer; hosts, bridges, and routers may all be interconnected transparently across ATM LAN segments in the same manner as existing IEEE 802 LAN segments. Communication between an existing IEEE 802 LAN segment and an ATM LAN segment may also be supported by bridging or routing. A bridge, router, or an end station (e.g. a

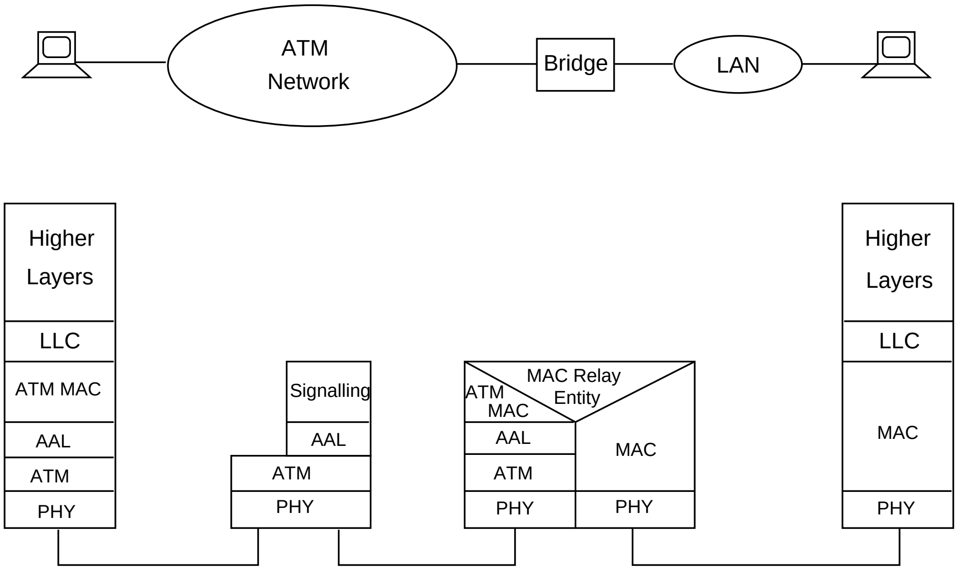

Figure 8: Local bridging of LAN and ATM LAN segments.

server) may be connected to multiple ATM LAN segments. Each interface to an ATM LAN segment has a separate (software) ATM MAC entity with its own MAC address but requires only a single physical connection to the ATM network. An example of the protocol structure of a bridged connection between stations on an ATM LAN segment and an IEEE 802 LAN is given in fig. 8.

A router may be used to selectively interconnect ATM LAN segments. It may be programmed to permit only certain machines or protocols etc., to communicate between specific segments. As such it permits the network administrator to implement certain security measures. However, once two end stations on different ATM segments have exchanged packets across one or more routers using a common network layer protocol they can discover each other's ATM address. They may now establish a direct ATM connection in order to transfer data. A direct ATM connection will offer much greater performance than is available at the network layer across one or more routers. Thus we see that inside an ATM network, the role of the multiprotocol router is likely to evolve to that of an address resolution service at the network layer because the forwarding of data will migrate to the specialized hardware of the ATM switch.

The function that the router performs is very similar to the function that signalling in the ATM network performs. Both employ an underlying datagram transfer service. Both require a routing protocol to route traffic across the network between two end-points specified by a particular address. An ATM signalling service only understands ATM addresses whereas a multiprotocol router can understand the network layer addresses of each protocol it supports. It is possible to combine both functions within the ATM switch such that connection requests to the signalling service may specify any of the supported network layer addresses or an ATM address. However, this will significantly complicate the signalling service. A better approach may be to combine both functions in the ATM switch but to maintain the separation of functionality for ease of maintenance, upgrade, and testability.

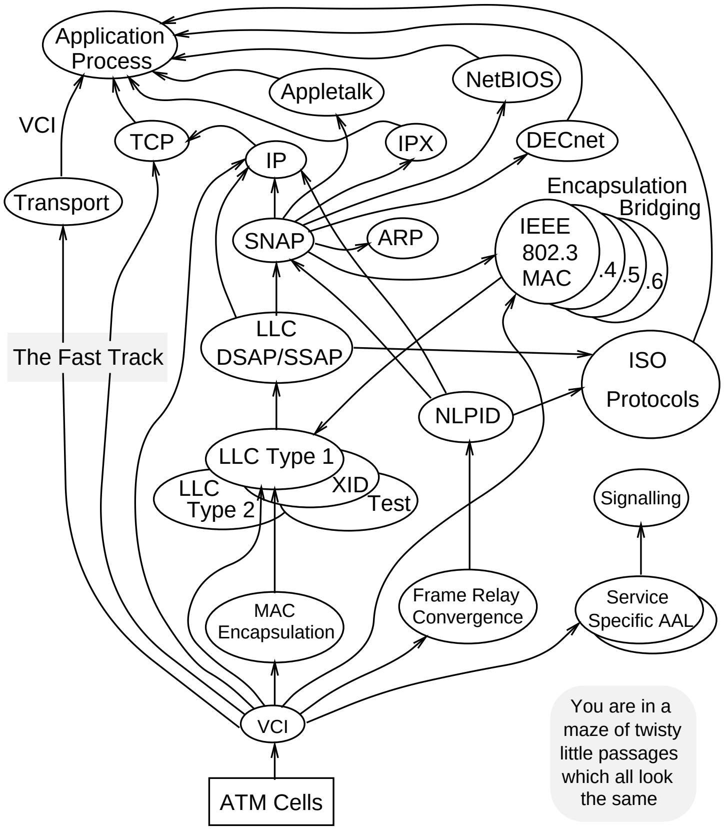

Figure 9: Artist's impression of the current protocol demultiplexing tree.

4.5 The Fast Track

A major advantage of LAN emulation at the MAC sublayer is that it is common to all higher layer protocols. By implementing a single ATM MAC sublayer we are able to support all higher layer protocols. With a single common interface we have compatibility with the existing installed base of 'legacy networks'. However, compatibility with the legacy of protocol development is not without cost. Fig. 9 gives an artist's impression of the maze of demultiplexing paths that have developed due to the rapid growth in local area networking. Encapsulation, demultiplexing, and address resolution in the data link layer reduce the packet processing performance at the network interface. So it may be worth interfacing popular network layer protocols directly from the network layer to the ATM adaptation layer, cutting out the conventional data link layer altogether. However, any protocol that we interface directly to ATM from the network layer can no longer be bridged because it has no MAC sublayer. Such a protocol would require a router to interconnect an ATM LAN and an IEEE 802 LAN.

To connect a network layer protocol directly to ATM we may use a similar approach to LAN emulation at the MAC sublayer. The major difference is that the address resolution mechanism must translate from the network layer address to the ATM address directly [3]. Each network layer protocol would have to be interfaced to ATM individually, so this approach makes sense for the most popular protocols while others may use LAN emulation at the MAC sublayer. Both approaches may happily co-exist in the same manner in which current commercial networks choose to route some protocols and bridge others.

Now that we have begun to prune the protocol demultiplexing tree, why stop at the network layer? It has long been argued that offering multiplexing at every layer in the protocol stack significantly reduces throughput [42]. Multiplexing need only be offered at a single layer in the protocol stack. For an ATM end station this layer should be the ATM layer. Ultimately we may use the ATM VCI as the demultiplexing identifier all the way up to the application process. But this implies a radical departure from the existing protocol stack. Compatibility with existing implementations may be retained by employing the full protocol stack for connection setup. Once the connection is established, the full protocol stack may be switched out and for the data transfer phase the VCI may be used to demultiplex directly from the AAL through the transport layer to the application process (assuming an all ATM end-to-end connection).

5 Switch Requirements for ATM LANs

The design of an ATM switch to support an ATM LAN is likely to differ in a number of important respects from a switch designed to offer B-ISDN service in the public network.

5.1 Best-Effort Service

One of the most fundamental concepts in packet switching is that users contend dynamically for access to a pool of shared bandwidth. This is statistical multiplexing — but it implies the probability that at times more traffic will arrive than can be serviced by the available bandwidth. If the overload is short lived it is sufficient to buffer the excess traffic. If it is possible for the excess traffic to exceed the available buffer capacity a congestion control mechanism is required to share the available network resources (bandwidth and buffer memory) dynamically between all contending users. This class of service is generally referred to as 'best-effort' in that the network offers no specific performance guarantee to the user. The term 'available bit-rate' (ABR) is also used to describe this service suggesting that the network attempts to share the bandwidth currently available for this class of service between active users.

Two fundamental classes of service are being considered for ATM networks: guaranteed, and best-effort. For the guaranteed service the traffic characteristics of the source are specified at call setup and the network either guarantees a particular quality of service for the duration of the call (in terms of delay and cell loss probability) or rejects the call. For a constant bit-rate (CBR) guaranteed connection, only the required peak cell rate is specified (and possibly the cell jitter). The network ensures that bandwidth for this connection is always available. This is similar to circuit switching but using cells instead of octets. It is more flexible than traditional circuit switching in that any requested bit-rate may be supported up to the link capacity. For a variable bit-rate (VBR) guaranteed service the required traffic characteristics are specified at call setup using some statistical definition (e.g. peak rate, sustainable rate, and maximum burst length). Again the network guarantees the quality of service or rejects the call [21, 16, 43, 39]. This service permits statistical multiplexing but requires the statistical characteristics of the source to be known in advance. For the best-effort (ABR) service only the peak rate of the source is specified at call setup but users of this service are expected to adjust their rate in response to feedback received from the network. The best-effort service makes use of the bandwidth remaining after serving the guaranteed traffic.

An alternative reservation approach for bursty traffic is to reserve bandwidth for each burst using a fast reservation protocol [5, 6]. To operate at the burst rate this technique must be implemented in hardware in each switch and requires specific traffic management cells to request and release bandwidth for each burst. This method becomes increasingly inefficient as the peak transmission rate of each burst approaches the link capacity [31, 12], which is exactly the mode of operation a LAN would prefer.

In a LAN, each end station wishes to transmit at the full line rate of its network interface to achieve as low a latency as possible. So the allocation of constant bit-rate guaranteed bandwidth, for each connection, will severely restrict the amount of bandwidth available to each connection. It will result in increased latency and inefficient utilization of bandwidth since data traffic is very bursty. Bandwidth could be allocated on a statistical basis, for each connection, if the traffic characteristics of each source were known. However, data applications cannot predict their bandwidth requirements or source traffic characteristics in advance of transmission. This is because their access to the network is controlled by the operating system and the operating system concurrently schedules a number of applications, each with differing traffic characteristics. Clearly a best-effort service is the most natural fit for bursty data traffic in the local area.

5.2 Multiple Traffic Classes

It is most likely that an ATM local area network will be required to offer a guaranteed service to handle real-time traffic (e.g. voice, video), components of multimedia services, or circuit emulation traffic, in addition to a best-effort data service. Such traffic is able to specify its traffic characteristics and may therefore request a quality of service guarantee from the network. The switch hardware needs to ensure that at no time will the quality of service of the guaranteed traffic be adversely affected by best-effort traffic.

The simplest approach to ensure this is to separate the cell buffering in the ATM switch into at least two traffic classes implemented in separate physical or logical queues. Guaranteed traffic is placed in one queue and best-effort in the other. The queue service algorithm always serves the guaranteed traffic in preference to the best-effort traffic [22, 34]. More complex queueing structures and service algorithms are investigated in [24, 37] but it is not clear that the enhanced performance justifies the increased complexity.

5.3 Burst Buffering

The dimensioning of the cell buffers for the guaranteed, constant bit-rate traffic in an ATM switch is related to the jitter in cell arrival for traffic of known (and enforced) characteristics. This may be achieved with relatively small buffers — typically some hundreds of cells, either per port or shared across a number of ports. To offer a best-effort service for statistical traffic, buffering needs to be provided in relation to the burst (or packet) arrival statistics of the traffic. This will require much larger buffers — several megabytes of buffering is common in current bridges and routers. Thus burst buffering cannot be implemented in static RAM within the switching elements that form the switch fabric of many current switch designs. Burst buffering may require DRAM or video RAM and is more likely to be implemented in the port cards or between groups of port cards and the switch fabric.

5.4 Congestion Control

A best-effort service must permit stations to contend dynamically for access to a pool of shared bandwidth. In an IEEE 802 LAN the shared medium provides the shared bandwidth and the MAC sublayer provides arbitration. In an ATM switch, each output port is a pool of shared bandwidth, regardless of switch capacity.

If the best-effort service is carrying data protocols that use a window flow control mechanism there will be a limit on the amount of data that any connection can inject into the network. This will permit the burst buffers to be dimensioned according to the number of active connections that may be supported [13]. As link bandwidths increase, the size of the window must also increase to maintain high throughput which will reduce the number of active connections that may be supported for a given buffer size. Also, it would be unwise for the network to rely on correct sizing of the user's window to maintain an acceptable quality of service. Some form of congestion control scheme must be implemented in an ATM local area network to support the statistical sharing of bandwidth between competing stations without prior bandwidth reservation.

Three fundamental approaches are available for congestion control: over-provisioning in terms of bandwidth or buffers; loss mechanisms; and delay mechanisms.

Over-Provisioning: When Ethernet was first introduced, 10 Mbits/s shared between all stations seemed like an infinite amount of bandwidth. Yet currently a bandwidth of 10 Mbits/s per end station is a popular goal. Clearly bandwidth over-provisioning has its limits as a congestion control mechanism. Over-provisioning in terms of buffering also has finite limitations determined by delay. If the delay through the buffer exceeds the retransmission timeout of the higher layer protocols, additional retransmission traffic will be inserted into the network during a period of congestion [26, 33].

Loss Mechanisms: Loss mechanisms discard traffic during periods of congestion. One approach uses the cell loss priority bit in the cell header to discard low priority traffic in preference to high priority traffic when the buffer length exceeds a threshold [30, 2, 28]. This can be useful for sources that can code their information into multiple priority levels, such as the high and low definition components of a video signal. But it is difficult to see how data traffic could be coded into two loss priorities at the MAC sublayer to make use of this mechanism. If the loss mechanism simply discards traffic when the buffer overflows, each discarded cell is likely to belong to a different packet. Therefore many packets will require retransmission and bandwidth is wasted by the onward transmission of the remaining cells from corrupted packets. Simulation studies of TCP over ATM with a simple cell discard congestion mechanism have shown significant throughput degradation and high levels of packet retransmission even though TCP has an internal congestion control mechanism [38, 17].

An alternative loss mechanism requires that the peak transmission rate of each virtual connection be declared when it is established and that the end of each burst be marked in the cell header. Each output port maintains an indication of the instantaneous sum of the peak traffic rates arriving at the output port. If an arriving burst causes the sum to exceed the link capacity the burst is discarded [43]. This scheme has the advantage of discarding entire packets but requires each buffer to maintain a count of the number of cells it contains for each virtual connection.

A similar scheme is to randomly select active virtual connections for discard when the buffer exceeds a threshold, e.g. [18]. Such a scheme will also discard complete packets rather than random cells and may have a simpler implementation. Loss schemes that drop entire packets have a much better performance than random cell loss schemes as fewer packets require retransmission and the remaining cells from corrupted packets are not transmitted beyond the point of congestion.

Delay Mechanisms: Delay mechanisms use negative feedback from the point of congestion back towards the source to reduce the amount of traffic entering the network. Forward explicit congestion notification (FECN) sends a congestion indication along the forward data path to the destination [1, 46]. The destination then takes some action to cause the source to reduce its transmission rate such as closing a window in a higher layer protocol or sending an explicit signal. Backward explicit congestion notification (BECN) sends the congestion indication directly back to the source along a return path [35]. On receipt of this indication the source reduces its transmission rate directly. BECN can respond to congestion much more rapidly than FECN but requires congestion notification cells to be inserted into the network whereas FECN can simply mark a bit in the cell header as it passes through the point of congestion. Other delay mechanisms have been proposed that use credit or backpressure on each virtual connection on a link-by-link basis between switches and ultimately back to the source, e.g. [27, 29, 8]. This approach offers much tighter flow control but is considerably more complex to implement.

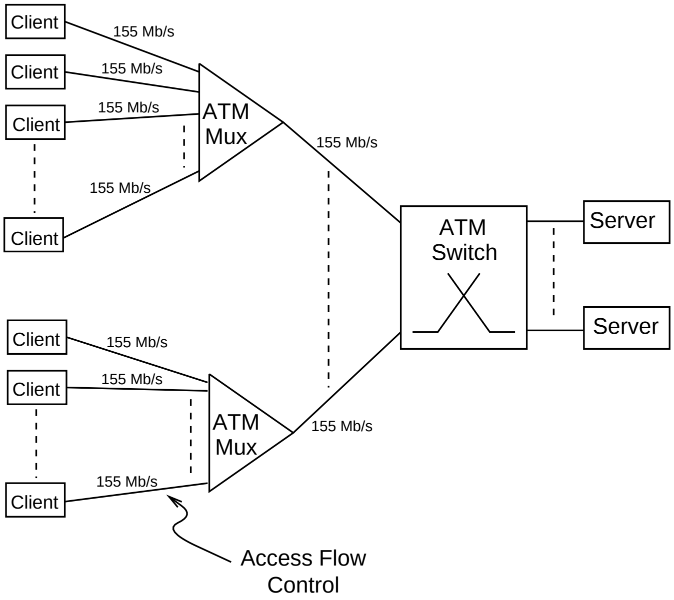

Figure 10: Access flow control.

5.5 Access Flow Control

ATM is only one of a number of technologies capable of offering high performance local area networking. To achieve general acceptance in the commercial world, ATM must attain a cost per port comparable with competing technologies. One simple technique to reduce the cost per port is to multiplex multiple stations onto the same ATM switch port, fig. 10. This is not unreasonable. An FDDI ring, for example, multiplexes all of the stations into a shared medium of 100 Mb/s capacity. Offering the capability to multiplex up to 8 or 16 stations onto each 155 Mb/s switch port allows the user a flexible degree of concentration to adjust the cost of each ATM interface to the performance required. Data traffic is extremely bursty. The proportion of time each user requires the full 155 Mb/s is likely to be very small. So for the majority of commercial data and interactive applications a moderate degree of concentration at the access ports is unlikely to be noticed by the user.

To support the ability to multiplex multiple 155 Mb/s ATM access ports onto a single 155 Mb/s ATM switch port we require an access flow control mechanism. A simple start/stop mechanism for the entire best-effort traffic class on each access link is all that is required. The guaranteed traffic does not require access flow control since concentration is taken into account during the call acceptance process and the entire call will be rejected if insufficient bandwidth is available. (Access flow control for the guaranteed class of traffic may, however, be useful to coordinate traffic arrival and to prevent buffer overflow in small input buffers.) The generic flow control (GFC) function in B-ISDN is designed to offer flow control across an access link. At present, the standards bodies seem to be converging on a minimum functionality that will permit flow control of the best-effort traffic class using two of the GFC bits in the cell header.

Access flow control and congestion control are different functions. Both are required. Access flow control operates on each entire traffic class on the access link as a single unit (or at least on the best-effort traffic class). It is not selective between different virtual connections within a traffic class because all connections pass through the same bottleneck — the multiplexed switch port. Congestion control however, attempts to share resources within the network between competing users. To achieve this it must operate on a per virtual connection basis because only the virtual connections passing through the point of congestion require flow control.

5.6 Multicast

There are two classes of multicast connection: one-to-many; and many-to-many. (Many-to-one connections may be regarded as a subset of many-to-many connections.) LAN emulation requires the provision of many-to-many ATM connections. A many-to-many connection is a multipoint connection, with a single group address, on which any member of the group may transmit and all members receive. Some implementations require the use of a multicast server and others may use the multicast capabilities of an ATM switch directly. In the server implementation each group member establishes a unicast connection to the server, and a one-to-many connection is established from the server to the group.

The server implementation must be employed if AAL 5 is used on the multicast connection. This is because AAL 5 can only perform reassembly on virtual connections originating from a single source. If multiple sources transmit simultaneously on the same multicast virtual connection, their cells will become interleaved and AAL 5 provides no mechanism to reassemble such a cell stream. The multicast server is used to resequence the cell stream so that cells from different sources are not interleaved on any multicast connection [44]. If AAL 3/4 is used for multicast connections the multiplexing identifier (MID) field may be used to identify the source thus no server is required to act as a relay. The use of AAL 3/4 for multicast connections will limit the number of members of an ATM LAN segment to less than 1024 and will require the network to assign MID values. An alternative that permits the use of AAL 5 without requiring a server is to establish a one-to-many connection from each member of the group to all others. This approach rapidly consumes VCIs and is much more difficult to manage because all of the one-to-many connections need to be updated whenever a member joins or leaves the group.

6 Conclusion

ATM technology is currently being applied to local and campus area networking to offer greatly increased bandwidth and to support broadband services. However, it must interwork with the existing installed base of LANs, bridges, routers, and protocols. While an interface to ATM could be offered from the transport layer or the network layer of the OSI model, such interfaces would be protocol specific. To offer general compatibility, regardless of the network and upper layer protocol stack, and to support transparent MAC bridging, an interface at the MAC sublayer is required. To avoid the requirement to modify the protocol stack in every end station, this ATM MAC sublayer should emulate the service offered by an IEEE 802 LAN. Thus the ATM MAC sublayer should offer a best-effort, connectionless, datagram transfer service.

While a connectionless service may be offered by a connectionless server, this approach denies the full benefits of ATM and requires substantial hardware in addition to the ATM switch. A solution based upon a full mesh of semi-permanent virtual connections may be adequate for a very small network but becomes increasingly difficult to manage as the number of end stations increases. An approach based upon switched virtual connections offers the most flexible solution.

LAN emulation using switched virtual connections introduces the requirement for address resolution to locate the destination end station followed by connection establishment to the resulting ATM address. Address resolution may be implemented by a broadcast technique or by an address server. The broadcast method is the simplest for small networks but a combination of both techniques is probably the better approach. The use of the MAC address to identify an end station, with dynamic binding to its current physical location in the ATM network, allows the ATM LAN segment to be viewed as a virtual LAN. The virtual LAN model permits end stations to move and change physical location while maintaining connection to the same ATM LAN segment. This greatly simplifies the management of large data networks.

The support of LAN emulation imposes certain requirements on the design of an ATM switch not necessarily found in switches designed for the public B-ISDN service. The bandwidth instantaneously available to the best-effort service on each output port must be dynamically shared between all contending users. This requires a substantial amount of burst buffering to absorb short bursts of traffic arriving in excess of the available output port capacity. Longer bursts require a feedback control mechanism. A switch designed to offer LAN emulation will also be required to support concentration for data traffic to offer the user flexibility in matching the performance requirements to the ATM port cost. This will require an access flow control mechanism.

Many technical and administrative issues remain to be solved and agreed upon in the application of ATM technology to the public network (e.g. traffic management, best-effort service, tariffing). However, there is a growing demand for high bandwidth networking in the local and campus area that can be well satisfied using ATM technology for LAN emulation. Solutions exist to the technical problems introduced by LAN emulation and the issues of standardization, interworking, and vendor interoperability are currently being addressed in the standards bodies and by the ATM Forum.

Acknowledgement

The architecture of N.E.T. Adaptive's LAN emulation service was designed by John Burnett to whom I am grateful for many patient and detailed explanations of the design issues.

References

[1] O Aboul-Magd and H Gilbert. Incorporating congestion feedback in B-ISDN traffic management strategy. In *Proc. Int. Switching Symp.*, volume 2, pages 12–16, Yokohama, Oct. 1992.

[2] K Bala, I Cidon, and K Sohraby. Congestion control for high-speed packet switched networks. In *Proc. IEEE Infocom*, volume 2, pages 520–526, San Francisco, Jun. 1990.

[3] E Biagioni, E Cooper, and R Sansom. Designing a practical ATM LAN. *IEEE Networks Mag.*, pages 32–39, Mar. 1993.

[4] J-Y Le Boudec, E Port, and H L Truong. Flight of the FALCON. *IEEE Commun. Mag.*, pages 50–56, Feb. 1993.

[5] P Boyer. A congestion control for the ATM. In *Proc. 7th ITC Seminar*, Morristown NJ, Oct. 1990.

[6] P E Boyer and D P Tranchier. A reservation principle with applications to the ATM traffic control. *Computer Networks and ISDN Systems*, 24:321–334, 1992.

[7] J Burgin and D Dorman. Broadband ISDN resource management: The role of virtual paths. *IEEE Commun. Mag.*, pages 44–48, Sep. 1991.

[8] J Cherbonnier, J-Y Le Boudec, and H L Truong. A GFC protocol for congestion avoidance in the ATM connectionless service. In *Proc. EFOC/LAN*, Paris, Jun. 1992.

[9] G H Clapp. LAN interconnection across SMDS. *IEEE Network Mag.*, pages 25–32, Sep. 1991.

[10] P Crocetti et al. ATM-based SMDS for LANs/MANs interconnection. In *Proc. Int. Switching Symp.*, volume 1, pages 87–91, Yokohama, Oct. 1992.

[11] P Crocetti, G Gallassi, and M Gerla. Bandwidth advertising for MAN/ATM connectionless internetting. In *Proc. IEEE Infocom*, volume 3, pages 1145–1150, Miami, Apr. 1991.

[12] S Crosby. In-call renegotiation of traffic parameters. In *Proc. IEEE Infocom*, volume 2, pages 638–646, San Francisco, Mar. 1993.

[13] S P Davis, T Worster, and W Fischer. Resource allocation for packet data traffic on ATM: Problems and solutions. *Submitted to Int. Switching Symp., Yokohama*, Oct. 1992.

[14] D Deloddere, P Reynders, and P Verbeeck. Architecture and implementation of a connectionless server for B-ISDN. In *Proc. Int. Switching Symp.*, volume 1, pages 224–228, Yokohama, Oct. 1992.

[15] ITU-TS draft Recommendation Q.93B. B-ISDN User-Network Interface Layer 3 Specification for Basic Call/Bearer Control, May 1993.

[16] H Esaki. Call admission control methods in ATM networks. In *Proc. IEEE Int. Conf. Commun.*, volume 3, pages 1628–1633, Chicago, Jun. 1992.

[17] C Fang, H Chen, and J Hutchins. Simulation results of TCP performance over ATM with and without flow control. ATM Forum/93-932, Stockholm, Nov. 1993.

[18] S Floyd and V Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Trans. Networking*, 1(4):397–413, Aug. 1993.

[19] ATM Forum. ATM user-network interface specification. Version 3.0, Sep. 1993.

[20] V S Frost and M T Mullen. Dynamic bandwidth allocation for B-ISDN based on end-to-end delay estimates. In *Proc. IEEE Int. Conf. Commun.*, volume 1, pages 225–231, Chicago, Jun. 1992.

[21] R Guerin, H Ahmadi, and M Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE J. Select. Areas in Commun.*, 9(7):968–981, Sep. 1991.

[22] T-C Hou and A K Wong. Queueing analysis for ATM switching of mixed continuous bit-rate and bursty traffic. In *Proc. IEEE Infocom*, volume 2, pages 660–667, San Francisco, Jun. 1990.

[23] D-P Hsing, W Leland, and D V Wilson. On the sustainable cell rate and burst tolerance traffic parameters proposed in the ATM Forum. ANSI T1S1.5/93-113, Somerset NJ, Jun. 1993.

[24] J M Hyman, A A Lazar, and G Pacifici. Real-time scheduling with quality of service constraints. *IEEE J. Select. Areas in Commun.*, 9(7):1052–1063, Sep. 1991.

[25] S Isaku and M Ishikura. ATM network architecture for supporting the connectionless service. In *Proc. IEEE Infocom*, volume 2, pages 796–802, San Francisco, Jun. 1990.

[26] R Jain. Congestion control in computer networks: Issues and trends. *IEEE Networks Mag.*, pages 24–30, May 1990.

[27] M G H Katevenis. Fast switching and fair control of congested flow in broadband networks. *IEEE J. Select. Areas in Commun.*, 5(8):1315–1326, Oct. 1987.

[28] H Kröner et al. Priority management in ATM switching nodes. *IEEE J. Select. Areas in Commun.*, 9(3):418–427, Apr. 1991.

[29] H T Kung. The FCVC (flow-controlled virtual channels) proposal for ATM networks. In *Proc. Int. Conf. Network Protocols*, San Francisco, Oct. 1993.

[30] B A Makrucki. A study of source traffic management and buffer allocation in ATM networks. In *Proc. 7th ITC Seminar*, Morristown NJ, Oct. 1990.

[31] B A Makrucki. An analysis of the fast reservation protocol. ANSI T1S1.5/91-404, Dallas, Nov. 1991.

[32] L Mongiovi, M Farrell, and V Trecordi. A proposal for interconnecting FDDI networks through B-ISDN. In *Proc. IEEE Infocom*, volume 3, pages 1160–1167, Miami, Apr. 1991.

[33] J Nagle. On packet switches with infinite storage. *IEEE Trans. Commun.*, 35(4), Apr. 1987. (Also available as Internet RFC 970).

[34] P Newman. A fast packet switch for the integrated services backbone network. *IEEE J. Select. Areas in Commun.*, 6(9):1468–1479, Dec. 1988.

[35] P Newman. Backward explicit congestion notification for ATM local area networks. In *Proc. IEEE Globecom*, volume 2, pages 719–723, Houston, TX, Dec. 1993.

[36] S Ohta and K-I Sato. Dynamic bandwidth control of the virtual path in an asynchronous transfer mode network. *IEEE Trans. Commun.*, 40(7):1239–1247, Jul. 1992.

[37] J M Peha and F A Tobagi. Implementation strategies for scheduling algorithms in integrated-services packet-switched networks. In *Proc. IEEE Globecom*, volume 3, pages 1733–1740, Phoenix, AZ, Dec. 1991.

[38] A Romanow. TCP over ATM: Some performance results. ATM Forum/93-784, Ottawa, Canada, Jul. 1993.

[39] H Saito. Call admission control in an ATM network using upper bound of cell loss probability. *IEEE Trans. Commun.*, 40(9):1512–1521, Sep. 1992.

[40] K Sato, H Ueda, and N Yoshikai. The roll of virtual path cross-connection. *IEEE Lightwave Telecom Systems Mag*, pages 44–54, Aug. 1991.

[41] S L Sutherland and J Burgin. B-ISDN interworking. *IEEE Commun. Mag.*, pages 60–63, Aug. 1993.

[42] D L Tennenhouse. Layered multiplexing considered harmful. In H Rudin and R Williamson, editors, *Protocols for high-speed networks*. Elsevier Science Pubs. B.V. (North-Holland) IFIP, 1989.

[43] J S Turner. Managing bandwidth in ATM networks with bursty traffic. *IEEE Networks Mag.*, pages 50–58, Sep. 1992.

[44] L Wei et al. Analysis of a resequencer model for multicast over ATM networks. In *Proc. 3rd Int. Workshop on Network and OS Support for Digital Audio and Video*, San Diego, Nov. 1992.

[45] K Yamazaki and Y Ikeda. Connectionless cell switching schemes for broadband ISDN. In *Proc. Int. Switching Symp. (ISS '90)*, volume 6, pages 5–10, Stockholm, May 1990.

[46] N Yin and M G Hluchyj. On closed-loop rate control for ATM cell relay networks. Submitted to IEEE Infocom, 1994.