

Protocols for High Speed Networks: Life After ATM?

James P. G. Sterbenz

+1 617 466 2786

[jpgs@{acmliee}.org](mailto:jpgs@acmliee.org) jpgs@gte.com

<http://info.gte.com/jpgs/jpgs.html>

GTE Laboratories Incorporated

40 Sylvan Road, MS-61

Waltham, MA 02254 USA

Abstract

This paper will provide a brief history of networking, noting the transition from disjoint network infrastructure and media to the emerging integrated broadband networks. The current state of affairs and significant challenges in deploying a broadband Global Information Infrastructure based on ATM and B-ISDN will be discussed. Even if we solve all the hard technical problems and practical challenges associated with this, all we have provided is an integrated network infrastructure that is capable of transporting high bandwidth. We still have not solved the latency problem (bandwidth \times delay product) for WAN applications, or provided for the delivery of this bandwidth to the applications through the host architecture and operating systems. The sorts of emerging applications we must support, and some of the challenges, issues, and areas of research that remain at the end system to support these new applications will be described. Finally, some of the research directions to be pursued will be briefly touched upon.

Keywords: Protocols for High Speed Networks, ATM, B-ISDN, Gigabit Networks

1. A Brief History of Networking

The history of networking and communications can be divided it into three generations, which have distinctly differing characteristics in terms of the scope of users and applications, and the integration of differing applications and media.

1.1 First Generation

The first generation lasted through roughly the 1970's and is characterised by three distinct categories: voice communication, entertainment, and data networking, each of which was carried by a different infrastructure. Voice communication was either analog circuit switched wireline telephony or analog radio transmission. Entertainment was carried by free space broadcast radio and television. Data communications was the latest entrant, and provided only a means to connect terminals to a host, either by serial link local communications (such as RS-232 or BSC) or by modem long haul connections over telephone lines for remote access.

1.2 Second Generation

In roughly the 1980's a dramatic jump in the types and scope of networking occurred, but the three categories of communication (voice, entertainment, and data) remained relatively distinct. This period took us from the experimental ARPANET to the ubiquitous Internet.

While the end user of the voice network generally continued to use analog telephone sets, the internal network switches and trunks became largely digital. Additionally, there was widespread deployment of digital PBXs (private branch exchange telephone switches). Mobile communications emerged in form of cellular telephony.

There were two significant additions to the entertainment category: the wide scale deployment of cable networks for entertainment video, and the emergence of BBS (bulletin board systems) and consumer online services (such as America Online, CompuServe, and Prodigy).

The growth in data networking during this period was significant, incorporating file transfer, remote login, and electronic mail. In the local area, shared media Ethernet and token ring networks allowed clusters of workstations and PCs to network with file and compute servers. In the wide area, store-and-forward packet routers formed the backbone of the Internet, with wide scale research and university access.

At the same time, but relatively distinct from this, connection oriented corporate networks using protocols such as BNA, DECNET, and SNA were deployed, along with the deployment of public X.25 networks (used primarily as corporate virtual private networks). Thus, even within data networking there were multiple incompatible architectures and poorly interconnected networks.

1.3 Third Generation

The first distinguishing characteristic of the third generation, which we are just beginning to enter, is an integration of the voice, entertainment, and data networking categories. Integrated media communication (data, voice, and video) runs over the same fast cell switched networks in LANs, MANs, and WANs, which will further enable applications utilising mixed media.

The second characteristic is in the scope of access: universal consumer access to the GII (global information infrastructure). This is beginning to occur with the emergence of public access internet service providers, as well as the gatewaying of the consumer online services to the Internet. The Internet is currently seeing an explosion in the number of users due to this, and some applications, such as Usenet news are becoming visibly stressed.

2. B-ISDN, ATM and Broadband Networking

The broadband network infrastructure will be constructed using B-ISDN and ATM as the core technology. We will briefly examine what B-ISDN and ATM are, what some of the desir-

able and undesirable characteristics are, and why they will succeed as the underlying technology for the broadband Global Information Infrastructure.

2.1 B-ISDN and ATM

B-ISDN (broadband integrated services digital network) is relatively easy to define: emerging ITU and ANSI standard compliant networks which support rates in excess of the N-ISDN (narrow band ISDN) primary rate of 64 kbps data channels, and use asynchronous transfer mode cell relay as the underlying network technology. B-ISDN tends to imply emphasis on public networks.

ATM (asynchronous transfer mode) is technically just the statistical multiplexing scheme of packets or fixed size cells, but in common use refers to the emerging switches, networks, protocols, and control mechanisms that use ATM relay of small fixed size cells for the network layer transport of data. In particular, ATM is used in the context of cell relay LANs in addition to wide area public networks.

So in general, B-ISDN and ATM are similar in use, and really refer to broadband networking based on ATM cell relay. Due to the over use and misuse of "ATM", the term "broadband networking" will be generally used after this section.

2.2 Desirable Characteristics of B-ISDN and ATM

There are a number of desirable characteristics of B-ISDN and ATM, based on fast cell switching technology. The network permits an arbitrary mesh topology of point-to-point links between switches. This allows the network to be flexible, scalable, and fault tolerant. The switches can be designed to be highly scalable, and growing the network consists of scaling the switches and corresponding links without the constraints of shared medium networks, such as rings or busses. We can use an arbitrary topology to construct giganode networks using the hierarchical organisations that will be required to reasonably address and manage them.

B-ISDN and ATM networks are connection oriented, allowing QoS (quality of service) guarantees (which is not to say either that this is easy to do in practise, or that QoS guarantees are impossible in connectionless networks). The routing overhead is performed *via* connection setup, allowing the *per* cell routing to be done easily in hardware as a simple table lookup. Finally, datagram service can be provided over a permanent connection set overlay for the purpose.

A significant advantage of B-ISDN and ATM networks is that the standards have been designed to scale (upwards) in data rates. SONET OC-3c (155 Mbps) and OC-12c (622 Mbps) are standard now, and OC-48c (2.4 Gbps) and beyond (OC-192c, OC-768c, *etc.*) will be trivial to standardise. This should be contrasted with the difficulty in arriving at agreement to a single order of magnitude increase in the Ethernet standards to 100 Mbps.

2.3 Undesirable Characteristics of ATM and B-ISDN

There are, however, a number of *undesirable* characteristics of B-ISDN and ATM, primarily due to the current standards. The biggest offender is the small cell size of 48 octets payload. This was a compromise between two small proposed cell sizes of 32 and 64 bytes, chosen by the voice community to avoid the need for echo cancellation. The fact that time scales with data rate seems to have escaped the decision making process, and we have an *almost* unworkably small cell size. In fact we may have been better off in the long run if the 16B proposal would have initially won; it would have been so small to really be unworkable and cell size could have been revisited early on.

The problem usually cited with small cell size is header overhead, which although not optimal is not the main problem. The worst effect of a small cell is the resultant short cell time (681ns at OC-12c), which require fast hardware to make decisions about the cell payload. If the cell time were somewhat longer (128+16B cell resulting in 1.8 μ s for OC-12c, for example), more could be done in microcontroller software or cheaper hardware. This has pushed the availability of ATM chips, particularly for the host-network interface segmentation and reassembly, back at least two years for higher data rates. Furthermore, network interface memory design and buffering strategies would be easier to implement if the payload length were a power of two.

Another problem is the decision to require sequenced delivery and omit a sequence number from the header (necessary, given the small cell size; the AAL3/4 sequence number is intended only for error detection but not for resequencing). While it is possible to design switches and routing algorithms to enforce all cells in a stream to follow the same path, this considerably restricts the design space. Furthermore, there are a number of error scenarios which result in effective missequencing, such as lost cells and path reconfiguration. These would be easier to deal with if sequence was not required. Finally, there are a number of cases where higher bandwidth can be obtained by striping of cells across parallel links (such as when multiplexed OC-12 but not OC-12c circuits are available). This requires a mechanism for preserving cell sequence across the parallel lines, but maintaining skew within a cell time is difficult, even per hop, and thus makes this scheme hard to implement in practice.

2.4 Why B-ISDN and ATM Will Succeed

In spite of the undesirable characteristics imposed by ATM and B-ISDN standards, it seems clear that "ATM" (as loosely defined above) and B-ISDN will be the basis of the future broadband infrastructure. In spite of particular technical problems with the B-ISDN and ATM standards, the fact that there *is* a standard that had been agreed on for the physical, link, and network layers is significant. The whole point of networking is interoperability, and this outweighs technical compromises and defects. A wide range of network component manufacturers and service providers are doing, or planning to do, B-ISDN and ATM, in many cases prematurely. This is really a double-edged sword; it is this same premature embracing of "ATM" and perception of agreement that is driving us to agree on the final solution and stan-

dards. The Global Information Infrastructure means the end of proprietary network technologies to any large extent.

Having said this, we should note again that there is much to still be defined. The easy problems such as cell format, segmentation and reassembly, and physical layer protocols have been defined; while the difficult issues of multipoint virtual connection routing and traffic management are far from settled. We also shouldn't be surprised to see some revisions in the current standards. The problems with a small cell size can be overcome by a large cell definition, cell size that scales with data rate or cell groups. The connection oriented nature allows this to be a connection parameter. Similarly, a larger cell payload would allow a larger header to be used for a sequence number without further sacrificing header efficiency.

There are also RISC-like arguments that come to bear. Some of the proposed traffic management and connection routing schemes are incredibly complex (and will get worse when multipoint-to-multipoint connections are fully worked out) with the goal of obtaining optimal bandwidth usage. While bandwidth will not be free, it will become sufficiently cheaper that we need to consider what effort is worth taking to conserve it in the broadband environment, rather than basing our efforts on the cost of bandwidth in the current narrowband environment. We may see far more over engineering with simpler control mechanisms than many would now like to admit.

We also should observe that although the connection *vs.* datagram and ATM *vs.* IP arguments have frequently taken the proportions of a "holy war", there will eventually be peace. In a connection oriented "ATM" world, datagrams will have to be supported; this is currently being pursued as ABR (available bit rate) traffic. In the IP world, state will be introduced in the packet routers to allow QoS (*e.g.* [ZhDe93], [ClSh92]). Both camps are trying to solve the same problem, and the solutions will converge. The Global Information Infrastructure will be based on a B-ISDN/ATM substrate with much of the flavor of existing IP networks and applications (which is *not* to say that it will be *current* IP over *current* ATM).

3. Challenges in Deploying the Broadband Infrastructure

Even though we have general agreement on ATM cell relay as the core transport technology for the emerging broadband infrastructure, there are a number of problems that have yet to be solved and worked into the standards. These technical and practical challenges pose non-trivial barriers to the deployment of a broadband Global Information Infrastructure.

3.1 Technical Problems

There are a number of very hard problems to be solved to deploy B-ISDN and ATM networks on a large scale. Virtual connection routing is difficult due to the need to find a path with available resources to meet the users requested quality of service parameters, while optimising network resources and load balancing. The general case for dynamic multipoint-to-multipoint

connections is very challenging indeed, and will require routing heuristics. There are also a number of open issues regarding signalling between switches within the network (the NNI: network node interface). Added to this are the problems of mobile addressing and the associated dynamic routing and rerouting, which will be increasingly important.

Whereas using a rate or flow specification to ensure an application's compliance to the traffic contract makes it easier to design network interfaces and to engineer switches and buffer sizes, it places the burden of determining a reasonable set of parameters on the application or user. Although this may be a simple matter for some well behaved or well understood applications such as voice or uncompressed video, it will be exceptionally difficult for other applications, such as for general case multimedia interactive applications or distributed computing.

The ABR (available bit rate) traffic class is targeted toward this, providing an IP-like datagram service over ATM networks. Unfortunately, there are a number of cases where this will be inadequate. As an example, a user requesting a file transfer using FTP may tolerate some variance in the response time it takes to get the file, and thus in the value of the average rate. On the other hand, if the file transfer is due to a click on a WWW (world wide web) browser (*e.g.* mosaic) hyperlink, then we would like to have subsecond response time (actually the first window should be available in $O(100\text{ ms})$), and thus a strict bound on the rate can be computed based on size of the first window. So even in the current IP world, communication that has been best effort will be demanded with sub-second response times by applications such as the WWW.

3.2 Unexpected Behavior

We should note that networks are extremely complex *systems of systems*, and frequently do not exhibit the behavior we would like. A significant example is the fractal or self-similar nature of network traffic that has recently been observed [LeTa94]. The assumption in broadband networking has been that many bursty sources would aggregate to relatively uniform traffic, with the corresponding benefits in reduced network resources. The observation of self-similar traffic that is bursty over several orders of magnitude in time questions that assumption. Another problem can arise in network control algorithms, an example being the systemic self-synchronisation of routing messages in IP networks [FlJa94]. This is a case of control mechanisms being difficult to predict and tune in very large scale networks. Furthermore, we will face problems as we attempt to glue implementations of existing control mechanisms together and their interaction is hard to predict (*e.g.* [MoGu95]).

3.3 Practical Challenges

There are a number of additional challenges of a more practical nature, for which the technical solutions are not easy. Network management has always been a difficult problem, but will become much more difficult in multiservice multirate giganode broadband networks.

Security, privacy, and authentication services will have to be provided. This is a standard concern in networks which trades against the desire to share information. This is not a new problem, but it is of increasing concern as the number of users having access to networks (including crackers) increases, as does the desire to conduct commerce and business across an integrated network infrastructure. Electronic fund transfers will no longer have the (perceived) physical security of private networks controlled by the banks with switches controlled by telephone carriers.

The architecture of the network and switches to allow the provision of content and the deployment of new services and application, some within the network, and some by external service providers will become an increasingly challenging problem. The current efforts in IN (intelligent networks) targeted toward the telephone network will not be sufficient for emerging integrated broadband networks, and many pieces of the current IN architecture will be subsumed by broadband signalling and multipoint connection routing. There will be a greater need for the structural and architectural aspects of IN to determine where in the network and switches function resides, and to allow the rapid deployment of new services and applications without replacing infrastructure.

These practical issues are complicated by the wide range of host and terminal equipment intelligence we should expect for a long time to come. Some of this will be legacy, but a typical home may have an large HDTV set in the living room whose primary interface is not a keyboard, a PC which also can display entertainment video, and a number of relatively dumb telephone sets.

3.4 Compatibility, Standards, and the Seamless Network

The ability to maintain backward compatibility with existing protocols and host software is critical, but can impose technical compromises on new protocols and architectures. We will have to live in a world of multiprotocol and legacy networks for a long time to come. Existing transport protocols, such as TCP and TP4 will be required to operate over new network infrastructure just to support the existing applications which use them. IP has served as a powerful unifying layer for heterogeneous sub-networks, and will also have to be supported for the foreseeable future (and indeed there will be an IP flavor to the ultimate meld of the "ATM" and "IP" worlds). Any solutions will have to balance the ability to enhance the performance and services of these existing protocols, while providing the new performance and functionality desired for emerging applications.

Similarly, standards are particularly necessary in the network world, since they provide the means to allow interoperation, but the premature adoption of poor or overly rigid standards can do more harm than good. It is a very tricky business to go from research and experimentation, to preliminary standards, to solid standards to which products can be designed.

The B-IDSN and ATM standards originally assumed a strict and relatively simple hierarchical network structure, with a public backbone network at the top. The current internetworking

reality is far from this, due to the post-divestiture structure of the US telephone network, and the structure of the current Internet. The vision of a seamless LAN-MAN-WAM ATM infrastructure will not be realised for some time to come. Furthermore the need for policy based routing complicate the original assumptions.

3.5 Economic Issues

There are certainly some components of the network infrastructure that are a natural monopoly, *e.g.* we expect controls on who can dig up the street with a backhoe to provide the last mile links. On the other hand, we are moving to a more competitive world in the provision of host and terminal equipment, as well as services and network content. In the middle there is the network infrastructure, and to what degree it should be regulated.

The emerging network infrastructure should make it easier on the user, but it appears that delivery to the home will be available *via* telephone lines, CATV cable, cellular and PCS communications, satellite feeds, and possibly fiber in the power transmission lines. The question arises of how many distinct network outlets should be “in the wall”. We will have to decide if the connectivity between the various local access networks is sufficient or if multiple feeds will be required with a local gateway or switch in the home. There are also a number of standard issues here relating to new technology, including providing reasonable incentives to deploy infrastructure, particularly given regulatory constraints.

Billing is considerably more difficult for broadband integrated networks than for the case of constant rate or leased line communication. It must be based on the resources required for the network, which means that in addition to the amount of data transferred and the peak rates required, it will have to be based on traffic characteristics such as burstiness. Furthermore, for broadband networks to be a success, they will have to be affordable, and the pricing will have to be non-linear, but this means that providers of bandwidth will face the reselling of fractional bandwidth. Incremental pricing of excess network capacity (*e.g.* to improve performance by prefetching or provide greater video resolution) may be a useful concept, but difficult to implement.

3.6 ATM Hype

One of the biggest problems with deploying the broadband network infrastructure is really a meta-problem. There has been so much hype with ATM, that everyone seems to want to be involved, whether or not they have anything to contribute. Although the deployment of test-beds should be encouraged, there will be overdeployment of products and networks that are not really ready for production use due to the immaturity of the standards and problems that have yet to be solved. We would have been much better just talking about how to deploy broadband technology, without creating this thing called “ATM” that everyone wants to sell and deploy, frequently prematurely. There are certainly niches where ATM *can* be deployed now, such as in some LAN environments as in backbone networks with permanent connections, but this is different from wide scale deployment in all environments. Furthermore, the

users don't really care what the underlying protocols and technologies are, but rather only that the applications work and perform as desired.

4. What We Are Trying to do

Now that we have discussed the history of networking, and the current trends in B-ISDN and ATM development, we should consider what the real goals of high speed networking are:

Deliver high bandwidth at low latency to the applications that require it.

Assuming that we are able to deploy a broadband infrastructure, and have solved all of the technical challenges described, where has that gotten us? All we have accomplished is to be able to deliver high bandwidth at the network layer. Specifically we still have to deal with the speed of light latency (the bandwidth- \times -delay product), and we have to deal with the transport and higher layer protocols to deliver the bandwidth to applications. This leaves us plenty to accomplish with high speed protocol research *even* if ATM/B-ISDN is a solved problem.

4.1 Application Requirements

There are two reasons to need high bandwidth: aggregation and *per* application bandwidth.

The current network infrastructure has been estimated by several sources to be $O(1 \text{ Tbps})$. As an example, the aggregate bandwidth required during prime time in the US for interactive HDTV video on demand is $O(10 \text{ Pbps})$ [NuPa94]. Thus it is clear that we will need a broadband infrastructure for emerging applications in aggregation. This affects the internal network design and network layer protocols more than it does the end nodes or higher layer protocols.

For the higher layers and end systems, we should consider how much bandwidth a single application can use. At the extreme end, a multi-sensory, distributed, projected three-dimensional, dynamic virtual reality session ("holodeck") would require extremely high bandwidth (It would be a useful exercise to compute the bandwidth required for this).

A more practical example is the use of the WWW (world wide web) and associated browsers (such as mosaic), which is likely to be the killer application we've been waiting for [To94]. While the network bandwidth currently supported is rather low, so is the corresponding response to files and images that are fetched by clicking on hyperlinks. It is not hard to imagine scenarios in which bursts of $O(100 \text{ Mbps})$ would be desired by a single user, particularly when large blocks of data are fetched to render into high resolution images [Pa94].

We also need to ask what latency is tolerable. The WWW currently behaves as a request/response system, and the standard sub-second response time arguments apply. Thus, from the time a user clicks on a hyperlink to the time that the screen is full of whatever was requested should not exceed $O(100 \text{ ms})$. Various estimates on the network topology and switch latency, coupled with the distance at most half way around the world give us a network latency of

$O(10\text{--}100\text{ ms})$. Thus we are potentially close to the $O(100\text{ ms})$ limit, without the transmission time of the data and the processing time on the local and server hosts.

So, we must at *least* deliver the right piece of information to the user with low latency at the end systems. At best, we should predict and prefetch data that will likely be requested next (*e.g.* breadth first using all the hyperlinks in the current page [To94]). We need to recognise that like memory, bandwidth is a resource, and idle bandwidth can be used to our advantage to effectively beat the speed of light.

The WWW also gives us a preview of some of the other higher layer issues that will need to be addressed. For example, as useful as URLs (uniform resource locators) are, we are already seeing dangling pointers when the resource they locate moves or goes away. Furthermore, the sheer volume of information available (relatively) instantaneously has created the need for a number of new navigation and search tools, such as web robots and spiders, which have protocol implications on their own.

5. End-to-End Issues and Principles

Now let us examine some of the issues important in the delivery of high bandwidth at low latency to applications. They assume that the broadband network infrastructure provides a high bandwidth, low latency pipeline between end hosts.

5.1 Host and Network Interface Architecture

Even though things are improving, there are still significant bottlenecks in current host systems and their network interface. There are several reasons for this, but one of the main ones is insufficient bandwidth between the network link and memory. Although researchers are beginning to realise this, it has not yet translated into many workstation architectures. Most current high performance network interfaces attach to the system bus and use DMA to transfer between the network interface and memory. In the ideal case, the network should look much more like memory than I/O [DeFa91], and data should be pipelined between memory and the network with zero store-and-forward hops [StPa93].

For integrated media, we need to consider where the streams merge and split. If data, voice, and video split at the host–network interface, transmitting them as a single stream through the network has not gained us much over using three individual streams. If the streams are to remain integrated through the host to the human interface, we need to understand how voice and video fit into the host architecture and memory hierarchy.

We should consider the distinction between the processor–memory interconnect, peripheral attach, and the network. There certainly are similarities, and the DAN (desk area network) efforts [Fi91], [HaMcC91] attempt to exploit this. But if the internal host interconnect becomes part of the seamless ATM fabric, then every processor, memory module, and peripheral must have the intelligence of an ATM UNI (user network interface) or NNI (network node inter-

face), or there must be a separate signalling entity. This raises issues of where and when is segmentation and reassembly done, and where raw cell (AAL0) interfaces reside.

5.2 Operating Systems

A number of the bottlenecks in high speed communications are operating system related, and they are being well documented (*e.g.* [NOSSDAV '95]). These primarily relate to avoiding unnecessary context switches, and avoiding extra data copying. In the ideal case, it is possible to have only a single context switch pair: one when the application requests the communication operation, and one when the operation is successful and the application can resume (in the case of a request for information this is when the first complete piece of information the application can use is in place on the host and available in the application user space) [StPa93].

In considering integrated communications, we note that we know how to engineer operating systems for data reasonably well. So although we know how to deal with data in the host memory hierarchy (cache, main store, backing store), what to do with audio and video objects is very much an open question. We need to understand how they fit into the memory hierarchy, and what *is* the hierarchy when we consider frame buffers and the like. Furthermore, we know how to schedule conventional processes reasonably well, but work on how to deal with the real time constraints of mixed media and stream synchronisation is just beginning.

5.3 Protocols

First, we should consider what we *don't* need in protocols for high speed networks: new protocols just for the sake of different protocols. There was a flurry of activity in new transport protocol design a few years ago, but the protocols are only one piece of the end system bottleneck. This does not mean that existing high layer protocols in their current form are adequate for high speed networks and the corresponding host and network interface architecture, but rather that when we create or modify protocols it should be for the right reasons. We should also note that the difference between a "new" protocol (*e.g.* TCP-like replacement) and a highly modified "old" protocol (*e.g.* TCP++) may be in name only. Furthermore, hacking a new protocol from an old one by simply omitting what can't be done fast doesn't get us a useful protocol.

Second, we should point out the distinction between what has to be done in support of higher data rates, and what has to be done in support of applications that tend to be enabled by these higher data rates [Pa94]. There is nothing inherently high speed about multimedia applications, particularly since we have existence proofs of such applications over current $O(10\text{Mbps})$ LANs. So a purist view of high speed protocols would be to consider what to do to an existing protocol to scale to higher data rates; if everything scales, nothing needs to be done [Pa90].

On the other hand, the advent of the broadband network infrastructure will enable the wide scale deployment of a number of new applications (including multimedia) to a degree not pos-

sible before. Thus it is reasonable to consider what must be done by the protocols to support these emerging applications on the new broadband infrastructure, so long as we understand that we are considering a number of things that are not purely “high speed” issues. Thus we wish to provide support for new integrated media (data, audio, video, and other sensory) applications over the broadband network infrastructure and the delivery of the high bandwidth directly to the application.

So what do we need in protocols for high speed networks? Generic research on high speed protocol architecture is of course important (*e.g.* [ClTe90], [Fe93], [Zi91]) as is the optimisation of protocol control mechanisms (*e.g.* [Ja88]). Furthermore, in the spirit of not inventing new protocols and interfaces just for the sake of it, the extension of existing protocols (or protocol implementations) such as TCP for high speed should be done when possible (*e.g.* [JaBr92], [Br92]).

Additionally, protocols which provide support for multicast and multipoint applications will be particularly important in the emerging broadband environment. B-ISDN and ATM networks provide this support natively at the network layer with multipoint connections. IP networks are beginning to provide this with the addition of multicast support [De89] and associated higher layer protocols and tools (*e.g.* [Ja], [Fr94]).

There are two performance aspects to consider here: bandwidth and latency. Since the broadband network infrastructure will support high bandwidths, the protocols must extend this end-to-end to the applications. This means:

- support for high bandwidth host and network interface architecture and operating system constructs
- sufficiently low protocol processing overhead to sustain the bandwidth and not steal excessive application CPU cycles or memory
- provision of required functionality to the application

So the protocol is really not an end in itself, but rather a “glue” that provides the functionality necessary to the application, while sitting on the host and network interface architecture (in concert with the operating system) to provide the high bandwidth pipe.

The second performance aspect is latency (really the bandwidth– \times –delay product). Since we are already on the margin of sub-second latency for long-haul networks due to the speed of light (and to some degree the queueing delay in the switches), the only other thing we can do is to effectively beat the speed of light, and the only way to do this is to predict what piece of information the application or user will need so that when it is referenced it is already cached locally. By treating bandwidth as a resource, information can be prefetched, either using idle network bandwidth as available, or by using additional bandwidth that the user is willing to pay for to guarantee the enhanced response time. There are various strategies for doing this, *e.g.* by prefetching related memory objects [StPa90] or building a tree of possible future requests and preloading them in parallel [To89]. It is the bandwidth– \times –delay product that will

cause us the *most* problems in the long run, because latency it is the one thing that *cannot* scale with data rate.

Finally, given the infancy of B-ISDN and ATM standards and the open issues previously discussed, as we move to giganode gigabit networks we will continue to need research at the network layer and below on signalling, routing, network engineering, and network management and administration.

6. How to Get There

Broadly, then, we need to pursue several areas: In support of the sorts of applications enabled by high speed networks and utilise end-to-end integrated communications we need to understand multimedia operating systems and computer architectures. For WANs we need to effectively violate the speed of light, by exploring prefetching mechanisms and ways of exploiting bandwidth as a resource.

Vendors of workstations need to recognise that it is important to design high performance communication ability into the system, in particular that a high bandwidth path to memory is essential. The traditional limit on the cost of a network interface is about ten percent of the cost of the entire workstation. In an environment where networking is one of the most important thing, this fraction may well be is too small.

The network is a *system of systems*, far more complex than the computer systems we have seen in the past. Interdisciplinary approaches will be required using mathematics, traditional networking, circuits and systems engineering, computer architecture, operating systems, and applications. Furthermore the entire spectrum from theory and analysis, modelling and simulation, development and design engineering, to implementation testing and measurement will be required.

While traditional performance analysis remains important, we will have to place more emphasis on new theory and models, particularly chaos and fractals, general system theory, and game theory (*e.g.* [Sh94]).

We also will need the infrastructure to allow us to do this research. The seeding of testbeds (*e.g.* the Gigabit Testbeds in the US [CNRI94]) aim to help solve the classic “chicken and egg” problem; the chicken being the network infrastructure and the egg being the applications that run on them and enable them (or the other way around). In the future we will need to consider even more aggressive programs to encourage a sense of community among the varied disciplines necessary to develop high speed networks, and to allow reuse and interoperation of the various data, tools, and models [KlSt93]. The WWW is certainly a step in this direction, particularly in enabling the sharing of information, but doesn’t provide us the mechanisms to do things like run distributed simulations and emulations, and glue together diverse tools and models.

7. Summary

We can expect the emerging broadband network infrastructure to be based on ATM cell relay technology, but there are significant technical and practical challenges to be solved in deploying a broadband Global Information Infrastructure. The easy problems have been standardised; the difficult ones, particularly multipoint signalling, connection routing, and traffic management are still open.

The ATM and IP communities are trying to solve the same problems, and the two will merge into a single community. The resultant network will have the flavor of both, with more over engineering and less predictability than currently expected in the ATM community, and more state and user QoS demands than currently expected in the IP community. Even when this is accomplished, all we have provided is an integrated network infrastructure that is capable of transporting high bandwidth. We still have not solved the latency problem (bandwidth \times delay product) for WAN applications, or provided for the delivery of this bandwidth to the applications through the host architecture and operating systems.

So, the answer to whether high speed protocols have a life after ATM is a resounding “yes”. More work needs to be done in implementing host architectures, operating systems, and network interfaces that deliver the high network bandwidth to applications that need it. The only solution to the latency problem is to beat the speed of light by predictive prefetching or pre-loading of information. If we are to gain much by an integrated media network, we need to extend this to multimedia host architectures and operating systems. The higher layer protocols are the “glue” that provide the necessary functionality to the application, while running on the broadband infrastructure.

Acknowledgments

Joe Touch (ISI), Greg Lauer (GTE Laboratories), Peter O'Reilly (GTE Laboratories), and Gurudatta Parulkar (Washington University in St. Louis), provided extensive helpful comments to the original presentation foils and to the text of this paper. The stimulating discussion at PfHSN'94 further contributed to the content of this paper, and I wish to thank all the participants. Discussions with Per Gunningberg (SICS) and Joe Touch (ISI) had particular influence. Finally, Gerald Neufeld and Mabo Ito provided a particularly fine venue for PfHSN'94 at the University of British Columbia, and their encouragement and understanding in the face of the turmoil associated with a relocation during this work are most appreciated.

References

Braden, R., *Extending TCP for Transactions – Concepts*, Internet RFC 1379, Nov. 1992.

Clark, David D., Scott Shenker, and Lixia Zhang, “Supporting Real-Time Applications in an Integrated Services Packet Network: Architecture and Mechanisms”, *SIGCOMM'92 Conference Proceedings*, ACM, New York.

Clark, David D. and David Tennenhouse, “Architectural Considerations for a New Generation of Protocols”, *SIGCOMM'90 Conference Proceedings*, 1990, ACM, New York.

Corporation for National Research Initiatives, A Brief Description of the Gigabit Testbed Initiatives, <http://www.cnri.reston.va.us:4000/public/overview.html>

Deering, S, *Host Extensions for IP Multicasting*, Internet RFC 1112, 1989.

Delp, Gary S., David J. Farber, Ron G. Minnich, Jonathan M. Smith, and Ivan Tam, “Memory as a Network Abstraction”, *IEEE Network*, Vol.5 #4, July 1991, pp. 34–41.

Feldmeier, D.C., “A Framework of Architectural Concepts for High Speed Communications Systems”, *IEEE Journal of Selected Areas in Communications*, Vol.11 #4, May 1993.

Frederick, Ron, *Experiences with Real-Time Software Video Compression*, available by anonymous FTP from <ftp.parcftp.xerox.com:net-research/nv-paper.ps>, 1994.

Finn, G., “An Integration of Network Communication with Workstation Architecture”, *SIGCOMM'91 Conference Proceedings*, Oct. 1991, ACM, New York, pp. 18–29.

Floyd, Sally and Van Jacobson, “The Synchronization of Periodic Routing Messages”, *IEEE/ACM Transactions on Networking*, Vol.2 #2, April 1994, pp. 122–136.

Hayter, M. and D. McCauley, “The Desk Area Network”, *SIGOPS Operating Systems Review*, Oct. 1991, ACM, New York, pp. 14–21.

Jacobsen, Van, *et al*, *sd* (session directory), *vat* (visual audio tool), and *wb* (distributed whiteboard), available by anonymous FTP from ftp.ee.lbl.gov:conferencing/*/

Jacobsen, Van, “Congestion Avoidance and Control”, *SIGCOMM'88 Conference Proceedings*, 1988, ACM New York, pp. 314–329.

Jacobson, V, R. Braden, and D. Borman, *TCP Extensions for High Performance*, Internet RFC 1323, 1992.

Kleinrock, Leonard, James P.G. Sterbenz, Nick Maxemchuck, Simon S. Lam, Henning Schulzrinne, Peter Steenkiste, and the HPNE group, *The National Exchange for Networked Information Systems: A White Paper*, UCLA Computer Science Technical Report CSD-930039, November 1993.

Leland, Will E., Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson, “On the Self-Similar Nature of Ethernet Traffic (Extended Version)”, *IEEE/ACM Transactions on Networking*, Vol.2 #1, Feb. 1994, pp. 1–15.

Moldeklev, Kjersti and Per Gunningberg, “Deadlock Situations in TCP over ATM”, *Protocols for High Speed Networks IV*, Gerald Neufeld and Mabo Ito, ed., 1995, Chapman and Hall, London.

NOSSDAV '95: *5th International Workshop on Network and Operating System Support for Digital Audio and Video*, Durham, New Hampshire, April 1995.

Nussbaumer, Jean-Paul, Baiju V. Patel, Frank Schaffa, and James P.G. Sterbenz, "Networking Requirements for Interactive Video on Demand", *Gigabit Networking Workshop '94*, INFOCOM'94, Toronto, June 1994; also available as <http://info.gte.com/jpgs/paper/gbn94.ps>

Partridge, Craig, "How Slow is One Gigabit per Second?" *ACM SIGCOMM Computer Communications Review*, Vol.20, #1, Jan 1990, ACM, New York, pp. 44–53.

Partridge, Craig, ed., *Report of the ARPANet Workshop on Research in Gigabit Networking*, Washington, D.C., July 1994, available by anonymous FTP from <ftp.std.com:pub/craigp/report.ps>; also available as <http://info.gte.com/jpgs/doc/arpansf94.ps>

Shenker, Scott, "Making Greed Work in Networks: A Game-Theoretic Analysis of Switch Service Disciplines", *SIGCOMM'94 Conference Proceedings*, Oct. 1994, ACM, New York, pp. 47–57.

Sterbenz, James P.G. and Gurudatta M. Parulkar, "Axon: Network Virtual Storage for High Performance Distributed Applications", *Proceedings of the Tenth IEEE International Conference on Distributed Computing Systems (ICDCS 10)*, May 1990, pp. 484–492; also available as <http://info.gte.com/jpgs/paper/icdcs90.ps>

Sterbenz, James P.G. and Gurudatta M. Parulkar, "Design of a Gigabit Host Network Interface", *Journal of High Speed Networks*, Vol.2 #1, 1993, IOS Press, Amsterdam, pp. 27–62; also available as <http://info.gte.com/jpgs/paper/jhsn93.ps>

Touch, Joseph D and David J. Farber, "MIRAGE: A Model for Ultra-high-speed Protocol Analysis and Design", *Protocols for High Speed Networks*, North-Holland, 1989, pp. 115–134.

Touch, Joseph D., "Defining High Speed Protocols: Five Challenges and an Example that Survives the Challenges", *Gigabit Networking Workshop '94*, INFOCOM'94, Toronto, June 1994.

Zhang, Lixia, Stephen Deering, Deborah Estrin, Scott Shenker, and Daniel Zappala, "RSVP: A New Resource ReSerVation Protocol", *IEEE Network*, September 1993, IEEE New York, pp. 8–18.

Zitterbart, Martina, "High Speed Transport Components", *IEEE Network*, Vol.5 #1, January 1991, IEEE, New York, pp. 54–63.