The Evolution of Campus Networks Towards Multimedia *

Ciro A. Noronha Jr. Fouad A. Tobagi†

Department of Electrical Engineering
Stanford University
Stanford, CA 94305

Abstract

The first step in providing multimedia services over computer networks is to insure that enough bandwidth is available to support them. This paper considers the problem of providing bandwidth for multimedia services in a campus network. The campus network is composed by a number of subnetworks, interconnected by a backbone. We first study each subnetwork in isolation, indicating what kinds of topologies should be used as a function of the bandwidth requirement, and then discuss the connection of the subnetwork to the backbone. We also touch upon issues other than the network infrastructure, commenting on the effect of providing multimedia communications on the other layers of the OSI model. Finally, we consider the Stanford Campus Network (SUNET), presenting measurements to characterize its current performance.

1 Introduction

One of the characteristics of the emerging multimedia applications, specially video, is the large volume of data they generate. In spite of advances in compression techniques, the data rates resulting from communication of multimedia information are still significant compared to the capacity of widely deployed data communication networks. For example, the data rate for one MPEG-I stream is 1.5 Mb/s; MPEG-II will go to 4-8 Mb/s. On the other hand, the capacity of an Ethernet is 10 Mb/s, sufficient only for carrying 6 MPEG-I streams.

The current network infrastructure might be sufficient for supporting multimedia applications in a small scale. For example, a single Ethernet can be adequate for a small workgroup retrieving stored video from its local server, or for very small scale videoconferencing. However, as the number of multimedia-capable users grows, the aggregate bandwidth of the network must also grow to support their communication requirements. Moreover, applications like videoconference are not likely to stay restricted to the individual workgroups; extensive use of the campus backbone is expected. This paper studies how to provide this bandwidth in a campus network environment.

Campus networks are subdivided into *subnetworks* due to geographical, administrative and addressing reasons. The subnetworks are interconnected by a *backbone*. In this paper, we consider that the traffic generated within a subnetwork is divided into two fractions: the intra-subnet traffic, which is uniformly addressed to all the users within that subnetwork, and the inter-subnet traffic, which is addressed to the other subnetworks on the campus network, and is proportional to their sizes. The traffic model is discussed in detail in section 2. This model, although simplified, is realistic and makes it possible to evaluate the required bandwidth for the campus network.

Subnetworks are composed of *segments*; a segment is a shared channel interconnecting a number of users. Segments could be Ethernets, Token Rings, FDDIs, etc, and are interconnected by bridges, which isolate the traffic local to the segment, repeating only the traffic to/from the other segments in the subnetwork.

As we have indicated before, if the bandwidth requirements are low, a subnetwork might be composed of a single segment. However, as bandwidth requirements grow and exceed the capacity of a single segment, two options exist to provide the service: (i) increase the capacity of the segment; and (ii) increase the concurrency by using multiple segments. Option (i) can be exercised up to a certain limit, and is usually not very attractive, because all the nodes must be retrofitted with interfaces for the higher speed. With

*This work was supported by in part by NASA under grant NAGW-419, by NSF under grant NCR-9016032, by Pacific Bell and by AT&T. Ciro Noronha is supported by a graduate scholarship from FAPESP under grant 89/1658.

†On leave at Starlight Networks Inc., Mt. View, CA 94041

option (ii), no modification is needed in the user stations, and the increase in bandwidth is realized by increasing the number of segments, thus allowing multiple concurrent transmissions to take place. However, some sort of switching function between those multiple segments must be provided. The simplest way to implement this switching function is by interconnecting the segments with bridges, but for uniformly-addressed traffic, this technique cannot lead to large improvements over the capacity of a single segment.

The reason why interconnecting the segments using bridges leads to limited improvement in the aggregate throughput is that the some of the segments are being used to implement the switching function between the concurrent channels, and become the bottlenecks. It is clear that a higher-capacity channel must be provided to carry the traffic between segments; this higher-capacity channel might be either another network (for example, a FDDI ring interconnecting a number of Ethernet segments) or a high-speed bus inside a switch (for example, a switching hub or an ATM switch). The topology of the subnetwork then becomes a star, with this high-speed channel in the center. This way, the individual nodes do not have to be able to access the network as its aggregate data rate; the segment's data rate becomes the access data rate to the network, and the current interfaces can still be used.

The evolution of the campus backbone follows the same path as the one described for the subnetwork, but in a larger scale. Depending on the requirements, a FDDI ring or a switching hub might suffice as the campus backbone; if not, an ATM switch is an alternative. In section 4 we give a number of examples to illustrate some of the possible backbone configurations.

The addition of multimedia to computer communication networks raises issues at levels other than the network infrastructure as well. Following the OSI model, we comment on the relevant issues related to the MAC layer, to the network layer, to the transport protocol, and to the higher layers. Finally, we present some measurements performed on the Stanford Campus Network in order to characterize its current performance, and illustrate the issues discussed in this paper.

2 The Traffic Model

In this section we define the traffic model used in the paper. We use the following symbols:

- D : Number of subnetworks in the campus network
- N^{CAM} : Number of users on campus
- N_i^{SUB} : Number of users in subnetwork i, $i = 1, \ldots, D$
- M_i : Number of segments in subnetwork i, $i = 1, \ldots, D$
- N_{ij}^{SEG} : Number of users in segment j of subnetwork i, $j = 1, \ldots, M_i$, $i = 1, \ldots, D$
- T_i^{SUB} : Traffic being generated in subnetwork i, in bits/second
- T^{BACK} : Traffic in the backbone
- W_{STR} : Data rate of the multimedia stream, in bits/second
- α : Fraction of the traffic generated in the subnetwork which is local (i.e., destined to users in the same subnetwork)
- δ : Fraction of the users in the campus network which are multimedia-capable

Since multimedia will be introduced gradually, we assume that only a fraction δ of the users in the campus network will be generating multimedia traffic. Each multimedia-capable user generates a stream of data rate W_{STR}:

$$T_i^{SUB} = \delta N_i^{SUB} W_{STR} \tag{1}$$

A fraction α of T_i^{SUB} remains inside the subnetwork, and the remaining $(1-\alpha)$ is transmitted onto the backbone. The traffic in the backbone is:

$$T^{BACK} = \sum_{i=1}^{D} (1-\alpha) T_i^{SUB} \tag{2}$$

The traffic from subnetwork i to subnetwork j is considered to be proportional to the size of subnetwork j, as follows:

$$\alpha_{ij}^{ext} \triangleq \text{Fraction of the traffic generated in subnetwork } i \text{ destined to subnetwork } j$$

$$\alpha_{ij}^{ext} = \begin{cases} \dfrac{(1-\alpha) N_j^{SUB}}{N^{CAM} - N_i^{SUB}} & \text{if } j \neq i \\ \alpha & \text{if } i = j \end{cases} \tag{3}$$

The traffic from subnetwork i into the backbone, T_i^{in}, is given by:

$$T_i^{in} = T_i^{SUB}(1-\alpha) \tag{4}$$

The traffic from the backbone into subnetwork i, T_i^{out}, is given by:

$$T_i^{out} = \sum_{\substack{j=1 \\ j \neq i}}^{D} \frac{(1-\alpha) N_i^{SUB}}{N^{CAM} - N_j^{SUB}} T_j^{SUB} \tag{5}$$

It is possible to show that T_i^{out} is:

$$T_i^{out} = \Delta_i T_i^{SUB}(1-\alpha) \tag{6}$$

where the factor Δ_i is given by:

$$\Delta_i = \sum_{\substack{j=1 \\ j \neq i}}^{D} \frac{N_j^{SUB}}{N^{CAM} - N_j^{SUB}} \tag{7}$$

Finally, if a subnetwork is composed of multiple segments, the traffic between segments is given by:

$\alpha_{ij}^{int} \triangleq$ Fraction of the traffic generated in segment i destined to segment j

$$\alpha_{ij}^{int} = \frac{N_j^{SEG}}{N^{SUB}} \tag{8}$$

3 Design of the Subnetwork

In this section, we consider the design of the subnetwork, as a function of the bandwidth it has to provide to its users. In other words, we study the following issue: given a certain bandwidth requirement, find the number of segments in the subnetwork and their interconnection to satisfy this requirement.

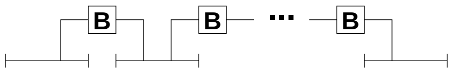
As we have discussed in section 2, we assume that the intra-subnetwork traffic is uniformly addressed, i.e., a user in the subnetwork is equally likely to communicate with any other user in the subnetwork. In the following discussion, we will assume that no traffic will be flowing to/from the campus backbone ($\alpha = 1$); the inter-subnetwork traffic will be considered in section 4.

In figure 1 we show the topologies considered for the subnetwork. We consider basically three classes of topologies: (i) a tandem made out of 2-port bridges (figure 1a), (ii) a tree made out of R-port bridges (figure 1b), and (iii) a star configuration with a central high-speed channel (figure 1c).

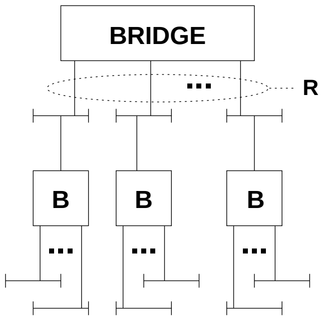
It can be shown that the aggregate throughputs achieved for each of the topologies shown in figure 1 for M segments of capacity C each is:

Tandem (figure 1a):

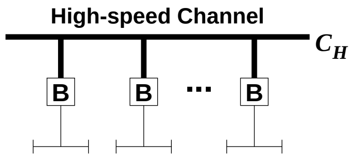
$$T = \begin{cases} \dfrac{2M^2C}{M^2+2M-1} & M \text{ odd} \\[2ex] \dfrac{2M^2C}{M^2+2M-2} & M \text{ even} \end{cases} \tag{9}$$



(a) Segments in Tandem, 2-port bridges



(b) Segments in Tree, R-port bridges



(c) Star Topology with a high-speed channel

Figure 1: Subnetwork topologies considered

Tree made out of R-port bridges (figure 1b):

$$T = \begin{cases} \dfrac{R^2C}{2R-1} & M = R \\[2ex] \dfrac{R^2C}{2(R-1)} & M = R^2 \end{cases} \tag{10}$$

Star topology, with a high-speed channel of bandwidth C_H (figure 1a):

$$T = \min\left\{ \frac{M^2C}{2M-1} \quad , \quad \frac{MC_H}{M-1} \right\} \tag{11}$$

It can be shown that:

- If one uses segments of the same capacity, the tandem in figure 1a can achieve a higher aggregate throughput than any other regular tree made out of two-port bridges;
- For the trees made of R-port bridges, the maximum throughput is achieved for the depth of 2 shown in figure 1b; the users are connected only to the leaves of the tree, and the bottleneck is the

capacity of the segments connected to the root bridge.

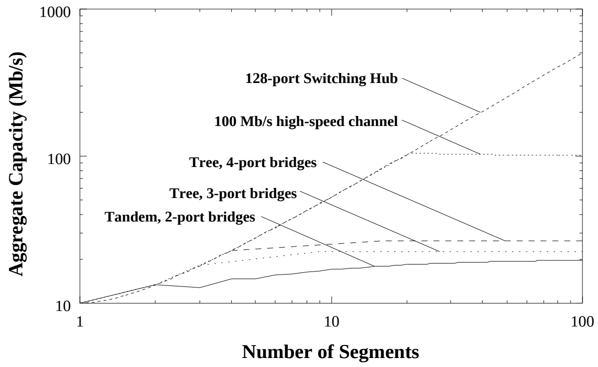


Figure 2: Aggregate throughput for the various subnetwork topologies

Figure 2 shows a comparison of the aggregate throughput achieved by 2-, 3- and 4-port bridges, a star configuration with a 100 Mb/s channel, and a 128-port switching hub, when the segments are 10 Mb/s Ethernets. The star configuration can outperform the tree topologies because it has more bandwidth exactly where it is needed, in the interconnection between the segments. The high-speed channel can be either another network (e.g., a 100 Mb/s FDDI ring), or the backplane bus inside the switching hub.

4 The Backbone

The campus backbone carries the aggregate traffic from the subnetworks. From a theoretical point of view, the design of the backbone is no different from the design of the subnetworks themselves; the steps taken in section 3 can be repeated here, with the difference that the network being designed is the backbone, and its "users" are the subnetworks. The difference is one of scale.

In this section, we will briefly comment on the effect of the backbone traffic on the aggregate throughput of the subnetwork (which was ignored in section 3), and then compute the backbone traffic under a variety of scenarios.

4.1 Connecting the Subnetwork to the Backbone

The main difference between the traffic to/from the backbone and the internal subnetwork traffic is the fact that former is concentrated in a single point (the backbone connection); in other words, all the traffic from the the subnetwork to the backbone is directed to the backbone connection, and all the traffic from the backbone to the subnetwork originates in the same location. In the other hand, the sources and destinations of the intra-subnetwork traffic are spread over the subnetwork.

The effect of adding the backbone connection in the aggregate subnetwork bandwidth[1] depends on its topology. In subnetworks organized as trees, the congestion in the branches near the location of the backbone connection might limit the aggregate throughput to a value lower than what is attained when no backbone traffic is present. In the star configuration, however, the high-speed channel is the ideal location for attaching the backbone connection, and the aggregate throughput in this case is always *higher*[2] than its value when there is no traffic from the backbone. The reason for this is simple: the internal subnetwork traffic uses bandwidth both at the source and at the destination segments, but the traffic to/from the backbone uses bandwidth only at the source/destination segments, leading to a higher aggregate throughput.

The actual increase or decrease in the aggregate bandwidth will be a function of α, the fraction of the traffic that is intra-subnetwork, and Δ, the ratio between the traffic from the backbone to the traffic into the backbone (defined in section 2. The increase is higher for small values of α (most of the traffic is inter-subnetwork).

4.2 Dimensioning the Backbone

Combining equations 1 and 2, we find the required backbone capacity:

$$T^{BACK} = N^{CAM} \delta W_{STR}(1-\alpha) \qquad (12)$$

Equation 12 shows that the traffic in the backbone is a function of:

- the number of users on campus;
- the fraction of the user population that is video-capable;
- bandwidth of the multimedia streams in use; and

[1] In the presence of traffic from the backbone, we define the aggregate subnetwork bandwidth as the sum of the local traffic from the subnetwork plus the traffic from the backbone.

[2] Provided that the capacity of the high-speed channel, C_H, is such that it does not become a bottleneck.

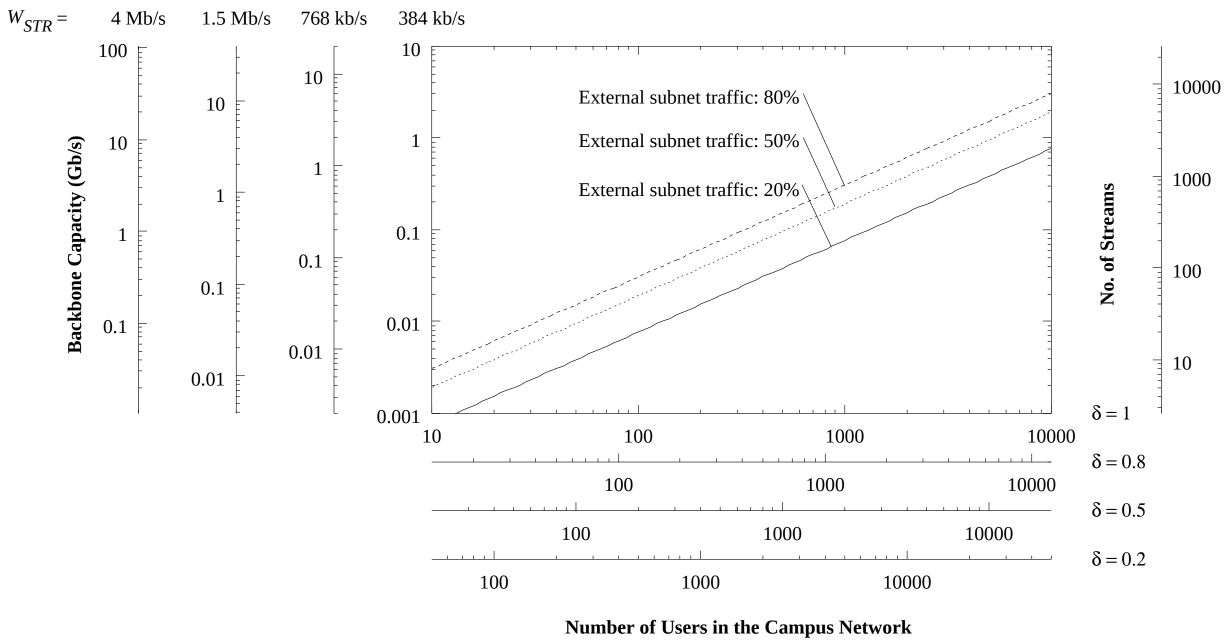


Figure 3: Backbone capacity in Gb/s

- the external subnet traffic, i.e., the fraction of the traffic generated in the subnetwork that is destined to other subnetworks and thus must be carried by the backbone.

Equation 12 is plotted in figure 3, for various values of δ, W_{STR} and $(1-\alpha)$. To use figure 3, one selects the appropriate x-axis according to the value of δ, and reads the required backbone capacity in the y-axis corresponding to the stream bandwidth in use.

Since the traffic in the backbone is composed by the aggregate traffic from many sources, it might be possible to take advantage of the statistical multiplexing of those sources and provide less bandwidth in the backbone than the amount requested by equation 12. Of course, the price to pay is measured by some tolerable blocking probability. The same reasoning could have been applied to the design of the subnetwork, but there the population is typically much smaller, and one basically has to provide a enough bandwidth to accommodate for the peak traffic, or the blocking probability will be unacceptable. However, if the traffic is low, some statistical multiplexing gain can be realized with a number of streams as low as 10.

The plot in figure 4 shows the reduction in the required backbone capacity as a function of the blocking probability. To compute the blocking probability, we have considered that the blocked calls are dropped [1]. The x-axis contains the maximum number of multimedia streams generated by the subnetworks; each

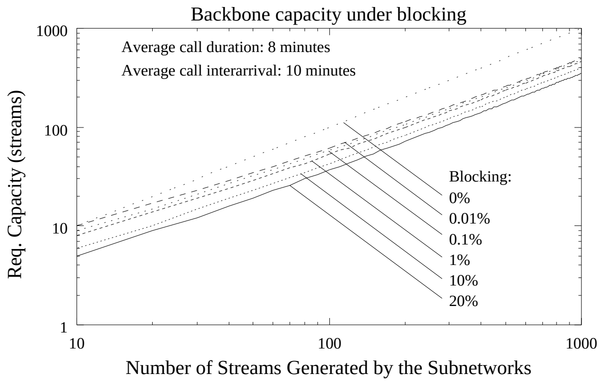


Figure 4: Required backbone capacity when blocking is acceptable

stream stays ON for an average of 8 minutes, and the call interarrival time for the streams that are OFF is an average of 10 minutes. The y-axis contains the backbone capacity in streams, order to achieve the specified blocking probability. For example, if the combined traffic from the subnetwork offered to the backbone is 100 MPEG-I streams (1.5 Mb/s), then the backbone capacity under no blocking has to be 150 Mb/s; however, if a blocking probability of 0.1% is acceptable, then one has to provide capacity for only 58 streams, or 87 Mb/s.

Table 1: Backbone Structures

N^{CAM}	W_{STR}	$(1-\alpha)$	δ	**Bandwidth**	**STRUCTURE**
60	1.5 Mb/s	0.2	0.5	9 Mb/s	Single Ethernet
260	1.5 Mb/s	0.5	0.5	97.5 Mb/s	FDDI Ring
1000	384 kb/s	0.8	0.5	150 Mb/s	30-port Switching Hub
5000	1.5 Mb/s	0.5	0.5	1.88 Gb/s	12-port ATM switch

4.3 Structure of the Backbone

In this section, we will give a number of scenarios that illustrate how the structure of the backbone can change as a function of the size of the campus network; we will use figure 3 to determine the required capacity. The results are shown in table 1; note that the solutions for the structure of the backbone are not unique, and we have listed one of such solutions for each case.

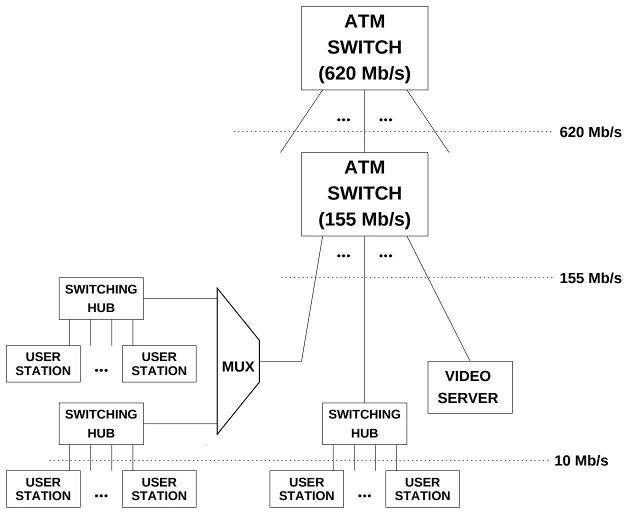


Figure 5: A possible backbone topology

In figure 5 we show one possible backbone topology which has all the main elements we have been discussing. In the figure we show the subnetworks being served by switching hubs. The switching hubs serving subnetworks with little external traffic can be grouped by an ATM multiplexer before reaching the ATM switch; this multiplexer can perform some local switching or rely completely in the central switch. Hubs generating larger traffic can be connected directly to the ATM switch. The switch is also the best location to connect centralized shared resources, such as video servers. The ATM switches themselves can be arranged in a hierarchical fashion, depending on the traffic requirements. It is important to note that there are some Ethernet switching hubs which have internal buses with higher bandwidths than the ones found in the available ATM switches. ATM, however, has been recognized as the the universal backbone technology, and it is likely to be the next step in the evolution of the campus network, as bigger and faster models become available.

5 Adding multimedia to the other OSI layers

In this section we discuss the issues raised by the addition of multimedia services on the other layers of the OSI model.

5.1 The Medium Access Control (MAC) Layer

The synchronous nature of audio and video traffic dictates that data be delivered to its destination within strict timing constraints. Failure to deliver data on time results in a discontinuity of the video or a degradation in the quality. The bursty nature of data traffic, on the other hand, means that a station transmits data in an unpredictable fashion, and the amount of data that the station transmits in a burst is also random.

Both types of traffic are to be supported by the same local network. Indeed, multimedia applications naturally involve both types of traffic simultaneously; existing local area networks are expected to carry audio and video traffic alongside existing data applications. Mixing the two types of traffic on the same network requires special attention, particularly where shared resources are contended to by both types. The bandwidth available on local area network segments is one such shared resource, and is the subject of this section. It is shared by all traffic originating at all stations, and accessed by means of the Media Access Control (MAC) protocol. Ideally, one would like the MAC protocol to include techniques which reserve the bandwidth necessary for synchronous traffic, so as to remove any effect of bursty traffic.

The CSMA/CD protocol (IEEE 802.3) does not differentiate among the different traffic types. In essence, it operates as follows. A station with a packet ready for transmission transmits the packet as soon as the channel is sensed idle. If it collides with another packet, then it attempts again but after incurring a random rescheduling delay. The mean rescheduling delay is doubled with each collision incurred. This is clearly inadequate for real-time traffic. Not only is the delay incurred by packets nondeterministic, but its variance is quite large, owing to the exponential backoff algorithm. However, since at low loads this protocol works well, it can still be used if one segments the network to insure that this condition is met.

IEEE 802.5 is a token protocol devised for a ring network. It uses a single token, in the sense that a station that has completed transmission will not issue a new token until the busy token returns to it. Since in a ring network the connection to the medium is active, a priority scheme can be implemented by assigning dynamically a priority value to the token, and by restricting access to the ring to packets of priority equal or higher than the currently assigned value. This scheme can be appropriately used to integrate the two types of traffic on the same channel simply by giving synchronous traffic higher priority than the bursty data traffic.

In FDDI (ANSI X3T9.5), another scheme with multiple priority levels is employed, by the use of a set of timers, that regulate for how long the station can transmit traffic of each priority.

In summary: the assignment of different types of service for real-time and data traffic at the MAC layer is very important when communicating multimedia information. This is not offered in the widely deployed IEEE 802.3 Other networks, such as the IEEE 802.5 token ring and FDDI can provide this functionality, although many vendors have chosen not to implement it. Given the large base of deployed networks, it is not reasonable to expect that changes in the MAC layer to support multimedia would be implemented. Therefore, one must seek other options (such as making sure that the network operates with low enough load that the problems described in this section will not be seen).

5.2 The Network Layer

As the changes described before take place in the network, the route to use when moving information from segment to segment becomes increasingly important. In simple topologies, such as star with the switching element in the center, routing is trivial because there is only one route from the source to the destination. However, in practice multiple routes might be provided between a given pair of stations, for reliability and/or increased throughput.

Currently, the path taken by a packet from its source to its destination through the network is determined by the bridges and routers that define the network topology. Transparent bridges use the spanning tree algorithm, i.e., the bridges, working together, identify a subset of the network topology that constitutes a spanning tree, and direct all traffic through this spanning tree. Links not on the spanning tree are not used and kept in reserve, to be activated in case of failure of a link in the spanning tree. Moreover, the spanning tree identified for a topology is essentially chosen at random, and it is not optimum in any sense. This has two consequences: first, extra capacity on the redundant links is never used; and second, bottlenecks may develop due to traffic concentration in some segments, as a result of the particular spanning tree chosen. We have discussed this traffic concentration in section 3.

Routers, on the other hand, have the ability to use multiple spanning trees; as a matter of fact, each router forwards the packets over the spanning tree having itself as the root. Many algorithms exist to compute those routes in a distributed fashion (RIP, OSPF, IS-IS, etc), but they all employ the cost of the link as the metric used when computing the spanning tree (the cost of a link is an arbitrary function, inversely proportional to the link's bandwidth). Routers can recover from network failures, and can make use of the bandwidth of all links available. However, in most cases, all the traffic between a source and a destination will be sent over the same route. Moreover, current routers have no provision for taking into account real-time requirements (such as delay) when computing routes, and the system of priorities implemented is clearly inadequate for multimedia communications.

Of special importance is the routing of multicast traffic (video-conference is an example of a situation where this kind of traffic is generated). For multicast traffic, two kinds of routes can be identified: minimum delay routes and minimum cost routes. Current multicast routers use the minimum delay criterion to compute the routes, but this might lead to unacceptable network loading. Therefore, new routing algorithms with the following characteristics are needed:

- Use of multiple spanning trees.

- Use of multiple routes between sources and destinations.

- Use of different criteria for routing the streams (delay, cost or a combination thereof).
- Use of a system of priorities that takes into account the real-time nature of the traffic, and not only different link costs for each priority.

5.3 Transport Protocols

Existing transport protocols have been designed and implemented to support data applications in which the traffic is usually bursty. Flow control procedures are embedded in the protocols to properly pace the flow between end users and achieve efficient utilization of network resources. For data applications, reliability is an absolute goal and is achieved by an error detection and retransmission scheme. Such transport protocols, however, are not appropriate for the delivery of video data which is rather steady in nature, is time-constrained, and in many cases can tolerate a certain degree of data loss. As a result, new transport protocols, or additions and modifications to existing transport protocols are essential to support stream traffic [2]. More precisely, it is important to include at the transport level means to appropriately package video, audio and other data for transmission. Among others, it may be important to give an indication of the logical grouping of data (which has meaning to the application, such as the beginning and end of a video frame), and to multiplex and demultiplex by media source, by media encoding, or by conference. Also of great importance is the ability to embed in the data stream timing information to permit the timely delivery of data belonging to a stream, as well as the synchronization of multiple streams at the destination. Rate control at the source in order to avoid congestion at the destination and in intermediate routers, also proves to be extremely useful. Finally, efficiency of the protocol, low processing overhead and ease of implementation become of utmost importance when the traffic rate is high, and the user desktop is not a high end workstation.

5.4 Higher Layers

For video services, the application layer plays an important role in managing video communication among nodes on the network, and in securing the necessary network resources to support effectively that communication. Accordingly, the development of video specific application layer protocols is a clear necessity. In this respect, at the present time, little has been done in that area. The desired functionality includes:

- Conference control and management.
- Managing different encoding techniques.
- Identification and use of resources.
- Encryption.

6 Measurements in the Stanford Campus Network

In this section we present a series of measurements performed in the Stanford Campus Network (SUNET), to characterize its current performance and illustrate some of the issues discussed in section 5.

6.1 Description of the Stanford Campus Network (SUNET)

The Stanford Campus Network (SUNET) is composed of a set of IP subnetworks[3], interconnected by a combination of FDDI and Ethernet backbones. Figure 6 shows the organization of the backbone.

There is a basic Ethernet backbone, which is composed by a number of segments interconnected by bridges, forming a mesh topology. This backbone connects to all subnetworks, and can be used to carry the traffic in event of a failure of the main FDDI backbones.

There are two FDDI networks in the backbone. The first one, shown in the left in figure 6, is basically "in parallel" with the Ethernet backbone; there are no routers connected to it. Ethernet-to-FDDI bridges are used to form a single logical network. The second backbone, at the right in figure 6, has only routers attached to it, and is logically distinct from the first one. All the IP subnetworks are connected to the Ethernet backbone via routers; some are also connected to the FDDI network in the right in figure 6, using the routers shown there. Averaged over the day, the traffic on each of the FDDI networks is relatively small: around 30 kb/s. In the busiest hour, the busiest IP subnetwork sends an average of 1.2 Mb/s into the backbone. Measurements indicate that the busiest hour varies from subnetwork to subnetwork; for example, the busiest hour on the subnetwork that houses the main campus systems was 10:00PM on Dec. 14, 1992, while the busiest hour in the Medical Center subnetwork was 5:00PM in the same day.

[3] The IP subnetworks usually connect all the research groups in a building; they are not exactly equivalent to the subnetworks we have been considering in this paper.

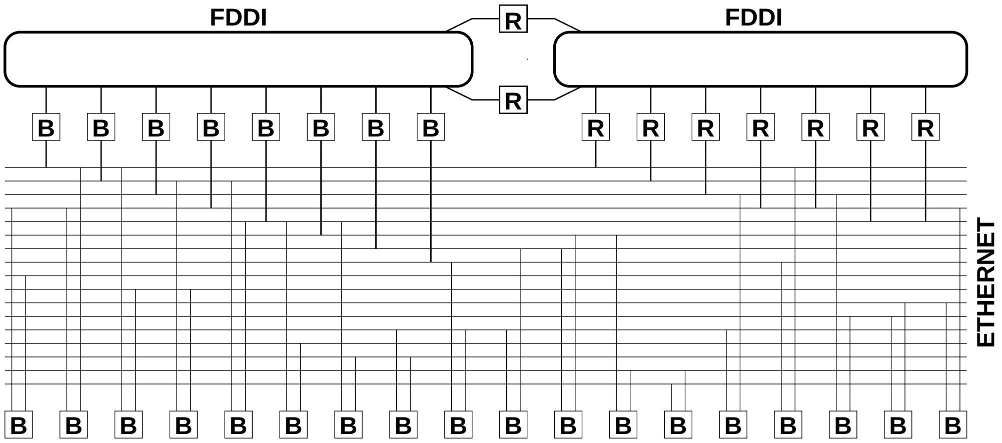


Figure 6: The Stanford Campus Network Backbone

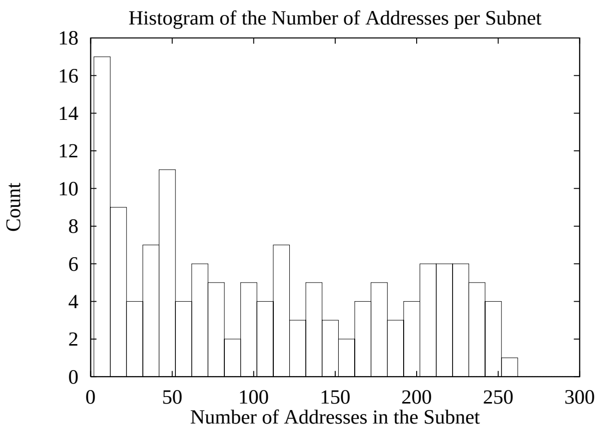


Figure 7: Histogram of the number of addresses per IP subnetwork in SUNET

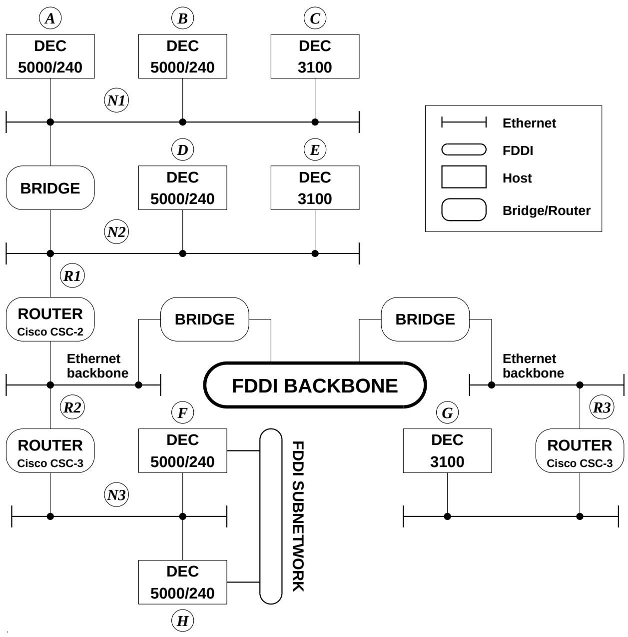


Figure 8: The subset of the Stanford Network used in the measurements

There are 124 IP subnetworks on campus, each one with an average of 119 addresses. Figure 7 shows the histogram of the number of addresses in the subnetworks. Note that some nodes, such as AppleTalk gateways, have multiple IP addresses, which are dynamically allocated to other nodes using this protocol.

6.2 Measurements

We have conducted a number of measurements over the Stanford Campus Network to determine what ranges of throughputs are achievable in practice. For all the measurements described in this section, we will consider the subset of the Stanford Campus Network depicted in figure 8.

Figure 9 shows the result of throughput measurements, under TCP/IP, on the Stanford Campus Network. The labels in the plot correspond to the hosts indicated in figure 8. All measurements consisted of opening a TCP connection, transmitting 4 Mbytes of random data, and closing the TCP connection. The plots give the throughput as a function of the TCP transmit buffer size. The measurements were taken on a weekday afternoon.

Curve 1 was represents communication from ***F*** to ***H*** using a two-node completely unloaded FDDI ring. The maximum throughput was in the order of 17 Mb/s, and was limited by the protocol processing

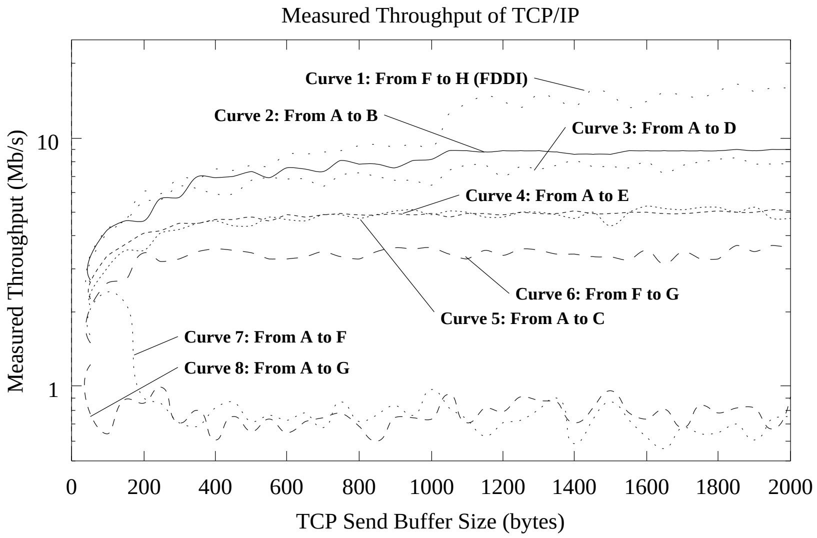


Figure 9: Measured throughput in the Stanford Campus Network

in the hosts. Curves 2 and 3 represent the throughput between two DEC 5000/240 workstations using Ethernet. For curve 2, the workstations were in a short unloaded Ethernet segment; for curve 3, they communicated across a bridge. We can see that the maximum throughput is quite high (9 Mb/s for curve 2, and 8 Mb/s for curve 3). The difference is due to the fact that network ***N1*** is lightly loaded, while network ***N2*** is not.

Curves 4 and 5 represent communication between system ***A*** and systems ***C*** and ***E***. Since systems ***C*** and ***E*** have slower CPUs, a lower throughput is achieved due to protocol processing. However, the two curves are virtually indistinguishable, indicating that the bridge really did not have any impact on performance, as far as throughput is concerned.

Curve 6 represents the throughput from ***F*** to ***G***, while curve 8 has the throughput from ***A*** to ***G***. This pair of curves shows the effect of processing power in the router. Since the source and destination systems are of the same kind in both cases, and topologically are placed in similar positions in the network, any difference in performance has to be explained by the connection to the backbone. And, indeed, figure 8 indicates that router ***R1***, which provides connection to the backbone for node ***A***, is a Cisco CSC-2, while router ***R2***, which provides service to node ***F***, is a Cisco CSC-3. The difference in performance is further confirmed by curve 7, which gives the throughput from ***A*** to ***F***.

In summary, our measurements have indicated that:

- The bridges in use do not introduce any degradation in the available throughput.
- Some of the current routers pose severe limitations to the throughput achievable in practice, due to their limited processing capabilities.
- Protocol processing in the hosts is still an important factor in determining the throughput achieved by the less-powerful machines.
- In spite of the data traffic, large end-to-end throughputs can be obtained, it the hosts have fast enough CPUs to do the protocol processing. For high-end machines, TCP/IP processing is not a bottleneck when communicating over Ethernet, but it still is the determining factor when using FDDI.

References

[1] L. Kleinrock, *Queueing Systems, vol. I*, John Wiley & Sons, 1975.

[2] C. Topolcic, Editor, "Experimental Internet Stream Protocol, Version 2 (ST-II)", *RFC 1190*, CIP Working Group, October 1990.