NEMESIS: A Multigigabit Optical Local Area Network

Adrian Popescu † and Ian F. Akyildiz ‡

† Royal Institute of Technology, Dept. of Teleinformatics, KTH - Electrum/204, 164 40 Kista, Stockholm, Sweden, adrian@it.kth.se
‡ Georgia Institute of Technology, School of Electrical and Computer Eng., Atlanta, GA 30332, U.S.A., ian@ee.gatech.edu

ABSTRACT A new architecture is developed for an integrated 20 Gbps fiber optic Local Area Network (LAN) that supports data rates up to 9.6 Gbps. The architecture does not follow the standard, vertically-oriented Open System Interconnection (OSI) layering approach of other LANs. Instead, a horizontally-oriented model is introduced for the communication process to open up the three fundamental bottlenecks, i.e., opto-electronic, service and processing bottlenecks, that occur in a multi-Gbps integrated communication over multiwavelength optical networks. Furthermore, the design follows also a new concept called Wavelength-Dedicated-to-Application (WDA) concept in opening up the opto-electronic and service bottlenecks. Separate, simplified and application-oriented protocols supporting both circuit- and packet-switching are used to open up the processing bottleneck.

1 Introduction

An urgent need exists to design and implement a multi-Gbps LAN capable of supporting a wide range of applications generating different isochronous and nonisochronous traffic with arbitrary bit rates. The increasing demand for high bandwidth networking, under increasingly strict performance constraints, has posed fundamental challenges to LAN design and implementation. In particular, the introduction of fiber optic technology has resulted in a substantial increase in the amount of potentially usable bandwidth. Due to this, the performance bottleneck is no longer the transmission channel, but rather the network nodes. Three fundamental bottlenecks exist: the *opto-electronic, service and processing bottlenecks*. In a multi-Gbps LAN environment, these bottlenecks must be handled in order to achieve the optimal performance.

Although the optical fiber technology may support traffic with capacities up to Terabps, the electronic components at the network nodes typically operate at rates only up to about 1 Gbps, which drastically limits the total throughput. For instance, new systems being designed and developed to take advantage of the lightwave technology, such as the FDDI, DQDB, HIPPI and ATM in combination with the SONET, are of architectures that do not take advantage of the lightwave technology capabilities. These networks are inherently limited by the use of electronic components at stations, resulting in performance hindrance and inefficient resource utilization.

Many approaches exist for defeating the *opto-electronic bottleneck*. One approach is the "multi-hop" architecture, which makes use of a new network architecture to achieve high capacity with existing devices. This architecture is used by several proposed networks, including the Manhattan Street Network, Shufflenet and Lightnet [1]. A drawback of this, however, is the so-called "deloading factor" that occurs in the case of nonuniform load patterns and fixed routing. The load imbalance causes throughput per station to be reduced by a factor of 30 - 50% as compared to the balanced load situation [2]. Alternatively, different adaptive (self-routing) mechanisms can be used in the spectral domain to compensate for variable traffic intensities and/or patterns [3]. Nevertheless, such architectures pose difficult problems of routing to reduce wasting of bandwidth resource and to provide the requested Quality of Service (QoS).

Another approach (so-called Wavelength-Dedicated-to-User) is to use a pure Wavelength Division Multiplexing (WDM) technique for transmission in combination with some form of WDM, TDM or SDM for switching [4]. Examples of networks using this include Lambdanet, Hypass, Bhypass switches, the photonic knockout switch and the star track switch [1]. Although these networks support multi-Gbps traffic rates, they have fundamental difficulties when using distributed control. These networks have generally a star topology with a central (passive) node to and from which all communication must occur, and each user transmits its information on an unique wavelength. When setting up a connection, the transmitter and the receiver must be tuned to the same wavelength. Several solutions are proposed in the literature, but they are time-consuming and cause high access delays [1, 4]. Moreover, the wavelength switching is a time-consuming process as well. Thus, there is an urgent need for a better WDM architecture to effectively share the optical bandwidth on optical fiber between various users with diverse and conflicting traffic requirements.

The *service bottleneck* occurs between the MAC and the higher layers, and refers to the difficulties in providing the requested Quality of Service (QoS) for all traffic classes, which compete for common transport resources. This bottleneck arises from the fact that the incoming traffic is highly heterogeneous in its characteristics, having performance requirements with different and conflicting bandwidth demands, holding times and call arrival rates. Moreover, each traffic type may differ in their admissable access and transit delays, end-to-end loss sensitivities, etc. Apart from the heterogeneity in bandwidth demands, the main network task expected by an isochronous service is the isochronicity in the data transport, i.e., bounded delay and jitter for isochronous (real or virtual) connections. On the other hand, the variety of computer applications models leads to a large number of performance requirements on the underlying transport network for nonisochronous traffic. The communications models for such traffic are based on different delay/throughput trade-offs, and they must generally provide an error-free data delivery.

Ideally, "uniform control mechanisms" should be used for the transport of traffic across the optical media, which is difficult to realize. For example, different types of problems must be solved when mixed broadband traffic is carried over ATM networks. These are mainly connected with the complexity of statistical multiplexing, network congestion under peak traffic loads and interference among user-traffic classes.

One of the most severe performance bottlenecks in computer communications is caused by the slow (software) processing versus high-speed transmissions, which we call the *processing bottleneck*. This bottleneck has two components in terms of the increased ratios of (software) processing time to cell/packet transmission time and of propagation delay to cell/packet transmission time. The first component means that there is not enough processing power available to handle the immense amounts of data on optical fiber, whereas large bandwidth-delay products mean that critical problems of management (connection establishment, resource allocation, etc.), flow control and error handling are introduced. The overall effects are significant bandwidth waste and a degrading of network throughput and latency [5]. Therefore, simpler and better control mechanisms for connection management, flow control, error handling, etc., are required in high-speed networks. They must be capable of managing large data flows in transit over the network, so that the throughput is restricted only by the source capacity or the sink ability of the end hosts. Furthermore, the network latency should be minimized for delay-sensitive applications such as used in parallel and distributed environments.

In this paper we propose a novel architectural solution called NEMESIS that is an integrated multi-Gbps optical LAN and is aimed to opening up *all bottlenecks* mentioned before. The NEMESIS architecture is the extended and enhanced version of the SUPERLAN architecture presented in [6, 7].

The remainder of this paper is organized as follows. In section 2 we present the fundamental design issues of the new architecture. In section 3 we introduce the architecture of the NEMESIS. In section 4 we present the Media Access Control (MAC) protocols for NEMESIS and together with preliminary performance results based on analytical modeling. In section 5 we conclude the paper.

0-7803-1820-X/94 $4.00 © 1994 IEEE

2 Design Principles of NEMESIS

In the design of NEMESIS we do not follow the standard layering approach of other LANs. Our preliminary investigations indicated that the vertically-oriented OSI model standard may not be suitable for a multi-Gbps network environment. Therefore, we use a so-called *horizontally-oriented communication model*. Our network model is capable of reducing the effects of the bottlenecks mentioned above. We use a WDM architecture for NEMESIS that is based on the concept of *Wavelength-Dedicated-to-Application (WDA)*. Such an architectural model offers the choice of performance restricted by optics, and less by electronics or processing. We use separate, simplified and application-oriented protocols to support both packet- and circuit-switching.

2.1 Horizontally-Oriented Protocol Layering Concept

We advance an alternative solution to the traditional model of carrying diverse services through separate networks. We propose a new model for multiwavelength-based optical networks in which time-synchronous channels placed in different wavelengths are dedicated to different applications and control mechanisms according to their traffic characteristics. The main impact of this novel architecture is the change of the seven-layer, vertically-oriented OSI model to a four layer horizontally-oriented model, as shown in Fig. 1.

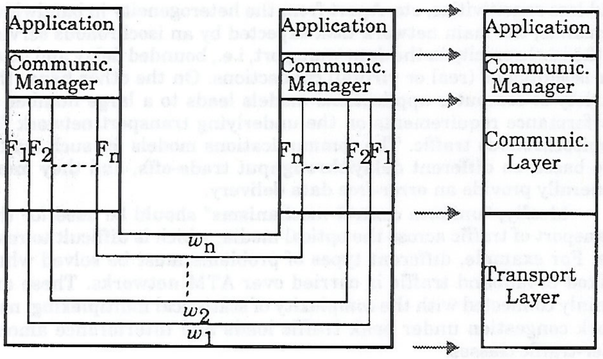

Figure 1. Parallel Communication Model

The protocol functions involved in a communication process are combined into a single layer (Communication Layer), which is horizontally structured. Examples of horizontal functions F_1 - F_n involved in the communication process are encryption, multiaccess mechanisms, flow control, error handling, session management, etc. Generally, some functions are mutually independent, i.e., they can be performed independently of each other, and thus, they can be executed simultaneously. This concept leads us to a parallel model for integrated communication over multiwavelength optical networks. The structural parallelism results in a better latency performance of the communication process, as it is determined now by the slowest function rather than the sum of the processing delays for each layer, as in standard OSI model. As a result, throughput is improved accordingly. Our new model has other advantages, such as the choice of removing redundant functions, to reduce processing overhead and increase processing rates, implementing the protocol stack in customized hardware/multiprocessor architectures, etc.

2.2 The Wavelength-Dedicated-to-Application Concept

We use this concept to open up the opto-electronic interface bottleneck. We no longer dedicate wavelengths to different users, but instead to different applications, such as isochronous traffic, nonisochronous traffic, dedicated control mechanisms for media access, flow control, network clock distribution, etc. The users have permanent access to all wavelengths of interest on fiber. We dedicate at least two wavelengths to each application or traffic class, i.e., a very high capacity channel to carry the user information traffic and one (or more) low capacity channel(s) to carry control information according to different functions F_i. Time synchronization is provided across the different data streams belonging to the same application, but placed in different wavelengths. Through this concept we can develop a network architecture that is restricted by optics and less by electronics or protocol parameters. We can easily integrate diverse traffic classes and not have any performance degradation of other models proposed before to open up the opto-electronic bottleneck. For instance, there is no need for wavelength agility as in other networks based on the Wavelength-Dedicated-to-User concept. This greatly expands the area of applicability. Fast packet switching services are easily provided without requiring technological breakthrough. Additionally, there is no need for adaptive routing to compensate for variable traffic intensities and/or patterns, as in networks based on the multihop concept. Low latencies and jitter can therefore be provided for the information transport that is advantageous for services with strict time constraints. Finally, this network model has the advantage to dedicate diverse network entities, e.g., functions F_i, to a specific application without interfering other applications.

2.3 The Separation Concept

This concept states that separate and distinct resources are provided for the transport of different classes of traffic and control traffic in order to open up the service and processing bottlenecks.

We provide separate resources, in different wavelengths, for the transport of different user traffic classes. Due to the large amount of resources available in a multi-Gbps network, the traditional policy, where all user traffic classes compete for a common resource, in time or in wavelength domain, is no longer appropriate. The advantages of this policy, expressed in terms of service flexibility and network efficiency, do not justify the increased complexity of the protocol processing, congestion control and switching. Instead, we suggest to partition the bandwidth resource among user traffic classes and to employ less complex protocols and switches that work under performance constraints based entirely on specific applications.

We also provide separate transmission resources, in different wavelengths, for the transport of information and control traffic. These two data sets have different requirements. The former generally requires high speed and high throughput links, whereas the latter has a higher priority and requires more reliable transmission, e.g., congestion-free, minimum error rate, etc. The traditional method for integrating control and signaling traffic together with information traffic in the same TDM frame/cell is no longer sufficient. The old solution suffers mainly because of the limited bandwidth available for the control channels. For instance, the maximum achievable throughput in a WDM network based on the traditional model is limited by the product of the packet/cell length of information and the bit rate of the control channel (electronic limits), and not by technological limitations on optical fibers (optical limits) [8].

2.4 The Synchronization Principle

This principle states that temporal relationships are provided across different data streams belonging to the same application, but placed on different wavelengths. The feasibility of the Wavelength-Dedicated-to-Application concept depends heavily on the synchronization issue. Synchronization is also a significant requirement for future multimedia applications. Generally, there are three levels of synchronization identified for ultimate presentation of multimedia objects to the users. The temporal composition of two (or more) concurrent objects occurs on the application and window manager layer (user interaction level), on the object composition level (presentation and scheduling/session of objects), and on the transport level (end-to-end synchronization capabilities). The synchronization concept specifies that this multilayered synchronization model is mapped onto the integrated network model shown in Fig. 1, for the provision of needed synchronization capabilities across the horizontal functions F_1 - F_n. On a low level, the transport system must provide a local synchronization (station-to-station) and a global synchronization (network synchronization). On a high level, temporal relationships must be provided between different objects.

2.5 The Principle of Performance Optimization

This principle is proposed to open up the service and processing bottlenecks. It states that the network entities supporting the communication process need to be application-specific designated to optimize different user performance criteria of interest. Examples of network entities involved in a communication process are (virtual) topology, resource management, diverse control tasks (MAC, flow control, error handling), transmission and switching schemes, etc.

3 The Architecture of the NEMESIS

NEMESIS is an integrated multi-Gbps LAN where data rates up to 9.6 Gbps are provided in every data channel (isochronous or nonisochronous), for a variable number of stations (up to at least 50) and a total network throughput of about 20 Gbps. Electronic logic speeds of 100 Mbps and processing speeds up to 20 - 30 MIPS are considered in design. The total user traffic on the network is separated into two classes, isochronous and nonisochronous, each of which being allocated two, respectively more wavelengths.

A WDM architecture is considered in design that is based on the Wavelength-Dedicated-to-Application concept. Circuit-switching services are considered for isochronous traffic. Also, packet-switching services are considered for nonisochronous traffic, which are based on diverse delay-throughput trade-off. Switched services at different rates, up to 1.2 Gslot/s, are considered for each traffic class, for a number of 8 bits/slot, or more. Communicative and distributive multimedia services are also considered. These may include voice, audio, images, video and data signals, and may need point-to-point and/or multi-point communications among a variable number of stations with a variable number (up to at least 10 for every traffic class) of substations connected to each station. Each substation generates different types of traffic, so every station provides a flexible interconnection to different devices, such as multi-media workstations, high-performance computers, high-capacity storage systems, PBXs, diverse audio, image and video devices, with throughputs independent of the network data rates.

The network model used for study has a physical ring configuration with a number of *(n+1)* stations {S_0, S_1, ..., S_i,..., S_n}. The network has a master station (S_0) and a number of *n* ordinary stations denoted by (S_1 - S_n). A number of *m* substations {SS_{i1}, SS_{i2},..., SS_{ij},..., SS_{im}} are connected to each ordinary station. Here SS_{ij} represents the substation *j* connected to station *i*. A single optical fiber (unidirectional link) is used for station-to-station interconnection. A number of eight wavelength channels are used on fiber.

The master station S_0 provides functions for network supervision, e.g., determination of temporal frames sizes, clock and frame generation and total loop-length adjustments, as well as functions for network operation control, e.g., resource allocation for isochronous traffic and network management. In contrast, the ordinary stations S_1 - S_n have functions in providing communication channels for their (local) traffic. Any ordinary station may transmit and receive simultaneously on both data channels and control channels.

The network is composed of eight logically separate subnetworks, but provides users with the functionality of a single, integrated multi-Gbit/s network. It makes use of eight parallel, wavelength-separated channels with time synchronization provided among subnetworks belonging to the same user-traffic class, as described below:

The *w_cl* Subnetwork This subnetwork provides the distribution of the basic network clock. A logical bus topology is used, where the master station generates a basic clock signal of 100 Mhz, and the other stations extract this clock signal and use it for proper network operation.

The *w_iso_d* Isochronous Data Subnetwork This subnetwork has a *maximum* capacity of 9.6 Gbit/s and provides the transfer of isochronous (user) traffic. It has a virtual ring topology. The transmission form is by "fictive" TDM frames of constant temporal length (125 μsec), and with *variable-capacity, fixed-time* slots. An integer number of frames is provided on ring with the help of a variable-delay register (elastic buffer) in the master station. There are 12500 time slots in a 125 μsec frame. There is no frame synchronization header in frames. Slots are 10 nsec long, and every slot provides a variable capacity ranging from 8 kbit/s to 768 kbit/s (so-called universal time slots). The total number *n* of (fixed-time) bits in each time slot can vary. However, the actual number of bits in each time slot is decided only in terms of the bandwidth need of the particular call allocated to that slot. There is no explicit reading of addresses in the block of user information data, since they are implicitly contained in the slot positions in each frame, which are decided through a call setup procedure. Also, a full-duplex scheme is used in allocation of data slots in the *w_iso_d* channel to a call.

The *w_niso_d* Nonisochronous Data Subnetwork This subnetwork has a *fixed* rate of 9.6 Gbit/s and provides the transfer of nonisochronous user information. It is a subnetwork similar to the *w_iso_d* subnetwork, but with *fixed-capacity, fixed-time* slots. The slot length is of 10 nsec duration, but each slot carries now a fixed number of bits corresponding to the subnetwork bit rate (i.e., $n = 96$ bits). Furthermore, the temporal frame size in *w_niso_d* and, accordingly, the number of time slots in a frame is adjustable and it is decided according to delay-throughput needs of different nonisochronous traffic carried. However, constant-length TDM frames are used in operation. Also, a destination removal scheme is used for removing transmitted data from *w_niso_d.*

The *w_iso_c* Isochronous Control Subnetwork This subnetwork has a *fixed* capacity of 100 Mbit/s and supports resource allocation (MAC) information flows and connection management (signaling) flows for isochronous traffic. It is a subnetwork running in a time-synchronous fashion with the *w_iso_d* subnetwork. A virtual bus topology is used. A synchronous transport mode with constant-length TDM frames is used. A train of 125 μsec TDM frames with 100 Mbit/s bit rate and *(a + 2)* time slots in a frame flows continuously along the subnetwork, originating and ending at the master station. The first slot serves as a frame synchronization header. The next *a* time slots contain various network control information for actual communication among the network users. They provide the transport of special dedicated control units, called *cells*, and the number *a* is chosen according to the media access mechanism for isochronous traffic. The last time slot, the trailer, is used for auxiliary functions (data stream synchronization, reservation bits, etc.). Because of fairness reasons, each station has (free) access to only a limited number of time slots in a frame.

The *w_niso_c* Nonisochronous Control Subnetwork This subnetwork has a *fixed* capacity of 100 Mbit/s and supports resource allocation (MAC) information flows and connection management flows for nonisochronous traffic. It runs in a time-synchronous fashion with the *w_niso_d* subnetwork. A virtual ring topology is used. A synchronous transport mode with constant-length TDM frames is used, where the frame size is exactly the same as the *w_niso_d* frame. Also, the form of the *w_niso_c* frame is similar to that of the *w_iso_c* frame. It contains *(b + 2)* time slots. The first time slot is used for frame synchronization, the next *b* time slots are used for the transport of control information (cells) and the last trailer slot is used for auxiliary functions. Similar to *w_iso_c*, the number of time slots that can be accessed by each station in a frame is limited. This number is decided according to the fairness criterion adopted for *w_niso_c*. Also, a source removal scheme is used for the removal of transmitted cells from *w_niso_c.*

The *w_err* Subnetwork This is used for the error handling of nonisochronous data. A redundant transmission scheme is used for error correction, and where the error handling is done "on the fly" at the slot arrival.

The *w_flc* Subnetwork This subnetwork supports flow control functions for nonisochronous data. Preventive, (open-loop) flow control mechanisms are dedicated to individual applications to prevent long-term overflow of the end buffers. Also, short-term overflow is handled by using a double-threshold configuration for the end buffers. Accordingly, the *w_flc* channel is used to provide the transport of flow control messages on a collision-free basis.

The *w_mng* Subnetwork This subnetwork is used for network management purposes, with particular reference to fault management. In addition to fault management, other functions incorporated in the network management entity are performance management, configuration management, diverse auxiliary functions, etc. The most important functions, however, are those connected with the "real-time" detection and correction of faults, as well as activation, deactivation and setting of diverse parameters needed for a proper operation of the network.

All these subnetworks are integrated at the transmission level, i.e., they share common transmission facilities between switching nodes, but maintain separate facilities for media access and buffering resources inside the stations. Ordinary stations are connected, at a minimum, to the *w_cl* and *w_mng* subnetworks and, according to their service needs, they may be connected to the other subnetworks as well.

4 Media Access Protocols

A specific solution is advanced for the MAC protocol in a WDA architecture. According to this, each traffic class/application is provided with its own simple, low-speed, application-oriented MAC protocol, with no interference from other applications. The MAC protocols are of type reservation and they are separated in wavelength domain. Their main parameters are chosen based entirely on application needs of interest. Two control channels, *w_iso_c* and *w_niso_c*, placed in two distinct wavelengths, are dedicated to multiaccess mechanisms for isochronous, and nonisochronous traffic.

4.1 Media Access Protocols for Isochronous Traffic

We develop a Connection-Oriented (CO) procedure with a centralized MAC protocol. We provide a contention resolution of type blocking. The three phases in a CO procedure are supported by different subnetworks in NEMESIS, i.e., the connection and the termination phases by the *w_iso_c* control subnetwork, and the data transfer by the *w_iso_d* isochronous data subnetwork. In the first phase, we consider unknown statistics for isochronous traffic. We use admission control on the call level, based on the use of peak rate for different isochronous traffic classes, both CBR and VBR. We partition the isochronous bandwidth resource available in the *w_iso_d* data subnetwork (up to 9.6 Gbps) into separate bandwidth pools, dedicated *to different isochronous traffic classes, so as to provide equalization of* the blocking probabilities (i.e., fairness) among different traffic with different offered loads and bandwidth requirements. We achieve this on the basis of a multidimensional Engset loss formula [9]. Dynamic resource allocation policies are considered in order to provide a better blocking performance. We also consider separation of CBR from VBR traffic (in time or wavelength) and investigate admission control mechanisms based on statistical multiplexing.

The ordinary station S_i^{iso} is modeled by a multiqueue system with a single cyclic server for the transmission side, and a buffer and two servers for the receiving side. A head-of-line (HOL) non-preemptive M/D/1 model with three queues is used to model the transmission side. These queues are dedicated to disengagement requests (priority 2), signaling messages (priority 3) and requests for call setup (priority 4). The highest priority in the transmit multiqueue is given to the incoming upstream traffic, i.e., the incoming control cells from the *w_iso_c* channel that are not addressed to that station, and therefore go further to the next station.

The master station S_0 has a fork-join model, where the different incoming cells are differentiated, processed, and joined for further transmission onto the *w_iso_c* channel. Four servers and four queues are used to model the master station. The join model has a HOL non-preemptive M/D/1 form, with two queues. These queues are dedicated to signaling messages (priority 1) and MAC messages (priority 2).

The *call setup delay time* for an isochronous call includes [9]: the queueing delay at the ordinary station; the access time at the ordinary station, i.e., the time taken for the first element/cell in queue to get its first free (temporal) slot in *w_iso_c*; the transmission times at the ordinary and master stations; the transmission delays to and from the master station, including the propagation delay on optical fiber and cross-station delays through the intermediate stations; the queueing delay at the master station; the processing time at the master station; and the synchronization latency between the data channel *w_iso_d* and the control channel *w_iso_c*.

Figure 2 shows the variation of the expected call setup delay time with the processing time in the master station T_{MS} (expressed in terms of the time-slot unit in *w_iso_c* m_{ctr} = 740 nsec), for different numbers *n* of stations available on the ring and a control cell size of 74 bits/cell. A long distance of 1.5 km is assumed between stations.

Good performance results are obtained for call setup delay times such that they do not exceed 1.5 msec even for extreme conditions such as assuming long distance between stations, maximum number of stations, high arrival rates for call setup requests, and large processing times in master station to serve the requests for call setup. Also, it is seen from this figure that there is no congestion in the *w_iso_c* channel, and delay requirements are well fulfilled. A large reserve of processing capability is available in the master station to develop specific mechanisms for resource partitioning according to different blocking requirements [9].

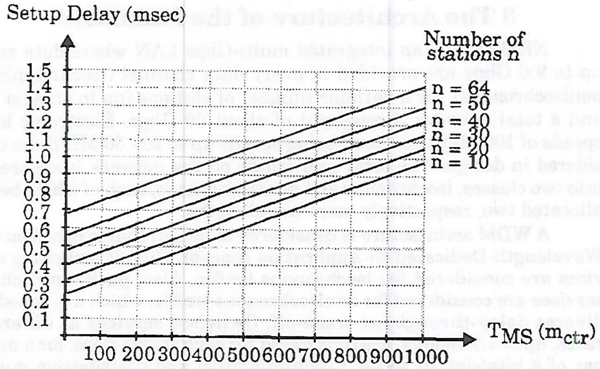

Figure 2. Call Setup Delay Time for Isochronous Traffic

4.2 Media Access Protocols for Nonisochronous Traffic

For the nonisochronous traffic we use a Connectionless (CL) procedure with a distributed MAC protocol working at a packet level, and with a congestion-tolerant mechanism based on dynamic resource management. The two phases of a CL communication procedure are supported by different subnetworks, i.e., the connection establishment (MAC decision) by *w_niso_c* channel, and the data transfer by *w_niso_d* channel.

To provide minimum access delay, a fast reservation scheme is used for the MAC protocol. Also, in order to provide bounded delay and jitter requirements for loss-free nonisochronous traffic, we suggest a new approach for congestion control and bandwidth allocation, which is of the Rate Controlled Admission with Priority Scheduling Service type [9]. We call this *Dynamic Time Sharing (DTS)*. This approach is based on guaranteeing specific traffic parameters (bandwidth requirements) through a policer/shaper unit, and then *optimizing the bandwidth assignment, within the network, for* specific parameters of interest (like delay or jitter, and loss). The optimization process is based on the parameters guaranteed by the shaper. A batch admission policy is used at the edges of the network according to a specific framing strategy to follow the traffic characteristics (e.g., peak bandwidth) of different traffic classes. On the other hand, another framing (congestion control) strategy is used within the network, which is based on different (delay/loss) requirements of the traffic classes. Proper management of bandwidth and buffer resources is provided in every (switch) node of the network, such as to guarantee the diverse performance of interest, regardless of traffic statistics, network size and transmission speed. We also consider segregation of the nonisochronous traffic classes in the wavelength domain, and with distinct multiaccess mechanisms.

A nonisochronous station S_i^{niso} has different, and distinct, multiqueue systems that are dedicated to serve cells in *w_niso_c*, and data messages in *w_niso_d*.

A HOL non-preemptive M/D/1 model, with a single cyclic server and *l* queues, is used to model the transmit multiqueue system in the control channel. These queues are dedicated to request cells for data transport of different priorities, as requested by the local hosts. The highest priority in this multiqueue is given to incoming upstream traffic, i.e., incoming control cells from the *w_niso_c* channel.

A HOL preemptive M/G/1 model, with a cyclic server and a number of *2l* queues, is used to model the transmit multiqueue system in the data channel. Every subclass of traffic is dedicated two specific queues, one for local traffic, and the other for transit traffic.

Besides these, there is one more set of buffers, the receive multiqueue system, to serve the incoming data messages that are addressed to hosts connected to that station.

The end-to-end delay, for a message/batch of class *k* (where *k = 1 to l*), has two components, i.e., the MAC delay and the transport delay in data channel *w_niso_d* [9]. The MAC delay includes: the queueing delay for the control cell; the access time at the control transmit multiqueue, i.e., the time taken for the first cell in queue to get its (free) temporal slot in *w_niso_c*; the transmission delay for the control cell; the synchronization latency between the control and data channels; and the access time for the data message (including the transmission delay for the specific data message). The transport delay has two components related to the propagation delay on optical fiber and the cross-node delays through the intermediate stations.

Figures 3 to 5 show the variation of the expected end-to-end delay for three classes of nonisochronous traffic. These classes of traffic correspond to applications of type Remote Procedure Call - RPC (class of priority 1), Demand Paging - DP (class of priority 2), and (bulk) data files (class of priority 3). Distributions of the type (truncated) exponential are considered for the batch sizes. Variable number of stations (limited to 50) and substations (limited to 10, for one station) are also considered. Furthermore, a long distance of 2 km is considered between stations.

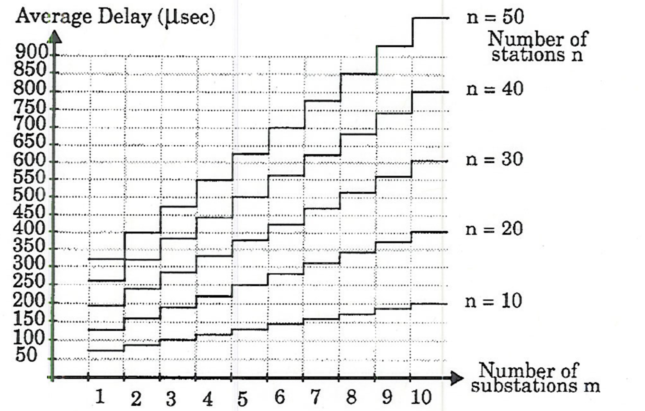

Figure 3. End-to-End Delay Time for Class 1 Nonisochronous Traffic

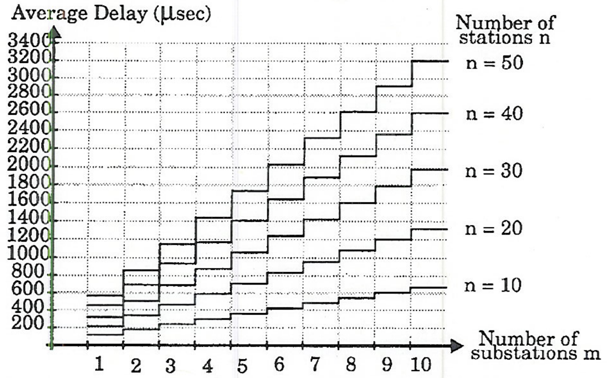

Figure 4. End-to-End Delay Time for Class 2 Nonisochronous Traffic

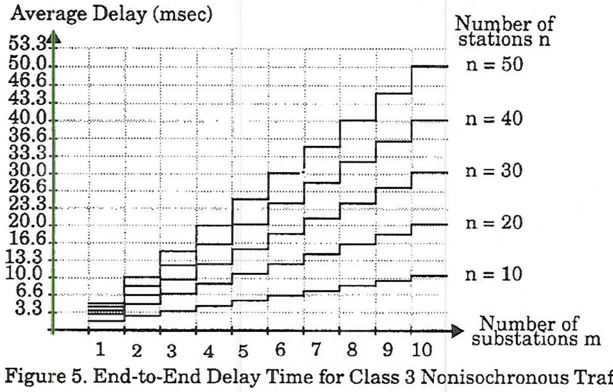

Figure 5. End-to-End Delay Time for Class 3 Nonisochronous Traffic

Good performance results are obtained for the MAC and end-to-end delays. These delays do not exceed hundreds of µsec for class-1 traffic, units of msec for class-2 traffic and tens of msec for class-3 traffic. These good performance results are mainly due to the proper network resource allocation (according to traffic and network conditions), with the consequence that the congestion within the network is practically eliminated, both in data and in control channels.

5 Conclusions

A generic architecture for multi-Gbit/s optical integrated LANs based on the Wavelength-Dedicated-to-Application concept has been motivated and presented. A horizontally-oriented communication model is advanced to open up the three fundamental bottlenecks identified in a multi-Gbit/s integrated communication over multiwavelength optical networks. The distinguishing features of this novel architecture include: horizontally-oriented communication model; WDM architecture with Wavelength-Dedicated-to-Application; utilization of a coarse WDM to alleviate the opto-electronic bottleneck; separation of different user-traffic, and of information traffic from control traffic, to open up the service bottleneck; and separate, simplified and application-oriented protocols supporting both packet- and circuit-switching to open up the processing bottleneck;

The novelty of NEMESIS is given mainly by the concept of allocating different wavelengths to different applications according to their traffic characteristics, thereby making use, in a better way, of the abundant bandwidth available in the fiber. The performance results clearly indicate the feasibility of this new concept, as a candidate for the future multi-gigabit communication over multiwavelength optical networks. Such an architectural model offers the choice of performance restricted by optics, and less by electronics or processing. It offers the choice of reducing the latency in communication to limits which are commensurable with those given by the propagation delay on fiber. Real-time services, with specified delay and bandwidth requirements, can be easily provided. Simple and low-processing application-oriented protocols are used, with no interference among user-traffic classes. It is therefore very suitable for service integration, irrespective of traffic and network conditions. There is no need for technological breakthroughs, no need for wavelength agility, and LAN environments can be easily developed that are capable of supporting large number of supercomputers. The disadvantage of this approach is given mainly by the hardware replication that is needed in every node for each traffic class.

References

[1] Brackett, C.A.,"Dense Wavelength Division Multiplexing Networks: Principles and Applications," *IEEE Journal on Selected Areas in Communications*, Vol. 8, No. 6, pp. 948 - 964, Aug. 1990.

[2] Eisenberg, M. and Mehravari, N.,"Performance of the Multichannel Multihop Lightwave Network Under Nonuniform Traffic," *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 7, pp. 1,063 - 1,078, Aug. 1988.

[3] Acampora, A.S. and Shah, S.I.,"Multihop Lightwave Networks: A Comparison of Store-and-Forward and Hot-Potato Routing," *IEEE Transactions on Communications*, Vol. 40, No. 6, pp. 1,082 - 1,090, June 1992.

[4] Green, P.E., *Fiber Optic Networks*, Prentice Hall, Englewood Cliffs, New Jersey 07632, USA, 1993.

[5] Doeringen, W.A., Dykeman, D., Kaiserswerth, M., Meister, B.W., Rudin, H. and Williamson, R.,"A Survey of Light-Weight Transport Protocols for High-Speed Networks," *IEEE Transactions on Communications*, Vol. 38, No. 11, pp. 2,025 - 2,039, Nov. 1990.

[6] Popescu, A., Ismailov, Y., Rajaei, H. and Ayani, R.,"Modeling and Performance Evaluation of Multiaccess Mechanisms at SUPERLAN," in *Proceedings of MASCOTS'93*, San Diego, California, pp. 176 - 182, Jan. 1993.

[7] Popescu, A. and Singh, R.P.,"An Alternative Solution to the Electro-Optic and Service Bottlenecks in Integrated Multi-Gbit/s LANs: the SUPERLAN Architecture," *Computer Networks and ISDN Systems*, Vol. 25, pp. 1,089 - 1,105, May 1993.

[8] Brackett, C.A.,"On the Capacity of Multiwavelength Optical-Star Packet Switches," *IEEE Journal of Lightwave Telecommunications Systems*, Vol. 2, No. 2, pp. 33 - 37, May 1991.

[9] Popescu, A., *A Parallel Approach to Integrated Multi-Gbit/s Communication over Multiwavelength Optical Networks*, Ph.D. Dissertation, TRITA - IT - 9306, Stockholm, Sweden, May 1994.