*Journal of Network and Systems Management, Vol. 10, No. 3, September 2002 (© 2002)*

# Design and Analysis of a Backbone Architecture with TDMA Mechanism for IP Optical Networking

**Chi-Yuan Chang[1] and Sy-Yen Kuo[1,2]**

This paper presents a new technique for constructing IP over photonic systems. The use of lable switching is assumed in the IP routers, while a new routing architecture is introduced to transport IP packets across an optical backbone network. The architecture is based on a two-level TDMA structure with wavelength division multiplexing (WDM). Many IP-based network applications such as high-resolution image, distributed database, and real-time video/audio service generally require high-speed transmissions in WAN/LAN. The network traffic in these applications usually exhibits traffic locality. As a result, traditional TDMA is not efficient for such traffic. Consequently, based on the traffic parameters such as locality and loading, an architecture named a PG (Partition-Group) Network is proposed. Furthermore, the interleaved control slot (ICS) with cross-group section (CGS) or non-cross-group section (NCGS) for reducing collisions is also presented. The slot reuse can be easily achieved by using the ICS scheme, and the slot utilization of the network can be improved within the high traffic locality.

**KEY WORDS:** Wavelength division multiplexing (WDM); time division multiplexing access (TDMA); optical networks; IP over WDM.

## 1. INTRODUCTION

Many new IP-based networks such as high-resolution image, distributed database, and real-time video/audio service with high performance needs, require more bandwidth/quality than ever before. These applications generally require the data to be transmitted very fast under heavy traffic in a wide area network (WAN) or local area network (LAN). Optical fiber has been widely adopted for satisfying these requirements. The wavelength division multiplexing (WDM) technique [1] is widely employed to fully utilize the huge bandwith available on optical fibers, and has contributed significantly to high-speed communications.

[1]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan.
[2]To whom correspondence should be addressed at Department of Electrical Engineering, Room 444, National Taiwan University, Taipei, Taiwan. E-mail: *sykuo@cc.ee.ntu.edu.tw*



1064-7570/02/0900-0383/0 © 2002 Plenum Publishing Corporation

### 1.1. IP over WDM

The diffusion of Internet traffic, IP-based networks, and applications is growing at an exponential rate in both private and public networks while IP is becoming the dominant protocol for information communication technology. Thus, IP over WDM has become a very important area of study. The conventional IP network, based on a two-tier architecture, has a lower tier of edge routers and an upper tier of big routers. The lower tier of edge routers operates for the traffic within the region, and the upper tier of big routers operates for the traffic among remote region. However these routers are operated based on the store-and-forward principle, need optical/electrical/optical (O/E/O) conversion, and so forbid the all-optical operations in the two-tier architecture. To avoid these drawbacks, much research effort focuses on developing an elegant solution to the mismatch between the transmission capacities offered by the WDM optical layer and the processing power of routers.

### 1.2. Related Works

Recent developments in multiprotocol label switching (MPLS) open new possibilities to address some of the limitations of the traditional IP systems. MPLS switches use a simple label-swapping algorithm replacing the standard destination-based hop-by-hop forwarding paradigm to quickly forward packets [2–4], enabling easy scaling to terabit rates. The MPLS technique uses a label to save significant processing time by avoiding network layer label analysis at each hop. In addition, the MPLS can provide many of the same advantages of a connection-oriented network while still retaining the underlying efficiency and operation of a datagram network.

Current applications of WDM as a networking technique focus on relatively static utilization of individual wavelength channels. These wavelength routing approaches are still basically variants of the circuit-switching paradigm, resulting in very inefficient optical bandwidth usage. A technological breakthrough in this direction is represented by optical packet switching [5], enabling fast allocation of the WDM channels and their utilization as shared resources in an on-demand fashion with very fine granularities. Recently, variable-length optical packet switching, such as optical burst switching (OBS) [6], has been proposed as an optical switching paradigm to combine the best of optical circuit and packet switching. OBS could achieve high bandwidth utilization with lower average processing and synchronization overhead than pure packet switching. All these researches are all focus on designing a technique to improve the parameters in terms of transmission scheduling, synchronization issues, contention resolution, and switching strategies, but never take the traffic characteristic into consideration.

On the other hand, for the foreseeable future, the Internet will continue to be an interconnection of routers/gateways sitting on top of a transport. These

routers/gateways are connected to many topologies. There are several basic physical network topologies for multiaccess WDM network: the bus, the ring, and the mesh. Other types of topologies can be easily extended by these basic topologies. Here, we propose and analyze a new routing architecture for IP operating on a WDM backbone network with different topologies. We also investigate the relationships between the traffic characteristic and performance. The proposed routing architecture is based on the multiplexing approach, with WDM addressing the number of regional exchanges (REXs) and time-division switches communicating among the backbone-hubs, which is essentially an optical crossconnect featuring a classical time-division space switch.

Several multichannel bus/ring lightwave networks [7–11] had been reported. In the past, researchers [12] have taken advantage of TDMA (time division multiplexing access) technique in IP over WDM backbone environment. TDMA is a direct solution to achieving fair transmission. TDMA corresponds to a fixed partitioning of the channel bandwidth among all the possible transmitters. But it does not allow a node/hub to grab more than the portion of the channel bandwidth assigned to it even if no other nodes/hubs are transmitting on the same channels. In addition, for a network with a large number of nodes over distinct areas, the slot utilization of network is proportional to the performance of network. For a lightly loaded traffic, the delay performance of the TDMA is very poor because a node must wait for the assigned slot before transmitting a newly generated packet.

### 1.3. Traffic Locality

The 80%-20% rule, 80% income from 20% of customers, can be found everywhere. It is also adopted and named as *traffic locality* in the network environment [5, 13–15]. Claffy *et al*. [15] plot the cumulative distribution of messages sent from and to the $n$ busiest source and destination networks within the NSFNET. Over 50% of the traffic is generated by the busiest 31 of the 4254 site networks (0.7%), over 50% travels to the 118 most popular (2.8%) destinations and 46.9% of the total traffic on the backbone travels between 1500 (0.28%) of the 560,049 site-pairs.

Many network applications such as multimedia and video conferencing, usually exhibit traffic locality (which means most traffic between the transmitter node and the receiver node is located at some specific areas) [13, 14]. This type of traffic uses relatively high bandwidth on a continuous basis for a long period of time. Besides, for the traffic with higher priority, the network can also route the traffic as soon as possible by sacrificing the lower priority traffic. Although networks with WDM and TDMA are popular, traditional TDMA is no longer suitable for such traffic because of inefficiency and lack of multilevel priority. Hence, based upon traffic locality and loading (light or heavy) characteristics, an architecture named as PG (Partition-Group) is proposed to partition/reconfigure the network into several control groups. The PG consists of $r$ different groups of stations. All

stations in each group exhibit traffic locality and have the same control channel. The control channel, named as the interleaved control slot (ICS), is arranged by overlapping half cycle. By partitioning/reconfiguring a network as much as possible and using ICS as control protocol, the slot reuse can be easily achieved and the slot utilization of the network can be imporoved within the high traffic locality.

This paper is organized as follows. In Section 2 we discuss the architecture and functional blocks for IP-based WDM networks. Section 3 describes the PG architecture with the interleaved control slot concept and presents the assignment of wavelengths. Section 4 discusses the traffic types and presents an algorithm to reconfigure the network for dynamic loading balance. Section 5 depicts a simulation model and the simulation results. In Section 6 we conclude the paper.

## 2. NETWORK ARCHITECTURE

### 2.1. System Level Architecture

Recently, there has been an increasing interest in the implementation of IP over photonic networks by using optical networking techniques [2–4, 16]. Consistent with this trend, IP with MPLS [2–4] over a WDM-based wavelength switching packet network is proposed as a solution to IP over photonic network. Figure 1 illustrates the general network architecture. The architecture consists of (1) an Internet access part (IAP), (2) a regional exchange (REX), and (3) an optical-hub

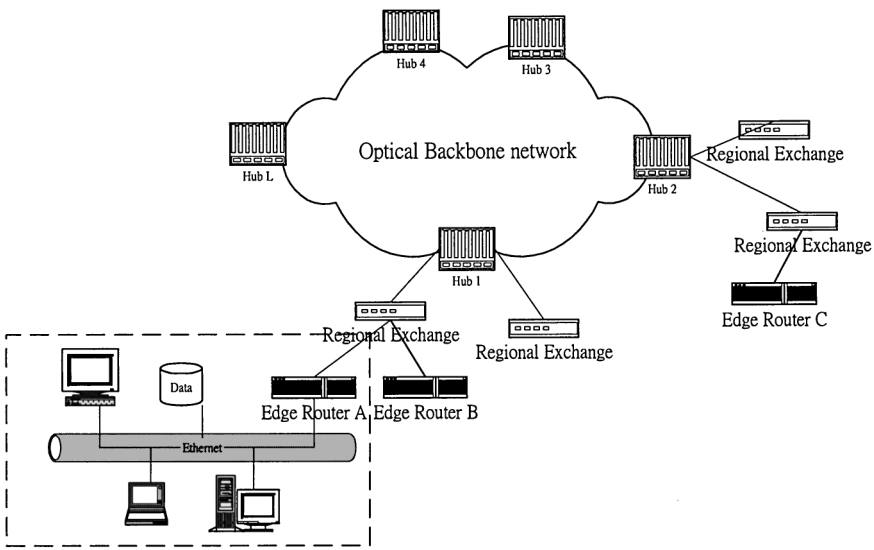


**Fig. 1.** The general network architecture.

switching backbone (OSB). The IAP, as shown in Fig. 1, consists of IP hosts connected to corporate serves via a local-area network (LAN) for offices or connected to an Internet services provider (ISP) via phone-line modems, digital subscriber line (xDSL), cable modems, or wireless access. The REX and OSB from the optical switching system for the IP network.

The network can be divided into two independent levels, the switching-hub level and the REX level. The former performs IP packet exchange in the optical domain within hub, while the latter performs the same task in the electrical domain. The switching-hub level and the REX level are described below respectively.

*2.1.1. Switching-Hub Level*

The connections between hubs can be a bus/dual bus, a ring/dual ring, a mesh (fully connected network), or other interconnection networks. We choose the fully connected network for its simplicity and its robustness. Specifically, $N$ (the number of hubs) switching states can easily be generated, and there are $N-1$ alternate paths to use when the direct path is down. The link utilization is limited to $1/(N-1)$ for this network. For nonfully connected interconnection between hubs, complicated switching patterns need to be designed so as to emulate a fully connected network with end-to-end path for all connections between REXs within the network. Let us begin by examining the proposed architecture at the hub level. The hub shown in Fig. 1 is for the exchange of packets between regions within the same hub or across distributed hubs. For example, if a source transmits data to the destination in the same region, the data will not be transmitted to other hub region (e.g., edge router A→edge router B in Fig. 1). On the contrast, the data will be transfer to distributed hub of remote region if the source and destination pair is in different region (e.g., edge router A→edge router C in Fig. 1). The architecture of the hub switching network shown in Fig. 2 is implemented by a time-division

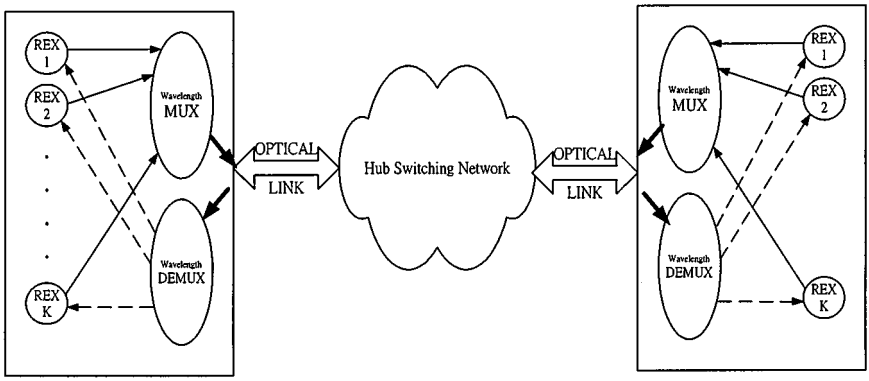


**Fig. 2.** Network with K connected REXs.

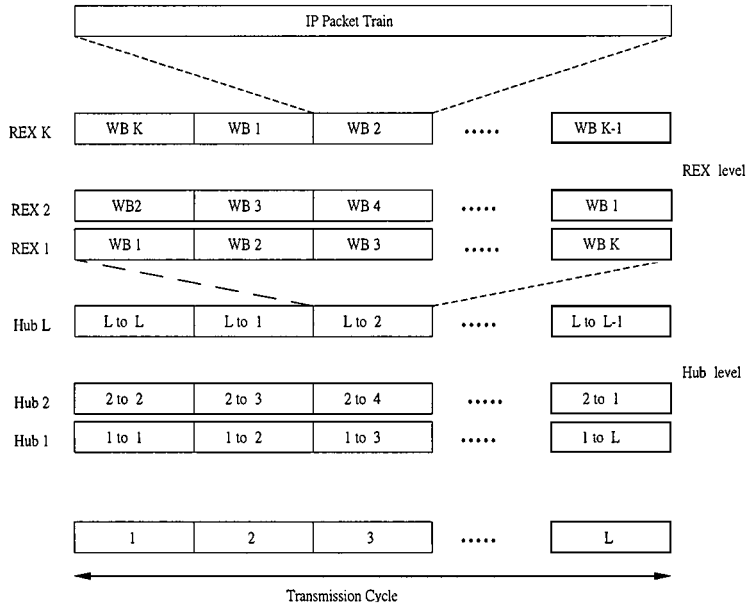


**Fig. 3.** The transmission cycle.

optical switch, as well as the switching states are cycled through in each period. The resulting transmission orders for the hubs are shown in the hub level of Fig. 3. In each time slot of transmission cycle, the hub transfers its packets to predefined hub. For example, in the time slot 2 of transmission cycle, the hub 1 to hub 2, to hub 2 to hub 3, . . . , and hub $L$ to hub 1, etc.

*2.1.2. REX Level*

At the REX level, let us first focus on intra-REX communication. The exchange of IP packets between edge routers in the same region is performed in the electrical domain with the local REX switch playing the role of a "switching" router. To allow higher throughput, label switching [2–4] is used. As we are concerned with the transport of IP packets, the label can be included as part of the header of Layer 3 (i.e., by using the Flow Label field in the IPv6 with appropriately modified semantics [17]). This label indexing is done at the edge router to ease the processing load of the REX switches. Since a label is bound to an IP address prefix, IP packets heading to a group of destinations (i.e., to edge routers connected at a specific region on a specific hub) can share the same label. For IP packet routing within a region, the destination router addresses on the labels are resolved at the local REX switch. IP packets destined for edge routers at a remote region are

assigned to different transmission queues according to their respective labels. The transmissions from each queue onto the optical backbone network are then organized into transmission cycles, with each cycle subdivided into wavelength burst (WB) periods using WDM. The WB periods consist of a series of wavelengths to different destinations. To illustrate, let us focus at Hub $L$ in period 3, as shown in the REX Level of Fig. 3. Here, Hub $L$ is connected to Hub 2. During this period, REX 1 of Hub $L$ transmits WB $\lambda_1, \lambda_2, \ldots, \lambda_k$ to REX $1, 2, \ldots,$ REX $K$ in Hub 2 (assuming there are $K$ REXs in both Hub $L$ and Hub 2). In the same period, REX 2 of Hub $L$ transmits WB $\lambda_2, \ldots, \lambda_k, \lambda_1$ to REX $2, 3, \ldots,$ REX 1 in Hub 2. Other REXs use cyclic permutations in the wavelengths so that no two wavelengths are used at the same time. These transmissions are merged at a coupler in the hub before being sent out to another hub. This is illustrated in Fig. 2. Note also that the flow of intra-REX traffic is decoupled from the flow of inter-REX traffic. Figure 3 shows the transmission cycle at different levels of switching hub level, REX level and IP packets level. Here we assume that a network has $L$ hubs and each hub contains several REXs. There are $K$ REXs in a network.

## 2.2. Functional Block of Edge Router and REX Switch

Figure 4 shows a functional diagram of the edge router and the REX switch. A station is equipped with an edge router and its local REX switch. In the edge router, the Input Dispatcher sorts IP packets from connected servers. Those destined for servers attached to the local router are buffered and sent to the Output Dispatcher. Those destined for remote routers are sent to the label-indexing buffer with the MPLS technology [2–4]. The MPLS is used to achieve fast and efficient packet forwarding and perform traffic engineering. Conventional IP protocols such

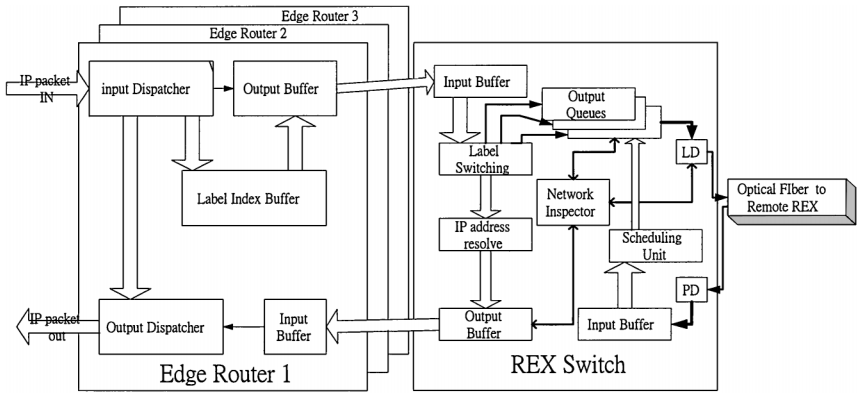


**Fig. 4.** The Functional block of edge router and REX switch.

as OSPF, RSVP, and PIM can be used for label distributions. IP packets in the label-indexing buffer are labeled according to their destination-router addresses and placed on the output buffer. These are then sent to the input buffer of the local REX switch where IP packets from individual edge routers are switched to output queues, according to their respective label, to form IP packet trains for inter-REX routing. In other words, IP packets from the same train are all destined to the same remote region. The destination addresses of those IP packets labeled for the local region will be resolved, and the packets are rerouted to their respective destination edge routers.

On the input side of the optical links, optical signals carrying labeled IP packet trains from the hub are converted to electrical signals before breaking up into individual IP packets. The REX switch then resolves the IP destination addresses of these packets and passes them to their respective destination edge routers. For example, as shown in Fig. 1, a label is used to identify a traffic flow from local edge router A $\rightarrow$ local REX switch $\rightarrow$ local edge router B or between local edge router A $\rightarrow$ local REX switch $\rightarrow$ hub 1 $\rightarrow$ hub 2 $\rightarrow$ remote REX switch $\rightarrow$ remote edge router C. The network inspector shown in Fig. 4 monitors all the packets of the outpout queues in the REX switch. It monitors all the transmission conditions from the edge router and inspects the traffic from the remote REX switch. In this way, we can know the load distribution along the optical backbone and decide on the switching reconfiguration in the network. The details of network inspector will be discussed in Section 4.1.

### 2.3. Modified Transmission Cycle

The transmission cycle shown in Fig. 3 lists the benefits for combination of WDM and TDMA under a full-connected hub environment. The hub $L$ can transmit to other $L$-1 hubs simultaneously by different wavelengths of independent optical paths on the transmission slot $L$. If the hubs in the network are not fully connected, such as a ring or a bus, the transmission trains for hub L in the $L$th slot must be sequential instead of parallel. Hence the length for a transmission cycle becomes very long. It means that each REX in a hub needs to wait for a longer time to transmit. Therefore, an architecture named as PG (Partition-Group) with ICS (Interleave Control Slot) is proposed to partition/reconfigure the network into several control groups. The number of control group is denoted as $r$.

Figure 5 shows the modified transmission cycle by PG, $r = 2$ at the switching hub level, the REX level and the IP packets level. The PG consists of $r$ different groups of stations. All the stations in a group exhibit traffic locality and have the same control channel. Figure 5 shows the slot-interleaved control slot for $r = 2$, where stations 1, 3, 5, 7 are together as group 1 and stations 2, 4, 6, 8 are group 2. The details will be discussed in Section 3.1.

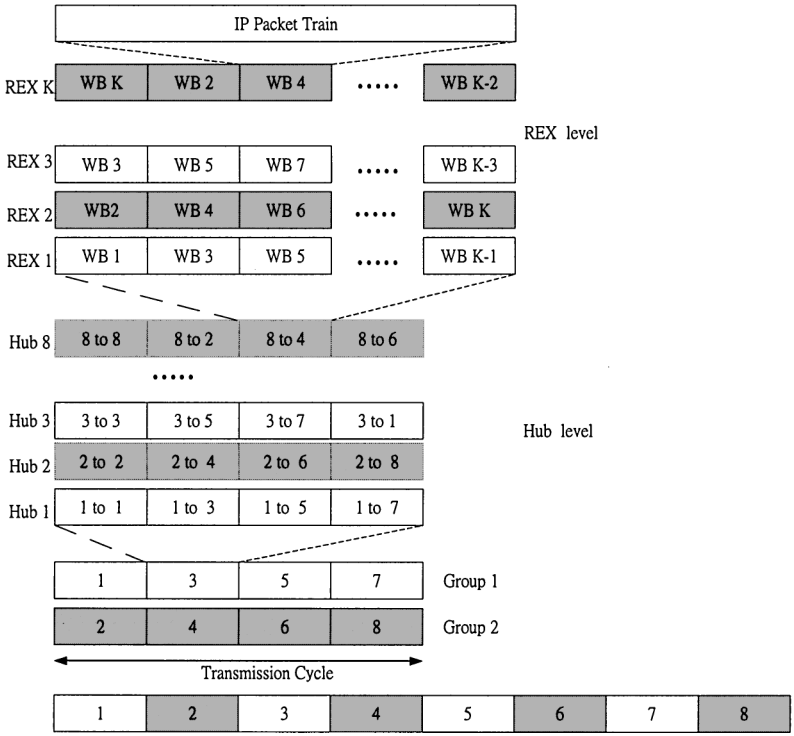


**Fig. 5.** The modified transmission cycle for PG, $r = 2$.

## 3. PARTITION-GROUP ARCHITECTURE

The connections between hubs can be a bus/dual a ring/dual ring, a mesh (fully connected network), or other interconnection networks. We choose the dual bus connected network for its simplicity and its robustness. The dual bus topology can be easily transferred to other network topology such as ring/dual ring or mesh by changing the station architecture. In fact, a Dual Bus network can be physically arranged according to the looped bus physical topology [9, 18], so that the two buses are closed in a dual ring fashion, using a special node to realize the enclosure connection.

In Figure 6, the dual bus lightwave network (DBLN) is considered and each station is equipped with some tunable transmitters (TTs) and a fixed receiver (FR). The DBLN consists of two optical buses, Bus A and Bus B, respectively. Each bus has $k + 1$ wavelengths/channels ($\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_k$) by employing the WDM technology. The DBLN consists of $N$ stations equipped with hub, REX, and LAN. The stations are labeled from left (upstream) to right (downstream) as 1 to $N$ with

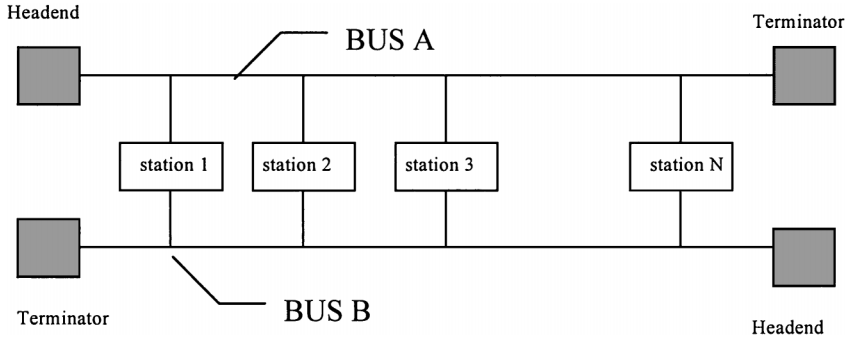


**Fig. 6.** The architecture of a dual bus lightwave network.

respect to Bus A. The traditional approach uses a wavelength to be the control channel to schedule a station to transmit data [9, 10, 13]. Hence, each station has one fixed receiver for receiving TDM-based control channel. The pseudo cycle with a length of $N$ control slots is generated on the control channel, and only one control slot is allocated for each station in a cycle. On the other hand, a tunable transmitter is used for transmitting data to destination and a fixed/tunable receiver is used for receiving data form source. For the sake of simplicity, we assume that the number of wavelengths is larger than the number of stations, that is $k \geq N$. This implies that a conflict case can be avoided and is temporarily not taken into consideration now. The other cases such as $k < N$ are discussed in Section 3.2. Figure 7 shows the channel assignment of a network with $N = 8$ and $k = 8$. The station has a fixed receiver tuned to $\lambda_0$ as the access control channel, a tunable transmitter *Tx* which can tune wavelengths from $\lambda_1$ to $\lambda_8$, and a fixed receiver *Rx*. When a station $i$ wants to transmit data to station $j$, there are several steps to set up before transmitting. First, the tunable transmitter *Tx* needs to tune its wavelength to $\lambda_j$ and then wait for transmitting data until the $i$th control slot is accessed by the control channel. Station $i$ can only transmit in the interval of the $i$th control slot. If station $i$ needs to transmit data again, it must wait until the $i$th control slot

| Station | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| cont ch. | $\lambda_0$ | $\lambda_0$ | $\lambda_0$ | $\lambda_0$ | $\lambda_0$ | $\lambda_0$ | $\lambda_0$ | $\lambda_0$ |
| Tx | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ |
| Rx | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | $\lambda_6$ | $\lambda_7$ | $\lambda_8$ |

**Fig. 7.** The channel assignment for nodes in a network.

| Station | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| cont ch. | $\lambda_0$ | $\lambda'_0$ | $\lambda_0$ | $\lambda'_0$ | $\lambda_0$ | $\lambda'_0$ | $\lambda_0$ | $\lambda'_0$ |
| Tx | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ | $\lambda_1,\cdots,\lambda_8$ |
| Rx | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | $\lambda_6$ | $\lambda_7$ | $\lambda_8$ |

**Fig. 8.** The channel assignment for nodes in a PG, $r = 2$.

of next control cycle. It is clear that each station must wait for $n$ control slots to transmit data again. If station $i$ has a great deal of data to transmit to station $j$, the TDM-based control channel is very inefficient. As a result, a new architecture called partition-group (PG) is proposed to alleviate this drawback and improve performance.

The PG is a network that consists of $r$ groups of stations. All stations in a group have the same wavelength for control channel. Figure 8 shows the channel assignment of stations in a PG, $r = 2$, in which stations 1, 3, 5, 7 and stations 2, 4, 6, 8 are grouped together using different control wavelengths $\lambda_0$ and $\lambda'_0$, respectively. Obviously, Fig. 7 is a special case of PG with $r = 1$. The control channels of different groups can adopt the subcarrier technique [8, 19] if the same control wavelength $\lambda_0$ is used. For instance, the wavelength $\lambda_0$ with subcarrier $f_1$ and $f_2$ is used for groups of stations (1, 3, 5, 7) and (2, 4, 6, 8), respectively. Since there are only $N/2 = 4$ stations in a group, the average waiting time of a station for transmitting data in the PG is less than that in the DBLN as shown in Fig. 6. If we assume that the traffic in the same group exhibits locality, the slot utilization is almost twice than that of the original DBLN. However, this is an upper bound of slot utilization for PG with $r = 2$ if full traffic locality exists in each group.

### 3.1. Interleaved Control Slot Mechanism

In this section, the control slot arrangement is proposed in order to get good slot utilization in networks. There are many solutions to resolve the collision problem. For example, the subcarrier technique can be used to avoid the collisions. That is, the receivers of all stations can do the channel inspection [14]. Apart from channel inspection, we propose a slot-interleaved mechanism as shown in Fig. 9 to prevent the collisions. Figure 9 shows the slot-interleaved control slot for $r = 2$, where stations 1, 3, 5, 7 are together as group 1 and stations 2, 4, 6, 8 are group 2. In Fig. 9 each control slot consists of two parts: cross-group section (CGS) and non-cross-group section (NCGS), where the CGS allows a station to transmit data to destinations in different groups and the NCGS can not transmit data across different groups. For example, for stations of group 1, if station 3 wants to transmit data to stations in the same group such as station 1, 5, *or* 7, it can transmit the data during the full control slot containing the CGS and the NCGS. However, if station

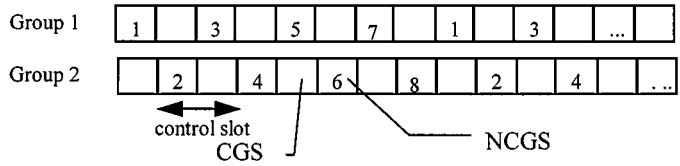


**Fig. 9.** An example for slot-interleaved control slot.

3 wants to transmit data to stations in group 2, such as station 2, 4, 6 *or* 8, it can only transmit the data during the CGS, that is a half of control slot. Owing to the traffic locality of the stations in the same group, the cross-group traffic is less than the traffic within the same group. As a result, the probability of collisions will be very small and collisions should occur at most one half of each control slot. In order to detect the collisions, a special unit known as the collision manager will be described here. The collision manager that is a tunable receiver in each station monitors the transmitter wavelength to decide whether to transmit data or not. If the collision manager detects the wavelength that has been used by other stations, then wait and try again in the next control cycle.

### 3.2. Assignment of Receiver Wavelength

In this section, the assignment of wavelengths is described. We consider a network of $N$ stations with $r$ control channels and $k$ data channels. For the sake of easy analysis, we only take the data channel and the number of stations into consideration. In general, there are three different cases as described below.

1. $N = k$: In this case, each node has the sole wavelength for receiving data. For instance, in Fig. 7 the receiving wavelength of node $i$ is assigned to $\lambda_i$. When a packet is ready for transmission from node $i$ to node $j$, the tunable transmitter of node $i$ is tuned to $\lambda_j$ for transmitting data until slot arrives.
2. $N > k$: Since the number of logical channels $k$ is less than the number of nodes in a network, wavelength sharing is used. First, for simplicity, we restrict to network topology in which the number of channels $k$ is a divisor of the number of nodes $N$. Therefore, the number of destination nodes in each channel is equal, i.e., $N/k = r$ and the $r$ destinations share the $i$th channel, $0 \leq i \leq k$. The receiving wavelength of node $i$ is $\lambda_{|j|_M}$ (The notation $|\cdot|_M$ indicates the modulo operator). When a packet is ready for transmission from node $i$ to node $j$, the tunable transmitter of node $i$ is tuned to $\lambda_{|j|_M}$ for transmitting data until the next slot. Figure 10 shows the channel assignment with $k = 4$, $N = 8$ and $r = 2$. Because of wavelength sharing, collisions will occur. Some mechanisms such as subcarrier [8, 19],

| Station | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Control ch. | $\lambda_0$ | $\lambda_0'$ | $\lambda_0$ | $\lambda_0$ | $\lambda_0'$ | $\lambda_0'$ | $\lambda_0$ | $\lambda_0'$ |
| Tx | $\lambda_1,\cdots,\lambda_4$ | $\lambda_1,\cdots,\lambda_4$ | $\lambda_1,\cdots,\lambda_4$ | $\lambda_1,\cdots,\lambda_4$ | $\lambda_1,\cdots,\lambda_4$ | $\lambda_1,\cdots,\lambda_4$ | $\lambda_1,\cdots,\lambda_4$ | $\lambda_1,\cdots,\lambda_4$ |
| Rx | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ |

**Fig. 10.** The channel assignment for nodes in a network with $N = 8, k = 4$.

TDMA technique [10, 11] and interleaved control slot described here can be adopted to solve the collision problem.

3. $N < k$: As already mentioned, each node has a fixed wavelength reciver for receiving data. Hence, this case in which the number of node is less than number of channels has a single wavelength for receiving data.

## 4. TRAFFIC AND LOAD BALANCE

In this section the traffic representation, the network manager, and the network loading balance are discussed. Certain types of traffic can be justified by traffic matrix. Based on the above information, the network manager issues the reconfiguration command to stations to change the network configuration based on dynamic loading balance.

### 4.1. Traffic Representation and Network Manager

In Fig. 4 the network inspector monitors all the transmission conditions from output queues. The output queue depicts the destination of output packets form the edge router and the source from the remote REX. Based on label switching shown in Fig. 4 the same destination packet will be put together in the same output queue. Hence the network inspector can easily understand the traffic bandwidth requirements in every REX switch. We represent the bandwidth requirements of source-destination pairs by a traffic demand matrix $T = [t_{ij}]$. The value of $t_{ij}$ is a measure of the traffic originating at node $i$ and terminating at node $j$. However, the traffic originating and terminating at the same node is assumed not existing in this paper and therefore, the diagonal elements of $T$ are all zero, i.e. $t_{ij} = 0$ *for* $i = j$. For example, a traffic demand matrix $T$ is shown in Fig. 11, in which the values of $t_{23}$ means 4 demands of the traffic originating at node 2 and terminating at node 3.

The upper triangular matrix of $T$ denoted as $U$ is the traffic demand matrix of Bus $A$ and the lower triangular matrix of $T$ denoted as $L$ is that of Bus $B$. Hence, the total bandwidth requirement $B_j$ of receiver $j$ in Bus $A$ is the sum of

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | | 13 | 1 | 2 | 5 |
| 2 | 2 | | 4 | 22 | 3 |
| 3 | 13 | 1 | | 0 | 2 |
| 4 | 0 | 22 | 7 | | 1 |
| 5 | 11 | 11 | 6 | 11 | |

**Fig. 11.** A traffic demand matrix T.

the elements of the $j$th column of $U$ and the total transmission requirement $T_i$ of node $i$ in Bus $A$ is the sum of the elements of the $i$th row of $U$. Similarly, the total bandwidth requirement $B_j$ of receiver $j$ in Bus $B$ is the sum of the elements of the $j$th column of $L$ and the total transmission requirement $T_i$ of node $i$ in Bus $B$ is the sum of the elements of the $i$th row of $L$. As a result, the following equations are derived:

$$B_j = \sum_{i=1}^{j-1} t_{ij}\, j = 1, \ldots, N \text{ for Bus } A \tag{4.1}$$

$$B_j = \sum_{i=j+1}^{N} t_{ij}\, j = 1, \ldots, N \text{ for Bus } B \tag{4.2}$$

$$T_i = \sum_{j=1}^{j-1} t_{ij}\, i = 1, \ldots, N \text{ for Bus } A \tag{4.3}$$

$$T_i = \sum_{j=i+1}^{N} t_{ij}\, i = 1, \ldots, N \text{ for Bus } B \tag{4.4}$$

In general, the traffic matrix represents the bandwidth requirements of source–destination pairs. The traffic status in lightly loaded condition or in heavily loaded condition can be obtained from the traffic matrix by $B_j$ or $T_i$, respectively. Moreover, the matrix can easily justify the parameters in terms of traffic locality and network loading condition. In the following, two types of traffic are considered. (1) General traffic (random): If the type of traffic is random, the values of $t_{ij}$ should be randomly distributed. (2) Special traffic: The existence of special traffic such as multicasting traffic, video conferencing or voice traffic can be observed from the traffic matrix. Because these types of traffic use relatively high bandwidth on a continuous basis for a long period of time and usually exhibits traffic locality, the values in the traffic matrix are symmetrical or are larger than peripheral elements. For example, in Fig. 11 the values of $t_{24}$ and $t_{42}$ may be video conferencing or

voice traffic. On the other hand, the stations 1, 2, 4 and 5 in bus $B$ may exhibit multicasting traffic because of equal values ($t_{51} = t_{52} = t_{54}$).

From the above discussion, we know there are many questions to be answered in a network. For example, how to get the traffic load of network, when to switch (change to different control channels) and how to change it. These can be implemented by a network manager. The network manager evaluates the load condition of the network in order to decide when to activate the reconfiguration. In fact, a Dual Bus network can be physically arranged according to the looped bus physical topology [9, 18], so that the two buses are closed in a dual ring fashion, using a special node to realize the enclosure connection. This node can play the role of the network manager, being able to monitor all the transmission conditions on two buses. In this way, we can know the load distribution along the buses and decide when to reconfigure the network. The network manager will issue the switching command including switching and grouping information after a fixed period of time. When the network manager issues the switching command to all nodes via control channel (e.g., setting a bit to '1' when the switch is enabled and to '0' otherwise), following events take place. First, all the stations stop transmitting data. Second, the stations tune the control channel according to the grouping information on control channel, and third the Stations retransmit data. Because the control slot usually has unused bits to make them available for future advanced applications, this switching command information can be placed in the reserved bit of control field. On the other hand, the network manager can also have programmable group capability. For stations with special traffic such as multicasting traffic, video conferencing or voice traffic, the network manager can assign a dedicated group for special application to get better performance.

Slowing down the less important communication sessions, interrupting them completely in order to preserve the entire available bandwidth for the highest priority level is adopted in general. Hence, reconfiguring the network topology [18] based on dynamic load balance is a very important issue. It includes how to assign the nodes to different control channels in order to achieve the highest throughout and the bandwidth balance, i.e. traffic is spread across various channels as evenly as possible. In the following, dynamic load balance is discussed.

### 4.2. Dynamic Loading Balance

In this section, a network reconfiguration mechanism based on dynamic load balance is presented. An algorithm named as CGA (Control Group Algorithm) for implementing dynamic load balance, is proposed. The CGA performs load balance in each control slot and is shown below.

CGA algorithm: load balance in control slot
Input: the traffic matrix $T$

Output: $r$ group control sets, $r = 2$
*Begin*{

1. *Search for all remaining $T_{ij}$ in U/L*
2. *Select the biggest value of $T_{ij}$ in U/L and remove ith row and jth column in U/L, then insert station i and j to group 1*
3. *Select the biggest value of $T_{ij}$ in U/L and remove ith row and jth column in U/L, then insert station i and j to group 2*
4. *Search for all remaining $T_{ij}$ in U/L, repeat step 2 to 3 until all stations are selected.*

}*End*

## 5. PERFORMANCE ANALYSIS

In the following, we analyze the network performance by comparing the relationship between the slot utilization and the number of groups. Each bus has $k$ wavelengths/channels ($\lambda_1, \lambda_2, \ldots, \lambda_k$) by employing the WDM technology, where channels $\lambda_1, \lambda_2, \ldots, \lambda_r$ are dedicated for control and others $\lambda_{r+1}, \lambda_{r+2}, \ldots, \lambda_k$ are for data.

### 5.1. Upper Bound on Slot Utilization

The slot utilization in each group is determined by the collision probability. If the source and the destination stations of the traffic are located in the same group, i.e. the traffic is fully local, the collision will not occur and the slot utiliation is improved. As a result, the slot utilization is proportional to the traffic locality and the number of groups. In order to derive the upper bound of slot utilization, the collision is assumed to be nonexistent. In addition, we assume that the slot rate is $k$ slots/s on each control channel and the total number of stations is $N$. Hence, the $N$ stations need $Nk$ slots/s to provide service. On the other hand, if the network has $r$ groups (control channels) where the traffic of stations in the same group is distributed totally locally, the slot rate is still $k$ slots/s. Obviously the $N$ stations need only $Nk/r$ s to provide service. Figure 12 shows the upper bound of slot utilization in different groups, in which the number of stations in each group is assumed equal.

Figure 12 shows the ideal case in which there is no collision. However, this is not a real case in practical situation. Hence, the results in Fig. 12 are the upper bound of slot utilization. For example, when the number of groups is 2, the number of nodes in each group is $n/2$ and the average waiting time of a node for transmitting data needs only $n/2$ time slots. As a result, we can find that the upper bound of slot utilization, compared with one control channel, is $r$ times if the network is partioned into $r$ groups.

| Group number | 1 | 2 | 3 |
|---|---|---|---|
| Node number in each group | N | N/2 | N/3 |
| Upper bound of slot utilization | 1 | 2 | 3 |

**Fig. 12.** The upper bound of slot utilization vs. the number of nodes in different groups.

## 5.2. Simulation Models and Results

In this section, we assume the message arrival rate of each station $i$ follows the Poisson distribution with a mean $\lambda$, and the message length follows the exponential distribution with a mean $L$. The station load ($SL$) for station $i$ can be defined as

$$SL_i = \lambda \times L \tag{5.1}$$

The network load (denoted as $NL$) can be defined as

$$NL = \sum_{1}^{N} SL_i \tag{5.2}$$

We assume $P_{ij}$ is the probability that station $i$ is transmitting to station $j$. Since traffic locality usually exists in real networks, the source–destination traffic distribution is derived from the following equality [8]:

$$P_{ij} = \begin{cases} 0 & j \le i \\ \dfrac{(1-p)^{j-i-1}p}{1-(1-p)^{N-i+1}} & j > i \end{cases} \tag{5.3}$$

Equation (5.3) represents a normalized geometric distribution where $p(0 \le p \le 1)$ determines the level of traffic locality. In addition, we assume that the reconfiguration time of network is regarded as a uniform distribution (i.e., reconfiguration of a topology is completed within a specified time period). The measures of interest are the relationship among the slot utilization, the traffic locality and the network loading. The simulation data are collected from the 100000th to the 200000th slot times. Other assumptions are listed as follows:

1) $N = 20$ stations in the network
2) $R = \{1, 2, 3\}$
3) NL $= \{5$ (light loading), 30 (heavy loading)$\}$
4) $p = \{0.0$ (uniform distributed), 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 (full locality)$\}$

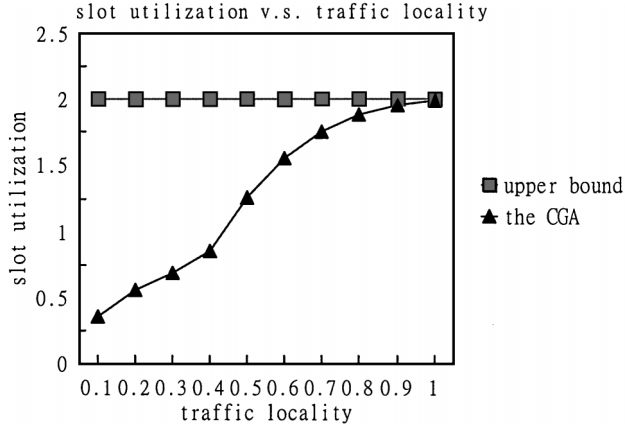


**Fig. 13.** The slot utilization in PG with $r = 2$.

Figure 13 shows the slot utilization in the PG network with $r = 2$ and the traffic locality obtained by the CGA algorithm. The increased traffic loclality could increase the slot utilization. Figure 13 illustrates that the slot utilization is large than 1 if the traffic locality $>0.5$. If the traffic locality is equal to 1, i.e. full locality, the slot utilization almost reaches its upper bound. However, if the traffic locality is less than 0.5, the slot utilization is below 1 due to the increasing collision probability.

In order to observe the relationship between the slot utilization and the number of groups under different traffic locality, we compare the cases of $r = 1, r = 2, r = 3$ of PG under different network loadings in Fig. 14. Figure 14(a) shows the slot utilization as a function of different groups under light network loading ($NL = 5$). The traditional TDMA is a special case of PG with $r = 1$. The simulation results show the slot utilization is very poor with $r = 1$ compared with $r = 2, 3$ for a lightly loaded traffic. We can conclude that the increased traffic locality could increase the slot utilization under light network loading as shown in Fig. 14(a). Fig. 14(b) shows the slot utilization as a function of different groups under heavy network loading (NL = 30). The simulation results show that the traditional TDMA case of PG with $r = 1$ has the best utilization. The increased number of groups has no benefit in slot utilization.

On the contrary, because of the greatly increased collision probability, the slot utilization is below 1. Hence, the simulation results in Fig. 14 show that the traditional TDMA technique does not allow a node to grab more than the portion of the channel bandwidth assigned to it even if no other nodes are transmitting on the same channels. However, under light traffic, the slot utilization of the TDMA is very poor. A group control slot can be adopted to improve the slot utilization. Therefore, we can conclude that TDMA is a better choice if the network is heavily loaded without traffic locality, otherwise PG is better in traffic with high locality.

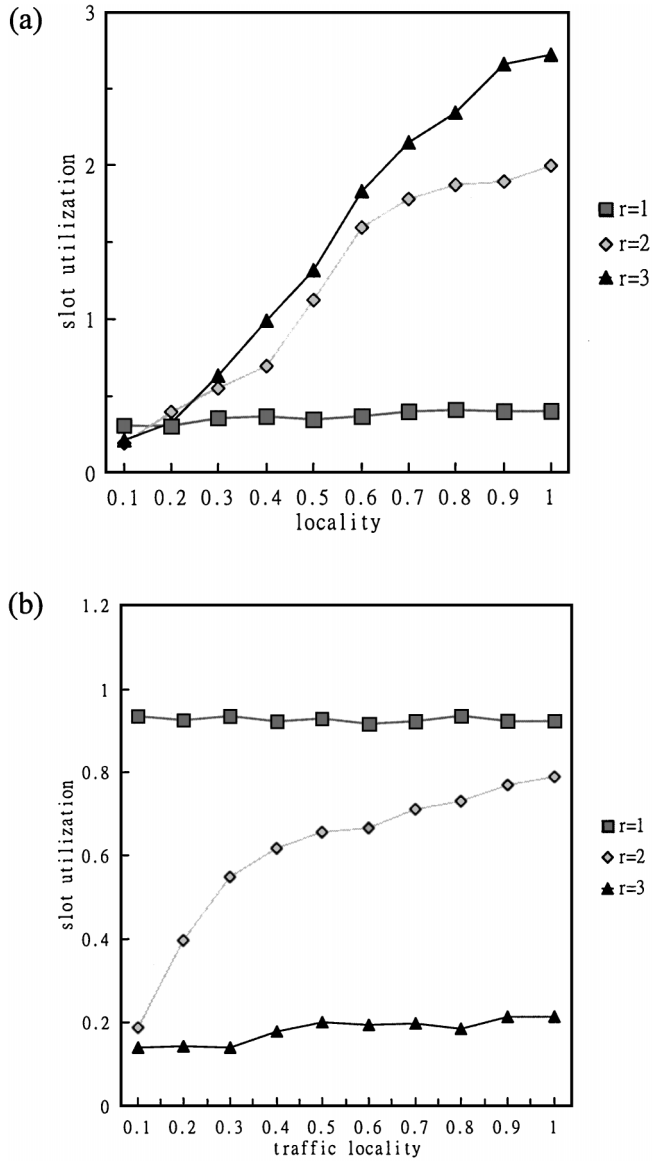


**Fig. 14.** Slot utilization vs. traffic locality under $r = 1$, 2, and 3.

### 5.3. Discussions

In Ref. 15 we can find that a network traffic loading is dependent on many parameters. Generally speaking, the growth curve of network traffic is from low to high and there exists different loading in different time. This means that there are some intervals in heavy loading, but sometimes not. We proposed the PG architecture to improve the network performance with high traffic locality. It seems to be not a realistic in a volatile network environment. But based on the results of Ref. 15 within NSFNET backbone, this case occurs repeatedly. The traffic distribution can be investigated by a long period of time to get the traffic characteristic. According the data obtained from network manager,we can adopt the better architecture to improve the networking performance. Of course, it can return to original architecture if the circumstance is changed. Hence, the results obtained from this research should present new and exciting opportunities for further theoretical as well as experimental work.

## 6. CONCLUSIONS

In this paper, we have proposed an architecture named as PG that is based on a two-level TDMA structure with wavelength division multiplexing (WDM) addressing the number of regional exchanges and optical switching communicating among the hubs. This solution integrated three technologies: IP addressing, label switching, and WDM network based on TDMA technology. Each station in the PG has one tunable transmitter tuned to a proper wavelength for transmitting data, one fixed receiver for receiving data, one tunable receiver for receiving proper control signals, and one tunable receiver for detecting the collision. Each control group has the same wavelength for the control channel based on the slot-interleaved structure. Each control slot consists of CGS and NCGS.

The virtual topology of the networks, i.e. control groups, is reconfigured by the traffic locality in the networks. Because of the traffic locality of the stations in the same group, the cross-group traffic is less than the traffic within the same group. As a result, the probability of collision will be very small and collision should occur at most one half of each control slot. Besides, in order to detect the collision, a special unit known as the collision manager was proposed. The collision manager monitors the transmitter wavelength to decide whether to transmit data or not. If the collision manager detects a wavelength that has been used by other stations, then wait and try again in the next control cycle. We also proposed a CGA algorithm to support loading balance for improving the network performance.

The simulation results show that the increased traffic locality could increase the slot utilization in light traffic environment. Because the slot utilization of the traditional TDMA is very poor in light traffic, the results suggest an alternative to choose. On the other hand, the simulation results also indicate that the traditional

TDMA has the best slot utilization than PG under the heavy load. Therefore, the TDMA is a better option in heavily loaded networks without traffic locality. Otherwise the PG is a good option in networks with high traffic locality.

Because the Internet traffic is more rigorous in fashion, we will strive to set up a stochastic model to justify our proposed system that is visible in the future. Not only the traffic locality will be considered, but also others parameters in terms of packet length, synchronization and component characteristic, etc. will be included.

## REFERENCES


1. C. A. Brackett, Dense Wavelength division multiplexing networks: Principles and applications, *IEEE J. Select. Areas Commu.*, Vol. 8, No. 6, pp. 948–964, 1990.
2. T. Li, MPLS and the evolving internet architecture, *IEEE Communications Magazine*, Vol. 37, No. 12, pp. 38–42, 1999.
3. E. C. Rosen, A. Viswanathan, and R. Callon, Multiprotocol label switching architecture, IETF draft-ietf-mpls-arch-06.txt, Aug. 27, 1999.
4. R. Callon, P. Doolan, N. Feldman, K. Fredette, G. Swallow, and A. Viswanathan. A framework for MPLS, IETF draft-ieft-mpls-framework-05.txt, Sept. 22, 1999.
5. P. Cadro, A. Fravey, and C. Guillemot, Performance evaluation of the KEOPS wavelength routing optical packet switch, *ETT*, Vol. 11, No. 1, pp. 125–132, 2000.
6. C. Qiao and M. Yoo, Optical burst switching (OBS): A new paradigm for an optical Internet, *Journal of High Speed Networks*, Vol. 8, No.1, pp. 69–84, 1999.
7. IEEE standards Board, IEEE standards for Local and Metropolitan Area Networks: Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN), IEEE std. 802.6-1990, 1990.
8. M. A. Rodrigues, Erasure nodes: Performance improvements for the IEEE 802.6 MAN, In *Proc. IEEE INFOCOM'90*, San Francisco, CA, pp. 636–643, 1990.
9. Wonhong Cho, Cheul Shim, and Sang-Bae Lee, MCDQDB (Multi-channel DQDB) using WDM, In *Proc. IEEE INFOCOM'90*, San Francisco, CA, pp. 2205–2209, 1995.
10. M. A. Marsan, A. Bianco, E. Leonardi, and S. Toniolo, An almost optimal MAC protocol for all-optical WDM multi-rings with tunable transmitters and fixed receivers, In *International Communication Conference* (*ICC*), Montréal, Quebec, pp. 437–442, 1997.
11. J. C. Lu and L. Kleinrock, A WDMA protocol for multichannel DQDB networks, *International Journal of Satellite Communications*, Vol. 9, pp. 23–35, 1991.
12. T. S. Peter Yum, Frank Tong, and K. T. Tan, An architecture for IP over WDM using Time-Division switching, *IEEE Journal of Lightwave Technology*, Vol. 19, No. 5, pp. 589–595, 2001.
13. N. F. Huang and S. T. Sheu, An efficient wavelength reusing/migrating/sharing protocol for dual bus lightwave networks, *IEEE Journal of Lightwave Technology*, Vol. 15, No. 1, pp. 62–75, 1997.
14. M. Ajmoni Marsan, A. Fumagallim, E. Leonardi, and F. Neri, R-Daisy: An all-optical packet network, *EUROPTO European Symposium on Advanced Networks and Services*, Amsterdam, Holland, March 1995.
15. K. Claffy, H.-W. Braun, and G. Polyzos, Traffic characteristics of the T1 NSFNET backbone, In *Proc. IEEE INFOCOM'93*, San Francisco, CA, Jan. 1993.
16. C. R. Giles and M. Spector, The wavelength add/drop multiplexer for lightwave communication networks, *Bell Labs Tech. Journal*, pp. 207–228, 1999.
17. S. A. Thomas, *The TCP/IP Protocols: Implementing the Next Generation Internet*, Wiley, New York, 1996.

18. Ilia. Baldine and George N. Rouskas, Reconfiguration in rapidly tunable transmitter, slowly tunable teceiver single-hop WDM network, *Technical Report TR-96-10*, North Carolina State University, Raleigh, NC, 1996.
19. Chen-Ken Ko and Sy-Yen Kuo, Multiaccess processor interconnection using subcarrier and wavelength division multiplexing, *IEEE Journal of Lightwave Technology*, Vol. 15, No. 2, pp. 228–241, 1997.

**Chi-Yuan Chang** received the BS in Electronic Engineering form National Taiwan University of Science and Technology, Taipei, Taiwan, in 1991 and the MS in electronic engineer form National Central University, Tao-yuan, Taiwan in 1993. He is currently a doctoral candidate at the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan. His current research interests include fault-tolerant WDM networks and optical Internet.

**Sy-Yen Kuo** received the BS in Electrical Engineering from National Taiwan University in 1979, MS in Electrical & Computer Engineering from the University of California at Santa Barbara in 1982 and the PhD in Computer Science from the University of Illinois at Urbana-Champaign in 1987. He is currently a professor and the Chairman of the Department of Electrical Engineering, National Taiwan University. He was the Chairman of the Department of Computer Science and Information Engineering. National Dong Hwa University, Taiwan, and a faculty member in the Department of Electrical and Computer Engineering at the University of Arizona. His current research interests include mobile computing and networks, dependable distributed systems, software reliability, and optical WDM networks. He is an IEEE Fellow.