An Architecture for IP Over WDM Using Time-Division Switching

T. S. Peter Yum, Frank Tong, and K. T. Tan

***Abstract*—This paper proposes an architecture for routing Internet protocol (IP) packets directly on optical networks. The use of label 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-tier multiplexing approach, with wavelength division multiplexing (WDM) addressing the number of regional exchanges and time-division switching communicating among the hubs. Such an architecture not only has the advantages of simple network management and high efficiency with low latency; it also is scalable by addition of regional exchanges, hubs, and fibers.**

***Index Terms*—Internet protocol (IP) network, optical routing, time-division switching, wavelength division multiplexing (WDM), wide-area network (WAN).**

I. Introduction

WAVELENGTH division multiplexing (WDM) is probably the most powerful technique available today to unleash the bandwidth in optical fiber so as to meet the explosive demand from the ever-growing number of Internet users and services.

So far, most of the reported packet-switched WDM networks [1], [2] are designed for local-area and metropolitan-area environments with a limited number of nodes and link distances. Proposals on packet switched backbone WDM network remain relatively few [3]. Here, we propose a new routing architecture for Internet protocol (IP) operating on a WDM backbone network. The proposed routing architecture is based on a two-tier multiplexing approach, with WDM addressing the number of regional exchanges (REXs) and time-division switching communicating among the hubs. Label switching [4] is introduced to improve the throughput performance of the conventional (electrical) IP routers. Such architecture not only has the advantages of having simple network management, high efficiency, and flexible and low latency; it also is scalable by addition of regional exchanges, hubs, and fibers. Furthermore, it allows regional exchanges on different hubs to use the same wavelength.

The conventional IP network has a lower tier of edge routers and an upper tier of big routers. It has the advantage of statistical multiplexing. But routers operate on the store-and-forward principle and so forbid the all-optical operation in the proposed architecture. What the proposed architecture offers is an all-optical link on a specific wavelength on a specific time-division multiplexing (TDM) slot for each REX–REX source–destination pair. Therefore, either a tunable transmitter or a tunable receiver is sufficient. (We choose tunable transmitter here.) Wavelength conversion is not needed, as all wavelengths are fully utilized without conversion in the proposed architecture. As will be explained later, a hub is essentially an optical crossconnect featuring a classical time-division space switch.

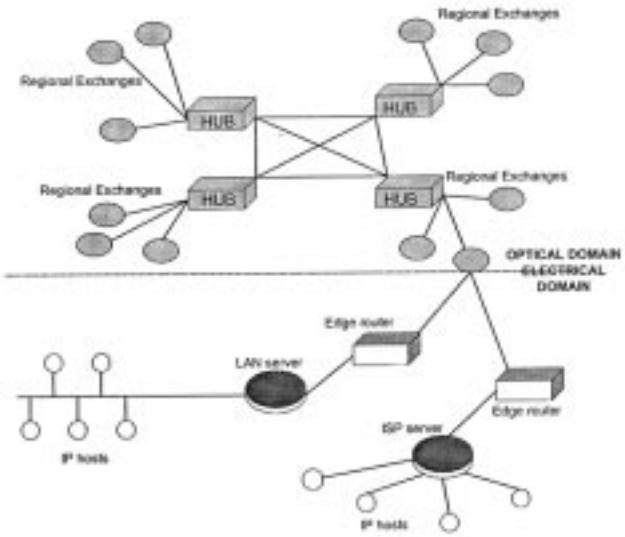

Fig. 1. Network architecture showing the fully connected hubs and the double star connections from routers to regional exchanges and to hubs.

Sections II and III present the network architecture and data payload organization, respectively. The intra- and interhub synchronization issues are presented in Section IV. The data packing efficiency is derived in Section V. The issues of network scalability and traffic management are discussed in Section VI.

II. Network Architecture

For the foreseeable future, the Internet will continue to be an interconnection of routers/gateways sitting on top of a transport network. The architecture consists of 1) an Internet access part (IAP), 2) a regional exchange, and 3) a hub. The IAP, as shown in Fig. 1, consists of IP hosts connected to corporate servers via 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. As part of IAP, these servers are in turn connected to high-speed edge routers for local communications and for Internet access.

Manuscript received March 15, 2000; revised January 23, 2001. This work was supported in part by RGC under Grants CUHK4157/98E, CUHK4159/98E, CUHK 4223/00E, and CUHK4371/99E.

T. S. P. Yum and F. Tong are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong, China.

K. T. Tan is with GoCDMA, 35/F Tower 2 Lippo Center, Central, Hong Kong.

Publisher Item Identifier S 0733-8724(01)04013-0.

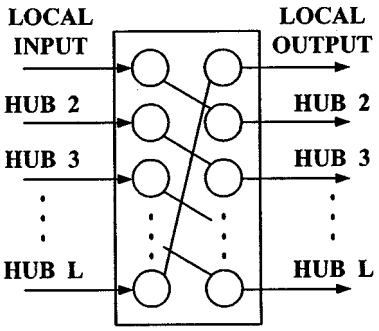

Fig. 2. A typical switching state of Hub 1 connecting $L-1$ distributed hubs.

The working premise of the proposed routing architecture is based on wavelength multiplexing and time-division switching. The REXs and the hubs form the optical backbone for the IP network. The former perform IP packet exchange in the electrical domain within a REX, while the latter perform the same in the optical domain between different regions on the same hub or across a fully connected hub network.

The connections between hubs can be a fully connected network, a ring, or other interconnection networks. We choose the fully connected network for its simplicity and its robustness. Specifically, L (the number of hubs) switching states can easily be generated, and there are $L-1$ alternate paths to use when the direct path is down. The tradeoff, however, is that the link utilization is limited to $1/(L-1)$ for this network. The remaining capacities, however, need not be wasted, as they can be used as high-usage bypass links (borrowing the term from the hierarchical telephone network architecture) between hub pairs. 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. The link utilization can be higher, but the alternate path protection is lower.

Let us begin by examining the proposed architecture at the hub level. The hub is for the exchange of packets between regions within the same hub or across distributed hubs. An architecture of a hub showing the instantaneous switching states of L distributed hubs is shown in Fig. 2. Here, a time-division optical switch is used. The L switching states are cycled through in L periods. The resulting transmission orders for the L hubs are shown in Level A of Fig. 5.

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.[1] To allow higher throughput, label switching [4] is used. As we are concerned with the transport of IP packets, the label can be included as part of the Layer 3 header (i.e., by using the Flow Label field in the IPv6 with appropriately modified semantics [5]). This label indexing is done at the edge router to ease the processing load of the REX switch. 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 periods using WDM. To illustrate, let us focus at Hub L in period 3, as shown in Level B of Fig. 5. Here, Hub L is connected to Hub 2. During this period, REX 1 of Hub L transmits wavelength bursts $\lambda_1, \lambda_2, \ldots, \lambda_K$ to REX 1, 2, ..., REX K in Hub 2 (assuming there are a full K REXs in both Hub L and Hub 2). In the same period, REX 2 of Hub L transmits wavelength bursts $\lambda_2, \lambda_3, \ldots, \lambda_K, \lambda_1$ to REX 2, 3, ..., REX 1 in Hub 2. Other REXs use other 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 sending out to another hub. This is illustrated in Fig. 3. Note also that the flow of intra-REX traffic is decoupled from the flow of inter-REX traffic.

For IP packet exchange within the same REX, no explicit synchronization is required. But for exchanges among REXs via the local hub, or across hubs, different wavelength sources transmitted in synchronism are required. Such a source can be a continuously tunable laser or a discretely tunable multiwavelength array laser [6]. Intra- and interhub synchronization issues will be discussed in Section IV. At the receiving end of the hub, a wavelength demultiplexer is used to separate the different wavelengths, each for a REX within the same hub (see Fig. 3). In other words, the routers in REX i always receive at λ_i. Note that the same set of wavelengths can be reused in another hub, as all the wavelength channels are switched together by the time-division optical switch in the hub.

Finally, let us look at the label switching architecture at the edge router and REX switch level. Due to the many functions that need to be performed, these parts would remain *electrical* except for the interface onto the optical backbone network at the transceiver end of the REX switch.

Fig. 4 depicts a functional diagram of 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. IP packets in the label indexing buffer are labeled according to their destination-router addresses and placed onto 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 address 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

[1]An edge router routes traffic between a number of IP hosts and servers and to outside; a switching router routes IP packets between source and destination edge routers.

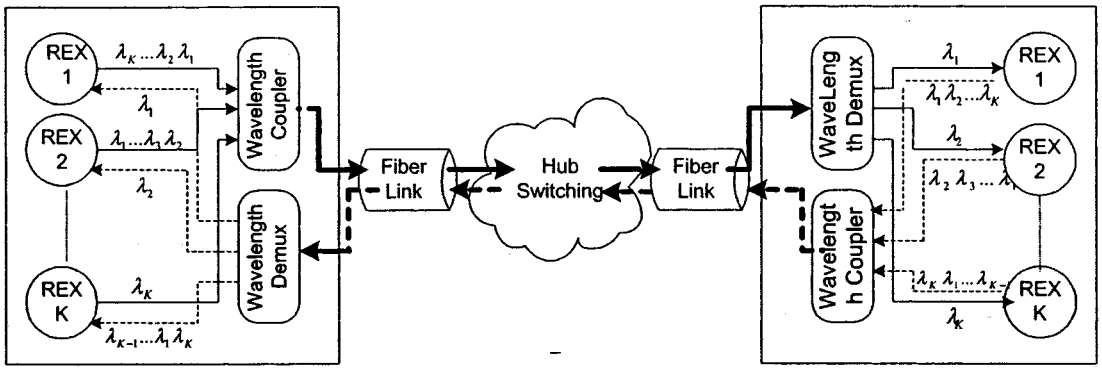

Fig. 3. Network with K connected REXs.

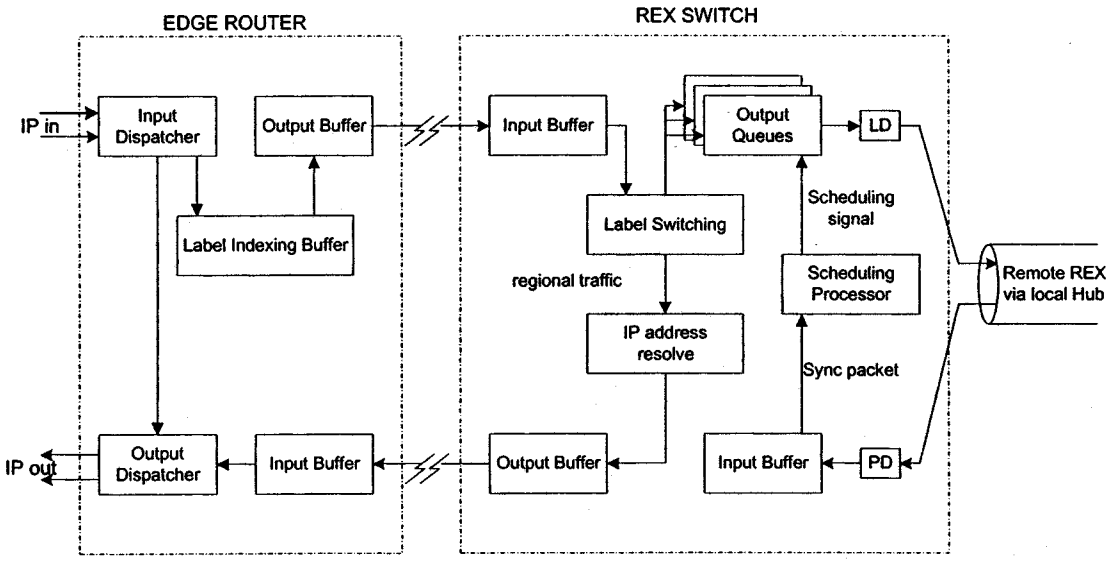

Fig. 4. Functional diagram of an edge router and its local REX switch. (LD: laser diode; PD: photo diode).

switch then resolves the IP destination address of these packets and passes them to their respective destination edge routers. The label switching used in this paper is for single hop only. Specifically, a label is used to identify a traffic flow between *local edge router A* $\rightarrow$ *local REX switch* $\rightarrow$ *local edge router B* or between *local edge router C* $\rightarrow$ *local REX switch*$\rightarrow$ *hub A*$\rightarrow$ *hub B*$\rightarrow$ *remote REX switch* $\rightarrow$ *remote edge router D*. This is reasonable because:

1) hubs and REXs do not perform store-and-forward function;
2) address pair (C, D) identifies a unique *REX* $\rightarrow$*Hub* $\rightarrow$ *Hub* $\rightarrow$ *REX* path for the traffic flow from edge router C to edge router D.

III. Payload Organization

IPv4 packets are of variable size with a nominal maximum of 64 Kbytes [7]. To improve the transmission efficiency, they can be concatenated into a train for fitting into a time slot of at least 64 Kbytes. At the beginning of every IP packet train, there is a guard time t_1 to accommodate the local REX clock jitter and the tuning time of the laser source from one wavelength to another. This is followed by the bit synchronization field of length $t_{\rm BS}$. These are illustrated in Level C of Fig. 5. Let the packet train payload size be U bits and the channel data rate be R bits/s. Then the labeled IP packet train size in seconds would be $U/R + t_1 + t_{\rm BS}$.

WDM is employed to multiplex the transmission from K REXs. K wavelength bursts are time multiplexed onto a transmission period for each hub switching state. A guard time t_2 is allowed for both the clock jitter and the circuit switching at the hub. In the hub, the time-division optical switch changes switching state at every interval of a transmission cycle, as illustrated in Level A of Fig. 5.

IV. Network Synchronization

The operation of the proposed network depends on the synchronization of the transmission periods in the hub level with the wavelength bursts at REX level. But these are only required to be loosely synchronized.

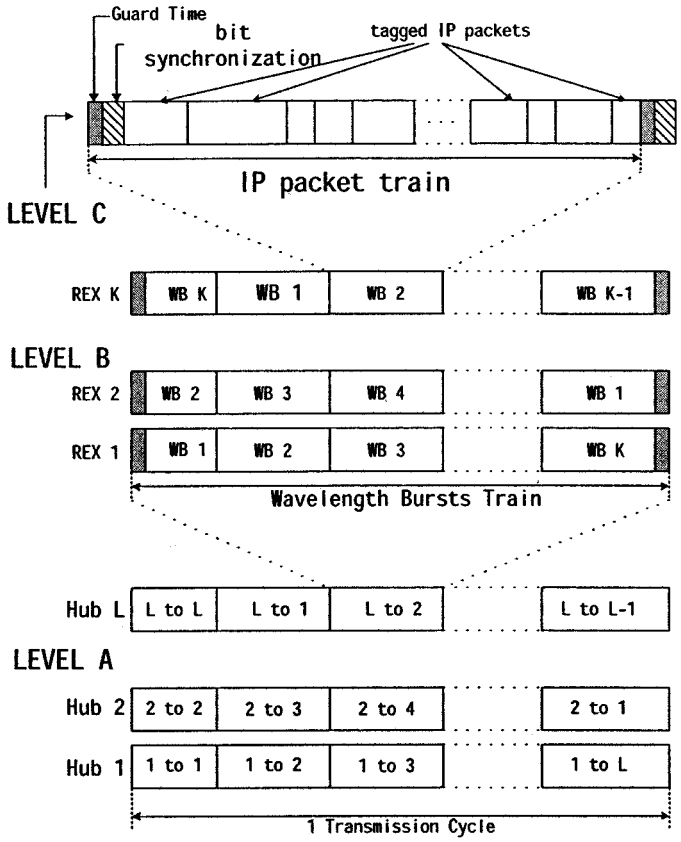

Fig. 5. Switching periods (Hub i to Hub j) at Level A, wavelength bursts (WBs) at Level B, and tagged IP packets at Level C.

A. Intrahub Synchronization

All REXs are synchronized with respect to its local hub. Let the end-to-end propagation delay τ_j from REX j to its local hub be measured by the hub and made known to REX j through a separate channel. In addition, let $\tau_{\max} = \max\{\tau_1, \tau_2, \ldots, \tau_K\}$, which is the propagation delay of the REX farthest away from the local hub, be computed by the hub and sent to all REXs within the hub. Note that aside from the very minor temperature variation, $\tau_{\max}$ is a constant and need not be adjusted if there is no change in the physical environment.

Each hub sends out sync pulses to all its REXs at the end of every Vth cycle, where V is determined by the accuracy of the REX clocks. Sync pulses can be transported either using a dedicated wavelength channel or inband using a small TDM slot. The proper choice depends on many factors and need not be elaborated here.

Consider the transmission of a sync pulse at T_0. Then REX j will receive the sync pulse at time $T_0 + \tau_j$. By $T_0 + \tau_{\max}$, all REXs would receive the sync pulse and can reset their local clocks simultaneously. From Fig. 6, it is clear that this is the same as REX j's resetting its local clock at $\tau_{\max} - \tau_j$ after receiving the sync pulse. In other words, the knowledge of T_0 by the REX is not needed for synchronization.

Next, in order for all transmissions to reach the local hub at the same time, the REX farthest away should transmit immediately after clock reset while REX j should delay transmission

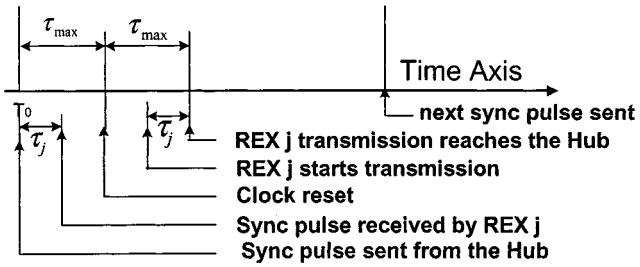

Fig. 6. Centralized intrahub synchronization at REX j.

by $\tau_{\max} - \tau_j$ so that its transmission can reach the hub at the same time as the hub farthest away. For all REXs to receive the sync pulse, a hub needs to insert a copy of the sync pulse on every wavelength of all the fiber links going to different regions within the hub. On the other hand, a copy of the sync pulse is also inserted into each fiber link going to other hubs for interhub synchronization. This will be elaborated in the next section.

B. Interhub Synchronization

The intrahub synchronization is closely coupled with the interhub synchronization to guarantee a network connection path between any two regions within the network in each transmission cycle.

At the interhub level, it is not practical to assign a centralized reference hub for network synchronization. This is because these hubs are envisioned to represent major nodes on the fiber-optic backbone across a wide area; and a centralized reference hub failure would result in the failure of the entire network. With this in mind, we proposed a simple *inband distributed global clock scheme*. Similar ideas are found elsewhere [8].

The objective of our scheme is to synchronize the distributed hubs so that the hub with the slowest clock determines the global clock time period of the entire network. The distributed hub synchronization algorithm is very simple.

1) The Algorithm: Consider a local hub, say, Hub A, sending periodic sync pulses to the other hubs (and its attached REXs), as shown in Fig. 7. After a sync pulse transmission, Hub A would wait for the arrival of sync pulses from all the other hubs. When the last sync pulse is received after an elapsed time Δ_A, Hub A simply resets its clock and waits for the nominal synchronization period T_A to expire before sending out the next sync pulse. The actual synchronization period, as shown in Fig. 7, is $T_A + \Delta_A$, as the sync pulse propagation delay also needs to be included.

Following this procedure, Hub A is actually synchronizing its sync pulse transmissions to that of the hub with the slowest clock. Fig. 7 depicts a timing diagram of this interhub synchronization scheme.

This inband distributed global clock synchronization scheme is fault tolerant to arbitrary hub failures. In the event that the failing hub is the hub with the slowest clock, the hub with the next slowest clock will automatically be used to set the sync pulse transmission period. Fig. 8 depicts one simple approach of implementing the synchronization circuit using a logic AND gate.

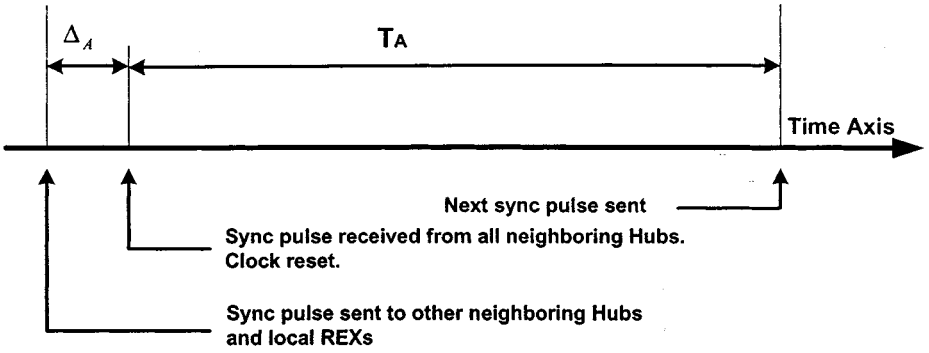

Fig. 7. Network-wide distributed synchronization at Hub A.

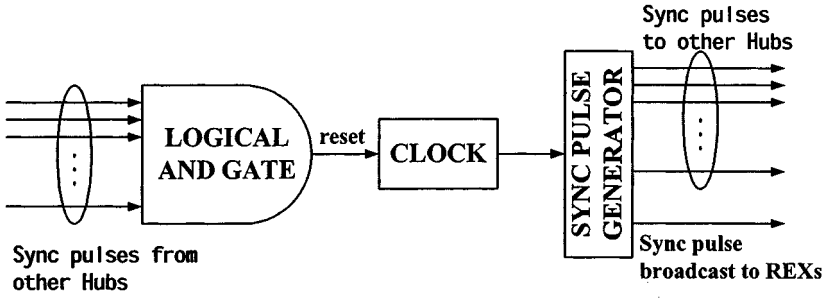

Fig. 8. Interhub synchronization circuit.

V. Data Packing Efficiency Analysis

As IP packets are of variable size, so are the IP packet trains. Let constant G be the slot size and random variable Y be the length of the IP packet train. Let $X_1, X_2, \ldots$ be independently and identically distributed random variables denoting the IP packet sizes with common density function $f_X(x)$ and distribution function $F_X(x)$. Then we have

$$Y = \begin{cases} X_1, & \text{if } X_1 \le G < X_1 + X_2 \\ X_1 + X_2, & \text{if } X_1 + X_2 \le G > X_1 + X_2 + X_3 \\ \sum_{i=1}^{q} X_1, & \text{if } \sum_{i=1}^{q} X_i \le G < \sum_{i=1}^{q+1} X_i \end{cases}.$$

For simplicity, we define

$$V(q) = \sum_{i=1}^{q} X_i$$

with density function $f_{V(q)}(x)$ obtained by convoluting $f_X(x)$ q times.

Taking the expectation of Y, we have

$$\begin{aligned} E[Y] &= \sum_{q=1}^{\infty} \int_0^G \int_{G-v}^{\infty} v f_{V(q),X_{p+1}}(v,x) dx\, dv \\ &= \sum_{q=1}^{\infty} \int_0^G v f_{V(q)}(v) \int_{G-v}^{\infty} f_{X_{p+1}}(x) dx\, dv \\ &= \sum_{q=1}^{\infty} \int_0^G v f_{V(q)}(v)[1 - F_X(G-v)] dv. \end{aligned}$$

The *packing efficiency* of wavelength bursts is simply

$$\rho_{\text{burst}} = \frac{E[Y]}{G}.$$

Hence the payload efficiency of a Level A period is

$$\rho_{\text{period}} = \rho_{\text{burst}} \times \left(\frac{1 - 2t_2}{GK}\right)$$

where t_2 is the guard time in level B.

VI. Network Scalability

The ability to scale up in size is one of the most crucial requirements of a backbone network. In our proposed architecture, we discuss this at the regional exchange and hub levels as follows.

First, at the REX level, we notice that traffic loading on different REXs is bound to be very different from time to time. The basic arrangement of assigning the same bandwidth to all REX pairs described herein is simple and fair in a sense. But the overall network throughput can be drastically improved if additional bandwidth (or wavelength) can be assigned dynamically. In the following, we describe a dynamic wavelength assignment algorithm for the optimal use of network resources. In case of insufficient network resources, our algorithm can also detect network bottlenecks identified by REXs in various hubs.

The traffic bottlenecks in such networks are most likely to appear at the fiber links connecting regions to the hub. Each of these links carries all the traffic to and from all other regions. But if we realize that regions are merely coverage areas that can be overlapped and a router can be attached to a second regional exchange, such bottlenecks can be easily removed by region splitting, just like cell splitting in cellular networks. Nevertheless, within a hub, how the total number of wavelength channels carried by the fiber link can be optimally shared among the REXs remains a problem.

Let there be K REXs in a hub and a total of W wavelengths available. As each REX requires one wavelength to begin with, there are $S = W - K$ spare wavelengths for sharing among the "big" REXs requiring multiple wavelengths.

The algorithm starts from the specification of the wavelength requirement matrix B of dimension $K \times K$ and the wavelength assignment matrix A of dimension $W \times K$. The requirement matrix $B = [b_{i,j}]$ is just like the traffic matrix, where $b_{i,j}$ is the number of required wavelengths from REX i in the local hub to REX j in the remote hub. The wavelength assignment matrix A has element $a_{i,j}$ indicating the source and destination REX pair using λ_i in time period j. As we assume a maximum of K REXs in both the local and the remote hubs, K periods are required to complete the transmission cycle. If there is no "big" REX around, the basic assignment scheme is used. Otherwise, A is augmented in rows by additional wavelength assignments until B is satisfied or all wavelengths are used up. When a "big" REX requires more wavelengths than what is available, it will be identified as a bottleneck by the algorithm with the information stored in vector C.

A pseudocode for the dynamic wavelength assignment (with bottleneck identification) algorithm is given in the Appendix. A simple example illustrating the finding of wavelength assignment and the identification of a bottleneck is also given.

The network bottleneck identified could be resolved either by reducing the traffic on the "offending" pair of REXs or by increasing the amount of wavelength capacity in the network by adding fiber-optic links. The merit of employing either technique in our proposed architecture is an issue for further research. Furthermore, the implementation of our wavelength assignment and bottleneck identification algorithm in the network and the required exchange of network control information is another issue for further investigation.

Finally, the scalability issue at the hub level for our architecture is addressed by the allocation of spare switching resources in the hubs from the onset with future network growth in mind. The hubs represent the biggest nodes in the network. There is little change in their total number, as evidenced in real-life backbone networks such as the SPRINT network, the European Optical Network, and MCI's vBNS [9]–[11].

VII. Concluding Remarks

This paper presented an architecture for routing IP packets over WDM networks. The architecture is based on wavelength multiplexing and time-division switching and has the advantages of scalability, being unbuffered at the core, and simple fault tolerant synchronization.

Appendix

Wavelength Assignment and Bottleneck Identification Algorithm:

Initialization:

$K = total\ number\ of\ REXs$
$W = total\ number\ of\ wavelengths$
B = *wavelength requirement matrix* $[b_{i,j}]$ *dimension* K *by* K
A = *wavelength assignment matrix* $[a_{ij}]$ *dimension* W *by* K
C = *vector of REX pairs causing network bottleneck*

Pseudocode:

1) *IF* $W < K$, *GOTO step 22;*
   ***Remark: W must be no smaller than K*
2) *SET* $i = 1, j = 1$;
3) $b(i,j) = b(i,j) - 1$;
4) *IF* $j + 1 = K$ *THEN* $a(i,j) = (K, i)$,
   ELSE $a(i,j) = ((i + j - 1)\ MOD\ K,\ i)$;
5) $j = j + 1$, *IF* $j <= K$ *THEN GOTO step 3,*
   ELSE $j = 1$;
6) $i = i + 1$, *IF* $i <= K$ *THEN GOTO step 3;*
   ***Remark: Basic assignment scheme executed*
7) *IF* $B = [0]$ *, GOTO step 22;*
   ***Remark: All wavelength requirements satisfied*
8) *SET* $col = 1, row = K + 1$;
9) *SET* $i = 1, j = 1$;
10) *IF* $b(i,j) = 0$ *THEN GOTO step 15;*
11) *IF* $col > K$ *THEN* $row = row + 1$;
    ***Remark: Bottleneck. Not enough wavelengths to satisfy requirements, identify the REX pair causing this*
12) *IF* $row > W$ *THEN GOTO 17;*
13) $col = 1$, $a(row, col) = (i, j)$;
14) $col = col + 1, b(i,j) = b(i,j) - 1$, *GOTO step 10;*
15) $j = j + 1$, *IF* $j <= K$ *THEN GOTO step 10;,*
    ELSE $j = 1$;
16) $i = i + 1$ *, IF* $i <= K$ *THEN GOTO step 10;*
17) $r = i, c = j$;
18) *IF* $b(r,c) = 0$ *THEN GOTO step 20;*
19) $C = [(r,c)C], b(r,c) = b(r,c) - 1$;
20) $c = c + 1$, *IF* $c <= K$ *THEN GOTO step 18,*
    ELSE $c = j$;
21) $r = r + 1$, *IF* $r <= K$ *THEN GOTO step 18;*
22) *END.*

To better illustrate the above algorithm, we have chosen a simple example. Consider $K = 4$ and

$$B = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 1 & 1 & 2 & 2 \\ 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}.$$

Let $S = 2$ and the two spare wavelengths be denoted as λ_5 and λ_6. After the basic assignment of

$$\begin{aligned} A(\lambda_1) &= [(1,1),(2,1),(3,1)(4,1)] \\ A(\lambda_2) &= [(2,2),(3,2),(4,2),(1,2)] \\ A(\lambda_3) &= [(3,3),(4,3),(1,3),(2,3)] \\ A(\lambda_4) &= [(4,4),(1,4),(2,4),(3,4)] \end{aligned}$$

the revised B matrix is

$$B = \begin{bmatrix} 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

As there are wavelength demands unsatisfied, λ_5 and λ_6 are used. The use of these wavelengths only needs to observe the rule that two or more REXs cannot use the same wavelength in the same period. Assigning λ_5 and λ_6 according to B row by row gives

$$A(\lambda_5) = [(1,2),(1,3),(1,4),(1,4)]$$
$$A(\lambda_6) = [(1,4),(2,3),(2,4),(3,1)]$$
$$C = [(4,4)].$$

Given that all wavelengths are used up, a network bottleneck is identified and the algorithm can stop.

REFERENCES

[1] A. Carena, M. D. Vaughn, R. Gaudino, M. Shell, and D. J. Blumenthal, "OPERA: An optical packet experimental routing architecture with label swapping capability," *J. Lightwave Technol.*, vol. 16, pp. 2135–2145, Dec. 1998.

[2] C. K. Chan, F. Tong, L. K. Chen, K. W. Cheung, and E. Kong, "Node architecture and protocol of a packet-switched dense WDMA Metropolitan Area Network," *J. Lightwave Technol.*, pp. 2206–2218, Nov. 1999.

[3] A. Watanabe, S. Okamoto, and K. Sato, "WDM optical path-based robust IP backbone network," in *Proc. Optical Fiber Communication Conf. 1999*, Feb. 1999, pp. 56–58.

[4] L. Andersson, P. Doolan, N. Feldman, A. Fredette, and B. Thomas. (1999) LDP Specification. IETF Networking Group. [Online]. Available: http://www.ietf.org

[5] S. A. Thomas, *IPng and the TCP/IP Protocols: Implementing the Next Generation Internet*. New York: Wiley, 1996.

[6] M. Zirngibl, "Multifrequency lasers and applications in WDM," *IEEE Commun. Mag.*, vol. 36, pp. 39–41, Dec. 1998.

[7] K. Thompson, G. J. Miller, and R. Wilder, "Wide-Area Internet Traffic Patterns and Characteristics," *IEEE Network*, pp. 10–18, Nov./Dec. 1997.

[8] C. S. Li and Y. Ofek, "Distributed Source-Destination Synchronization using Inband Clock Distribution," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 153–161, Jan. 1996.

[9] K. G. Laretto, "Sprint network survivability," in *Proc. Military Communications Conf.*, vol. 2, Oct. 1994, pp. 375–379.

[10] S. Baroni and P. Bayvel, "Key topological parameters for the wavelength-routed optical network design," in *Proc. 22nd Eur. Conf. Optical Communication*, Sept. 1996, pp. 277–280.

[11] J. Jamison, R. Nicklas, G. Miller, K. Thompson, R. Wilder, L. Cunningham, and C. Song, "vBNS: not your father's Internet," *IEEE Spectrum*, vol. 35, no. 7, pp. 38–46, July 1998.

T. S. Peter Yum was with Bell Telephone Laboratories for two-and-half-years. He taught at National Chiao Tung University, Taiwan, R.O.C., for two years before joining the Chinese University of Hong Kong, Shatin, in 1982. He has published original research on packet-switched networks with contributions in routing algorithms, buffer management, deadlock detection algorithms, message resequencing analysis, and multiaccess protocols. In recent years, he branched out to work on the design and analysis of cellular networks, lightwave networks, and video distribution networks. He believes that the next challenge is designing an intelligent network that can accommodate the needs of individual customers.

Frank Tong received the Ph.D. degree from Columbia University, NY, in 1987 with his thesis work completed at the Massachusetts Institute of Technology (MIT) Lincoln Laboratories, Cambridge.

After graduation, he joined IBM Almaden Research Center, where he worked on short wavelength lasers for optical recording. From 1989 to 1996, he worked on optical networking technologies under P. Green at IBM T. J. Watson Research Center, Yorktown Heights, NY. He is currently a Professor with the Information Engineering Department of the Chinese University of Hong Kong. He also served as a Senior Consultant to Lucent Technologies from 1999 to 2000.

Mr. Tong has received many awards from IBM including the IBM Outstanding Innovation Award in 1995.

K. T. Tan, photograph and biography not available at the time of publication.