Light-trails: an optical solution for IP transport [Invited]

Ashwin Gumaste and Imrich Chlamtac

Center for Advanced Telecommunications Systems and Services, The University of Texas at Dallas, 7601 N Floyd Road, Richardson, Texas 75083

aswing@ieee.org

RECEIVED 1 NOVEMBER 2003; REVISED 19 FEBRUARY 2004; ACCEPTED 20 FEBRUARY 2004; PUBLISHED 7 APRIL 2004

We present a solution for IP-centric communication at the optical layer through a combination of a hardware platform and algorithmic implementation. The presented approach, termed light-trails, is shown to yield a reconfigurable networking platform in which optical connections of arbitrary duration can be established and torn down flexibly in negligible time, accommodating the dynamic traffic requirements of the IP world. The hardware platform and protocol are evaluated with tractable mathematical models validated through detailed simulation. © 2004 Optical Society of America

OCIS codes: 060.4250, 060.4510.

1. Introduction

The exponential growth in data traffic has brought about the need for IP-based solutions for providing universal low-cost, effective communication. IP communication techniques have resulted in a paradigm shift in the way broadband and wide-area networks behave, leading to a concentrated effort in providing IP transport and making dynamic switching and real-time provisioning possible in all types of network [1–4]. Combined with WDM, an IP-based approach for broadband communication allows users to exploit the huge bandwidth offered by optical fiber—the key communication medium for backbone and regional networks—while providing the flexibility and granularity needed to yield a cost-effective optical networking architecture. The search for such solutions, i.e., for transporting IP over optical fiber, has resulted in a very large number of proposals over the years as summarized in Refs. [5–8], but more work is needed to find a solution that jointly addresses the optimization of networks for performance, price, and protocol (3P solution). One solution that attempts to do this is given by the *light-trails* approach [9, 10]. A light-trail is a generalization of a lightpath [5], in that a potential optical path is opened between any two nodes, such that multiple nodes along this path have access to the data in the path. Unlike lightpath communication, a light-trail allows any two nodes to communicate within a light-trail without the need for switching, allowing ultrafast communication and flexibility combined with multicasting. In this way the light-trails solution brings to the fore an architecture and a protocol that allows IP-centric communication. By being able to provide for dynamic reconfiguration of resources in an optical network, it avails users of the granularity of a subwavelength in the time domain, thereby creating a sub-lambda-capable architecture, essentially by time-differentiated intrawavelength multiplexing. In addition, light-trails' optical multicasting meets one of the key requirement for high-bandwidth killer applications such as video broadcasting and video on demand.

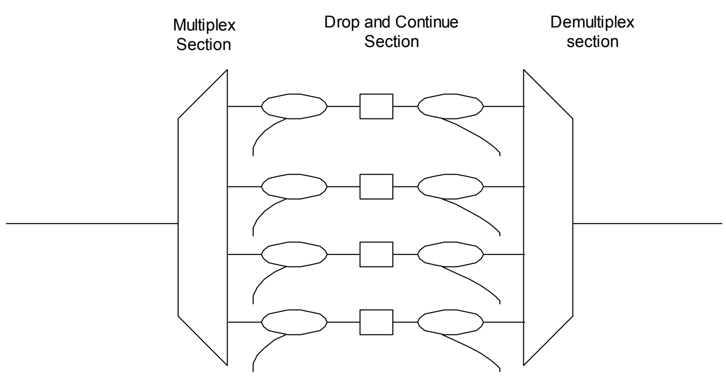
In this paper we consider the light-trail solution WDM rings, an important subset of topologies because of the massive deployment of ring networks worldwide. A general mesh deployment is treated in Ref. [10, 11].

2. Architecture and Hardware

To demonstrate the light-trails concept, we consider a 2-unidirectional fiber ring of N nodes. A light-trail of t nodes is a sequence of nodes utilizing a wavelength λ_i. In a light-trail $(N_1, \ldots, N_t)$ we call the first node, N_1, the *convener node* and the last node, N_t, the *end node*; the direction of communication is in the ascending order of nodal subscripts, i.e., from N_i to N_{i+k}. In this way a light-trail can be viewed as an optical bus between the convener and end nodes, with the characteristic that intermediate nodes can also access this bus, in contrast to a conventional lightpath or burst level path. Shown in Fig. 1 is a proposed node configuration for realizing light-trails. Shown here for a single fiber is a full demultiplex section that demultiplexes a composite WDM signal and feeds individual channels to a local access section. The local access section for each wavelength (channel) (Fig. 2) consists of two passive couplers separated by an optical shutter. The first coupler is called the drop coupler (DC, for dropping the signal), and the second coupler is called the add coupler (AC, for adding a local signal). The optical shutter is a contemporarily fast ON–OFF optical switch typically but not necessarily using Mach–Zehnder interferometer technology on lithium niobate substrates [12]. A light-trail is set up between two nodes by means of configuring the optical shutters on the desired wavelength at the convener and end nodes (in the OFF position) as well as by configuring the optical shutters (in the ON default position) at each of the intermediate nodes.

It is important to observe that in this configuration, once a light-trail of t nodes is set up, a connection between any source–destination pair within the light-trail can be established without requiring ON–OFF switching. The connection can be set up for a long duration (a lightpath) or very short duration (an optical burst). In other words, within each light-trail, lightpath or burst provisioning can occur without the need for fast ON–OFF switches. Specifically, if the optical shutter separating the two couplers is ON, then the system exhibits a drop and continue functionality. On the other hand, if the optical shutter is OFF, then wavelength reuse is obtained by virtue of spatial diversity (created by the OFF switch between add and drop couplers). The same structure therefore supports optical multicasting, too. Last, based on this realization, the light-trail network elements can be built virtually from available on-the-shelf technology.

The proposed configuration of the node shown in Fig. 3 supports a bidirectional light-trail–based ring, shown in Fig. 4.



A light-trail based solution,. Note that only the add dorp switches are replaced by a drop and continue section.

Fig. 1. Light-trail node configuration.

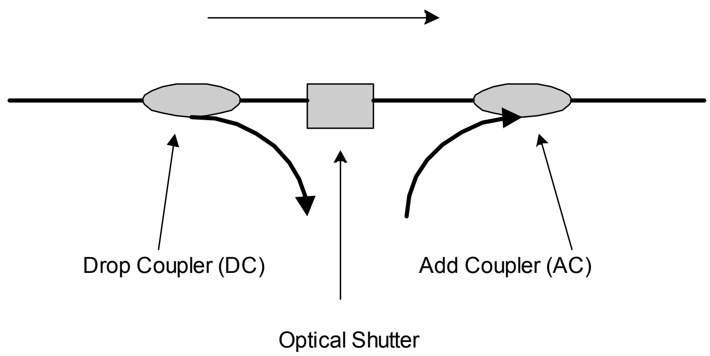


Fig. 2. Local access section of the light-trail architecture. The top arrow represents the direction of communication.

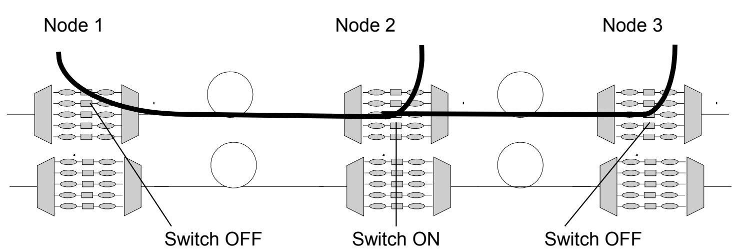


Fig. 3. Three nodes in a light-trail architecture.

3. Light-Trails Protocol

To establish light-trails and to provide optical connections within each light-trail, we define the following protocol. For obtaining the signaling required for setting up light-trails and for optical connections management within light-trails, we assume that there exists an out-of-band communication channel called the optical-service–supervisory channel (OSC), which is dropped and processed at each node. The OSC carries information about all the light-trails in the network and is responsible for provisioning connections within light-trails by providing a method for signaling. OSC is out of band and is separated in the frequency domain from the data wavelengths and hence is a self-sustained independent channel.

In light-trail communication, control packets are sent over the OSC ahead of the data by use of an offset that is a function of the sum of propagation delay from the ingress to the egress node and the control packet processing time at each node.

Five types of control packet are defined.

Setup packets (SPs): SPs are control packets used to set up light-trails. They contain the ingress and egress node information as well as the wavelength on which the light-trail is to be established and the direction of setup (clockwise or otherwise). Setup packets are deciphered at each intermediate node en route to the destination, hence adding and intimating each intermediate node to the light-trail.

Communication control packets (CCPs): CCPs are used to set up connections within a light-trail. CCPs carry information about the ingress and egress nodes, the light-trail number (a unique number that identifies the light-trail based on convener node, end node, and wavelength used) as well as the duration for which the connection is requested.

Dimensioning packets (DPs): DPs are used by the convener node to dimension light-trails, i.e., allowing new nodes to become members or eliminating nodes that are no longer active participants in the light-trail.

Global broadcast packets (GBPs): At regular intervals, on the OSC each light-trail sends GBPs throughout the network, informing all the nodes (and hence element management systems, EMSs) of the existence of the light-trails themselves and of member nodes. This avails each node of the information about existing light-trails, so a connection can be set up for data transmission within an existing light-trail, if one is available, instead of setting up a new one, thus minimizing switching and reducing setup time by orders of magnitude.

ACK packets: Acknowledgment (ACK) packets are used when a light-trail is set up. These are sent by the end node and acknowledged by intermediate nodes to the convener node, indicating the acceptance of the request to set up a light-trail.

Light-trail database: The network management system (NMS) for light-trails contains the necessary information regarding the current light-trails in the network. This database is updated by GBPs, and the database is assumed to be available to each node either locally or through a request scheme.

Local communication database: There is a single, time-sensitive pointer at each node, indicating whether the light-trail is occupied at that point through the node's local multiplex section (indicates whether data are flowing through the trail).

3.A. Setting Up Light-Trails

For creating a light-trail between nodes N_1 to N_t, node N_1 selects an available wavelength λ_t, based on the light-trails database information. Node N_1 sends a control packet (SP) through the OSC requesting opening of this optical path to N_t through the intermediate nodes $N_2, N_3, \ldots, N_{t-1}$. Node N_t, upon receiving the SP from N_1, replies through the OSC in the fiber (in the opposite direction) with an ACK. This ACK is received and recorded by intermediate nodes, too. If node N_t cannot allow the light-trail to be established, it so indicates by using a negative ACK. Nodes $N_2, N_3, \ldots, N_{t-1}$, upon receiving the control packet (SP), switch their optical shutters on the selected wavelength to the ON position while nodes N_1 and N_t keep the shutter in the OFF position. In this way a light-trail is created whose member nodes are $N_1, N_2, \ldots, N_t$, enabling downstream communication between them.

3.B. Communication within Light-Trails: Setting Up Optical Connections

If node N_j wishes to send data to downstream node N_k (both being in the same light-trail), it proceeds as follows: To avoid conflict in use of the same light-trail by an upstream node within the same trail, the initiating node N_j determines the availability of the light-trail by examining the local communication database to discover the occupancy of the light-trail at its own port (multiplex section). If no upstream node is using this light-trail for communication, then N_j can initiate communication to node N_k by sending a control packet (a CCP) to be followed by the data after an offset interval. Nodes $N_{j+1}, \ldots, N_{k-1}, \ldots, N_t$ now become aware of this communication. Note that after sending a connection set up control packet (a CCP), node N_j is guaranteed successful connection to a downstream node or multiple destination nodes (as in the case of multicasting), because all destination nodes can detect the data as part of the optical power is split through the multiplex section of each node locally.

If a connection in a light-trail is in progress, and an node upstream (with regard to the ingress node of the connection) wants to start another connection, then the previous connection (started by a downstream node) has to be cancelled to allow for the new connection initiated by the upstream node. Such connection overlapping is inherent to light-trails operation, and several solutions are possible for dealing with it. Node N_j, which is upstream of node N_q, can access the optical path, even when N_q is transmitting information. If such does happen, then N_j requests the opening of an optical connection in the light-trail. Node N_q, upon recognizing the possibility of conflict, inhibits its data and allows the data from node N_j to pass through. In this event the downstream node (N_q) either may send its data after the upstream node (N_j) has finished its data transmission or may send the data on another light-trail. However, for delay-sensitive applications, it may not always be possible to inhibit the local transmission. In that case the downstream node (N_q) may switch its optical shutter (for that light-trail) into the OFF position and collect the data from the upstream node (N_j) (through its drop coupler). It may then either buffer these data or send them on another light-trail, usually involving optical–electrical–optical (O-E-O) conversion from the initial light-trail to the new light-trail. In this process of retransmission (either over time or over different light-trails) the incumbent node does not restrain its data flow into the light-trail.

3.C. Dimensioning Light-Trails: Expanding and Contracting

As is seen above, light-trails can potentially support dynamic optical communication, such as bursty IP-centric traffic, as they do not require the resetting of optical switches along the light-trail for connections within the trail. If, however, a connection cannot use an existing light-trail, a new trail needs to be established, a procedure that involves reconfiguration of optical switches and therefore takes considerable time relative to the processing and setting up of a connection. For efficient light-trail communication, it is therefore important to dynamically optimize the light-trail set within the network to maximize the probability that a data transmission request will occur between a source and destination pair that share a light-trail. Seemingly it is desirable, from this point of view, to have the longest light-trail possible. However, since only one connection can be active per trail at a time, light-trails that are too long would lead to underutilization of wavelengths. By contrast, if light-trails are too short, optical communication deteriorates to that of a standard switching-based optical architecture. Therefore light-trails require dimensioning (expanding or contracting of trails) to optimize performance in terms of utilization, setup times, etc. The most efficient way of managing the light-trail length is naturally a self-regulating procedure in which a light-trail is extended or shortened so that it covers all active connections (bursts) that use the trail at any given time. Specifically, the leftmost active node in the trail defines the convener node of the trail, while the rightmost node of the existing connection defines the trail's end node. Since connections (bursts) come and go dynamically, to obtain this type of light-trail management it is necessary to have a distributed mechanism to move the trail's end nodes dynamically. Such a procedure can be realized through the following mechanism.

A node N_a upstream of node N_1, and not part of a trail, $\mathrm{LT}_1 = \{N_1, N_2, \ldots, N_t\}$, may request communication to a node in LT_1. The convener node, by virtue of its dominant status, may allow N_a to join the trail. If it does so, it shifts the convener status to node N_a and the new trail $\mathrm{LT}_{\mathrm{NEW}} = \{N_a, N_{a+1}, \ldots, N_1, \ldots, N_t\}$ is formed, and this information is broadcast through the network by a GBP. The act of expanding a light-trail is similar to setting up a light-trail, although we use dimensioning packets (DPs) as control mechanisms for this purpose.

Similarly, if the end node, through the local communication database, learns that it is no longer the recipient of information, it may use a DP sent to the convener requesting to be relieved (or that the group be relieved) from the light-trail. In that case the first node from the end node in the reverse direction that is still an active member of the light-trail now becomes the end-node of the light-trail by configuring its shutter accordingly.

Using this mechanism, light-trails thus can expand and contract over time, depending on the demands of traffic, allowing burst communication within optical connections (without switching) to occur, intuitively, for most of the traffic most of the time. On reconfiguration of a light-trail (expansion or contraction), the new light-trail information is broadcast throughout the entire network to facilitate nodes learning about preset optical paths.

4. Optical Characteristics of Light-Trails

The drop-and-continue functionality at each intermediate node in a light-trail allows the signal to be dropped to any desired node in the downstream, hence avoiding the need for high-speed optical switch reconfiguration, but it also involves splitting the signal into a suitable ratio at each node, thereby degrading system performance as compared with the optical budget achieved through a conventional WDM add–drop device as shown in Ref. [13]. For evaluating the feasibility of the drop-and-continue architecture for different networks, local, regional, and wide area, its optical properties thus need to be appraised, in terms of the depletion in the power budget and OSNR (optical signal-to-noise ratio) in the optical characteristics of the light-trail network. It is therefore necessary to verify that, despite this apparent cost in terms of optical budget, the budget remains within the conformity requirements for metropolitan and wide-area networks.

Power Budget: Especially over a wide-area network, high-speed signals need enough power (with and without amplification) to sustain a suitable bit error rate (BER). Hence we calculate the end-to-end power budget for an N-node light-trail. Assume that the power input (through a transponder) into a light-trail having nodes that possess the architecture shown in Fig. 1 is P_{in} dB. Further, let L_i km be the length of the ith span of fiber between the neighbor nodes of the ith span and let α be the attenuation constant for the fiber type used. Also let D dB be the pass-through loss across a multiplexer and be the same for a demultiplexer. Finally, let S be the pass-through loss through the optical shutter and let each coupler, assuming equal splitting ratios, have a 3-dB loss; then the power P_o dB at the receiver transponder at the destination node for an N-node connection across a light-trail is

$$P_o = P_{\text{in}} - 2\,(3 + D) - \alpha \sum_{i=1}^{N-1} L_i - (N-1)\,(2D + 6 + S)\,. \qquad (1)$$

For instance, for typical values in metro environments the per-span power loss for $P_{\text{in}} = 5\,\text{dB}$, $L = 20$ km, $D = 6$ dB, $S = 1$ dB, and $N = 6$—giving P_o without amplification as $-17\,\text{dB}$. This means that for these environments, using light-trail architecture, we need a system with optical amplifiers. In this case the optical characteristic for validation of light-trails now shifts from power budget to OSNR budget (considering noise related to amplified spontaneous emission). For simplicity, if we assume that we have amplifiers with sufficient gain (e.g., > 17 dB as in the numerical example above) and we also assume spans to be of the same length and hence same loss as before, the OSNR for an optical signal with input power P_{in}, amplifier noise figure NF, number of spans N, and the loss per span Γ is given as

$$\text{OSNR} = 58 + P_{\text{in}} - (\Gamma + 2D + 6 + S)\ \text{dB} - \text{NF dB} - (10\ \log_2 N)\ \text{dB}. \qquad (2)$$

For an input power of 5 dB, span loss of 4 dB, and NF of 4 dB for 6 spans, the OSNR at the last node in the light-trail is 29 dB. Note here that we are assuming standard photodetector values for the operating frequency band of a PIN-type photodetector (0.1 GHz). If we assume an avalanche photodetector (APD) or we use a Raman amplifier, we can get a much better OSNR. However, an OSNR of 29 dB is also excellent for 10 Gbit/s communication, and therefore we will have much more relaxed budgets. Hence light-trails can cater to optical metro networks and multiple line speeds.

5. Timing Consideration for Light-Trails and Light-Trail Connections

The ability of a single light-trail of n nodes to support up to $(n2)$ chronological connections within the light-trail allows light-trails to vary much more slowly than the variation of the traffic demands between source destination pairs. The speed with which such trails and connections can be established is thus central to the light-trail system.

5.A. Time Calculation in Setting up Light-Trails

If t_p is the propagation delay between two adjacent nodes, t_{pr} is the processing delay at any node for a control packet, and t_s is the time required for configuring an optical ON-OFF switch into the OFF position, then the time required for setting up an h-hop light-trail is

$$T_{\text{setup}} = 2h\left(t_p + t_{\text{pr}}\right) + t_s. \qquad (3)$$

Explanation: Fig. 5 illustrates the sequence of setting up light-trails. Assume node N_1 wants to set up a light-trail through nodes N_2, N_3, and N_4 to node N_5. Node N_1sends a control packet at time t_1 to node N_2, which processes the packet and sends it to node N_3 and so on. Upon receiving the control packet at time t_8, node N_5 begins to configure the optical shutter (ON–OFF switch) in the OFF position at time t_9. In parallel, an acknowledgment packet is sent back to the initiating source node via the desired route, informing the nodes of the acknowledgment marking the setting up of a light-trail. Once the acknowledgment packet reaches the source node N_1, it begins configuring its optical shutter in the OFF position, culminating the process of setting up light-trails.

The time required to set up a connection in a light-trail is given by

$$T_{\text{connection}} = h\left(t_p + t_{\text{pr}}\right). \qquad (4)$$

Note here that $h\left(t_p + t_{\text{pr}}\right) \ll t_s \ll t_{\text{sw}}$, where t_{sw} is the configuration time of an optical cross-connect switch as used in classical optical burst switching (OBS) solutions [14]. With

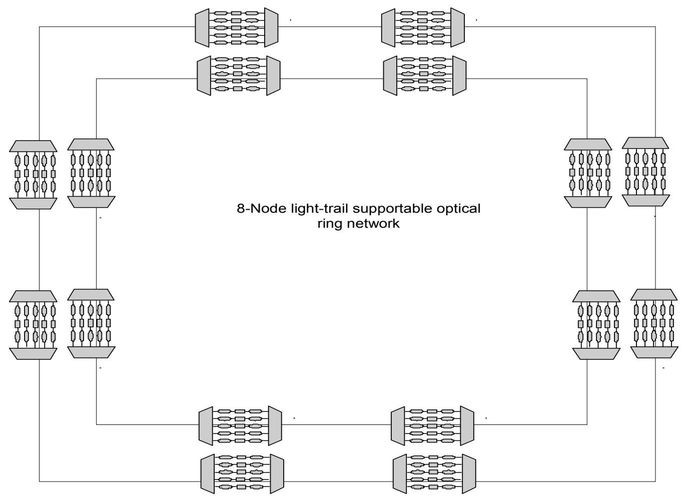


Fig. 4. Eight-node optical ring with ability to support light-trails.

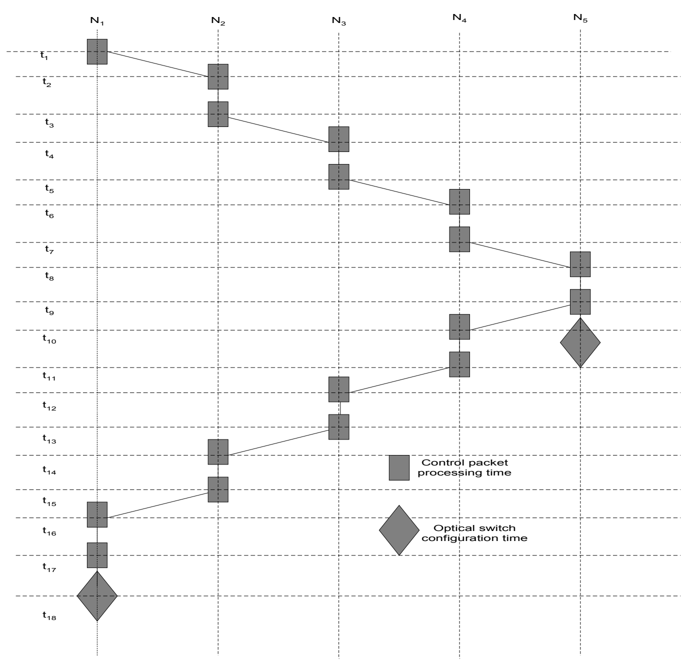


Fig. 5. Chronological sequence for setting up light-trails.

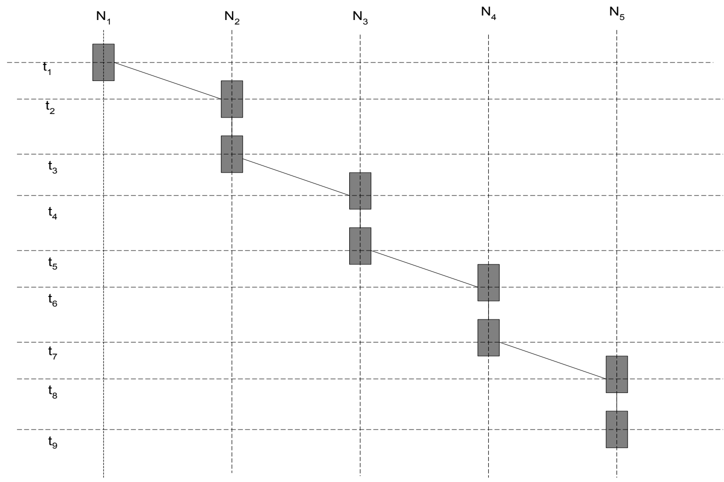


Fig. 6. Establishing connections within a light-trail.

the current maturity of optical technology, t_s will be in microseconds while $t_{\rm sw}$ can be in milliseconds. Further, to provide dynamic guarantees, $t_{\rm sw}$ can be as high as a few seconds as a result of management and other issues. This means that there is a major benefit obtained with respect to dynamic provisioning in light-trails as compared with other architectures. For optical cross connects the technology available today is either mechanical, bubble based, or a series of Mach–Zehnder interferometers (MZIs) leading to millisecond or longer configuration times. For a more detailed note on practical switches including the limitations of semiconductor optical amplifiers (SOAs) and related implementations, the reader is referred to Ref. [12].

Last, Fig. 6 shows the setting up of a connection within a light-trail. Assume that a light-trail exists from node N_1 to node N_5 as shown in the figure. To set up a connection between N_1 and, say, N_5, a connection setup control packet SP is sent from node N_1. As the control packet reaches each node it, informs that node of the subsequent transmission. After all the nodes gain knowledge about the transmission from N_1, node N_1 actually begins sending the data. In the case when another node, say N_i ($i \neq 1$), has a connection setup in progress, and in the course of this connection if a node N_k ($k < i$) wants to send its data, node N_k sends its SP and informs the downstream nodes of its desire to begin transmission; then the transmission from node N_i needs to be truncated. Node N_i, upon its receiving the SP control packet from N_k, truncates its transmission; to do so it has time $t_c = (N - i)\left(t_p + t_{\rm pr}\right)$.

6. Analysis of Light-Trails

In this section we consider the performance of light-trails as supporting optical burst transport, which can be considered to be a special case of very fast lightpath switching and a prelude to IP-centric communication over the optical layer.

6.A. Burst Blocking Probability

Consider an N-node light-trail $\mathrm{LT}_1 = \{N_1, N_2, \ldots, N_N\}$. We wish to evaluate the blocking probability of a node N_i within the light-trail LT_1. Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the arrival rate of connections at each of the N nodes. Let $S_1, S_2, \ldots, S_N$ be the average sizes of the bursts that arrive at nodes $N_1, N_2, \ldots, N_N$. For simplicity assume that $S_1 = S_2 = \ldots S_N = S$. Consider the diagram as shown in Fig. 7.

To calculate the blocking probability for node N_i when trying to transmit a burst into the light-trail of size S, we have to first find the probability of success for node N_i to transmit the same burst at any time. Two factors determine the success of transmission for node N_i: first, its position in the light-trail; second, depending on its position, whether any upstream node transmits during the interval $[0, S/C]$, where C is the transmission rate (line rate) of the channel. So for successful transmission

$$\begin{aligned} &P(\text{successful transmission for node } N_i \text{in duration } [0,S/C]) \\ &\quad = P(\text{no node upstream of } N_i \text{transmits in the same duration } [0,S/C]) \\ &\quad \times P(\text{no node upstream of } N_i \text{had any unfinished transmission at instant } t=0). \end{aligned} \tag{5}$$

In this discussion we neglect control packet processing delay, as it is assumed to be negligible compared with the burst transmission time S/C.

Assuming Poisson distribution with λ as the arrival process, S_j the number of bits queued up at the jth node, and C the line speed (in bits/second). The first term on the right-hand side of Eq. (5) is

$$= \prod_{j=1}^{i-1} \exp\left(-\lambda_j S/C\right), \tag{6}$$

when the burst should arrive for transmission at any of nodes $N_1, \ldots, N_{i-1}$ in the duration $[0, S/C]$.

The second term on the right-hand side of Eq. (5) is the same as Eq. (6):

$$= \prod_{j=1}^{i-1} \exp\left(-\lambda_j S/C\right). \tag{7}$$

This means that Eq. (7) gives the probability that nodes $N_1, \ldots, N_{i-1}$ have all completed their transmission (no further burst arrives in the interval $[-S/C, 0]$. In other words if any burst arrives in the interval $[-S/C, 0]$, then this burst would get precedence as compared to the burst at N_i.

Therefore we obtain the blocking probability

$$P_b^{N_i} = 1 - \prod_{j=1}^{i-1} \exp\left(2\lambda_j S/C\right). \tag{8}$$

6.B. Utilization in a Light-Trail System

The light-trail system can be considered as a best-effort system in which upstream nodes that have a higher probability of success than downstream nodes need to compete for a successful connection.

To evaluate the average utilization of a light-trail in this system, let λ_{ij} denote the arrival rate at node i, destined for node j. Let S_{ij} denote the size of the connection (in bits) at node i, destined for node j. Let h_{ij} denote the number of hops (links) from node i to node j. Then the total occupied capacity of the light-trail is

$$\begin{aligned} &= \frac{\lambda_{12} S_{12}}{C} h_{12} + \frac{\lambda_{13} S_{13}}{C} h_{13} + \cdots + \frac{\lambda_{1n} S_{1n}}{C} h_{1n} + \frac{\lambda_{23} S_{23}}{C} h_{23} + \cdots + \frac{\lambda_{2n} S_{2n}}{C} h_{2n} \\ &\quad + \cdots + \frac{\lambda_{(n-1)(n)} S_{(n-1)(n)}}{C} h_{(n-1)(n)}, \end{aligned} \tag{9}$$

simplifying to

$$U = \frac{1}{C} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\lambda_{ij} S_{ij} h_{ij}\right). \tag{10}$$

6.C. Burst Queuing Delays

If a node N_i generates a burst of size S at time t_o and it cannot transmit it into the light-trail because of ongoing connections in the light-trail, the burst needs to be queued. To evaluate the maximum burst queuing delay until a successful transmission, we designate λ to be the arrival rate of bursts at any node in the network and S to be the size of the bursts (and thus the transmission time, $t = S/C$, where C is the line rate). The probability that node N_i cannot send its burst in the light-trail because the trail is occupied can be obtained from the consideration that follows.

Considering the distributed nature and upstream priority accorded in light-trails, the burst from N_i has to wait till the time it receives a cumulative empty time period t s for transmission. Note that the transmission for N_i need not necessarily be in one continuous period of t s but can be in several disjoint time interludes, summing to t s. Note that here we neglect the connection setup time, which is negligible compared with the connection and burst duration time. Consider that node N_a, upstream of node N_i, had asked to begin transmission at the same time t_0 that node N_i asked to begin transmission. The next expected burst from node N_a would be after $1/\lambda$ s. In other words, the next burst at node N_a is expected at time $t_0 + 1/\lambda$. Generalizing, this means that node N_a can be expected not to disrupt transmission from node N_i for the next $1/\lambda$ seconds. In the worst-case situation we assume that node N_i is the penultimate node in the light-trail and hence has the worst blocking probability (hence the highest delay). That means node N_i can transmit its burst only when none of the $(i-1)$ nodes are transmitting. Hence in a fully loaded system (one where each node in the trail has a burst to send at time t_0) the penultimate node N_i would have to wait till all the nodes complete transmission. This leads us to the following inequality condition that relates the arrival rate, number of nodes (i) in the light-trail, and burst size as

$$\frac{1}{\lambda} \geq (i-2)t \quad \text{or} \quad \frac{1}{\lambda} \geq (i-2)S/C \tag{11}$$

related to the maximum queuing delay by node i as

$$\lceil t_{\text{queuing}} \rceil \leq \frac{1}{(i-2)\lambda}. \tag{12}$$

Note that Eq. (15) is valid only if Eq. (14) is valid.

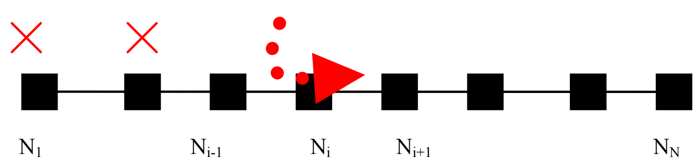


Fig. 7. Conceptualization for evaluating blocking probability.

In Fig. 7 we assume an eight-node light-trail, such that node N_1 is the convener node and node N_8 the end node. Note that node N_7 is the penultimate node whose performance is the most critical in our light-trail system. Assume that bursts of size S bits arrive at each node. If at instant t_1 a burst is aggregated at node N_3 for transmission and another burst is aggregated at N_2 upstream of N_3, then the burst from N_2 would be transmitted and that from N_3 would have to be queued. Subsequently, the burst at node N_3 could be sent through if none of the upstream nodes were transmitting bursts. In the worst case, assume that when some node upstream of N_3 finishes transmission of a burst, another node, also upstream of N_3, has a burst to send. Now, assuming the burst arrival process is governed by λ, we note that the node that first blocked transmission of N_3 (in this case N_2) is expected to have another burst only after $1/\lambda$ s from the initial reference point t_1. Further assuming that the burst at N_7, our penultimate node, would have to wait for the maximum time, equal to the cumulative transmissions of bursts from all the upstream nodes, the maximum delay is given by $(n-1)\,S/C$, where n is the number of nodes in the light-trail.

Bound on Maximum Delay is Tight: The worst-case situation is shown in the Fig. 7, where the penultimate node has to wait till all upstream nodes finish transmission. However, there may be cases in which one upstream node finishes transmission and there is no burst at any other upstream node for some time $< S/C$ (a fraction of the burst transmission time), following which a burst at an upstream node arrives. This type of variable arrival pattern does not invalidate the bound. The efficient utilization property of light-trails by itself preserves the bound: When there is a void in transmission from the upstream nodes, immediately the downstream node (N_7 in this case) begins transmission of its burst. Subsequently, when a burst arrives at any of the upstream nodes, N_7 curtails its transmission and allows the upstream node transmission and allocation of the light-trail bandwidth; N_7 resumes transmission upon completion of the burst from the upstream node. This means that the bound is tight, and as long as $(n-1)\,S/C < 1/\lambda$ the light-trail system performance is documented in inequality (11) and is valid. If inequality (11) is not valid, then the queuing delay is infinite and is explained below; also, the invalidity of inequality (11) would signify the failure of effective low-delay-based transmission for downstream nodes, and in general the invalidation of inequality (11) would mean that more light-trails need to be created (taking into consideration the downstream nodes) for optimum performance in terms of achieving a permissible delay for transmission.

The criteria under which the equation for maximum queuing is valid is

$$\sum_{N-1} t \leq \frac{1}{\lambda}, \tag{13}$$

where t is the transmission time for any of the bursts from N nodes. If the burst sizes are not same, then the relationship is as follows:

$$\frac{1}{C}\sum_{j=1}^{i-1} S_j \leq \frac{1}{\lambda}. \tag{14}$$

Hence for a tractable bound on the maximum delay in an N-node light-trail, we have to satisfy the inequality

$$\frac{1}{\lambda} \geq (N-1)\,\frac{S}{C}. \tag{15}$$

6.D. Average Queuing Delay in Light-Trails

In the previous subsection we calculated the worst-case bound for queuing delay for the worst affected node—the penultimate node in a light-trail. We now calculate the performance of a light-trail system in terms of the average delay for uniformly distributed traffic (when the probability of a request for a connection between any source destination pair has equal probability).

In a light-trail (in rings) of n nodes with convener i, the possible connections in the light-trail are

$$\begin{aligned} Y = [&(i,i+1)\,(i,i+2)\,(i,i+3)\,(i,i+4)\,(i,i+n)\,(i+1,i+2)\,(i+1,i+3)\\ &\cdots(i+1,i+4)\cdots(i+1,i+n)\,(i+2,i+3)\,(i+2,i+4)\cdots(i+2,i+n)\\ &\cdots(i+n-1,i+n)]_{(n-1,n-1)}; \end{aligned} \tag{16}$$

hence the number of connections (maximum) that can be set up by using a node are

$$Y_{\text{node}} = [n-1 n-2 n-3 n-4 \cdots 1]. \tag{17}$$

This means that the maximum number of distinct (source–destination) connections that can possibly be set up in a light-trail are $(n2)$.

For successful transmission of a burst of time duration t, the success probability matrix in a light-trail (each row corresponding to a node in descending order from the convener to the penultimate node) is given by

$$\begin{aligned} P_S &= \left[\prod_{i=1}^{r}\exp\left[-\lambda_i t_0\right]\prod_{i=1}^{r}\exp\left[-\lambda_i t_0\right]\prod_{i=1}^{r}\exp\left[-\lambda_i t_0\right]\cdots\prod_{i=1}^{r}\exp\left[-\lambda_i t_0\right]\right] \\ &= \left[1\exp\left[-\lambda t_0\left(1\right)\right]\exp\left[-\lambda t_0\left(2\right)\right]\exp\left[-\lambda t_0\left(3\right)\right]\cdots\exp\left[-\lambda t_0\left(n-2\right)\right]\right] \end{aligned} \tag{18}$$

where $r \rightarrow$ row denotes the number of the node (signifying the previous neighbor of the current node) and all arrival rates are assumed to be Poissonian and identical.

We then have the following delay matrix for the maximum delay that a node incurs when trying to transmit a burst of duration t_0 given that the inequality in Eq. (11) holds true:

$$\lceil\Delta\rceil = [0t 2t 3t \cdots (n-2)t]. \tag{19}$$

Given the upper bound on the transmission delay experienced by a node in Eq. (19), and also given the maximum number of distinct connections a node can successfully create, we can then postulate for that the average transmission delay in a light-trail is

$$t_{\text{ave}} = \frac{\Sigma_{k=1}^{n-1}\left(n-k\right)\left(k-1\right)}{\binom{n-1}{2}}t, \tag{20}$$

where t is the time for transmission of a single burst and inequality (11) is valid (necessary condition).

7. Network Performance

In the previous sections we studied internal light-trail behavior. We now turn to examine the effect of light-trail communication in the entire ring network. We assume that the traffic demands are bursty in nature and can be aggregated to form a burst. Let C denote the line speed or line rate in bits per second. Now assume that at an arbitrary node we aggregate a burst of S bits in Δ s, where Δ represents the maximum permissible delay that the first packet to arrive can wait while the burst is being aggregated. Assume that the flow of bursts is uniform; then we can say that the normalized flow from the node into the light-trail $f = S/C$. Here f indicates the flow from the node or hypothetically indicates the fraction of bandwidth that this node will need (on average) to fulfill its demand.

We can now build a traffic flow matrix called TF with dimensions of $N \times N$ for an N-node ring. TF_{ij} denotes the average flow from node i to node j, and $0 \leq \text{TF}_{ij} \leq 1$. Our objective is to find the minimum number of light-trails that can map all the demands denoted within matrix TF. Minimizing the number of wavelengths becomes a subproblem of the main problem of minimizing the number of light-trails. Before we formulate the optimization constraints, we define one more characteristic of nodes in a light-trail—active and passive nodes. From the protocol section and the analysis of intra-light-trail communication, it was shown that a node downstream has a smaller chance of successful transmission than a node upstream. If viewed from our normalized traffic flow model, we see that for an N-node light-trail, if the sum of flows of the first (from convener side) K nodes into the light-trail is 1, then the remaining $N-K$ nodes can, in principle, never succeed in sending in any data into the light-trail. That is, these nodes are pure destinations for this light-trail, and hence we term these nodes as *passive nodes*, as they can either be used for pass-through purposes or as destinations but cannot be used as traffic initiators (adders) as far as this light-trail is concerned. The other set of K nodes, which can act as sources as well as destinations (except the convener which is always a source node) are termed *active nodes*. These can send data into the light-trail, as well as serve as pass throughs or destination nodes, depending on the requirement.

7.A. Minimizing the Number of Light-Trails for a Ring Network

Our approach is to formulate a simple linear program whose solution gives the optimal number of light-trails required for mapping the set of demands. We start by defining a matrix P such that P has the dimensions of $N \times 2(N-1)$ and denotes the node-path matrix. That is, P_{ij} denotes the path j that is initiated from node i. The number of distinct paths possible from a node is $2(N-1)$ for a ring network of N nodes.

For a path p_i (note that there are $2N(N-1)$ paths in the ring), let $|p_i| = \sigma_i$ denote the number of nodes in path p_i. The number of possible combinations of active nodes in path p_i is given by

$$\psi_i = \sum_{j=1}^{\sigma_i} (\sigma_i j) . \tag{21}$$

We create a matrix d, such that d_{ij} denotes the jth combination of active nodes for the ith path. The size of d is $\sigma_{\max} \times \psi_{\max}$.

Let us create a flow matrix F such that F_{ij} is the flow from node i to nodes within path j. The size of matrix F is $N \times [2N(N-1)]$.

From F and d, we form matrix X by inspection, such that X_{ij} represents the flow in path i with the jth combination of active nodes.

Now we can formulate our minimization function as

$$\min L = \sum_{i=1}^{2N(N-1)_{\max}} \sum_{j=1}^{\psi_{\max}} y_{ij}, \tag{22}$$

where $y_{ij} = 1$ if a light-trail is assigned for theX_{ij} path and active node combination, and $y_{ij} = 0$ otherwise, subject to the constraint

$$\sum_{i=1}^{2N(N-1)_{\max}} \sum_{j=1}^{\psi_{\max}} X_{ij} y_{ij} = \sum_{a=1}^{N} \sum_{b=1}^{N} \mathrm{TF}_{ab}. \tag{23}$$

Note:

- $X_{ij} = 0$ if no path or no combination exists.
- $X_{ij} > 1$ signifies that more than one light-trail will be needed to satisfy F_{ij}.
- $X_{ij} < 1$ signifies that the cumulative flow of nodes in path i with $j$combination of active nodes is < 1 and hence a light-trail can exist.

By not keeping a bound on y_{ij} for all j, we keep all multiple light-trails on the same path and active node combination. This allows cases such as $F_{ij} = 1$, creation of a lightpath, or $2 > X_{ij} > 1$ creation of two similar light-trails.

8. Quantiative Evaluation

This section deals with the numerical evaluation of a light-trail network.

Simulation Model: For validating the analytical behavior and deriving quantitative results regarding the performance of a light-trail-based network we built a simulation model for optical burst switching. In the model we assume 12 nodes forming a metropolitan ring with 20 km of single-mode fiber (SMF) with 0.2 db/km of loss specification between two adjacent nodes. The channel rate is assumed to be 1 Gbit/s, and at each node we assume a 1-GHz 16-bit processor that processes control packet information. The network is laid on a 2-fiber ring with 40 data wavelengths and 1 control wavelength in each fiber, and each fiber is unidirectional in communication. Each fiber also has one channel (the forty-first channel) dedicated for control purposes. Optical bursts are assembled through an aggregation process described in Ref. [15]. In the process of assembling a burst, packets are collected from multiple service disciplines (voice, data, video, etc.). The burst size directly depends on the maximum allowable holding time of the initial set of packets that start forming the burst.

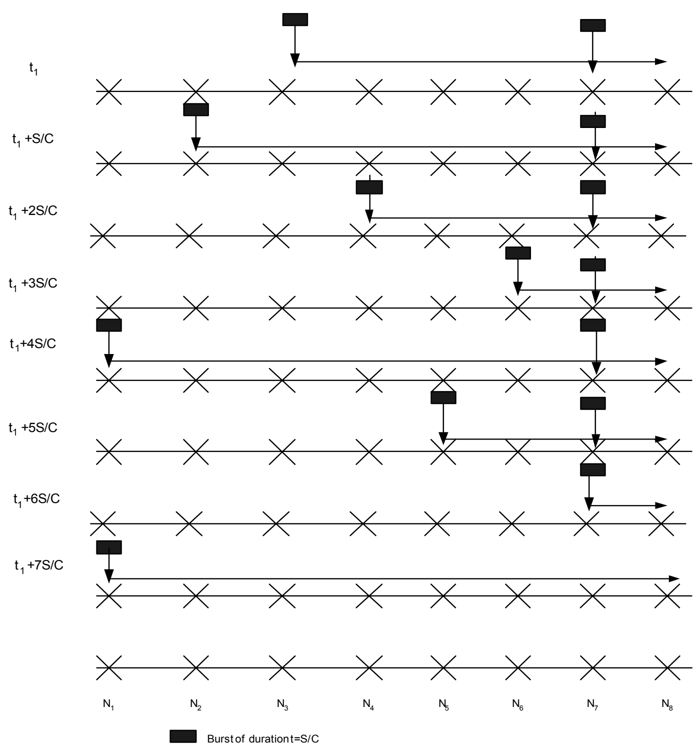


Fig. 8. Self-evidently shows the reasoning behind Eq. (12) and hence the upper bound on queueing delay.

8.A. Maximum Queuing Delay and Burst Size

Inequality (11) governs the three parameters affecting light-trail communication—namely, burst size, burst arrival rate, and number of nodes in the light-trail. These parameters in effect give rise to an interesting factor—the maximum waiting time a burst has to undergo at a node before being successfully transmitted. To evaluate the potential of light-trails for serving as a platform for packetlike (burstlike) dynamic communication, we conducted a simulation and compared the results with those achieved through the use of the analytical model of Section 5. Table 1 shows the analytical and simulated values of maximum delay at the penultimate node seen in a light-trail set up as a function of burst sizes for the same arrival rate λ. We consider the case of light-trails for different numbers of nodes and different burst sizes. Note that we take into account the fact that if maximum delay does not satisfy Eqs. (13), (14) then the system fails. We see that analytical and simulation results match with an error below 5%. Shown in Table 2 is the variation of queuing delay with respect to burst size.

We see that as burst size increases the delay increases accordingly. In fact the increase of delay in this section of the graph is almost linear, while for very small bursts (submillisecond duration) the delay seen was also exponentially decreasing.

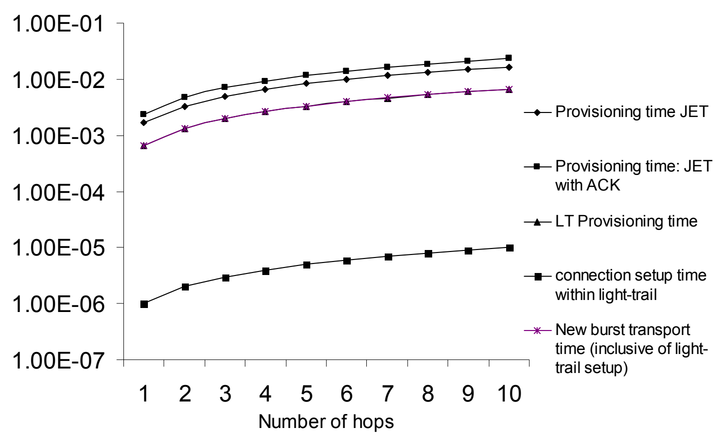


Fig. 9. Comparison of provisioning times for conventional burst-based communication algorithm JET and for light-trails.

8.B. *Provisioning Time in Light-Trails for Optical (Burst) Connections*

For efficient burst-type communication the ratio of burst setup time to duration is a key parameter for effective burst switching. Burst transport algorithms, because of the constraints on switching speed and the uncertainty of resource availability that is due to the distributed nature of a network, have relied on preallocation of resources to create an end-to-end optical path. JET [14] is a leading burst transport algorithm based on preallocation of resources in time, using an out-of-band signaling approach. In light-trails, because the optical connection does not need to reset switches, one can expect an advantage in provisioning time. Shown in Fig. 9 is the provisioning time for light-trails, including connections established within a light-trail and those that require establishing a new trail. For both light-trail and JET burst switching we consider the provisioning time to be a function of the hop length, switch configuration time, and control packet processing time. Optical bursts are generated by multiplexing different classes of traffic (namely, voice and data). In the simulation study we assume Poisson and Pareto distributions for burst aggregation. The scheduling policy for bursts that are delay sensitive is shown in Ref. [15]. In the simulation we use a line rate of 1 Gbit/s and bursts of 22 ms in length (worst case). In the simulation propagation delays are taken to be for 20-km links between consecutive nodes to emulate a typical metro area. Control packets are 20 kb in length, and we assume a 1-GHz. processor at each node to process the dropped control packets. For a collection of simulated rings of sizes varying from 10 to 16 nodes and 40 wavelengths, we have the speed of control channel to be 51 Mbit/s to avoid collision and to guarantee control packets to nodes as desired with a probability of 0.999. Figure 8 shows a significant benefit in the provisioning times for light-trail communication. Quantitatively we see that even if a contemporary fast switch [16] having configuration time of 0. 1 ms is assumed for JET we still see an approximate 3-order-of-magnitude decrease in provisioning time for single connection with light-trails. If we further assume that the light-trail is already set up and that burst transport only involves creating connections within a light-trail, we observe results such that there is on average an order of two advantage in provisioning as compared with provisioning by using JET, showing the importance of optimized light-trail length. This validates the light-trail architecture as a method for providing high bandwidth on demand to end users on a real-time basis.

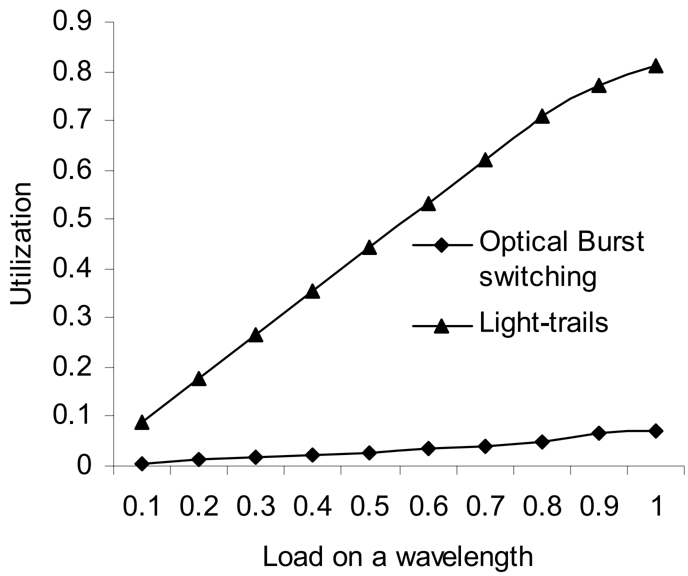


Fig. 10. Utilization of a wavelength using light-trails and compared with classical optical burst switching.

8.C. Network Utilization Benefits of Light-Trails

A major advantage of using light-trails for burst transport as compared with classic optical burst switching is the benefit observed in network utilization. In classic optical burst switching the data burst is sent upon the successful reservation of bandwidth in the path. That is, for every burst to be broadcast, a control packet has to be sent, and switches have to be configured. This procedure can be time consuming. By requiring an exclusive wavelength path to be set up for each burst, long voids are created within the channel, since there is no utilization of the channel while the control packet is in transit or while switches are being configured. In contrast, in light-trails we do not have to configure switches. This leads to excellent provisioning times as seen in the previous subsection. Moreover, this also leads to better utilization of the system. Shown in Fig. 9 is a comparison of utilization of a single wavelength for different loads for both light-trails as well as classical burst switching. We assume a 1-Gbit/s channel in both cases. Bursts are aggregated according to algorithm in

Ref. [15]. The average length of the path (for both light-trails and OBS) is 5 hops. Propagation delays are assumed to be for 20 km per hop, and processing at each node done by a 1-GHz processor (processing for control packet is 1.25 μm). Control packets were 20 kbit in maximum length. Load is computed stochastically in Erlangs. Utilization is defined as the ratio of capacity used over time for actual data transmission to the total capacity. Switching time for classical OBS switches when practical mechanical optical switches are used is 0.1 ms [16]. Faster implementations of optical cross connects are not feasible primarily because optical technology for fast switches (SOA or MZI) is not mature enough [12]. Owing to this long switching time requirement, OBS solutions result in poor network-wide utilization. On average, as seen in the simulation, the utilization of light-trails is an order of magnitude better than that seen in OBS under similar conditions.

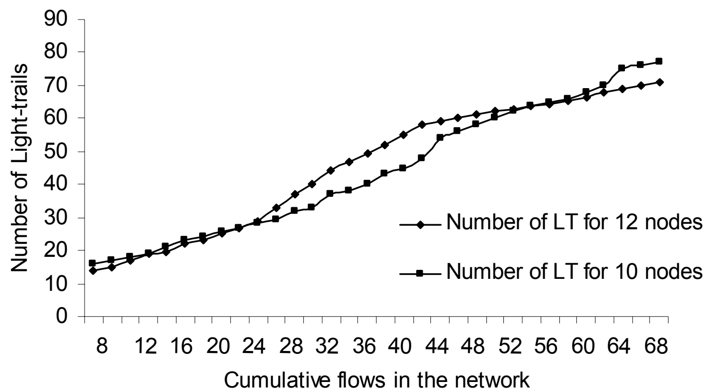


Fig. 11. Optimized number of light-trails for given traffic flows in a ring network.

8.D. Bound to the Number of Light-Trails in a Ring Network

From the optimization algorithm shown in Section 7, we simulate a ring network with various flows. These flows are between source destination pairs and can occupy fractional or whole wavelength granularity. Further, these flows are arbitrary in nature, with any source–destination pair being equally likely to result in a particular flow. These flows are normalized so that for every source–destination pair we get a number between 0 and 1, which denotes the net flow from the source to the destination normalized over the maximum delay requirement as well as the line rate. In Fig. 10 we show the number of light-trails required for creating a preset logical topology. This logical topology indicates the number of light-trails needed to facilitate the given cumulative flow (sum of TF) in the network in a way that the logical light-trails topology does not have to change, i.e., that the optimized logical topology can support all the flows in the traffic flow matrix TF without the need for further reconfiguration of the network. In Fig. 10 we show two cases, for 10 and 12 nodes, illustrating the behavior of the light-trail network. We see that the optimization algorithm binds the number of flows to the number of light-trails required, and toward the higher side of the graph (more flows) we note that the utilization of light-trails (number of light-trails/cumulative flow) is very good, of the order of 90% or greater. This also validates the lower bound given above.

Table 1. Performance of Light-Trails (delay)

Burst Size (ms)	Theoretical Delay for 6 Nodes	Simulation	Theoretical Delay for 8 Nodes	Simulation	Theoretical Delay for 10 Nodes	Simulation	Theoretical Delay for 12 Nodes	Simulation
6	12	11.181	16	15.933	20	19.692	24	23.928
8	16	16.008	21.33	21.085	26.67	26.892	32	32.115
10	20	20.424	26.67	26.714	33.3	33.254	40	40.178
12	24	19.982	32	31.992	40	39.925	48	48.024
14	28	27.924	37.33	37.284	46.67	46.821	56	55.938
16	32	31.99	42.67	41.555	53.3	53.299	64	63.878
18	36	35.423	48	48.181	60	59.201	72	71.944
20	40	40.887	53.33	53.877	66.67	66.841	80	80.13
22	44	44.092	58.667	58.908	73.3	73.845	88	87.894
24	48	48.489	64	64.69	80	80.127	96	96.381
26	52	52.038	69.33	70.442	86.7	86.619	104	104.892
28	56	55.803	74.67	75.103	93.3	93.456	112	112.451
30	60	59.231	80	79.927	100	100.1	120	120.131

Table 2. Performance of Light-Trails (different burst sizes)

Number of Nodes in Light-Trail	Theoretical Delay for $B = 10$ ms	Simulation	Theoretical Delay for $B = 20$ ms	Simulation	Theoretical Delay for 30 ms	Simulation	Theoretical Delay for 40 ms	Simulation
6	20	19.2	40	40.218	60	59.881	80	79.119
7	23.3	22.8	46.7	46.922	70	71.267	93.66	94.011
8	26.6	27.154	53.3	53.123	80	80.539	106.67	105.48
9	30	29.936	60	59.825	90	90.106	120	120.317
10	33.3	33.802	66.7	67.187	100	99.135	133.34	133.426
11	36.7	37.113	73.3	73.529	110	110.498	146.67	146.863
12	40	39.811	80	80.421	120	120.892	160	160.216
13	43.3	43.527	86.7	86.382	130	130.692	173.34	173.362
14	46.6	46.733	93.3	93.297	140	140.273	186.67	186.882

9. Conclusions

We introduced and evaluated the light-trail architecture. Light-trails were demonstrated as a potential solution for services requiring very fast provisioning. In addition, light-trails were shown to be capable of time differentiated sub-lambda provisioning—the key to future IP-centric optical networks for applications such as bandwidth on demand. The architectural benefits of light-trails include its characteristic drop-and-continue functionality, creating optical multicasting, the fundamental requirement for many of the so-called high-bandwidth killer applications driving the introduction of optical networks, such as video on demand. The proposed solution has the added advantage of being in principle realizable with off-the-shelf components.

References and Links

[1] A. Fumagalli, J. Cai, and I. Chlamtac, "The multi-token inter-arrival time (MTIT) access protocol for supporting IP over WDM ring network," in *IEEE International Conference on Communications* (IEEE Communications Society, New York, 1999), pp. 586–590.

[2] K. Dolzer, C. Gauger, J. Späth, and S. Bodamer, "Evaluation of reservation mechanisms for optical burst switching," AEÜ Int. J. Electron. Commun. **55**(1), 1–8 (2001).

[3] N. Ghani, S. Dixit, and T.-S. Wang, "On IP-over-WDM integration," IEEE Commun. Mag. (March 2000), pp. 72–84.

[4] L. Tancevski, S. Yegnanarayanan, G. Castanon, L. Tamil, F. Masetti, and T. McDermott, "Optical routing of asynchronous, variable length packets," IEEE J. Sel. Areas Commun. **22**(3), 2084–2093 (2000).

[5] I. Chlamtac, A. Ganz, and G. Karmi, "Lightpath communications: a novel approach to high bandwidth optical WANs," IEEE Trans. Commun. **40**(7) (1992).

[6] H. Zhang and A. Durresi, "Differentiated multi-layer survivability in IP/WDM networks," in *Network Operations and Management Symposium* (IEEE, New York, 2002), pp. 681–696.

[7] L. Sahasrabuddhe, S. Ramamurthy, and B. Mukherjee, "Fault management in IP-over-WDM networks: WDM protection versus IP restoration," IEEE J. Sel. Areas Commun. **14**(1), 21–33 (2002).

[8] K. Shrikhande, A. Srivatsa, I. M. White, M. S. Rogge, D. Wonglumsom, S. M. Gemelos, and L. G. Kazovksy, "CSMA/CA MAC protocols for IP-HORNET: an IP over WDM metropolitan area ring network," in *Global Telecommunications Conference, 2000* (IEEE, New York, 2000), pp. 1303–1307.

[9] A. Gumaste and I. Chlamtac, "Light-trails: a novel solution for IP-centric communication," in *Workshop on High Performance Switching and Routing* (IEEE, New York, 2003).

[10] I. Chlamtac and A. Gumaste, "Light-trails: a solution for IP-centric communication at the optical layer" (invited paper), in *Quality of Service in Multiservice IP Networks: Second International Workshop, QoS-IP 2003*, Vol. 2601 of Lecture Notes in Computer Science (Springer-Verlag, Heidelberg, 2003).

[11] A. Gumaste and I. Chlamtac, "Mesh implementation of light-trails: a solution to IP-centric communication," in *12th International Conference on Computer Communication and Networks* (IEEE, New York, 2003).

[12] R. Ramaswami and K. Sivarajan, *Optical Networks: A Practical Perspective*, 2nd ed. (Morgan Kauffman, Los Altos, Calif., 2001).

[13] G. H. Sasaki and C.-F. Su, "The interface between IP and WDM and its effect on the cost of survivability," IEEE Commun. Mag. (January 2003).

[14] S. Verma, H. Chasker, and R. Ravikant, "Optical burst switching: a viable solution for terabit IP backbone," IEEE Netw. Mag. (June 2000), pp. 48–53.

[15] A. Gumaste and J. Jue, "A scheduling procedure for control signaling in an optical burst switched network," in *Proceedings of the First International Conference on ICOCN 2002* (World Scientific, Singapore, 2002).

[16] A. Gumaste and T. Antony, *DWDM Networks Design and Engineering Solutions* (McMillan, New York, 2002).