Multiconfiguration Multihop Protocols: A New Class of Protocols for Packet-Switched WDM Optical Networks

Jason P. Jue, *Member, IEEE*, and Biswanath Mukherjee, *Member, IEEE*

***Abstract*— Wavelength-division multiplexing (WDM) local-area networks based on the optical passive-star coupler have traditionally been classified as being either single-hop or multihop. A single-hop network provides a direct connection between the source and the destination of a packet during the packet transfer duration, but may require some amount of coordination between the nodes which may involve tuning of the transmitters or receivers at each node. Since the time required to tune a tunable optical transmitter or receiver may be high, a single-hop network may incur significant overhead. On the other hand, a typical multihop network requires little or no tuning, but a packet may traverse a number of intermediate nodes between the source and destination nodes. Each hop incurs additional queuing delays at each node and also increases the overall load on each link and on the network. In this paper, we propose a new class of multiconfiguration multihop protocols (MMPs) which use tunable transmitters and receivers to cycle through a number of configurations which together make up a multihop logical topology. This class of protocols offers a trade-off between the tuning required in a single-hop network and the number of hops required in a multihop network. We present a generalized framework for comparing the proposed protocols with existing single-hop and multihop protocols, and we show that these protocols may offer significant performance gains for systems with high tuning delays and a limited number of transmitters and receivers at each node.**

***Keywords*— Optical Network, Wavelength-Division Multiplexing (WDM), Passive-Star Coupler, Packet Switching, Single-Hop, Multihop, Multiconfiguration.**

1 Introduction

A wavelength-division multiplexing (WDM) optical network based on the optical passive-star coupler (PSC) offers the potential for high bandwidth in a local-area network (LAN) environment. WDM divides the bandwidth on a fiber into a number of non-overlapping channels, each defined by an unique wavelength. In order to access the channels, a node is equipped with transmitters and receivers (transceivers), which may be either fixed or tunable. A fixed transmitter or receiver may only send or receive information on a single channel, while a tunable transmitter or receiver may tune to one of several channels. In a PSC-based WDM network, a transmission on any channel is broadcast to every node in the network. A node may receive the transmission by tuning its receiver to the appropriate channel (see Fig. 1 in which each of N nodes has a fixed-tuned transmitter operating on its own unique wavelength channel and a tunable receiver which can receive from any of the N channels numbered $\omega_1, \omega_2, \cdots, \omega_N$). Thus, communication between nodes requires some amount of coordination. This coordination is provided by the network protocol.

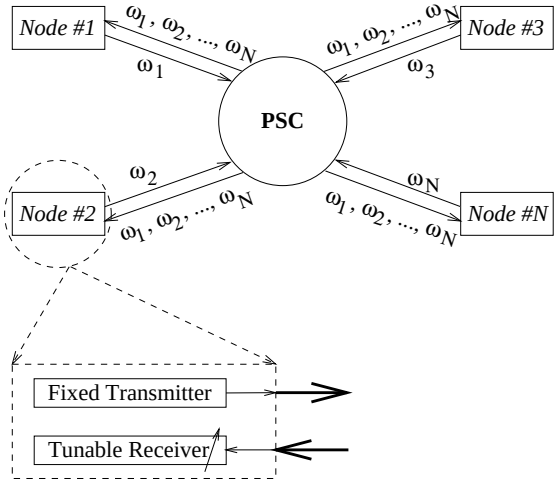

Figure 1: Passive-star-based optical WDM network.

Protocols for PSC networks are typically classified as either single-hop or multihop. The basic goal of the protocol is to provide a certain degree of connectivity among the nodes in order to enable communications. A single-hop protocol typically achieves connectivity through the use of tunable components. In a sense, a single-hop network achieves full connectivity (a fully-connected logical topology) among nodes, since each node can communicate *directly* with every other node by appropriately tuning its transmitters or receivers; however, each such connectivity is only meaningfully established for the packet transmission duration. A limitation of single-hop networks is that they may incur significant overhead due to the high tuning times of tunable components. Typical transceiver tuning times may range from tens of microseconds to a few milliseconds, depending on the technology employed; thus, a tuning period may be on the order of several packet transmission times. A multihop network achieves connectivity through the use of additional transmitters and receivers at each node and typically requires no tuning, instead relying on a

This work was supported in part by DARPA Contract Nos. DABT63-92-C-0031 and DAAH04-95-1-0487, and NSF Grant Nos. NCR-9205755 and NCR-9508239. A shorter version of this paper was presented at IEEE INFOCOM '98, March 1998.

J. P. Jue is with the Department of Computer Science and the Center for Advanced Telecommunications Systems and Services, University of Texas at Dallas, Richardson, TX 75083 USA (e-mail: jjue@utdallas.edu).

B. Mukherjee is with the Department of Computer Science, University of California, Davis, CA 95616 USA (e-mail: mukherje@cs.ucdavis.edu).

static logical topology. Since a multihop network doesn't rely on tuning, it doesn't incur the potentially high tuning overhead of a single-hop network. On the other hand, a multihop network's reliance on multiple transmitters and receivers, combined with the expense of such components, often results in low connectivity in the network, and, consequently, higher average hop distances.

In this paper, we consider a class of protocols called multiconfiguration multihop protocols (MMPs). These protocols are multihop protocols which not only rely on the number of transmitters and receivers to achieve a desired connectivity, but also *make use of tuning to increase connectivity*. These protocols use tunable components to cycle through a fixed number of configurations[1], and the logical topology of the network is then defined in the time domain over all of these configurations. Using tuning in this manner, we can increase the logical connectivity of the network, thus reducing the average hop distance at the expense of additional tuning requirements.

In the following section, we describe a framework for classifying and comparing various protocols for the passive-star network architecture, and we present some previously proposed single-hop and multihop protocols. Section 3 describes the proposed class of multiconfiguration multihop protocols and some of the design issues which must be considered when implementing such protocols. We apply the class of protocols to a specific network architecture in Section 4. Section 5 presents our illustrative numerical results, and Section 6 concludes this work and discusses areas for future study.

2 Framework for Protocol Classification

The most basic goal of any network is to provide connectivity. By providing connectivity, a network allows any of its nodes to communicate with any other node. One measure of the level of connectivity in a network is the nodal degree[2]. The nodal degree specifies the number of direct links that one node may have to other nodes, and is inversely related to the average hop distance in a network: i.e., a higher nodal degree (high connectivity) results in fewer hops between a source and a destination, while a lower nodal degree (low connectivity) results in more hops between a source and a destination. The exact relation between nodal degree and the required number of hops depends on the actual logical topology of the network.

In a WDM PSC network, there are three approaches to increasing the nodal degree, and thus, connectivity, in the network. These approaches are outlined below:

Additional transmitters and receivers: One approach to increasing connectivity is to increase the number of transmitters and receivers at each node. Each unit increase in the number of transceiver pairs per node increases the nodal degree by one. The practicality of this approach is limited by the high cost of optical transmitters and receivers.

Channel sharing: Another method for providing connectivity in a WDM PSC network is to reduce the number of channels in the network, and allow channel sharing [1] [2] [3]. By limiting the number of channels in this manner, multiple transmitters and receivers from different nodes can be made to occupy the same channel, thus creating more direct connections between nodes. As an example, consider a network with N nodes and one transceiver pair per node. Assume that transmitters and receivers are uniformly distributed across all channels. If N channels are used, then one transmitter and one receiver may be tuned to each channel; thus, the overall nodal degree is one. If two channels are used, then $N/2$ transmitters and $N/2$ receivers are tuned to each channel, resulting in a nodal degree of $N/2$. If only one channel is used, then we have a single-channel broadcast environment in which every node is directly connected to every other node (a fully-connected logical topology). In general, if w channels are used, then an average nodal degree of N/w can be achieved. The disadvantage of using shared channels is that the bandwidth (e.g., time slice) on each channel must be divided among all transmitters or receivers tuned to that channel. Thus, the transmitters and receivers will not be fully utilized at all times.

Multiconfiguration multihop approach: The final approach for increasing connectivity is to use tunable components *to increase the connectivity in the time domain*. At any given point in time, the nodal degree is determined by the number of transceivers at each node; however, if the transceivers are tunable, the nodal degree may be increased by considering connections made over a period of time. For example, consider a network with t transceivers per node. Suppose that time is divided into cycles which are repeated indefinitely, and each cycle is divided into two periods. During the first period, the tunable transmitters and tunable receivers are configured according to some configuration C_1. During the second period, the tunable transmitters and tunable receivers are reconfigured according to a different configuration C_2. Within each period, the nodal degree is equal to t, since each node can have a direct connection to only t other nodes. However, over an entire cycle, the nodal degree can equal $2t$, if each node connects to t nodes in the first period and t other nodes in the second period. In general, the tunability in a network may be used to achieve an arbitrary degree of connectivity by adjusting the number of periods or configurations per cycle. The cost of using tunable components to provide connectivity, however, is that tuning a component from one channel to another may incur some amount of delay. Also, tunable components generally cost more than non-tunable components depending on the technology used to

[1] We define a configuration as the topology of the network at a given point in time as determined by the state of the tunable components in the network.

[2] In this paper, the term *nodal degree* will refer to the *logical* nodal degree, or the number of connections each node has to other nodes in the logical topology, rather than the *physical* nodal degree which specifies the number of transmitters and receivers at each node. We will assume that the logical degree is the same for each node in the network, and that the physical degree is the same for each node in the network

build the components.

The concept of using a number of different network configurations over time has been previously studied for mesh networks [4] and for networks with centralized switches [5], but has not been applied to PSC networks in which nodes are equipped with tunable components.

In summary, a network may be characterized by its degree of connectivity, specified by the nodal degree and the methods by which it achieves this connectivity. We employ the following parameters to characterize such a network:

- d: Upper bound on average nodal degree.
- N: Number of nodes in the network.
- w: Number of wavelengths used in the network.
- t: Number of transceiver pairs per node. (The number of transmitters is equal to the number of receivers and is the same for all nodes.)
- c: Number of configurations or tunings required to achieve a desired logical connectivity.

From the above description, we observe that the number of transceivers per node, the level of tunability, and the number of channels in the network can all be related by the amount of connectivity they provide. This connectivity, expressed as an upper bound on the average nodal degree, can be represented as:

$$d = \frac{tc}{N}\left[(N \bmod w)\lceil\frac{N}{w}\rceil^2 + (w - N \bmod w)\lfloor\frac{N}{w}\rfloor^2\right], \quad (1)$$

where *mod* is the modulo operator. The t and c represent the nodal degree which can be achieved through additional transceivers and tuning respectively, while the rest of the equation represents the nodal degree which can be achieved through the sharing of channels. Equation (1) assumes that the transmitters and receivers in the system are equally divided among the available wavelengths. If N/w is not an integer, then $(N \bmod w)$ wavelengths will be shared by $\lceil\frac{N}{w}\rceil$ nodes, and $(w - N \bmod w)$ wavelengths will be shared by $\lfloor\frac{N}{w}\rfloor$ nodes. Each of the nodes sharing a given wavelength will be able to establish a direct connection with every other node on the same wavelength, thus if $\frac{N}{w}$ nodes are sharing a wavelength, then the nodal degree for nodes sharing that wavelength is upper bounded by $(\frac{N}{w})^2$. If we assume that N/w is an integer, then the equation simplifies to:

$$d = t \cdot c \cdot \frac{N}{w}. \quad (2)$$

A graph depicting this relationship for $d = N-1$ is shown in Fig. 2. The graph also illustrates a relation between single-hop and multihop networks. In a single-hop network, a nodal degree of at least $N-1$ is required; thus, a necessary condition for a single-hop network is:

$$t \cdot c \cdot \frac{N}{w} \geq N - 1$$

or,

$$\frac{t \cdot c \cdot N}{w \cdot (N-1)} \geq 1. \quad (3)$$

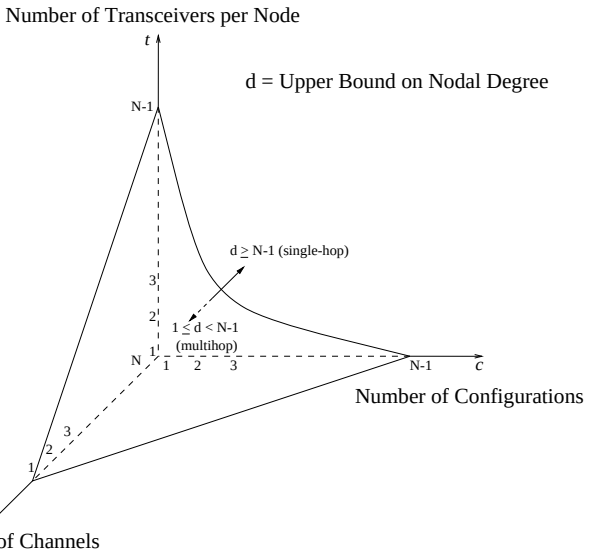

Figure 2: Achievable connectivity vs. t, c, and w.

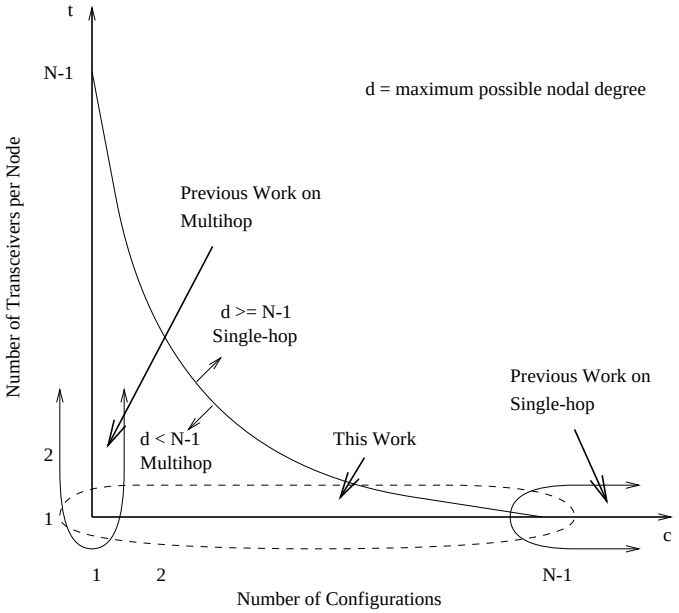

Figure 3: Achievable connectivity vs. t and c, with $w = N$.

If we assume that there are no multiple links between any pair of nodes and no self loops in the network, then Eqn. (3) is also a sufficient condition for a single-hop network.

For a multihop network a sufficient condition is:

$$\frac{t \cdot c \cdot N}{w \cdot (N-1)} < 1. \quad (4)$$

The origin of the graph in Fig. 2 indicates a nodal degree of one, in which case we would have a static ring network.

Figure 3 illustrates the relation of connectivity to the number of transceivers and the tunability requirements when the number of channels is equal to N. Previous works on single-hop protocols have typically relied on a single tunable transceiver pair at each node (not including transceivers for a control channel), and they have employed tuning to

achieve connectivity. Thus, previous single-hop protocols can be classified in the region $c \geq N-1$ of Fig. 3. On the other hand, previous works on multihop protocols have typically relied on multiple transceivers for connectivity, and can therefore be classified along the t-axis of Fig. 3. In this work, we wish to focus on the parameter c, which ranges from 1 (for a traditional multihop network) to $N-1$ (for a traditional single-hop network).

We now briefly describe previous work on single-hop and multihop packet-switched protocols. Reviews of single-hop and multihop protocols may be found in [6] and [7].

2.1 Single-Hop Protocols

In order to accommodate packet switching in a single-hop network, a protocol must overcome the potentially high tuning delays of transceivers and obtain the required level of coordination between nodes. We note that tunable receivers may be used for two purposes: 1) to provide increased connectivity in a network, and 2) to provide an adaptive, dynamic network in which the logical connectivity changes appropriately in order to accommodate varying traffic patterns. In a dynamic single-hop network, the network is often reconfigured on a packet-by-packet basis. In general, setting up connections on a packet-by-packet basis may be very inefficient. Also, if each node only has a single receiver, it may only listen to a single channel at any time. Thus, a node cannot obtain information regarding each of the other nodes' traffic demands without tuning to each channel.

Packet switching may be provided in a static manner in a single-hop network by using time-division multiplexing (TDM) [8] [9]. In the static WDM/TDM approach, time on each channel is divided into a number of frames, with each frame subdivided into slots. Nodes are equipped with a fixed transmitter and a tunable receiver (FT-TR); however, such protocols may be easily modified to accommodate a tunable-transmitter, fixed-receiver system (TT-FR). During a slot, each node's receiver is assigned to a particular channel, allowing any transmitter on that channel to transmit packets to a given receiver. Each TDM frame contains a sufficient number of slots which are assigned such that each node is able to communicate with every other node. In this situation, the tuning property of a transceiver is used only to provide connectivity and not used to provide adaptivity in the network.

A dynamic single-hop approach requires nodes to exchange information regarding packet transmission requests. In some cases, this control information is exchanged by equipping each node with an additional transmitter and receiver, and having these components fixed-tuned to operate on a control channel which is separate from the data channels, e.g, [10] [11] [12]. Alternatively, the control information may be exchanged over the data channels using in-band signaling [13]. The dynamic single-hop approach requires more tuning and overhead, but is able to accommodate varying traffic patterns. Both static single-hop protocols and dynamic single-hop protocols may perform poorly if tuning times are high.

2.2 Multihop Protocols

In a multihop protocol, a static logical topology is established over the physical topology. In previous works, the logical topology has been created by fixing each transmitter and each receiver to an appropriate channel [14]. In the case of a single transmitter and a single receiver per node, the only possible fully-connected topology would be a ring. A ring allows all nodes to communicate with one another, but requires each packet to traverse an average of $N/2$ hops, where N is the number of nodes in the network. For a large network, a ring topology may be very inefficient.

A number of dynamic multihop protocols have been proposed [15] [16]. These networks reconfigure their logical topology based on prevailing traffic conditions. However, each logical topology is still constrained by the number of transmitters and receivers at each node. In this case, tuning is used to create an adaptive network, but not to increase the connectivity of the network.

3 Multiconfiguration Multihop Protocols

Multiconfiguration multihop protocols (MMPs) employ tunable components to obtain a desired logical topology. Instead of defining the logical topology over a single configuration of the transmitters and receivers, these protocols define the logical topology over a number of configurations in the time domain. This approach results in a lower average number of hops than in a single-configuration (i.e., standard) multihop network; also it requires less tuning than in a single-hop network. In this section, we present a high-level description of the proposed MMPs and discuss the corresponding architectural issues that need to be considered when designing such protocols. The protocols may be implemented either statically or dynamically.

3.1 Static MMPs

In a static MMP, time is divided into frames, and in each frame the network cycles through a series of c configurations in a TDM fashion. The topology of each of the configurations is superimposed to determine the logical topology over the entire frame. For example, in Fig. 4, each node is equipped with a single fixed transmitter and a single tunable receiver (FT-TR). Two configurations, C_1 and C_2, are used to construct the logical topology of a bidirectional ring (even though each of the N nodes has only one transmitter and one receiver). The corresponding TDM frame for the protocol is given in Fig. 5. The ω's in the schedule represent the wavelengths to which transceivers must tune during each configuration. For example, in Configuration C_1, Node 1's transmitter and Node 2's receiver must both be tuned to channel ω_1. Between configurations, each

node must re-tune its transceivers appropriately to generate the next configuration. The time required to tune the transceivers is given by τ. The receivers cycle through the configurations specified by the TDM schedule indefinitely. In this manner, any arbitrary virtual topology may be embedded onto a network which has tunable components.

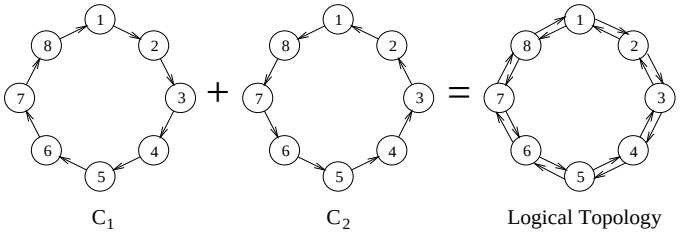

Figure 4: A bidirectional ring with two configurations.

	C_1	τ	C_2	τ
ω_1	1->2		1->8	
ω_2	2->3		2->1	
ω_3	3->4		3->2	
ω_4	4->5		4->3	
ω_5	5->6		5->4	
ω_6	6->7		6->5	
ω_7	7->8		7->6	
ω_8	8->1		8->7	

Figure 5: TDM frame for bidirectional ring topology.

Note that, due to the synchronous nature of the proposed protocols, it is possible that MMPs may be able to support circuit-switched traffic through the reservation of slots within configurations. In using MMPs to implement circuit-switching schemes, restrictions need to be placed on the length of the TDM frame in order to provide guarantees with respect to the data rate and delays.

3.2 Dynamic MMPs

A dynamic version of the MMP may also be implemented. The primary difference in the dynamic protocols is that the logical topology is reconfigured based on prevailing traffic patterns, while in the static protocol, the logical topology is selected based on known (or assumed) traffic patterns.

Among dynamic protocols, the time scale on which the logical topology can change may vary widely. The logical topology can change infrequently based on long-term traffic patterns, or it can change based on immediate traffic demands. In either case, some amount of coordination is required. This coordination may be provided either through a separate control channel, or through the use of in-band signaling over the data channels (see, for example, [13]).

The case in which the logical topology changes infrequently is a straightforward extension of the static protocol. It may be implemented by having each node accumulate traffic information over a period of time, and then having the nodes exchange this information at predetermined intervals. Once the information is exchanged, each node can determine the new logical topology. Rather than generating topologies on-line, an alternative implementation is to generate a number of logical topologies off-line and simply have the nodes switch from one logical topology to another depending on the traffic load. For example, under light loads, the network may simply maintain a ring topology which doesn't require any tuning. If the queues at the nodes start to build up, the network may then switch to the bidirectional ring topology, as shown in Fig. 4, in which the nodes cycle between two unidirectional ring configurations. The basic problem in the case of both static and dynamic MMPs is then to find an appropriate logical topology for a given traffic matrix.

3.3 Topology Selection

In general, when selecting a logical topology and the corresponding configurations, the objective is to minimize delay or maximize throughput for a given traffic matrix. In [14], the problem of finding an optimal logical topology for a given traffic matrix and a given physical topology is addressed. However, in our case, we are not constrained by a physical topology, nor is the logical topology constrained by the number of transmitters or receivers at each node. Instead, we are allowed to add as many logical links as we wish to the logical topology with the condition that each increase in the nodal degree of the logical topology also contributes to an increase in the average delay (due to additional tuning requirements). We can then modify the problem statement in [14] to accommodate networks in which multiple configurations are employed to create a logical topology. Such topological optimization problems are known to be hard to solve [14]. In this work, we will simplify the problem by assuming that the traffic matrix is uniform, and that each node is equipped with one fixed transmitter and one tunable receiver (FT-TR). Since the traffic matrix is uniform, we will consider only regular logical topologies. One class of regular logical topologies which is scalable and which is able to accommodate an arbitrary nodal degree is GEMNET [17]. GEMNET also compares favorably to other classes of topologies in terms of average hop distance.

Once a logical topology is selected, another issue to consider is how to partition the logical topology into individual configurations, and how long the network should remain in any given configuration. For any given logical topology, there may be a number of ways to partition the topology into individual configurations. For example, the bidirectional ring topology in Fig. 4 may also be partitioned into the configurations shown in Fig. 6. Although different sets of configurations may result in the same logical topology, they may affect the performance of the network in different ways, specifically with regard to multihop traffic. In some instances, a multihop packet may be able to traverse multiple hops within a single configuration. For example, in Fig. 4, a packet from Node 1 to Node 3 may traverse

the link from Node 1 to Node 2 and the link from Node 2 to Node 3 during Configuration C_1, if the duration of the configuration is long enough. If the configuration is not long enough, or if the packet arrived to Node 2 towards the end of the configuration, then the packet will have to wait for the network to cycle through Configuration C_2 and then back to Configuration C_1 before traversing the link from Node 2 to Node 3. On the other hand, if the logical topology is partitioned into the configurations as shown in Fig. 6, then the packet will only be able to traverse one hop in any configuration. Thus, the same packet will traverse the link from Node 1 to Node 2 during Configuration C_1, wait for the network to switch to Configuration C_2, and then traverse the link from Node 2 to Node 3. The difference between the two possible partitionings becomes greater when the average hop count is high and the duration of the configuration is longer.

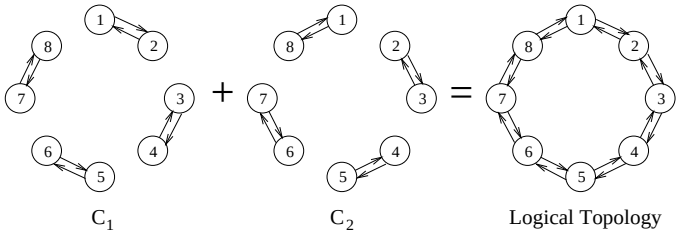

Figure 6: Alternative configurations for the bidirectional ring.

In the topologies that we will study, we will set each configuration to be of equal duration. However, there are situations in which it may be beneficial for some configurations to have longer durations than other configurations. Such situations include non-uniform traffic and cases in which the logical topology leads to unbalanced link loads. In extreme cases of link load imbalance, it may also be feasible to employ the same logical link in more than one configuration. We do not consider such situations in this paper.

3.4 Synchronization

In order to implement MMPs, we require some method of synchronization to ensure that all of the nodes are tuned to the appropriate channel at the beginning of each configuration. One approach to achieving synchronization is through the use of a separate control channel. Each node may be equipped with an additional transmitter and receiver which are fixed to a control channel, and one of the nodes may broadcast a global synchronization signal to all of the other nodes. If there is no separate control channel, then the synchronization must occur in-band, or over the data channels. One approach is to design the configurations such that they only consist of rings, as in Figs. 4 and 6. One node in each ring can send a control message over the ring to synchronize all of the nodes in that ring. When the control message returns to the originating node, then it knows that all of the other nodes in the ring are tuned to the correct channel.

3.5 Routing

Each node in the network maintains a separate queue for each outgoing link which is connected to the node in the logical topology. During a given configuration, only queues which correspond to links existing in the specified configuration are serviced. If a link exists in more than one configuration, a single queue is still maintained for that link, and the queue will be served during more than one configuration. When a packet arrives at a node, the node must choose the appropriate queue in which to place the packet. This decision is based on the routing mechanism. The routing mechanism may be dynamic, in which case the routing decision is based on the current network configuration as well as the network's logical topology, or it may be static, in which case the routing decision is based only on the logical topology.

A simple dynamic routing mechanism is one in which a packet is placed in the queue for the current configuration if the topology of the current configuration has a path to the packet's destination. A limitation of this type of mechanism is that, if a packet doesn't reach its destination by the end of a configuration, then the routing decisions for subsequent hops may be based on different configuration information. As a result, the packet may end up being routed back to nodes which have already been visited. For example, consider the network in Fig. 4. If a packet destined for Node 7 arrives at Node 1, and if the network is currently in Configuration C_1, then a dynamic routing mechanism may choose to send the packet to Node 2 with the expectation that the packet will traverse Nodes 3, 4, 5, and 6 before finally reaching Node 7. However, if the network switches to Configuration C_2 immediately after the packet has been transmitted to Node 2, then Node 2 may decide to send the packet back to Node 1, since a path exists from Node 1 to Node 7 via Node 8. An approach to avoid this "looping" problem is to have each packet maintain a list of visited nodes in order to prevent a packet from traversing a node more than once. Time-out mechanisms may also be employed to prevent packets from remaining in the network indefinitely.

If a static routing mechanism is used, the above problem can be avoided altogether. In a static routing mechanism, a fixed set of paths is specified for each source-destination pair. Each packet from a given source to a given destination will choose a single path from the available set of paths specified, and the packet will be routed along the chosen path (using source routing). If the links in each configuration are unique, then every link on the chosen path will exist in exactly one configuration, and thus correspond to a specific queue at each node. In this paper, we consider a static shortest-path routing mechanism in which all shortest paths from a source to a destination are considered, and the source will randomly choose one of these shortest paths for a given packet.

4 Analysis

To illustrate the trade-off provided by MMPs, we apply MMPs to an N node network based on a FT-TR PSC-network architecture, and we develop an analytical model to measure the delay and throughput characteristics of the network. Each node's transmitter is fixed to an unique wavelength channel. Thus, each node is uniquely identified with the channel on which it transmits. Receivers are tunable across all channels and have a tuning time of τ.

One example of the above architecture is the Rainbow-I network [18] developed at IBM. The tunable receivers in the Rainbow network have significant tuning delays which are on the order of tens of milliseconds. Due to these high tuning times, the Rainbow protocol described in [18] is suited only for circuit-switched traffic. However, by replacing the Rainbow protocol with an MMP while using the same optical hardware, the network may also be able to handle packet-switched traffic.

In our analysis, each node is modeled as shown in Fig. 7, and each outgoing link in the logical topology at a node corresponds to a queue at that node. Time is slotted with the slot length being equal to one packet transmission time (which may be, for example, approximately 10 μs, assuming a transmission rate of 1 Gbps and a packet size of around 10,000 bits). The following parameters are given:

- N: Number of nodes.
- w: Number of channels (assumed to be equal to N).
- t: Number of transceivers per node (assumed to be equal to one).
- c: Number of configurations.
- Logical topology for above N, w, t, and c.
- τ: Tuning time in slots.
- K_{tot}: Total number of data slots per frame.
- K: Total number of data slots per configuration (equal to $\frac{K_{tot}}{c}$ if configurations are of equal duration).
- λ: Arrival rate of new packets to the system across all nodes (measured in packets per slot).
- μ: Service rate of packets (assumed to be one packet per slot).
- p: Propagation delay in slots.

Given the logical topology, we can calculate the following terms:

- h: Average hop distance.
- L_i: Link load for link $i = 1, 2, \cdots, N \cdot c$ as a fraction of total load in the network.

In this paper, we calculate h and the L_i's using a direct approach. We assume one unit of traffic flow from each source to each destination, and shortest-path routing. The shortest path from each source to each destination is calculated, and the appropriate flows are assigned to each link in the shortest path. For the case of multiple shortest paths, the flow is equally divided among the shortest paths. (In the actual protocol, if multiple shortest paths exist, one path is chosen at random from among the shortest paths, and each shortest path has an equal probability of being chosen.)

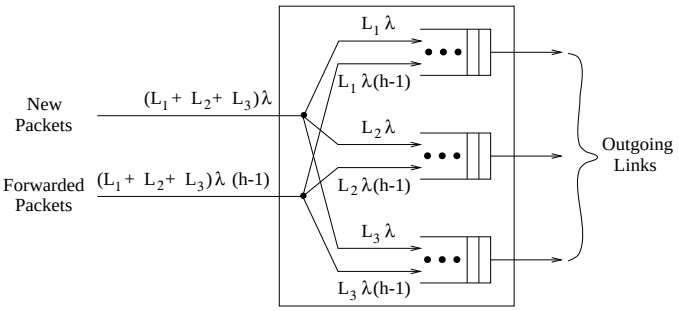

Figure 7: Model for a node with a logical nodal degree equal to three ($t = 1$, $c = 3$).

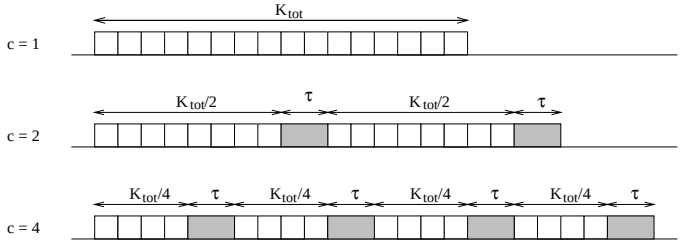

Figure 8: Frames for different number of configurations.

Since λ packets arrive to the system per slot and each packet traverses an average of h hops, the total arrival rate of both new and forwarded packets to the system is $\lambda \cdot h$ packets per slot. Each of the $N \cdot c$ links in the network is modeled as a single queue (Fig. 7). The packet arrivals to each queue i are assumed to be according to a Poisson process with rate $\lambda \cdot h \cdot L_i$ (this is an approximation since only the arrivals of new packets are Poisson, while the arrivals of forwarded packets are non-Poisson). The service discipline is conducted in a generalized TDM fashion, with the link transmitting one packet per slot for the duration of a configuration (K slots), and transmitting no packets during the remainder of the frame, where the duration of an entire frame is $K_{tot} + \tau \cdot c$ slots (see Fig. 8). The queueing model can then be solved to find the queuing delay, D_i, for each individual link using the generalized TDM model from [19].

The queueing delays, D_i, consist of three components. The first is the time required for an arriving packet to synchronize to the appropriate configuration. This term can be calculated as the expected number of slots that a packet must wait in order to reach the beginning of a slot of the desired configuration. If the packet arrives outside of the desired configuration, then it must wait an average of $\frac{K_{tot}+\tau \cdot c-K}{2}$ slots to reach the beginning of the proper configuration. If the packet arrives within the desired configuration, then it must wait an average of $\frac{1}{2}$ slot to reach the next slot in the same configuration. The second component of delay is the time required to transmit packets already in the queue. This component consists of a number of complete frames followed by a partial frame, and is calculated by first finding the expected number of packets in the queue. After finding the expected number of pack-

ets in the queue, we can then find the expected value for the number of full frames and the expected number of remaining packets in the partial frame. The final component of the delay is the transmission time of the packet itself, which is equal to one slot. The average link delay for the entire system can then be found by weighting the delay of each link according to its link load, summing the weighted delays, and dividing by the total number of links. The average packet delay is found by adding the propagation delay to the average link delay and multiplying by the average hop count, i.e.,

$$D_{avg} = h \cdot (p + \sum_{i=1}^{N \cdot c} D_i \cdot L_i). \quad (5)$$

The normalized throughput of the system is defined as the fraction of time spent transmitting new packets, i.e.,

$$S = \frac{\lambda}{N} \quad (6)$$

The maximum normalized throughput (or capacity) of the system is the throughput at which the system becomes unstable and can be calculated as follows. The fraction of time that link i is busy is given by:

$$\rho_i = \lambda L_i h \cdot \frac{K_{tot} + \tau c}{K} \quad (7)$$

In order to maintain stability, ρ_i must be less than one for all i; we then have

$$\lambda < \frac{K}{L_{max} h (K_{tot} + \tau c)} \quad (8)$$

where

$$L_{max} = \max_i L_i.$$

The maximum throughput is then given by

$$S_{max} = \frac{K}{N L_{max} h (K_{tot} + \tau c)}. \quad (9)$$

5 Illustrative Numerical Examples

In most of the examples presented below, we plot average packet delay versus throughput for various 32-node GEMNET logical topologies, with $\tau = 5$[3]. A topology consisting of a single configuration ($c = 1$) corresponds to a unidirectional ring, while a 32-configuration topology corresponds to a single-hop network. The remaining topologies correspond to one-column or two-column GEMNETs. GEMNET is a general class of shuffle-exchange networks which includes regular topologies such as ShuffleNet and de Bruijn graphs, and a GEMNET topology can be specified by three parameters: x, y, and z. An (x, y, z) GEMNET has x columns, y rows, and a nodal degree of z (see Fig. 9 for an illustrative example). We select one or two columns for the GEMNETs since these configurations yield the lowest average hop distance [17]. The details of each logical topology are given in Table 1. In these examples, we also assume that the propagation delay is negligible, since we are considering a LAN environment. The analytical results are verified through simulation.

[3]This value corresponds to a tuning time of 50 μs if we assume a slot length of 10 μs. The tuning time depends on the device technology being employed and may be higher or lower than this value.

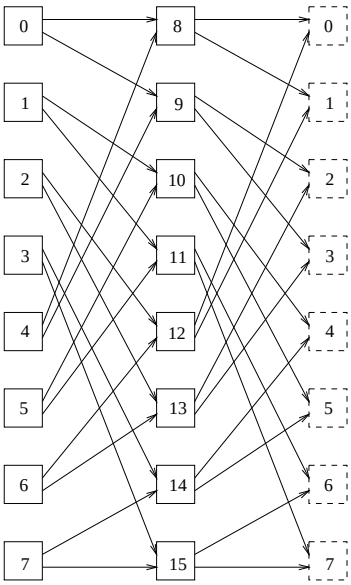

Figure 9: A (2,8,2) GEMNET topology.

c	Logical Topology	h	L_{max}
1	Unidirectional ring	16.0	0.0313
2	(2,16,2) GEMNET	3.564	0.0195
4	(2,16,4) GEMNET	2.258	0.0085
8	(1,32,8) GEMNET	1.75	0.0052
16	(2,16,16) GEMNET	1.484	0.0020
32	Fully-connected (Single-hop)	1.0	0.0010

Table 1: Logical topologies, average hop distance, and maximum link load for a system with $N = 32$ and an uniform traffic matrix.

5.1 Effect of Logical Topology

In Fig. 10 we plot the delay versus throughput for $K = 10$ and $\tau = 5$. We observe that, for low loads, topologies with fewer configurations perform the best, while for higher loads, topologies with more configurations perform best. In low-load situations, the queuing delays at intermediate nodes are small compared to the tuning times. Therefore, reducing the hop count by increasing the number of configurations doesn't improve network performance for low loads and may actually increase delays due to the additional tuning overhead. As the load increases, the delay for topologies with fewer configurations increases quickly. The high average hop count of these topologies leads to a larger number of forwarded packets, thus resulting in a higher overall load on the system. Figure 10 verifies the

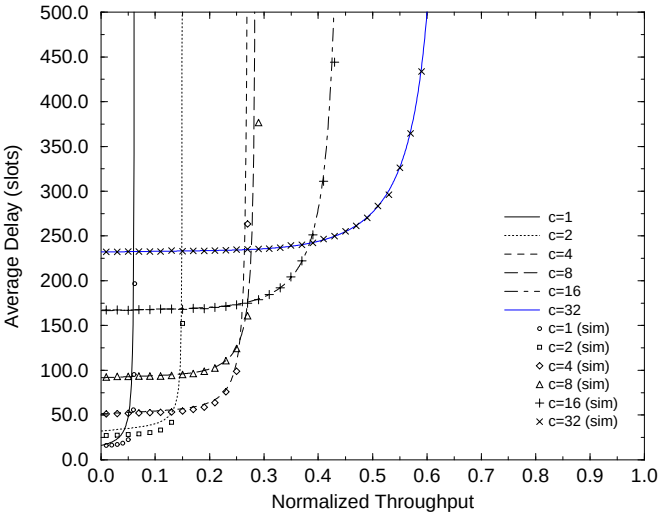

Figure 10: Average delay vs. normalized throughput ($N = 32, K = 10, \tau = 5$).

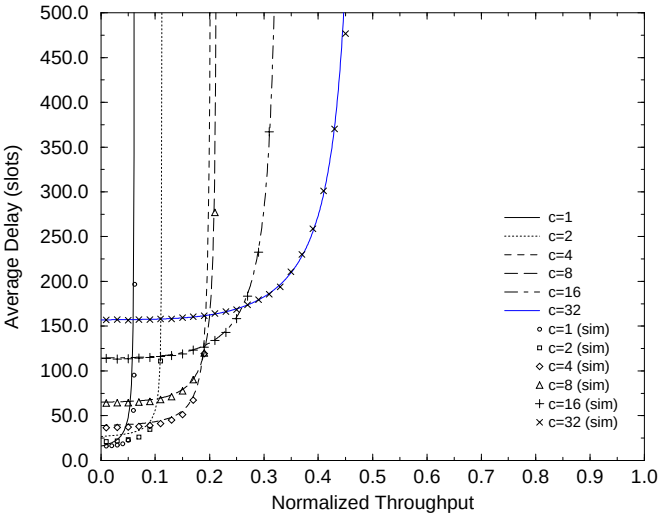

Figure 11: Average delay vs. normalized throughput ($N = 32, K = 5, \tau = 5$).

maximum throughput as derived in Eqn. (9). Also, we observe good agreement between the analysis and simulation, indicating that the Poisson arrival approximation for forwarded traffic is not very detrimental.

Note that, for any given load, there is an optimal topology which minimizes the average packet delay. Consequently, a simple, but effective dynamic protocol can be developed which switches between different topologies. In the above example, in order to achieve minimum packet delay, the network can switch between a (2,16,2) GEMNET ($c = 2$), a (2,16,4) GEMENT ($c = 4$), a (2,16,16) GEMNET ($c = 16$), and a single-hop logical topology ($c = 32$)

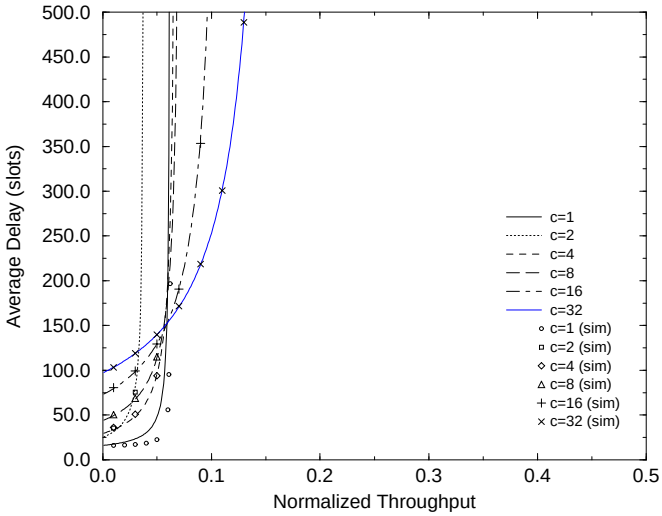

Figure 12: Average delay vs. normalized throughput ($N = 32, K = 1, \tau = 5$).

depending on the traffic load.

Figures 11 and 12 also plot delay versus throughput for $K = 5$ and $K = 1$. We observe that decreasing K increases the ratio of tuning slots to data slots in a frame and greatly reduces the maximum throughput of those topologies which require tuning ($c = 2, 4, 8, 16, 32$). At the same time, decreasing K also leads to lower delays at low loads, since packets which are waiting for the network to tune to the appropriate configuration do not have to wait as long.

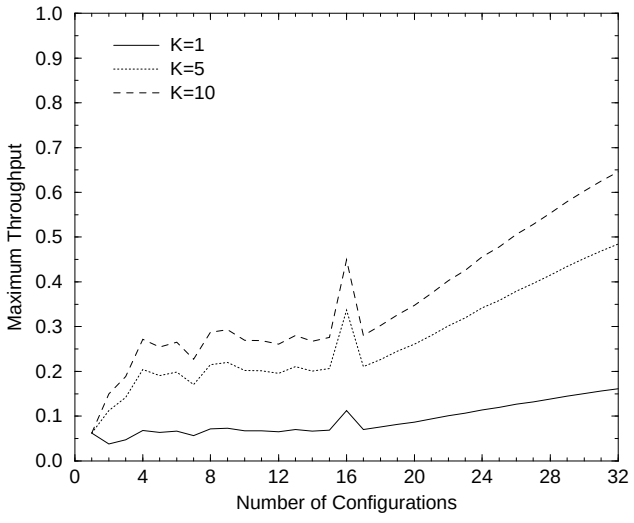

Figure 13: Maximum throughput vs. number of configurations ($N = 32, \tau = 5$).

Figure 13 shows the upper bound on the system throughput for topologies consisting of various number of configu-

rations. The single-hop topology ($c = 32$) is able to achieve the highest throughput for a given K, whereas the ring topology ($c = 1$) achieves the lowest throughput for $K = 5$ and $K = 10$. In general, a lower average hop distance results in a higher maximum throughput, assuming that the links in the logical topology have balanced loads. If the link loads are unbalanced, then the throughput is limited by the most heavily loaded link in the network. For the case of $c = 16$, the chosen topology results in evenly loaded links, resulting in a much higher maximum throughput than for surrounding points ($c = 15, c = 17$, etc.) which have unequally balanced link loads.

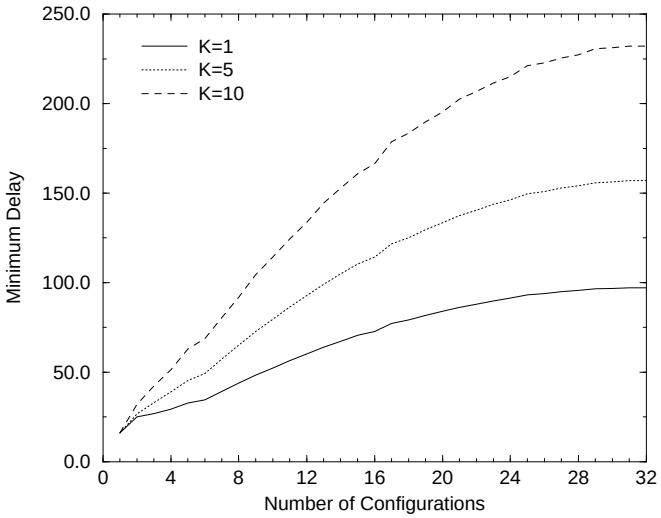

Figure 14: Minimum delay vs. number of configurations ($N = 32, \tau = 5$).

Although a lower average hop distance results in higher throughput, the tuning required to reduce the hop distance increases the delay for low system loads. In Fig. 14, we show the minimum delay for very light loads. We observe that the ring topology yields the lowest minimum delay, while the single-hop topology has the highest minimum delay. In a lightly loaded system, most of the delay is encountered when the packet must wait for the correct configuration before being transmitted. In general, if there are c configurations, a packet must wait an average of $c/2$ configurations per hop. In the case of a ring topology, there is only a single configuration; therefore, the delay is caused only by the transmission time of the packet at each hop.

5.2 Comparison with a Multiple-Receiver Architecture

In this section, we compare the performance of MMPs to a network in which there are two receivers per node and no tuning requirements. The two-receiver-per-node network is configured as a (2,16,2) GEMNET with each node transmitting on a separate fixed wavelength, and each node receiving two different wavelengths. Thus, each of the 32 channels is shared between two receiving nodes, but is unique to each transmitter.

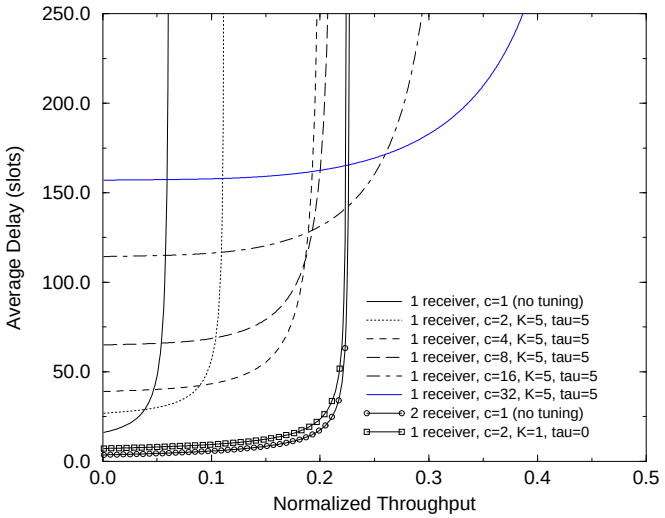

Figure 15: Average delay vs. normalized throughput ($N = 32$).

Figure 15 shows the delay versus throughput for the networks being compared. Only analytical results are shown. (For the remainder of the paper, only analytical results will be shown unless otherwise specified.) We also show the case for which $K = 1$ and $\tau = 0$ to illustrate the degree of performance difference which is due to tuning requirements. We observe that the two-receiver GEMNET performs better than the corresponding two-configuration GEMNET. However, we also note that, by increasing the number of configurations to 16 or 32 in the multiconfiguration network, we can achieve a higher maximum throughput than the two-receiver network. We observe that there is a small difference in performance between the GEMNET topology in which tuning is required with $\tau = 0$ and the two-receiver network in which no tuning is required. This difference results from the imbalance in traffic between pairs of links which are sharing the same wavelength. For the case in which $K = 1$ and $\tau = 0$, the bandwidth is equally shared between each of the two outgoing links from a node, whereas in the two-receiver case, a greater amount of the bandwidth can be allocated to the link which requires more bandwidth, resulting in higher utilization.

5.3 Effect of Configuration Length

In Fig. 16, we plot delay versus throughput for different values of K and $c = 2$. We observe that, by increasing the parameter K, we can achieve higher throughput at the expense of slightly higher delays under light loads. Increasing K reduces the overhead of tuning by allowing more packets to be transmitted per configuration. However, if the load is low, a packet may potentially have to wait a longer

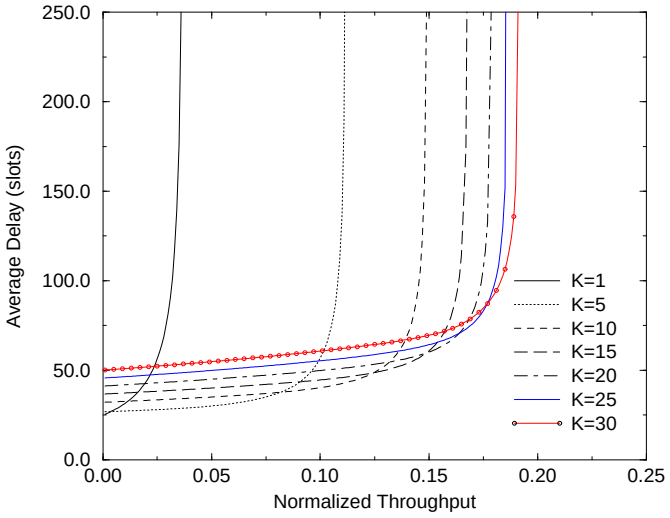

Figure 16: Average delay vs. normalized throughput ($N = 32, c = 2, \tau = 5$).

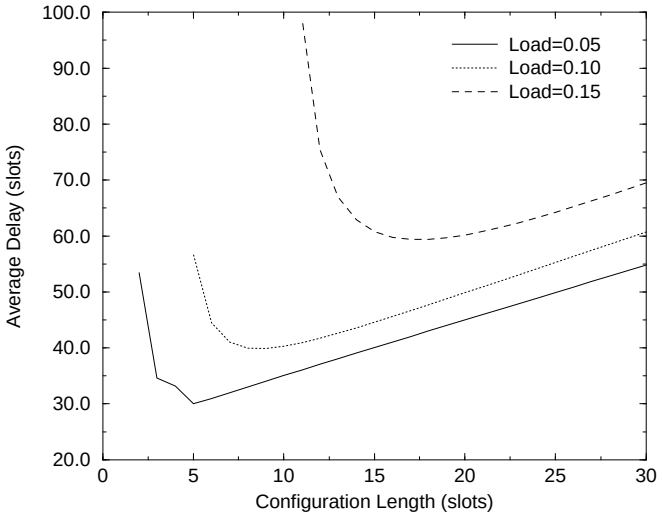

Figure 17: Average delay vs. configuration length ($N = 32, \tau = 5$).

time before the network reaches the appropriate configuration, since each configuration has a longer duration. We also note that, as K increases beyond a certain point, the corresponding increase in throughput becomes very small. Eventually, as $\frac{K}{K+\tau}$ approaches unity, the throughput will be more limited by $\frac{1}{h}$ as indicated in Eqn. (9). The results in this section also suggest another approach to implement a dynamic protocol. Instead of switching between different logical topologies to accommodate different loads, we can simply extend or reduce the length of each configuration as the load changes. In Fig. 17, we show that, for a given load, there is a value of K which minimizes the average delay. If K is too large, packets will have to wait longer for a desired configuration, and if K is too small, the network becomes bandwidth-limited, and the queueing delay at each node increases.

5.4 Effect of Tuning Time

In Fig. 18, we plot delay versus throughput for different values of τ, and $c = 1, 2, 4$, and 32. Note that, since the static ring topology ($c = 1$) requires no tuning, the curve for $c = 1$ remains unchanged for different values of τ. As expected, in a system with low tuning times (e.g., $\tau = 0$), topologies with a higher number of configurations offer better performance, while in a system with high tuning times, systems with fewer configurations perform better. We observe that, when $\tau \geq 50$, the static ring topology ($c = 1$) will provide the best performance for any offered load. However, we note that, by increasing K, the maximum throughput can be increased for $c = 2$ and $c = 4$ (as in Fig. 16).

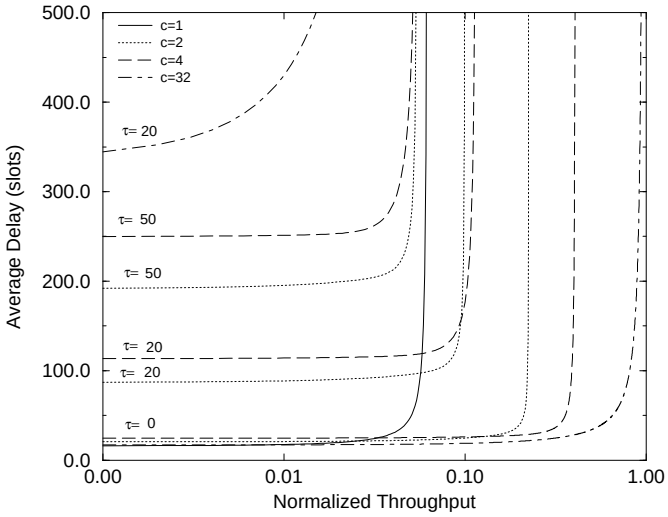

Figure 18: Average delay vs. normalized throughput ($N = 32, K = \frac{32}{c}$).

5.5 Effect of Configurations

In Fig. 19, we study the effect of implementing the same logical topology with different sets of configurations. We consider the two implementations of an 8-node bidirectional ring given in Figs. 4 and 6. In Fig. 4, each configuration consists of an 8-node unidirectional ring, while in Fig. 6, each configuration consists of four 2-node rings. Since the analysis doesn't model individual configurations, we present only simulation results. We observe that, for $K = 5$ and $K = 10$, the configurations consisting of two 8-node rings performs better at light loads. The reason for

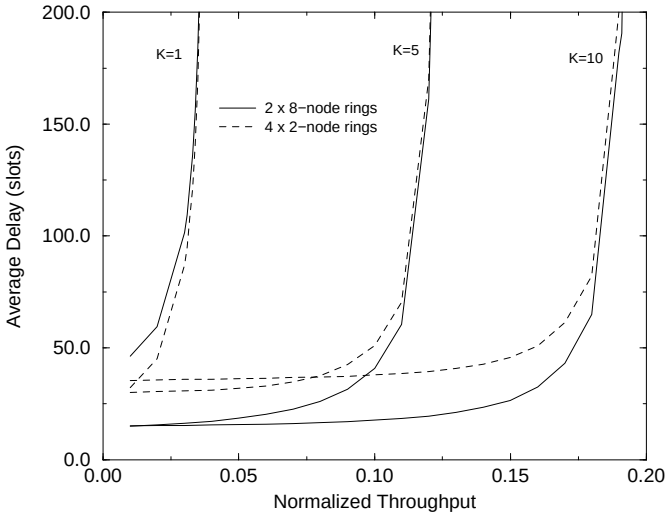

Figure 19: Average delay vs. normalized throughput for different sets of configurations ($N = 8, \tau = 10$).

this performance is that the 8-node rings allow packets to traverse multiple hops in a single configuration, whereas the 2-node rings only allow a single hop per configuration. The number of hops that a packet may traverse per configuration is not only limited by the topology of the configuration, but also by the length of the configuration. A packet is limited to at most K hops per configuration, where K is the duration of the configuration in slots. Therefore, when K is small, there is not as much of a difference between the two different sets of configurations. In fact, for $K = 1$, we observe that the 2-node ring configurations perform better. Although all packets are limited to a single hop per configuration, consecutive hops in the 2-node ring configurations occur in consecutive configurations (i.e., the first hop occurs during Configuration 1, the second hop during Configuration 2, etc.), while in the 8-node ring configurations, each consecutive hop must wait for the same configuration (i.e., the first hop takes place during Configuration 1, the second hop must wait for the network to cycle around back to Configuration 1, etc.).

It is important to note that selection of configurations is a secondary consideration, and that ultimately, performance will depend more heavily on the logical topology of the network. Thus, rather than selecting configurations with few partitions and constructing a logical topology from these configurations, it would be better to select a logical topology which minimizes the average hop distance and then determine the set of configurations which results in the least partitioning and which also yields the desired logical topology.

6 Conclusion

One of the key points in this work is that tunable components not only provide the capability to increase the connectivity of the network, but also provide dynamic capabilities. Many networks take advantage of only one of these benefits while ignoring the other. In this paper, we proposed a class of multiconfiguration multihop networks which takes advantage of tunable receivers to increase connectivity and to reduce the average hop distance, without requiring as much tuning as in a single-hop network. We also showed how these networks can be designed to accommodate dynamic traffic.

We have addressed some of the issues involved in finding a logical topology for an uniform traffic matrix and have shown that the offered load as well as the tuning time play a significant role in deciding on a logical topology. We have also shown that, by appropriately selecting the number of configurations and the duration of these configurations, we can obtain a trade-off between the amount of tuning required and the average hop distance. Reduced tuning requirements lead to lower delays at low loads, while lower hop distances result in a higher amount of network utilization at high loads.

Further work is needed to refine and solve the problem of selecting an optimal logical topology for a given non-uniform traffic pattern. An open problem is to find an appropriate logical topology, as well as the composition and duration of each configuration in the logical topology, such that the average packet delay is minimized.

Another extension to this work is to introduce channel sharing. With channel sharing, the number of channels, w, is less than the number of nodes. It has been shown in [1] that, by using fewer channels, the connectivity of the network can be increased. The study assumed a single fixed transmitter and a single fixed receiver per node. By replacing the fixed receiver with a tunable receiver, not only can we obtain very high connectivities through both channel sharing and receiver tuning, but we can also gain the ability to dynamically adjust the network to meet changing traffic demands. Furthermore, when the number of channels is less than the number of nodes, it may be possible to mask the tuning latency of the tunable receivers. For example, if there are $N/2$ channels and N nodes, then half of the nodes may be receiving data on the channels while the other half of the nodes are re-tuning their receivers. In order to ensure that the tuning time can be masked, the time to tune a receiver must be less than the length of a configuration.

References

[1] S. Tridandapani, B. Mukherjee, and G. Hallingstad, "Channel Sharing in Multi-Hop WDM Lightwave Networks: Do We Need More Channels?" *IEEE/ACM Transactions on Networking*, vol. 5, no. 5, pp. 719-727, Oct. 1997.

[2] S.-R. Tong, D. H. C. Du, and R. J. Vetter, "Design Principles for Multihop Wavelength and Time Division Multiplexed Optical Passive Star Networks," *Journal of High Speed Networks*, vol. 4, no. 2, pp. 189-200, 1995.

[3] M. Kovacevic, M. Gerla, and J. Bannister, "On the Performance of Shared-Channel Multihop Lightwave Networks," *Proc., IEEE Infocom '95*, pp. 544-551, Jun. 1995.

[4] Y. Yemini and D. Florissi, "Isochronets: A High-Speed Network Switching Architecture," *Proc., IEEE Infocom '93*, pp. 740-747, Mar. 1993.

[5] G. Liu, K. Y. Lee, and H. F. Jordan, "Time Division Multiplexed de Bruijn Network and ShuffleNet for Optical Communications," *Proc., IEEE Infocom '94*, vol. 3, pp. 1244-1251, June 1994.

[6] B. Mukherjee, "WDM-Based Local Lightwave Networks - Part I: Single-Hop Systems," *IEEE Network*, vol. 6, no. 3, pp. 12-27, May 1992.

[7] B. Mukherjee, "WDM-Based Local Lightwave Networks - Part II: Multihop Systems," *IEEE Network*, vol. 6, no. 4, pp. 20-32, July 1992.

[8] I. Chlamtac and A. Ganz, "Channel Allocation Protocols in Frequency-Time Controlled High Speed Networks," *IEEE Transactions on Communications*, vol. 36, pp. 430-440, Apr. 1988.

[9] M. S. Borella and B. Mukherjee, "Efficient Scheduling of Nonuniform Packet Traffic in a WDM/TDM Local Lightwave Network with Arbitrary Transceiver Tuning Latencies," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 5, pp. 923-934, June 1996.

[10] M.-S. Chen, N. R. Dono, and R. Ramaswami, "A Media-Access Protocol for Packet-Switched Wavelength Division Multiaccess Metropolitan Area Networks," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 6, pp. 1048-1057, Aug. 1990.

[11] D. Guo, Y. Yemini, Z. Zhang, and W. Guo, "Scalable High-Speed Protocols for WDM Optical Star Networks," *Journal of High Speed Networks*, vol. 4, no. 4, pp. 341-360, 1995.

[12] F. Jia, B. Mukherjee, and J. Iness, "Scheduling Variable-Length Messages in a Single-Hop Multichannel Local Lightwave Network," *IEEE/ACM Transactions on Networking*, vol. 3, no. 4, pp. 477-488, Aug. 1995.

[13] K. M. Sivalingam and J. Wang, "Media Access Protocols for WDM Networks with On-Line Scheduling," *IEEE/OSA Journal of Lightwave Technology*, vol. 14, no. 6, pp. 1278-1286, June 1996.

[14] J. A. Bannister and L. Fratta, "Topological Design of the Wavelength-Division Optical Network," *Proc., IEEE Infocom '90*, San Francisco, pp. 1005-1013, June 1990.

[15] G. N. Rouskas and M.H. Ammar, "Dynamic Reconfiguration in Multihop WDM Networks," *Journal of High Speed Networks*, vol. 4, no. 3, pp. 221-238, 1995.

[16] J.-F. P. Labourdette, G. W. Hart, and A. S. Acampora, "Branch-Exchange Sequences for Reconfiguration of Lightwave Networks," *IEEE Transactions on Communications*, vol. 42, no. 10, pp. 2822-2832, October 1994.

[17] J. Iness, S. Banerjee, and B. Mukherjee, "GEMNET: A Generalized, Shuffle-Exchange-Based, Regular, Scalable, Modular, Multihop, WDM Lightwave Network," *IEEE/ACM Transactions on Networking*, vol. 3, no. 4, pp. 470-476, Aug. 1995.

[18] F. J. Janniello, R. Ramaswami, and D. G. Steinberg, "A Prototype Circuit-Switched Multi-Wavelength Optical Metropolitan-Area Network," *IEEE/OSA Journal of Lightwave Technology*, vol. 11, no. 5/6, pp. 777-782, May/June 1993.

[19] R. Rom and M. Sidi, *Multiple Access Protocols: Performance and Analysis*, New York: Springer-Verlag, 1990.

Jason P. Jue (M '99) received the B.S. degree in electrical engineering (with honors) from the University of California, Berkeley in 1990, the M.S. degree in electrical engineering from the University of California, Los Angeles in 1991, and the Ph.D. degree in electrical and computer engineering from the University of California, Davis in 1999.

In 1999, he joined the faculty at the University of Texas, Dallas, where he is currently an Assistant Professor of Computer Science and a member of the Center for Advanced Telecommunications Systems and Services. His research interests are in the area of high-speed communication networks, with an emphasis on the design and performance evaluation of WDM optical networks.

Biswanath Mukherjee (M '83) received the B.Tech. (Hons) degree from Indian Institute of Technology, Kharagpur (India) in 1980 and the Ph.D. degree from University of Washington, Seattle, in June 1987. At Washington, he held a GTE Teaching Fellowship and a General Electric Foundation Fellowship. In July 1987, he joined the University of California, Davis, where he has been Professor of Computer Science since July 1995, and Chairman of Computer Science since September 1997. He is co-winner of paper awards presented at the 1991 and the 1994 National Computer Security Conferences. He serves on the editorial boards of the IEEE/ACM Transactions on Networking, IEEE Network, ACM/Baltzer Wireless Information Networks (WINET), Journal of High-Speed Networks, Photonic Network Communications, and Optical Network Magazine. He also serves as Editor-at-Large for optical networking and communications for the IEEE Communications Society. He served as the Technical Program Chair of the IEEE INFOCOM '96 conference. He is author of the textbook "Optical Communication Networks" published by McGraw-Hill in 1997, a book which received the Association of American Publishers, Inc.'s 1997 Honorable Mention in Computer Science. His research interests include lightwave networks, network security, and wireless networks.