IEEE
COMMUNICATIONS
SURVEYS
The Electronic Magazine of
Original Peer-Reviewed Survey Articles

MULTICASTING IN WDM NETWORKS

JINGYI HE, S.-H. GARY CHAN, AND DANNY H. K. TSANG
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

ABSTRACT

Wavelength-division multiplexing (WDM) networks are believed to be a promising candidate to meet the explosive increase of bandwidth demand in the Internet. In this article, we survey the problems of and approaches to multicasting in WDM networks. In particular, we address the issues in the context of three types of WDM networks: broadcast-and-select, wavelength-routed, and optical burst-switched (OBS) WDM networks. Broadcast-and-select WDM networks are typically for WDM LANs/MANs, and can be either single-hop or multihop. Various multicast scheduling algorithms (MSAs) are discussed for single-hop networks. For multihop networks, we discuss how channel sharing can be employed to effectively support multicast. In a wavelength-routed WDM network, supporting multicast leads to the multicast routing and wavelength assignment (MC-RWA) problem, which has been discussed for different scenarios, including sparse-splitting networks. We also discuss the problem of efficiently supporting multicast in optical burst-switched (OBS) networks, where the overheads due to control packets and guard bands need to considered.

Wavelength-division multiplexing (WDM) is an effective technique to exploit the large bandwidth of optical fibers to meet the explosive growth of bandwidth demand in the Internet. WDM networks therefore have large capacities to provide broadband and high-quality services, among which multicast services such as video conferencing and distance learning are becoming more and more prevalent. Multicast is the simultaneous transmission of information from one source to multiple destinations, i.e., one-to-many communication. It is bandwidth-efficient because it eliminates the necessity for the source to send an individual copy of the information to each destination, and it avoids flooding the whole network by broadcasting. Multicasting in WDM networks involves different issues, depending on the construction of the network. A WDM local area network (LAN) or metropolitan area network (MAN) is usually constructed based on a shared transmission media and operate in a broadcast-and-select manner, without switching (routing). In contrast, a WDM wide area network (WAN) is constructed with point-to-point WDM links interconnecting the network nodes, where switching (routing) is essential for data transmissions. According to the switching technique used, a WDM WAN could be further classified as circuit-switched (wavelength-routed), packet-switched, or burst-switched. As the optical packet-switched network still faces such technical difficulties as the lack of optical random access memories and stringent synchronization requirements, it is not believed to be practical in the near future and little work has been devoted to multicasting in such networks. Therefore, in this article we discuss multicasting in three types of WDM networks: broadcast-and-select, wavelength-routed, and optical burst-switched (OBS). Following is a brief introduction to the three types of WDM networks and the major issues of multicasting therein.

Broadcast-and-select WDM networks are usually based on a passive star coupler (PSC) [1]. A PSC is a passive optical device without any electronic component, which simply divides the incoming light from any port equally to all the other ports. Therefore, in a network composed of nodes connected with each other through a PSC, the information (for example, a packet) sent from any node is broadcast to all the other nodes through the PSC. Those nodes simply check whether they are the destination of the packet, and then either accept (select) the packet or ignore it. Due to the broadcast capability of the PSC, multicasting in broadcast-and-select networks is inherently very efficient. Problems exist, however, because the PSC is a shared media and many nodes may want to use it on the same wavelength at the same

This work was supported, in part, by the Areas of Excellence (AoE) Scheme on Information Technology funded by the University Grant Council in Hong Kong (AoE/E-01/99), and by the Competitive Earmarked Research Grant funded by the Research Grant Council in Hong Kong (HKUST6014/01E & HKUST6199/02E).

WDN networks		Application areas	Major issues for multicasting	Approaches
Broadcast-and-select		LAN, MAN	Contentions in the shared-media and shared-channel environment	Multicast Scheduling Algorithms (MSAs)
Wavelength-routed	Mesh	WAN	Limitations in • number of wavelengths • wavelength conversion capability • light splitting capability	Multicast Routing and Wavelength Assignment (MC-RWA)
	Ring	LAN, MAN, WAN	Limited number of wavelengths	
Optical burst-switched		WAN	Overheads of the control packets and guard bands	Sharing schemes

Table 1. *A taxonomy of WDM networks.*

time. Therefore, the challenges of multicasting in this type of network are mostly related to the design of some media-access protocol or so called multicast scheduling algorithm (MSA).

Wavelength-routed networks operate based on the concept of *lightpath*. A lightpath is an all-optical communication channel set up between two end nodes, which may span more than one fiber link and pass through some intermediate nodes. At each intermediate node, an optical switch (the so-called wavelength-routing switch) routes the incoming signal all-optically to its corresponding output port. To multicast in a wavelength-routed network, a *light-tree* (i.e., an all-optical multicast tree) needs to be built for each multicast request. However, challenges exist in building such an all-optical light-tree, because of the limitations in the number of wavelengths in the network, wavelength conversion capability and light splitting capability of the wavelength routing switches. The solution involves finding the routes (which may not be a single light-tree) from the source node to all the destination nodes and determining the wavelength(s) to be used on these routes, i.e., the so-called multicast routing and wavelength assignment (MC-RWA).

Although in general wavelength-routed networks assume mesh topologies, WDM ring networks can be regarded as a special type in this category, because a ring network is also composed of a set of individual point-to-point links, as in a mesh network, except that the set of links form a circle. We will include a separate discussion on the ring network, because it is an important form of optical network that has been playing an important role in local area networks (e.g., fiber distributed data interface (FDDI) networks) and in metropolitan and wide area networks (e.g., the synchronous optical network (SONET) rings). A WDM ring network can simply be thought of as multiple ring networks, with each network operating independently on a different wavelength. Multicasting in a WDM ring network also involves routing and wavelength assignment. However, due to the simple topology of a ring network, the routing problem is relatively simple, and light splitting capability is not required at the nodes. Therefore, the challenge is mainly imposed by the limited number of wavelengths, and MC-RWA in such networks is aimed at the efficient use of the wavelengths.

Wavelength-routed networks are basically circuit switched. For bursty traffic, the long duration of lightpaths or light-trees may result in low bandwidth efficiency. Packet switching is a good solution to this problem. However, as mentioned earlier, optical packet-switched networks still need major breakthroughs in technologies. Therefore, optical burst switching (OBS) has been proposed to make use of the strength of both circuit switching and packet switching [2]. In an OBS network, before the transmission of a burst of data, a control packet is first sent to set up a connection by configuring the switches along the path. As opposed to circuit switching, the burst does not wait for a connection acknowledgment, but is sent right after the control packet or after a certain period of time. In order to accommodate the possible timing jitters the burst may suffer at intermediate nodes, guard bands are needed for each burst. Therefore, in addition to the challenge of building a light-tree, reducing the overheads of the control packets and guard bands has been a major objective for multicasting in OBS networks. The general approach is to share the control packets and guard bands among multiple traffic sessions by allowing their traffic to be assembled in the same burst.

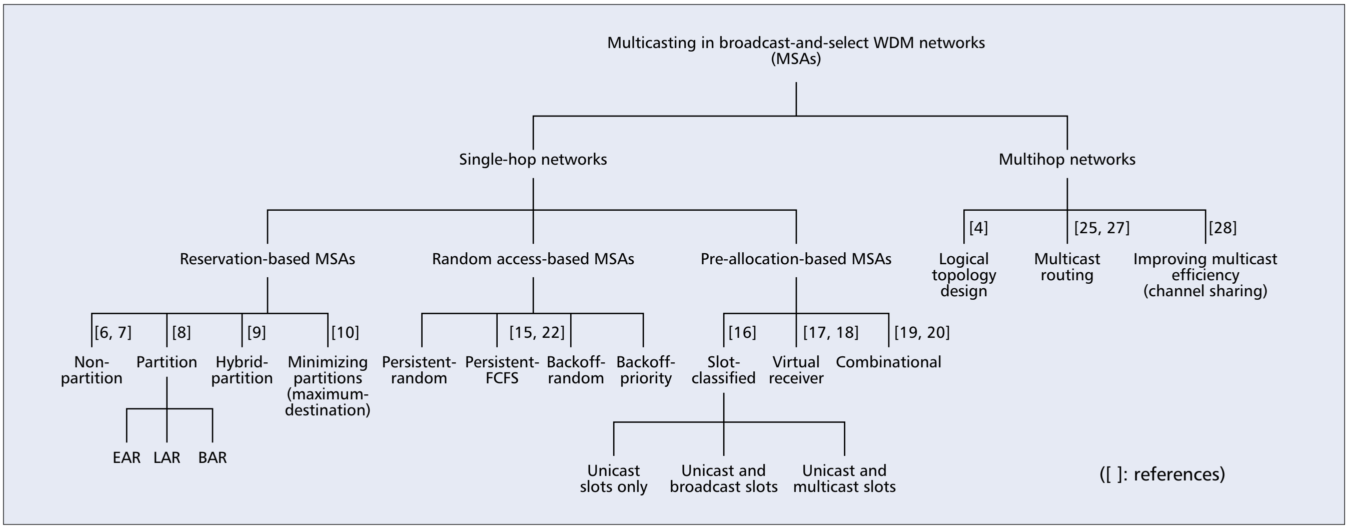
A summary of multicasting in these several types of WDM networks is given in Table 1. In the following sections, the multicast issues in the context of broadcast-and-select WDM networks will be first discussed, followed by the wavelength-routed WDM networks and OBS networks. The conclusions are drawn in the last section.

Multicasting in Broadcast-and-Select WDM Networks

A broadcast-and-select WDM network can be either single-hop or multihop [3, 4]. In a single-hop network, for a transmission to occur, the transmitter of the source (sending) node and the receiver of the destination (receiving) node must be tuned to the same wavelength during the period of the transmission. Data originate at the source node, pass the PSC, and finally reach the destination node, without passing any intermediate network node. In a multihop network, on the other hand, the transmitter and the receiver may not be tuned to the same wavelength, hence a packet sent on the sending wavelength may have to pass through some intermediate nodes and be retransmitted on different wavelengths before it finally gets to the destination node on the receiving wavelength. In this section, we first give a general system description, and then discuss multicasting in single-hop and multihop networks, respectively. A summary of the surveyed contents is given in Fig. 1.

System Description

We show in Fig. 2 a broadcast-and-select WDM network consisting of N network nodes connected via optical fibers to a passive star coupler (PSC). Each node is equipped with a number of transmitters and receivers, and is connected to the PSC by a pair of fibers, one for transmitting and the other for receiving. The transceivers may be either fixed to a wavelength or tunable over a number of wavelengths. In general, there are four possible combinations of a node's transceiver equipment: fixed transmitter(s) and fixed receiver(s) (FT-FR); fixed transmitter(s) and tunable receiver(s) (FT-TR); tunable transmitter(s) and fixed receiver(s) (TT-FR); and tunable transmitter(s) and tunable receiver(s) (TT-TR). An example



■ **FIGURE 1.** *Multicasting issues and approaches in broadcast-and-select WDM networks surveyed in this article.*

of the transceiver equipment of a node is shown in Fig. 3 [1]. The node is equipped with a pair of fixed transceivers and a pair of tunable transceivers. The fixed transceivers are used to access a control channel λ_0, which is for the purpose of coordinating the transmissions of all the network nodes. The tunable transceivers are used to access the W data channels$\{\lambda_1, \ldots, \lambda_W\}$. This example can be denoted by CC-TT-TR, where CC stands for control channel. In later discussions, we assume a network of N nodes and W wavelengths (as data channels), if not otherwise stated.

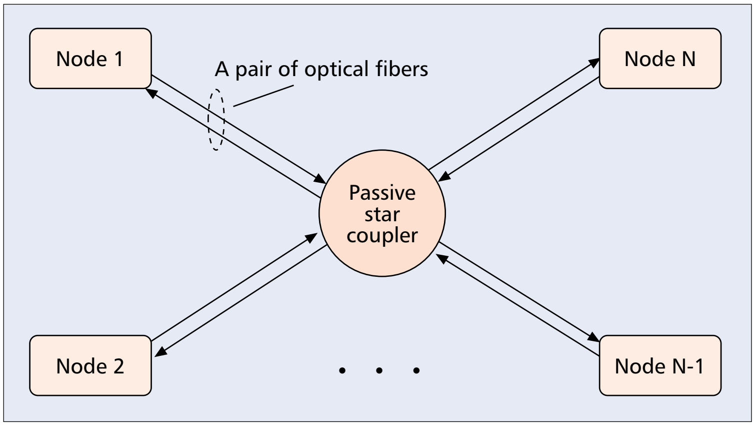
In a single-hop network, in order for each pair of nodes to communicate with each other, it is essential for a node to be equipped with at least one tunable transmitter or tunable receiver. Therefore, the possible node structures for single-hop networks are FT-TR, TT-FR, and TT-TR, with TT-TR being the most flexible node structure. A multihop network usually has a FT-FR node structure.

With the tunability of the transceivers, a single-hop network is more flexible than a multihop network. However, the system cost is relatively high because of the expensive tunable transceivers, and complex algorithms may have to be developed to coordinate the transmissions. Moreover, although all packets reach their destination in one hop, the tuning latency of the transceivers may unfavorably affect the system performance. On the other hand, a multihop network is less costly, while the delay of packets may be long since a transmission between two nodes may be possible only through multiple hops. Considering multicast, a single-hop network may be preferred since the FT-FR equipment of a multihop network may prevent a multicast packet from being transmitted to multiple destinations simultaneously.

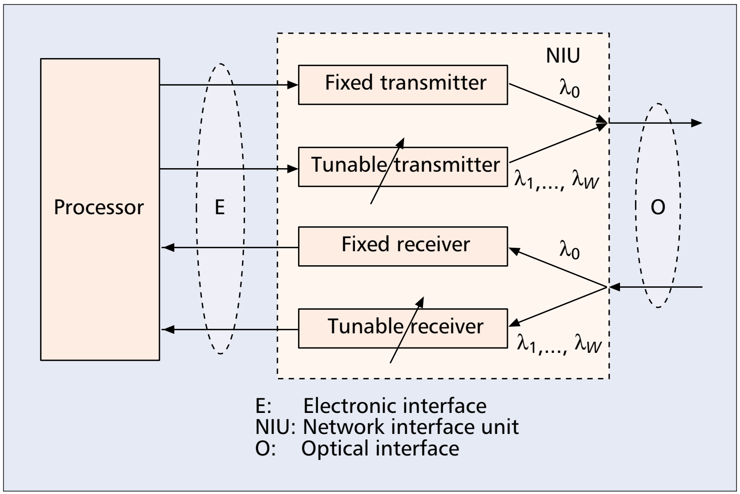
SINGLE-HOP WDM NETWORKS

In single-hop WDM networks, the major issue is the coordination (or scheduling) of the transmissions, because contentions may happen in such shared-media and shared-channel networks. One source of contention is so-called *collision*, when two or more transmitters want to transmit to the same wavelength channel at the same time. Another source of contention occurs when, in a system with tunable receivers, two or more transmitters want to transmit to the same destination node on different channels simultaneously. This situation is called a *destination conflict* [5].

The multicast scheduling algorithms (MSAs) can generally be classified as reservation-based (or pre-transmission-based) [6–14], random-access-based [15], and pre-allocation-based [16– 20]. In the reservation-based MSAs, each node sends a transmission request before it can actually transmit its data, and the transmission time is determined by the scheduling algorithm after its request is received. The random-access-based MSAs are proposed to reduce the complexity of the reservation-based algorithms, in which the nodes are coordinated to access the data channels in a random manner. The pre-allocation-based MSAs simply coordinate the transmissions according to some pre-determined schedule. In general, scheduling multicast transmissions is much more challenging than scheduling unicast transmissions, because the transmitter of the source node and the receivers of all the destination nodes in the multicast group need to be tuned to a common wavelength simultaneously. Hence, some researchers propose to partition a multicast transmission into multiple unicast and/or multicast transmissions [8]. In this way an earlier completion of a multicast transmission may be achieved at the cost of sacrificing the bandwidth efficiency of multicast. This introduces another classification of the MSAs, i.e., MSAs without partition [6, 12, 13, 16], and MSAs with partition [8–11, 14, 15, 17–20]. Among the latter some are actually hybrid schemes, where a multicast transmission is dynamically determined to be either partitioned or not partitioned depending on the average utilization of the data channels and the receivers [9], or the multicast session length and group size [19, 20].



■ **FIGURE 2.** *A broadcast-and-select WDM network based on passive star coupler.*



■ **FIGURE 3.** *A network node with a pair of fixed transceivers and a pair of tunable transceivers.*

Before we discuss the details of the MSAs, we first introduce some performance metrics of interest when designing an MSA. We summarize as follows.

- Transmitter throughput: Defined as the average number of packets transmitted by all the transmitters in the system per unit time (e.g., a time slot in a slotted system).
- Receiver throughput: Defined as the average number of packets received by all the receivers in the system per unit time. It is differentiated from the transmitter throughput to capture the fact that with multicast traffic the number of nodes receiving packets can be larger than the number of nodes transmitting [21].
- Multicast throughput: Defined as the average number of completions of multicast transmissions per unit time. This metric represents the actual throughput of the network for multicast traffic. When designing an MSA, we will want to maximize the multicast throughput.
- Average packet delay: Defined as the average amount of time from the arrival of a (multicast) packet into the system to the time when all the destination nodes in the group have received the packet. This metric shows how long it takes to make a multicast transmission. A small average packet delay is desirable.
- Average receiver waiting time: Defined as the average amount of time a receiver must wait before it begins to receive a packet. The waiting time is measured either from the point at which the receiver becomes available or from the instant at which the packet arrives in the system, whichever is later. This metric reflects the utilization of the receivers. A small average receiver waiting time is desirable.
- Average number of transmissions per multicast packet: Defined as the average number of transmissions needed to deliver a multicast packet to all its destinations. It reflects how much bandwidth efficiency of multicast has been sacrificed. When designing an MSA, we will want to minimize this metric.

Some of these metrics may conflict with each other. For example, for MSAs without partition, the average number of transmissions per packet is the smallest (equal to 1), but the throughput (both transmitter and receiver) may be low and the average packet delay and average receiver waiting time may be long, because of the difficulty in finding a time slot when all the receivers are available on a free wavelength channel.

We now discuss some representative MSAs.

Reservation-Based MSAs — In order for the network nodes to send their transmission requests, a shared control channel is usually used in single-hop WDM networks employing reservation-based MSAs. In the following discussion, we assume a CC-TT-TR system model, as shown in Fig. 2 and Fig. 3. All the nodes are assumed to be equal-distant from the PSC, which makes the propagation delay between all node pairs identical (denoted by R). This assumption can be realized for LANs and MANs by extending some fiber lengths or adding appropriate optical delays. The system operates as follows. When a packet arrives at a node, it is placed in that node's arrival queue. A control packet, which contains the information of the source node, destination node(s), and maybe the packet transmission time (in terms of time slots, for example, for a slotted data channel) if data packets can have variable sizes, will then be sent on the control channel to all the nodes. The control channel is slotted, and network nodes access it via round-robin time division multiple access (TDMA). In particular, the control channel is composed of control frames, each of which is subdivided into N control slots numbered from 1 to N. Node i can transmit its control packet only in the ith slot of each control frame. After the control packet is transmitted, the corresponding data packet is moved to a waiting queue of the node until it is transmitted to all its destination nodes. After a certain propagation delay, the control packet will be received by all the nodes, which then run the same MSA to schedule the transmission of the packet (i.e., to reserve the data channel(s) and time slot(s) for this packet). After the reservation, each node updates its record of the system state and gets ready for a new scheduling. What differentiates all the reservation-based schemes is the MSA run by all the nodes after they receive a control packet. In the following sections we review different reservation-based MSAs.

Non-Partition Scheme — A reservation-based MSA without partitioning the multicast transmissions is proposed in [6].[1] The MSA first examines the tunable receivers of all the nodes in the multicast group to determine the earliest time at which the multicast packet can be received simultaneously by all of them, denoted by *earliest_rec_time*. This suggests that the transmitting time *trans_time* of the packet cannot be earlier than *earliest_rec_time-R*, where R is the propagation delay. On the other hand, the MSA also checks the transmitter and the channels to determine the earliest time that the packet can be transmitted, denoted by *earliest_trans_time*. Then the transmitting time can simply be determined as *trans_time* = max(*earliest_trans_time*, *earliest_rec_time-R*). Since the packet size is assumed to be fixed, the MSA knows the amount of time the transmission takes, or equivalently the free time of the reserved channel and the receivers, and therefore can update the system state accordingly.

An interesting performance characteristic of this algorithm is that the achieved receiver throughput may have a local maximum with respect to the multicast group size. This is because scheduling a single transmission for each multicast packet may result in an inefficient use of the receiver resources. In particular, when the multicast group size is large, some receivers may have to wait for a long time without

[1] The system model in the article is not exactly a CC-TT-TR system. Each node is assumed to be equipped with multiple tunable transmitters and tunable receivers instead of one. However, the basic idea is the same as what we describe in the following.

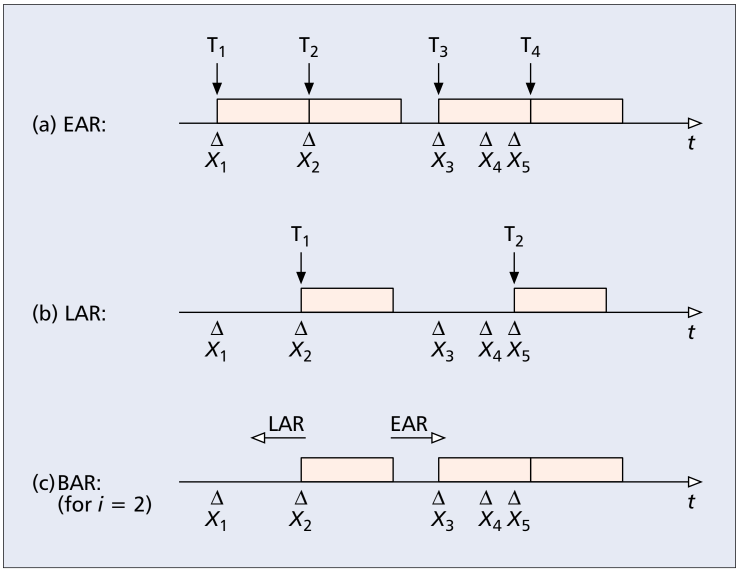


FIGURE 4. *Partition schemes for a multicast transmission.*

receiving anything because some other receivers in the same group are not available. An analytical study of this problem is given in [7].

Partition Schemes — Given that the non-partition scheme may waste the receiver resources, partition schemes are proposed to address this problem in [8]. Specifically, a multicast group is partitioned into subgroups and a separate transmission is scheduled for each subgroup, in order to minimize the average receiver waiting time. Three scheduling algorithms are developed: earliest available receiver (EAR), latest available receiver (LAR), and best available receiver (BAR). For convenience, let L denote the length of each multicast packet, M denote the number of destinations in (i.e., the size of) a multicast group, and X_i, $i = 1, 2, \ldots, M$ denote the time when the receiver at destination node i becomes available, where $X_1 \leq X_2 \leq, \ldots, \leq X_M$.

- EAR: Schedules the first transmission to the earliest available receiver. If any of the remaining receivers become available during the first transmission, the next transmission is scheduled immediately after the first one. Otherwise, the next transmission is scheduled whenever the next receiver becomes available.
- LAR: Begins by scheduling a transmission at the time when the latest receiver becomes available, i.e., X_M. Receiver $M - 1$ is considered next. If a transmission to this receiver would conflict with the previously scheduled transmission, then receiver $M - 1$ is placed in the same group as receiver M. Otherwise, a separate transmission is scheduled at time X_{M-1}. This process proceeds backward through all the receivers in the group until all of them have been scheduled.
- BAR: First schedules for each receiver as follows. For receiver i, a transmission is scheduled at time X_i. EAR is then used to schedule the transmissions after $X_i + L$, and LAR is used to schedule the transmissions before $X_i - L$. After all the receivers have been scheduled, BAR chooses among the M schedules the one having the minimum receiver waiting time.

The three algorithms are illustrated in Fig. 4. Among the three partition schemes, BAR yields the smallest average receiver waiting time under low loads, while for high loads LAR yields the smallest. Something to be noted is that LAR and BAR can always achieve a smaller average receiver waiting time compared with a non-partition scheme, while EAR may even result in a larger average receiver waiting time under high loads. Their performance in terms of the average packet delay shows similar trends. Regarding the average number of transmissions per multicast packet, among the three schemes LAR schedules the fewest transmissions while EAR schedules the most.

Hybrid-Partition Scheme — As shown above, an MSA that tries to partition a multicast transmission into multiple unicast or multicast transmissions (e.g., EAR) may not always produce smaller average packet delay than an MSA that does not partition multicast transmissions. This problem is studied in [9]. A greedy algorithm, which always tries to partition a multicast transmission into multiple unicast or multicast transmissions and schedules as many destination nodes as possible in the earliest data slot (basically the EAR), is compared with the non-partition algorithm under different traffic and channel conditions. It is shown that partitioning a multicast transmission requires more channel resources and may result in a non-optimal use of the data channels. In particular, a packet may have to be scheduled in a later data slot (or even partitioned into multiple slots) although both the transmitter of the source node and the receivers of the destination nodes are available at the mean time, simply because there is no available data channel in the slot due to a previously scheduled partial transmission. The following important observations can be made:

- When there are a sufficient number of available data channels, the receivers should be utilized as much as possible. In other words, the greedy algorithm (i.e., partition) should be used in such cases.
- When the channel resources become a potential bottleneck, the data channels should be used conservatively. In other words, the non-partition algorithm should be used in such cases.
- The utilizations of the data channels and the receivers are the key factors that determine the performance of a scheduling algorithm.

The authors of [9] propose a hybrid scheduling algorithm that dynamically chooses between the greedy and non-partition algorithms, depending on the utilization of the channel and receiver resources. Simulation results show that the hybrid algorithm produces the smallest average packet delay compared with the greedy and non-partition algorithms. As we have said that the greedy algorithm is basically the EAR algorithm, how the hybrid algorithm performs compared with the LAR and BAR algorithms is not addressed in this article.

Minimizing Partitions — Since partitioning a multicast transmission into multiple unicast or multicast transmissions sacrifices the bandwidth efficiency of multicast, a small number of partitions is desirable while some other performance requirements are satisfied. The problem of minimizing the number of transmissions for a multicast transmission under the condition that the packet delay is minimum is studied in [10]. Note that the problem is for the scheduling of a single multicast packet, and the number of transmissions and the packet delay stated here are not in terms of the overall average. The problem is proved to be NP-hard, and a heuristic *maximum-destination* scheduling algorithm is proposed. The algorithm works as follows.

When a transmission request is received, the algorithm finds the earliest data slot in which the last transmission of the multicast packet can be scheduled, by checking the available data slots of the transmitter and the receivers. This last slot

<table>
<tr><th>Performance metric</th><th>Under low loads</th><th>Under high loads</th></tr>
<tr><td>Average receiver waiting time</td><td>BAR &lt; LAR < EAR < non-partition</td><td>LAR &lt; BAR < non-partition &lt; EAR</td></tr>
<tr><td rowspan="3">Average packet delay</td><td>BAR &lt; LAR < EAR < non-partition</td><td>LAR &lt; BAR < non-partition &lt; EAR</td></tr>
<tr><td colspan="2">Hybrid-partition < EAR, non-partition</td></tr>
<tr><td colspan="2">Maximum-destination < EAR, non-partition</td></tr>
<tr><td rowspan="2">Average number of transmissions per multicast packet</td><td colspan="2">Non-partition (=1) < LAR < BAR &lt; EAR</td></tr>
<tr><td colspan="2">Maximum-destination &lt; EAR</td></tr>
</table>

Table 2. *Performance comparison of the reservation-based multicast scheduling algorithms.*

determines the minimum delay of the packet. The problem is then to find a partition of the destination nodes among the available data slots from the current time to the last slot just found. The algorithm first schedules as many destination nodes as possible in the last slot, and then repeatedly schedules the destination nodes in the data slot in which the number of the available destination nodes is the maximum, until all the destination nodes are scheduled. Compared with the greedy (EAR) and non-partition algorithm, the maximum-destination algorithm always produces a smaller average packet delay. It also achieves a smaller average number of transmissions per packet than the greedy algorithm.

A performance comparison of the above reservation-based MSAs is summarized in Table 2.

Random-Access-Based MSAs — Random-access-based MSAs are motivated by the demand for simple scheduling algorithms in the face of the huge number of runs per unit time of the scheduling algorithm in high bit-rate WDM networks. For example, in a single-hop network of N nodes operating at 10 Gb/s per WDM channel, assuming each packet has a fixed size of 1K bits, one time slot is then 0.1 μs (ignoring the tuning latency). If the packet arrival rate at each node is 0.1 per slot, the overall packet arrival rate to the system is then $N \times 10^6$ per second. If a reservation-based MSA is used, each node will run the MSA $N \times 10^6$ times per second. Obviously, the simpler the MSA the better.

Some random-access-based MSAs are proposed in [14]. The system employs a centralized scheduler that receives the transmission requests from the nodes on a control wavelength λ_C, then runs the MSA and informs the nodes of their transmission schedules on a separate control wavelength $\lambda_{C'}$. The nodes may send their requests to the control channel λ_C at any time according to some unslotted random access protocol. However, the centralized scheduler operates in a slotted fashion. It maintains a request queue for each node, and checks the request queues and makes appropriate scheduling in each slot. While the details of the system operation are presented in [22], [15] focuses on the MSA run in each slot. Two transmitter schemes and three receiver schemes are proposed.

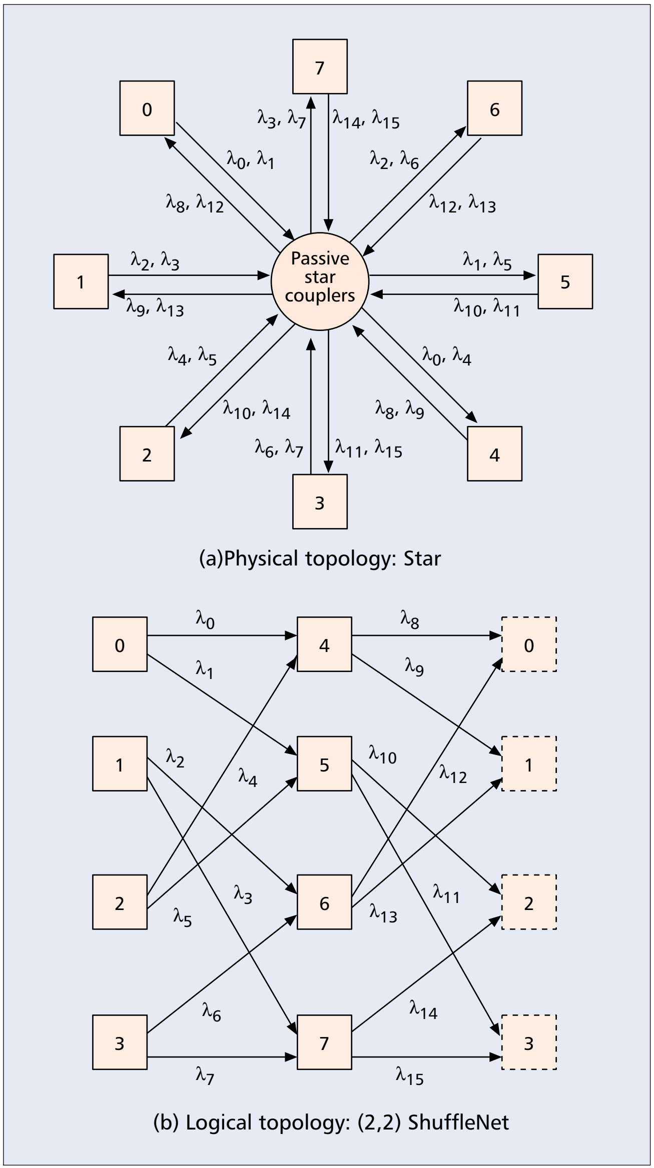
Transmitter Schemes — In the current scheduling slot, as long as there is a wavelength channel available a packet is chosen for transmission randomly from among the available nodes that have a new message to send (i.e., the nodes whose request queues are not empty). Since all the receivers of the destination nodes may not be available on this wavelength at the scheduled time, the packet may have to be transmitted multiple times. One scheme is to repeatedly transmit the packet on this wavelength until it has been received by all of its destination nodes. This wavelength channel will then be continuously occupied by this packet until the completion of its transmission. This scheme is called *persistent retransmission*. A problem with this scheme is a form of head-of-line (HOL) blocking due to the continuous occupation of the channel by a packet. To solve this problem, another *backoff retransmission* scheme is proposed, in which a random delay is introduced between retransmissions of the same packet.

Receiver Schemes — Since a node may have more than one packet addressed to it in a slot, it must choose only one of them to receive. Three schemes are proposed for making the choice. The first scheme is a *random* scheme, in which the receiver chooses one packet at random from among the packets addressed to it. The second scheme is a first-come-first-served (FCFS) scheme, in which the receiver selects the packet based on the time it was first transmitted in a FCFS order. If two or more packets were transmitted in the same slot, then the receiver chooses among them at random. The third scheme is a *priority* scheme, in which the receiver selects the packet with the smallest number of (remaining) intended recipients. The intuition behind this scheme is that, by selecting the message with the smallest number of intended recipients, the probability that a message will be released is maximized, thereby making way for the transmission of a new packet.

With two transmitter schemes and three receiver schemes, there are six possible combinations. Four of them, namely, Persistent-Random, Persistent-FCFS, Backoff-Random, and Backoff-Priority, are compared in terms of the multicast throughput. The results show that Persistent-Random yields the worst performance, Persistent-FCFS offers a small improvement, Backoff-Random can achieve further improvement, and Backoff-Priority is the best.

Pre-allocation-Based MSAs — As we have said, simple MSAs are desirable. The simplest MSAs may be those that pre-determine for each slot the active transmitters, the corresponding transmitting wavelengths, the active receivers, and the corresponding receiving wavelengths, i.e., the pre-allocation-based MSAs. In systems employing the pre-allocation-based MSAs, the overhead of the control messages can be avoided. However, a problem with pre-allocation-based MSAs is that they are static and may not always perform well for each single multicast transmission. The design problem of a pre-allocation-based MSA is usually focused on the long-term traffic demand and can be stated as follows: given the long-term traffic pattern in the network, determine a transmitting (and receiving) schedule, such that some performance metric (e.g., the multicast throughput) is optimized.

Since the schedule is pre-determined, systems employing such pre-allocation-based MSAs may not need the flexibility of both a tunable transmitter and a tunable receiver at each node. For example, we can use a FT-TR system, in which each transmitter is fixed at a certain wavelength and is allocated some data slots to transmit, while the receivers are tunable and are tuned to different wavelengths in different data slots such that communications between all node pairs are possible. TT-FR is not employed for two reasons. First, it is believed that TT-FR is less favorable than FT-TR considering the eco-



(a)Physical topology: Star

(b) Logical topology: (2,2) ShuffleNet

■ FIGURE 5. *The physical and logical topology of a multihop WDM network.*

nomics and performances of the devices [23]. More importantly, supporting multicast is less efficient in a TT-FR system than in a FT-TR system, since in a TT-FR system the nodes in a multicast group may have receivers fixed-tuned on different wavelengths, meaning that the source node has to send a packet multiple times each time on a different wavelength. In the following we discuss some pre-allocation-based MSAs for FT-TR systems.

Slot-Classified Schemes — The MSAs proposed in [16] are based on classifying (i.e., pre-allocating) the time slots into unicast slots, broadcast slots, and multicast slots. In a unicast slot, exactly *W* nodes are given permission to transmit, each on a different wavelength and to a different receiver. In a broadcast slot, only one node is allowed to transmit, which is called the owner of the slot, and all receivers have to tune to its transmitting wavelength. In a multicast slot, a number *m*, $1 \leq m \leq W$, of nodes can transmit, each on a different wavelength. One of the nodes, say *k*, may transmit to the receivers of a multicast group *g*, and is called the owner of the slot, while the other $m - 1$ nodes may transmit to exactly one receiver that is not a member of *g*. The following three MSAs, classified according to the types of slots the schedule consists of, are proposed for a mixture of unicast and multicast traffic:

- Unicast Slots Only — This MSA consists of only unicast slots, which means a multicast packet has to be transmitted to each of the destinations individually.
- Unicast and Broadcast Slots — This MSA consists of unicast slots and broadcast slots, where the former are used to transmit unicast traffic and the latter are used to transmit multicast traffic.
- Unicast and Multicast Slots — This MSA consists of unicast slots and multicast slots, where the former are used to transmit unicast traffic and the latter are mainly used to transmit multicast traffic while also available for unicast traffic as long as no conflict happens.

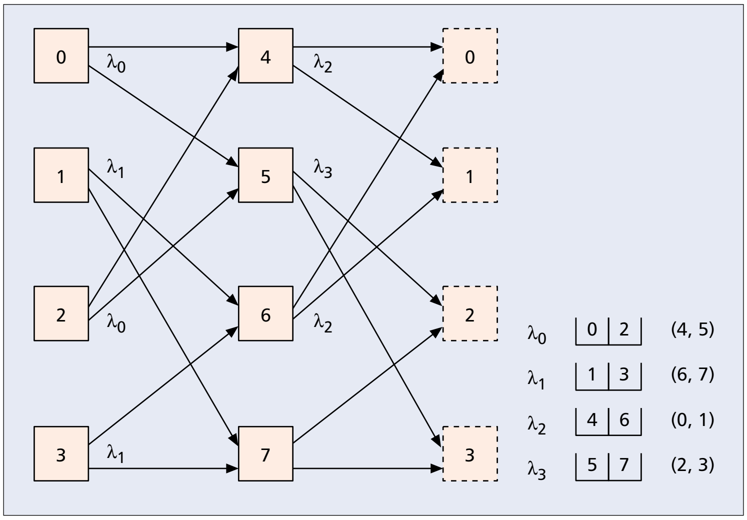
The three MSAs have almost identical performance in terms of receiver throughput. However, they perform differently in terms of the average packet delay. When the MSA with unicast slots only is used, the packet delay increases drastically as the multicast group size increases, because a multicast packet has to be transmitted multiple times. When the MSA with unicast and broadcast slots is used, the packet delay is independent of the multicast group size, because the multicast packets are transmitted in broadcast slots and received by all destinations at once. When the MSA with unicast and multicast slots is used, the packet delay is always close to that of the best static schedule. Basically, the three MSAs are suitable for different traffic patterns. Specifically:

- The MSA with unicast slots only is suitable for multicast traffic that has relatively short session length and few group members.
- The MSA with unicast and broadcast slots is suitable for multicast traffic that is also relatively short but has a large number of group members.
- The MSA with unicast and multicast slots is suitable for multicast traffic that has relatively long session length.

Virtual Receiver Schemes — A *virtual receiver* is defined as a set of physical receivers that behave identically in terms of tuning [17]. Accordingly, all the physical receivers in a network can be partitioned into a number of virtual receivers. The motivation for this partition is twofold. First, the previously-studied (reservation-based) partition schemes that partition a multicast group into subgroups consider each multicast packet independently of others. In contrast, partitioning the receivers can take the overall traffic offered to the network into account. Therefore, better performance (in terms of multicast throughput) may be achieved. Second, by partitioning all the physical receivers into virtual receivers, the original network with multicast traffic can be transformed into a network with unicast traffic. The multicast schedule can therefore be determined by taking advantage of the unicast scheduling algorithm developed in [24], which has proven optimal properties.

Having the unicast scheduling algorithm, the problem is then reduced to determine a partition of the physical receivers (i.e., a virtual receiver set) such that for a given multicast traffic demand matrix the multicast throughput is maximized. The problem is proved to be NP-hard, and some heuristics have been developed for the problem. It is shown that the heuristics can yield near-optimal performance in terms of multicast throughput. A Tabu-Search-based improvement to the above virtual receiver MSA can be found in [18].

Combinational Scheme — An MSA that combines a pre-allocated unicast schedule and a multicast slot reservation (MSR) algorithm is proposed in [19, 20]. The network is a CC-FT-TR system, which is different from all the systems we



λ_0	0	2	(4, 5)
λ_1	1	3	(6, 7)
λ_2	4	6	(0, 1)
λ_3	5	7	(2, 3)

FIGURE 6. *Channel sharing: a (2,2) ShuffleNet with four wavelengths and its TDM channel access assignment.*

have seen so far (i.e., CC-TT-TR systems employing reservation-based or random-access-based MSAs and FT-TR systems employing pre-allocation- based MSAs). The basic idea of the MSA is to first determine the transmitter schedule of the source node upon receiving a transmission request, and then obtain the receiver schedule of the nodes in the multicast group by modifying the pre-allocated unicast schedule. This MSA is very much like and may actually be classified as a reservation-based MSA, since multicast transmissions are scheduled upon requests, although by modifying a pre-allocated unicast schedule.

Another characteristic of the scheme is that a multicast packet may be transmitted by multiple unicasts, depending on a distance metric M that is defined for each multicast packet as

$$M = \sqrt{S^2 + |G|^2},$$

where S is the multicast session length (in terms of the number of time slots) and $|G|$ is the size of the multicast group G. A network-wide constant M_d is defined to classify the multicast traffic. If $M \leq M_d$, the multicast packet will be transmitted by multiple unicasts. If $M > M_d$, the packet will be transmitted by a single multicast transmission. It is shown that if M_d is properly chosen, the scheme can result in better performance tradeoffs between unicast traffic and multicast traffic (in terms of network throughput and packet delay) than either always multicasting ($M_d = 0$) or always unicasting ($M_d = \infty$, or a large number). The problem of how to choose M_d is not addressed in [16, 17].

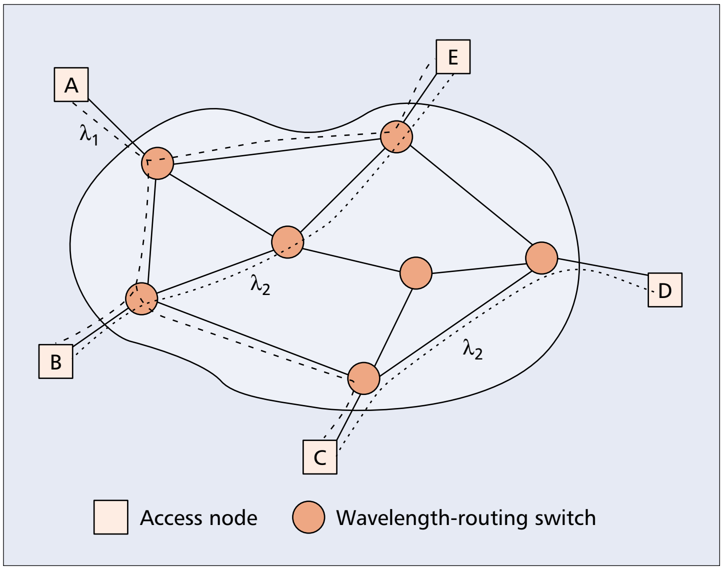
MULTIHOP WDM NETWORKS

Multihop WDM networks are generally based on the FT-FR node structure. This means that the connectivity between nodes in a multihop network is fixed, resulting in a fixed logical (virtual) topology. The major issues are logical topology design, multicast routing, and multicast efficiency. Each of these is discussed separately in the following paragraphs.

The logical topology of a multihop network can be either irregular or regular. The design of an irregular logical topology can usually be stated as follows: given a (long-term) traffic matrix, determine the transmitting and receiving wavelengths and hence the logical connections of the nodes, such that some performance metric is optimized. For example, the objective can be to minimize the average packet delay or to minimize the maximum link flow. While there are several pieces of work addressing this problem with unicast traffic as summarized in [1], we do not find any work addressing the problem with multicast traffic. Basically, the irregular multihop logical topology design does not seem to have raised much research interest. This may be because the routing complexity in irregular topologies can be high, while on the other hand there exists a number of well-studied regular topologies with simple routing schemes, such as the ShuffleNet, Manhattan Street Network (MSN), and Hypercube [4]. We show in Fig. 5 a multihop WDM network with (2,2) ShuffleNet logical topology built from a star physical topology with eight network nodes ($N = 8$) and 16 wavelengths ($W = 16$).

Given a logical topology, the routing problem then has to be solved. The routing in regular multihop networks is usually simple because of the regular connectivity pattern of such networks. The routing algorithms of a number of regular topologies can be found in [25]. For multicast, a multicast tree that is rooted at the source node and spans all the destination nodes has to be built. In general, there can be two objectives: one is to build a minimum-delay multicast tree in which the delay (distance) from the source node to each destination node is minimized, while the other is to build a minimum-cost multicast tree in which the sum of all edge costs is minimized. A minimum-delay multicast tree is basically a shortest path tree, which can be built based on the (unicast) shortest path routing algorithms in [25]. The problem of building a minimum-cost multicast tree is well known as the Steiner Tree problem [26], which is NP-hard for general networks. An algorithm that modifies the minimum-delay tree into a tree with lower cost but near-minimum delay for ShuffleNet is proposed in [27].

Knowing the multicast tree, each node can then determine whether a packet it receives should be "absorbed" or/and forwarded to some other outgoing links. However, forwarding a multicast packet in a multihop network may be inefficient. For example, to transmit a multicast packet from node 0 to a set of destination nodes {2, 3, 4, 5} in the network shown in Fig. 5, it requires four transmissions on four different wavelengths, which is essentially multiple unicasts. An improvement on this situation can be achieved by *channel sharing*, where a number $W < N$ of wavelength channels are used in the system and each channel is shared by one or more nodes in a TDM fashion. We show in Fig. 6 the same (2,2) ShuffleNet realized with only four wavelengths. In this system, each node is equipped with only one transmitter and one receiver, resulting in a total of 16 transceivers, as opposed to the system in Fig. 5 where a total of 32 transceivers are used. The channel sharing scheme is also shown in the figure. Each TDM frame consists of two time slots. The numbers in the slots indicate the transmitting nodes and the numbers in the parentheses indicate the nodes that can receive the packets on this channel. We can see that only two transmissions are needed to transmit the aforementioned multicast packet. Channel sharing was originally proposed to reduce the number of required wavelengths and the system cost, while the work in [28] shows that it is inherently effective in supporting multicast (as also shown by our simple example). In particular, an analytic model has been developed for the analysis of multicast traffic in shared-channel multihop WDM networks. The average packet delay in such networks has been compared with a number of other systems, including systems with dedicated channels, ring networks, and classical TDM systems. The results show that having a small number of channels (equivalently, channel sharing) is not only a technol-



■ **FIGURE 7.** *A wavelength-routed WDM network.*

ogy requirement, but may actually be desirable from a system performance perspective with the presence of multicast traffic.

Overall, there is little literature on multicasting in multihop WDM networks. This might be because multihopping actually wastes the broadcast (and hence multicast) capability of the PSC in a sense. However, supporting multicast in multihop WDM networks is indeed applicable.

MULTICASTING IN WAVELENGTH–ROUTED WDM NETWORKS

The two basic problems of multicasting in wavelength-routed networks are the routing problem and the wavelength assignment problem (abbreviated as MC-RWA). In this section, we first give a system introduction of the wavelength-routed networks, with an emphasis on the multicast capability of the wavelength-routing switches that is crucial in solving the MC-RWA problem. We next review different forms of the MC-RWA problem and various approaches to them in general mesh networks. After that, we give a separate discussion on MC-RWA in WDM ring networks. Finally, we summarize other research on MC-RWA for completeness.

SYSTEM DESCRIPTION

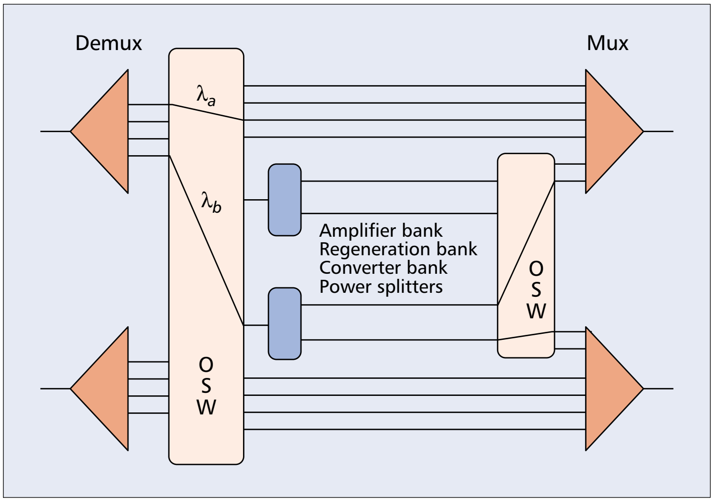
An example of a wavelength-routed network is shown in Fig. 7. The access nodes are where end users reside, and are equipped with a set of transmitters and receivers (which may be tunable). The wavelength-routing switches are responsible for routing any incoming light signal to its intended outgoing link(s). In the absence of wavelength converters, a lightpath (light-tree in the case of multicast) is required to be on the same wavelength throughout its path in the network. This requirement is the so-called *wavelength continuity constraint*. In Fig. 7, a light-tree with node *A* being the source and nodes *B*, *C*, and *E* being the destinations is set up on λ_1, and two lightpaths connecting node *B* with node *E* and node *C* with node *D* are set up on λ_2.

To support multicast in wavelength-routed networks, the wavelength-routing switches should be *multicast-capable*. By multicast-capable, we mean the optical signal from each incoming link is able to be replicated and forwarded to all the outgoing links all-optically. This can usually be achieved by using an optical splitter. A 2 × 2 multicast-capable wavelength-routing switch supporting four wavelengths on each link is shown in Fig. 8 [29]. The light from each incoming link is first demultiplexed (DEMUX) into separate wavelengths. The separate signals are then switched by an optical switch (OSW). Unicast signals are sent directly to OSW ports corresponding to their outgoing links, while multicast signals are sent to an OSW port connected to a splitter bank. (The splitter bank may be enhanced to provide optical signal amplification, wavelength conversion, and signal regeneration.) The splitter equally splits the optical signal into *n* parts, where *n* is the number of outgoing (as well as incoming) ports. Following the splitter bank is another OSW, which routes the replicated optical signals to the intended outgoing links of the multicast signal (some of the split signals may be blocked). In the figure, wavelength λ_a carries a unicast signal and λ_b carries a multicast signal.

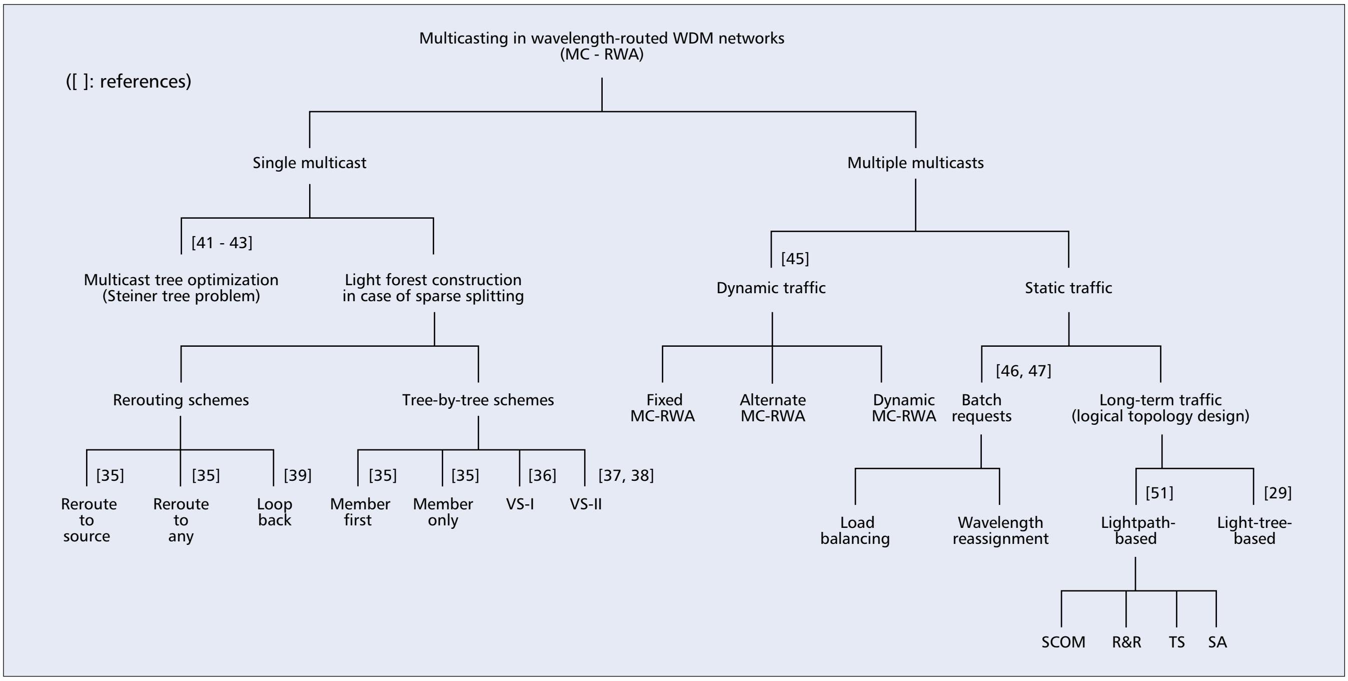
The light-splitting capability of the wavelength-routing switches is an important factor in supporting multicast in wavelength-routed WDM networks. A number of issues related to it need to be addressed, and two of them are summarized in the following.

Power Considerations: After an optical signal passes an *n*-way splitter, the power of the signals at each output port is only 1/*n* of the input signal. In order for a multicast signal to be eventually detected at the end users, either the power of the transmitting signal must be high enough or optical amplifiers need to be employed. This power consideration affects the design of multicast wavelength-routed networks [30, 31], as well as the construction of multicast trees [32].

Sparse Splitting: In a real network, all wavelength-routing switches may not have light-splitting capability. This situation is called *sparse splitting* [33]. In addition to the practical reasons such as the gradual evolution of the networks and economic considerations, sparse splitting is also justifiable from the performance perspective. It has been shown in [33] that in general only a fraction (e.g., 50 percent) of the switches need to be equipped with the splitting capability to obtain almost the same benefit of having the splitting capability at all the switches, and the allocation of splitting-capable switches is studied in [34]. In a sparse-splitting network, the generic algorithms (heuristics) for building a multicast tree cannot be applied directly, since the multicast tree built by them may not be supported by the physical network. Much research has



■ **FIGURE 8.** *A multicast-capable wavelength-routing switch.*



■ FIGURE 9. *Multitasking issues and approaches in wavelength-routed WDM networks surveyed in this article .*

been devoted to finding feasible routes to support multicast transmissions in a sparse-splitting network [32, 35–40].

In addition to the light-splitting capability, the wavelength-conversion capability is another important factor in supporting multicast in wavelength-routed WDM networks. With such capability, the wavelength continuity constraint can be relaxed, and setting up a multicast connection may be more easily achieved. Similarly, in general not all the wavelength-routing switches may have such capability, which is called *sparse wavelength conversion*. It has also been shown that not all the wavelength-routing switches need the wavelength-conversion capability; instead, in general only a fraction (e.g., 50 percent) of the switches need to be equipped with such capability to obtain almost the same benefit [33].

MULTICAST ROUTING AND WAVELENGTH ASSIGNMENT (MC-RWA)

The MC-RWA problem takes different forms for different scenarios. We first focus on a single multicast request. In this case, the core of the problem is to find a set of routes from the source to all the destinations that can be supported by the physical network to realize the transmission. If all the nodes in the network are multicast-capable and there is a number of multicast trees to choose from, the optimal one (in terms of tree cost, for example) should be chosen. We denote this kind of problem as *multicast tree optimization*. If the network is sparse-splitting, the multicast trees built by the generic algorithms may not be realizable, hence we have to find alternative routes to support the multicast transmission in such cases. Note that a single light-tree may not be found, and the set of alternative routes may be multiple subtrees on different wavelengths with each spanning a subset of the destination nodes, i.e., the so-called *light-forest*. We denote this kind of problem as *light-forest construction in the case of sparse splitting*.

We next look at the case where multiple multicast requests are to be accommodated. For a realistic scenario where multiple multicast requests arrive and leave dynamically, the main problem is to find routes and wavelengths for each arriving request, such that the blocking probability is minimized (for a system with a limited number of wavelengths). We denote this kind of problem as *MC-RWA for dynamic traffic*. As another scenario, we may be given multiple static multicast requests: either a batch of multicast requests that need to be supported simultaneously at the meantime or a long-term multicast traffic matrix. In the former case, we will want to maximize the number of multicast requests that can be supported for a limited number of wavelengths, or to minimize the number of wavelengths needed to accommodate all the requests. In the latter case, we are faced with the problem of designing an optimal (in terms of average hop distance, for example) logical topology for the given traffic pattern. We simply denote these two cases as *MC-RWA for static traffic*.

In the following, we discuss these four forms of problems respectively. A summary of approaches to these problems is given in Fig. 9.

Multicast Tree Optimization — The multicast tree optimization problem can usually be formalized as a Steiner Tree problem [26], which is stated as follows.

Given:

- a graph $G = (V, E)$, where V is the set of nodes and E is the set of links,
- a cost function that assigns a cost $c(e)$ to each link $e \in E$,
- a set of nodes $D \subseteq V$ that belong to the multicast group,

find a tree $T = (V_T, E_T)$ that spans D, such that its cost $C_T = \Sigma_{e \in E_T} c(e)$ is minimized. The Steiner Tree problem is NP-hard. A number of heuristics, such as the shortest path tree (SPT) heuristic and the minimum spanning tree (MST) heuristic, can be found in [26].

Note that the generic Steiner Tree formulation only addresses the tree-building (i.e., routing) problem. On which wavelength(s) the multicast tree can be supported (i.e., the wavelength assignment problem) is not considered. Without wavelength conversion, a single wavelength must be assigned to the multicast tree. If no wavelengths can be found available on all the links of the multicast tree, this multicast request will be blocked. If wavelength conversion can be employed, the multicast tree may be supported by multiple wavelengths, and an optimal wavelength assignment may be found such that the total wavelength conversion cost is minimized. The optimal wavelength assignment problem for a given multicast tree is

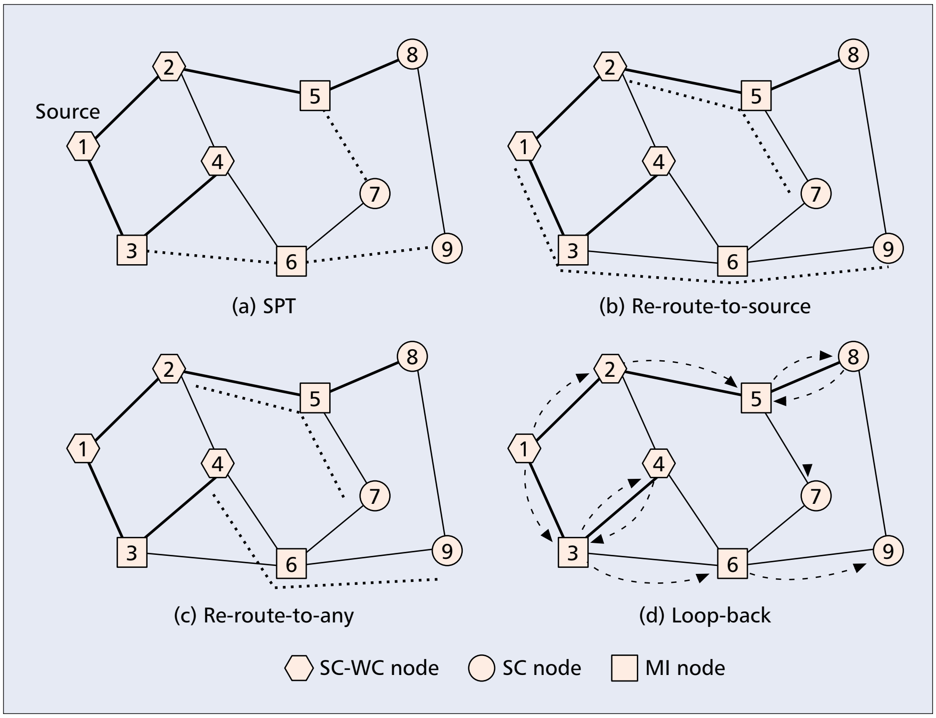


FIGURE 10. *Re-routing schemes in a wavelength-routed network composed of nodes with different multicast capabilities.*

proved not NP-hard and a polynomial-time algorithm is presented in [41].

The routing problem and wavelength assignment problem can also be combined into a single problem by generalizing the cost model of the generic Steiner Tree formulation [42]. In particular, a cost $c(e, \lambda_i)$ is defined for using wavelength λ_i on link e, and a cost $c_v(\lambda_i, \lambda_j)$ is defined for the wavelength conversion at node v from λ_i to λ_j. In this way, the total cost of the multicast tree includes both the wavelength-usage cost and wavelength-conversion cost, and solving the optimization problem solves the routing problem and wavelength assignment problem simultaneously. A further generalization is made by taking the queuing delay at the nodes into consideration [43]. All these generalized formulations can be reduced to the Steiner Tree problem on some auxiliary graph, and hence can be solved using the aforementioned heuristics. Since the Steiner Tree problem has been well studied, we will not go through the details any further.

Light-forest Construction in the Case of Sparse Splitting — There are in general two methods to construct a workable light-forest (with light-tree as a special case) for a multicast request in a sparse-splitting network: either modifying the multicast tree built by some generic algorithm by rerouting the unsupported paths, or developing some new algorithm to find the routes from scratch. In this process, the multicast capability of the nodes is the key factor. Let $M(v)$ be the multicast capability of node v in terms of the number of copies of an incoming signal that v can forward to other nodes. Assume the nodal degree of v is $N_d(v) + 1$ and the number of wavelengths on each link is W. We first classify the network nodes according to their multicast capability as follows.

Multicast-Incapable (MI) nodes:

- Drop-and-Continue (DaC) nodes: For these nodes, $M(v) = 1$. Such nodes can forward an incoming signal to only one outgoing link, in addition to receiving a copy of the signal locally.
- Drop-and-Continue with Wavelength Conversion (DaC-WC) nodes: For these nodes, $M(v) = 1$. Such nodes can forward an incoming signal to only one outgoing link, maybe on a different wavelength, in addition to receiving a copy of the signal locally.

Multicast-Capable (MC) nodes:

- Splitting-Capable (SC) nodes: For these nodes, $M(v) = N_d(v)$. Such nodes can forward an incoming signal to all the outgoing links (as well as receive the signal locally).
- Splitting-Capable with Wavelength Conversion (SC-WC) nodes: For these nodes, $M(v) = N_d(v) \times W$. Such nodes can forward an incoming signal to all outgoing links on any wavelength, including sending multiple copies of the signal to the same outgoing link on different wavelengths.

In general, all nodes are assumed to have at least the DaC capability. DaC-WC nodes do not affect the tree building, while it may affect the wavelength assignment. One thing to be noted is the assumption that a SC-WC node can send multiple copies of the signal to the same outgoing link on different wavelengths. The source node of a multicast is always assumed to be a SC-WC node in this part of the discussion, although such capability may be realized by using an array of transmitters tuned on different wavelengths instead of by wavelength conversion. A SC-WC intermediate node enables itself to be a root of rerouted subtrees, and avoids rerouting all the way back to the source, as will be made clear later.

When building a light-forest for a multicast request, we would like the forest to be cost effective, since a trivial but costly solution that sets up a lightpath from the source to each of the destinations can always be found as long as the group size is not larger than the multicast capability of the source $M(s)$. Specifically, three cost variables are of interest:

- The number of wavelengths: Represents the amount of resources needed. When multiple subtrees in the forest overlap on a link, a different wavelength has to be used for each subtree. Therefore, the number of wavelengths needed by the light-forest is simply the maximum number of overlaps of the links.
- The total number of branches: Represents the bandwidth consumed. Since each branch means that a wavelength channel is occupied on the corresponding link, the total number of branches is simply the total number of wavelength channels the light-forest uses.
- The average number of hops from the source to a destination: Represents the delay performance of the light-forest.

We now discuss the various schemes and their performance in terms of these metrics.

Re-routing Schemes — Schemes of this kind build a workable light-forest by rerouting the unsupported paths in the multicast tree built by some generic algorithm. Three such schemes, namely, Re-route-to-Source, Re-route-to-Any [35], and Loop-Back [39], are discussed in the following sections with illustrative examples shown in Fig. 10. In the figure, the sample network is composed of different types of nodes (the DaC and DaC-WC nodes are not differentiated but represented by MI nodes), where node 1 is the source of a multicast and all the other nodes are destinations. Figure 10a shows the multicast tree built by the SPT heuristic with the bold lines being the part that can be supported by the physical network and the dotted lines being otherwise, while Figs. 10b–d show how the three re-routing schemes build a workable light-forest from this part.

Re-route-to-Source: After the multicast tree is built using some generic algorithm assuming all the nodes are multicast-capable, its nodes are checked one by one (in the breadth-first order, for example). If node v has more than one child on the tree and is MI, all but one downstream branches from v are cut (which branch to keep is not specified by the algorithm). Node v is now included in the forest, and each of the affected children joins the forest at an MC node (more exactly, a SC-WC node, including the source node) along the reverse shortest path to the source. For example, in Fig. 10b node 6 joins the forest along the reverse path $\{6 \rightarrow 3 \rightarrow 1\}$ at node 1, while node 7 joins the forest along the reverse path $\{7 \rightarrow 5 \rightarrow 2 \rightarrow 1\}$ at node 2. After node 6 joins the forest, node 9 joins the forest automatically since it is the only child of node 6 on the original multicast tree. In other words, the light-forest consists of three subtrees: the subtree represented by the bold lines, the subtree $\{1 \rightarrow 3 \rightarrow 6 \rightarrow 9\}$, and the subtree $\{2 \rightarrow 5 \rightarrow 7\}$. Note that node 7 joins the forest at node 2 (which is SC-WC), resulting in a subtree not rooted at the source.

Re-route-to-Any: Instead of along the reverse shortest path to the source, an affected node joins the forest at a MC or leaf MI node already on the forest along any other path. As shown in Fig. 10c, node 6 (and hence node 9) now joins the forest at node 4, instead of two hops back at node 1.

Loop-Back: This scheme assumes that the link between any two adjacent nodes is composed of two unidirectional fiber links with each in one direction. When a signal reaches a DaC node having multiple downstream links, the signal is forwarded to only one of them at the mean time. However, after the forwarded signal reaches the leaf (member) node in this path, it can be forwarded back to this node again along exactly the same path (but on the other set of fibers). Then the signal can be forwarded to other downstream links, one by one. Eventually, the signal can reach all the destinations, at the cost of longer delays. This process is shown in Fig. 10d, with the loop-back happening at node 3 and node 5. It has been shown that any multicast session can be realized in a network of only DaC nodes in this way [31]. Obviously, in such a scheme only one wavelength is needed for each multicast request.

Scheme	Number of wavelengths	Amount of bandwidth	Average delay	Computational complexity
Re-route-to-Source	Large	Large	Small	Low
Re-route-to-Any	Medium	Large	Large	Medium
Loop-Back	1	—	Large	Low
Member-First	Small	Medium	Medium	Medium
Member-Only	Small	Small	Large	High
VS-I	Small	Small	—	High
VS-II	Small	Small	—	Low

Table 3. *Performance of the MC-RWA schemes for sparse splitting networks.*

Tree-by-tree Schemes — Schemes of this kind build a workable light-forest by building multiple multicast subtrees one by one, with each spanning a subset of the destination nodes. We discuss in the following several such schemes, namely, Member-First, Member-Only [35], and virtual-source-based algorithms [36–38].

Member-First: This algorithm starts by building a spanning tree of the network in a manner similar to Dijkstra's algorithm, except that when choosing a node to be added to the current tree from among the candidate nodes which have the same shortest path length from the source, a member node is chosen first. After adding a new node, some upstream links are cut if necessary to ensure the new member node can indeed be reached from the source. When the current tree cannot be expanded any more, the algorithm prunes from the tree those branches not leading to any member, and repeats the above procedure to build another subtree for the member nodes not yet included in the light-forest (if there are any).

Member-Only: Different from Member-First, where a spanning tree is first built and then pruned, this algorithm builds a multicast subtree by including member nodes one by one, thus eliminating the pruning stage in Member-First. The algorithm expands the current tree by adding a feasible shortest path from the current tree to the member nodes still not in the light-forest. When the current tree cannot be expanded any more, the algorithm restarts to build another subtree for the member nodes still not in the light-forest.

Virtual-Source-Based Algorithms: A *virtual source (VS)* is essentially a SC-WC node that we defined. A VS-based algorithm (denoted as VS-I) is proposed in [36]. It is basically an improvement to the Member-Only algorithm, and the idea is as follows. When adding a destination node to the current tree, if there are multiple shortest paths connecting the node to the MC (or leaf MI) nodes already on the tree, choose those nodes in the decreasing order of their multicast capability, i.e., SC-WC nodes first, followed by SC nodes, DaC-WC nodes, and last, the DaC nodes. The rationale is that nodes with larger multicast capability should be used more.

Another VS-based algorithm (denoted as VS-II) is proposed in [37]. The basic idea is as follows. Each node in the network finds a shortest path to the nearest VS and establishes a connection to it. In this way, the network can be partitioned into a set of trees (regions) each rooted at a VS. The connectivity between the VSes is pre-established, by reserving some number of wavelengths, for example. When a multicast request arrives, the source first establishes a connection to its VS. This VS then establishes connections to other VSes that have one or more destination nodes in their respective regions. These VSes then establish connections to the destination nodes in their regions. In this way, a multicast tree can be built. This algorithm differs from the previous schemes in that the multicast tree is not source-rooted but VS-rooted. The advantages of this algorithm include a shorter setup delay and a simpler procedure of dynamic addition or deletion of group members than the source-rooted schemes, while the limitations include the overhead due to the resources reserved for paths between VSes. A similar idea is also proposed in [38].

After a light-forest has been built, wavelengths are to be assigned to it. The wavelength assignment is based on the concept of *segment*. A segment is a collection of links on which the same wavelength has to be assigned. It corresponds to a link in the case of full wavelength conversion (i.e., all nodes are either DaC-WC or SC-WC), and a subtree in the case of no wavelength conversion (i.e., all nodes are either DaC or SC). In the case of sparse wavelength conversion (i.e., all four types of nodes are possible), segments are the connected parts after removing all the non-leaf nodes that have the wavelength conversion capability. After the segments of a light-forest are determined, wavelengths are assigned to the segments using a First-Fit algorithm [44]. The performance of Re-route-to-Source, Re-route-to-Any, Member-First, and Member-Only are compared in [35]. Among the four schemes, Re-route-to-Source results in the shortest average delay, and is the simplest to implement. However, it requires the largest amount of bandwidth and number of wavelengths. At the other extreme, Member-Only requires the least amount of bandwidth and number of wavelengths, but results in the longest delay and has the highest computational complexity (due to the need to compute all-pair shortest paths). Re-route-to-Any and Member-First achieve some balance between bandwidth and delay, with Member-First having a better overall performance than Re-route-to-Any. Loop-Back requires only one wavelength, but may result in long delays. VS-I always requires less wavelengths and bandwidth than Member-Only, while VS-II results in less wavelengths and bandwidth than Member-Only when the multicast group size is large and the number of VSes is small. The performance of these schemes is qualitatively summarized in Table 3.

MC-RWA for Dynamic Traffic — In a system with dynamic multicast traffic, we only need to deal with individual multicast requests at the times when they arrive. The algorithms we have just discussed for single multicast may therefore be used in this case. However, in the above study the number of wavelengths available for a multicast request is unlimited (minimizing the number of wavelengths to be used is an objective instead of a constraint), while for a realistic system with dynamic multicast traffic the number of wavelengths is limited and some of them may have already been used on some links by existing multicast sessions. Therefore, even if we can build a multicast tree (or forest) for a new request, we may not be able to find a wavelength (or multiple wavelengths) to completely support it. The objective of MC-RWA for dynamic traffic is then to minimize the blocking probability of the multicast traffic.

There are two different blocking policies: *full destination blocking (FDB)* and *partial destination blocking (PDB)*. With FDB policy, a multicast call is accommodated only when the corresponding multicast tree can be completely supported. This policy may be appropriate for applications such as distributed computing and video conferencing, where all destinations must be reached for the communication activity to take place. Under this policy, the appropriate performance metric is *session blocking probability*, which is defined as the probability that an arriving multicast request is blocked. For applications such as stored video services, however, it is more reasonable to use the PDB policy, with which part of the users in a group may be served instead of the whole group being blocked. In this case, the appropriate performance metric is *destination (user) blocking probability*, which is defined as the probability that a destination (user) in the group is blocked.

MC-RWA for dynamic traffic has been studied in [45]. In the system, all nodes are assumed to be splitting-capable (SC). Therefore, a multicast tree can always be built for each multicast request, and should be assigned a single wavelength. Three schemes, namely, Fixed MC-RWA, Alternate MC-RWA, and Dynamic MC-RWA, are proposed. We first discuss them under the FDB policy in the following sections.

Fixed MC-RWA: In this scheme, a multicast tree is pre-calculated for each possible multicast request using a certain algorithm. When a multicast call arrives, search the wavelength set in a fixed order to find a wavelength that is available on all links of the multicast tree for this call. Once an available wavelength is found, it is assigned to the call. The call is blocked if all the wavelengths are exhausted without success.

Alternate MC-RWA: As opposed to a single multicast tree for each multicast group in Fixed MC-RWA, a set of multicast trees are pre-calculated for each multicast group. When a multicast call arrives, the trees in the set for this group are checked sequentially. For each tree, the same wavelength searching process as in Fixed MC-RWA is performed. The first tree that is found supportable by a certain wavelength is chosen for the multicast call, and the corresponding wavelength is assigned to the tree. If none of the trees can be supported, the call is blocked. Fixed MC-RWA and Alternate MC-RWA can be summarized as Static MC-RWA, since the multicast trees are pre-determined.

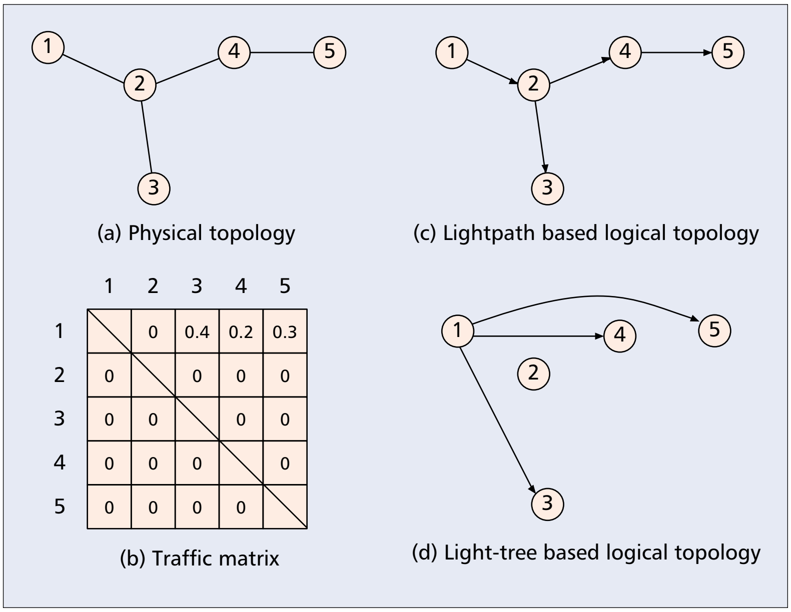
Dynamic MC-RWA: This scheme is also called Adaptive MC-RWA, and it proceeds as follows. When a multicast call arrives, the current network state (e.g., the wavelength usage on each link) is examined. A graph is then constructed for each wavelength by removing from the original network graph the links on which this wavelength is being used. These wavelength graphs are searched in a fixed order by executing some multicast tree building algorithm on each of them. The first wavelength on which there exists a multicast tree that spans all of the destinations is assigned to the call. If no complete multicast tree can be built using a single wavelength, the call is blocked.

Under the PDB policy, Fixed MC-RWA and Dynamic MC-RWA simply choose the wavelength on which the maximum number of destinations can be supported, and Alternate MC-RWA chooses the combination of multicast tree and wavelength on which the maximum number of destinations can be supported.

Among the three schemes, Dynamic MC-RWA performs the best, while Fixed MC-RWA performs the worst. The reason is as follows. The two static approaches divide the MC-RWA problem into two subproblems, namely, the multicast routing problem and the wavelength assignment problem, and solve them separately (or more exactly, sequentially). Although in each step an optimal solution may be found, the overall result may not be optimal. Dynamic MC-RWA solves the two subproblems in a coupled way. It can always make the best use of the wavelengths by adaptively building multicast trees according to the current wavelength usage on the links, and therefore achieves the best performance among the three. Alternate MC-RWA outperforms Fixed MC-RWA because it provides more choices in choosing the multicast tree and hence leads to a better usage of the wavelengths. On the other hand, Dynamic MC-RWA has the highest computational complexity, while Fixed MC-RWA is the simplest.

In the case of a sparse splitting network, the above schemes may not be used directly and PDB policy may have to be used, since it is essential for such networks that only a portion of the users in a multicast group can be served by a single multicast tree on a single wavelength. This area of work may need further research.

MC-RWA for Static Traffic — As mentioned earlier, the problem of MC-RWA for static traffic usually takes two forms: MC-RWA for batch requests, and logical topology design for a given long-term traffic matrix. The former case is important for such applications as near video-on-demand (near-VOD) services, where a batch of multicast requests need to be supported simultaneously [46]. Moreover, study of the problem may also provide some insights (or guidelines) on how multicast should be supported in a wavelength-routed network [47]. The latter case is important for the design of a wavelength-routed WDM WAN. We have mentioned the logical topology design problem in the context of multihop broadcast-and-select WDM networks, where the physical topology is based on a passive star coupler. In a WDM WAN, the physical topology can be an arbitrary mesh. However, the problem can be similarly defined: given a physical topology and a long-term traffic matrix, find logical connections between network nodes such that some objective function is optimized. The objective function can be, for example, minimizing the network-wide average packet hop distance, minimizing the maximum traffic flow on the connections, or minimizing the total number of transceivers in the network. Such a problem can usually be formulated as a mixed integer linear program (MILP), and has been extensively studied with unicast traffic [48–50]. With the increase of multicast traffic on the Internet, the optimal design of logical topologies taking multicast traffic into account is attracting more attention [29, 51].


1
2
3
4
5
(a) Physical topology
1
2
3
4
5
(c) Lightpath based logical topology
1
4
5
2
3
(d) Light-tree based logical topology

	1	2	3	4	5
1		0	0.4	0.2	0.3
2	0		0	0	0
3	0	0		0	0
4	0	0	0		0
5	0	0	0	0	

(b) Traffic matrix

■ **FIGURE 11.** *Lightpath-based and light-tree-based logical topologies for traffic from node 1 to node 3, 4, and 5 given one wavelength on all links.*

MC-RWA for Batch Requests — A system with all the nodes being multicast-capable and without wavelength conversion is assumed. A multicast tree can therefore always be built for a multicast request, and should be assigned a single wavelength according to the wavelength continuity constraint. Moreover, two multicast trees having shared links must be assigned different wavelengths to avoid wavelength conflict. The problem now is to assign wavelengths to the multicast trees, such that either the number of wavelengths needed to accommodate all the multicast requests is minimized or the number of multicast requests supported is maximized (or equivalently, the blocking rate is minimized) for a limited number of wavelengths. This problem can be transformed into a graph coloring problem, where each vertex in the graph represents a multicast tree to be supported and two vertices are adjacent (i.e., there is an edge between them) if and only if the corresponding multicast trees share a common link in the original network. Thus, the problem can be solved using any appropriate graph coloring algorithm.

The above scheme is a generic two-step approach: first determine the multicast trees and then assign the wavelengths. As we have just discussed for the dynamic traffic case, this kind of scheme may not result in the optimal solution. The problem of further reducing the number of wavelengths needed beyond the two-step approach is studied in [47]. The basic idea is that the multicast trees built independently in the two-step approach may be modified to make more efficient use of the wavelengths. Specifically, two approaches are studied.

- **Load balancing:** Load balancing is related to wavelength assignment in that the number of wavelengths needed is at least equal to the maximum link load in the system, where the load on a link is defined as the number of wavelength channels being used on that link. Therefore, modifying the multicast trees to minimize the maximum link load may reduce the number of wavelengths to be used.
- **Wavelength reassignment:** Given a wavelength assignment for a set of multicast trees, the least used wavelength (in terms of the number of multicast trees using it) might be freed by re-routing the multicast trees using it so that they can be assigned with other more-used wavelengths.

Let A denote the algorithm for building the multicast trees (for example, the shortest-path tree heuristic); let B denote the algorithm for wavelength assignment (for example, the aforementioned graph-coloring heuristic); let C denote the load balancing algorithm; and let D denote the wavelength reassignment algorithm. The generic two-step approach can then simply be denoted as AB, while the possible improving solutions include ACB, ABD, and ACBD. Simulation results show that ABD and ACBD perform much better than AB and ACB, while ACB and ACBD only offer a small improvement over AB and ABD, respectively. In other words, wavelength reassignment reduces the number of wavelengths more effectively than load balancing.

Full destination blocking (FDB) policy has been assumed in the above discussion. The problem under the partial destination blocking (PDB) policy is formulated as a nonlinear integer program and studied in the context of a single-source environment where all the multicast requests share a common source node (e.g., the video server) in [46].

Logical Topology Design — The logical topology embedded in the physical network can be either lightpath-based or light-tree-based. In a lightpath-based logical topology, each link represents a lightpath between its two nodes. Communication between two nodes may happen either through a single lightpath (i.e., a direct link in the logical topology) or through a number of concatenated lightpaths (i.e., multiple hops in the logical topology). In the latter case, electronic packet switches can be used at the intermediate node between two lightpaths. Previous research on logical topology design with only unicast traffic has been focusing on lightpath-based logical topologies. With the presence of multicast traffic, a light-tree-based logical topology may be more beneficial. In such a logical topology, a node has direct links to a set of nodes if it is the root (source) of a light-tree and the set of nodes are the destinations on the tree. Therefore, the transmission from the source to all the destinations takes only one hop and is done all-optically. Moreover, a light-tree-based logical topology may also require fewer transceivers than a lightpath-based solution, as will be shown later. However, it relies on the availability of multicast capabilities in the physical layer, which may not always be possible. The splitting of light power may also necessitate the use of optical amplifiers, which will increase the network cost. Moreover, the data transmitted on a light-tree will reach all the nodes on the tree, regardless of whether they are unicast or multicast. For unicast, this means a large

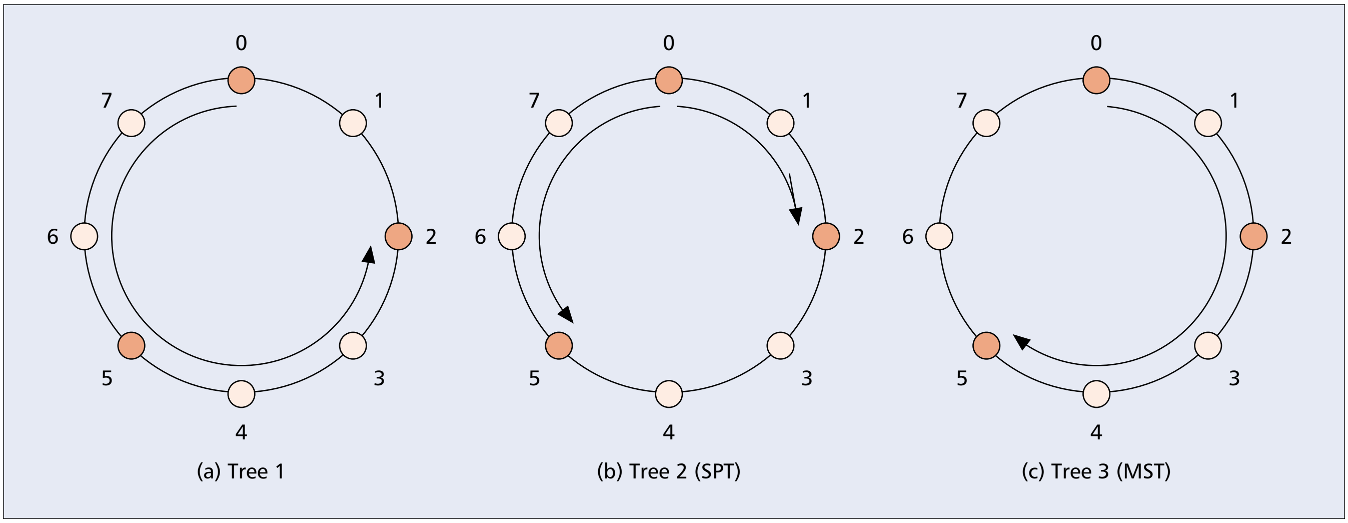


FIGURE 12. *Multicast routing in a bidirectional ring network.*

amount of unnecessary replications of the data in the network.

Lightpath-Based Logical Topology Design — This problem with the presence of multicast traffic is studied in [51]. In this case, a multicast tree is composed of multiple lightpaths, and the multicast traffic arriving at a branching node of the tree is replicated in the electronic domain before being forwarded to the downstream lightpaths. Given a physical topology and a traffic matrix (consisting of both unicast and multicast traffic), the objective is to minimize the maximum traffic flow on lightpaths. Since it is an MILP problem, four heuristics are proposed.

- Source Copy Multicast (SCOM): This heuristic simply transforms the traffic matrix into a pure unicast traffic matrix by assuming that each multicast is realized by multiple unicasts at the source, and then employs any existing unicast topology design algorithm to generate the logical topology.
- Route & Remove (R&R): This heuristic starts with a fully-connected logical topology, and then iteratively removes from the logical topology the least-loaded lightpaths, until the nodal degree constraints are satisfied.[2]
- Tabu Search (TS): TS is an iterative optimization approach. At each iteration, all neighbor solutions of the current solution are evaluated, and the best is selected as the new current solution. After a given number of iterations, the algorithm returns the best visited solution. For the logical topology design problem, the initial solution is obtained from the R&R heuristic, and a neighbor solution is obtained from the current solution by selecting two lightpaths (e.g., from node 1 to node 2 and from node 3 to node 4, respectively) and exchanging their destinations (obtaining two new lightpaths from node 1 to node 4 and from node 3 to node 2, respectively).
- Simulated Annealing (SA): SA is also an iterative optimization approach. However, at each iteration only one neighbor solution is visited and evaluated. If the new solution performs better than the current one, it is accepted as the new current solution, otherwise it is accepted with probability p, and discarded with probability $1 - p$.

TS and SA are called *metaheuristics*, while SCOM and R&R are greedy heuristics. Both metaheuristics perform better than the greedy heuristics, with TS outperforming SA by a small margin. SCOM and R&R yield performance in the same range.

Light-tree-Based Logical Topology Design — This concept is proposed in [29]. Two objective functions, i.e., minimizing the average packet hop distance and minimizing the total number of transceivers needed in the network, have been studied with unicast and multicast traffic. Using the NSFNET as an example, the numerical results by solving the MILP formulation show that an optimum light-tree-based logical topology has a lower value of average packet hop distance and requires fewer transceivers than an optimum lightpath-based logical topology. We give an illustrative example in Fig. 11. We assume that only one wavelength is available on all the links, and the bandwidth of a wavelength channel is one unit. We consider a unicast traffic, where node 1 wants to send 0.4 units of traffic to node 3, 0.2 units of traffic to node 4, and 0.3 units of traffic to node 5. The sum of all the traffic is less than one unit, so that they can be carried by one wavelength channel. A lightpath-based solution would consist of four lightpaths: 1 → 2, 2 → 3, 2 → 4, and 4 → 5. A total number of eight transceivers (one transmitter and one receiver per lightpath) are required, in addition to an electronic switch at nodes 2 and 4. On the other hand, a light-tree-based solution consists of a single light-tree, which requires a total number of four transceivers (one transmitter at node 1 and one receiver per node at nodes 3, 4, and 5) without electronic switches needed. In terms of the average hop distance, the light-tree-based solution is also better than the lightpath-based solution (one for the former, and more than two for the latter).

A Special Case: Ring Networks

A ring network can be either unidirectional or bidirectional. In a unidirectional ring network the traffic can only flow in one direction (either clockwise or counterclockwise), while a bidirectional ring network can simply be thought of as consisting of two unidirectional rings operating in opposite directions.[3] To multicast in a ring network, each node in the ring should be able to forward an incoming signal in addition to receiving a copy of the signal. Since each node in a ring has only one outgoing link, light-splitting capability is not needed in an optical ring network. In other words, the nodes should be DaC or DaC-WC nodes. A

[2] *The in and out degrees of a node in the logical topology cannot exceed the actual number of receivers and transmitters the node has, respectively. This constraint ensures that the output logical topology is realizable.*

[3] *We do not consider the protection rings in self-healing ring networks.*

study of multicast-capable access nodes for optical ring networks can be found in [52]. In the following sections we focus on the multicast routing and wavelength assignment (MC-RWA) strategies in WDM ring networks.

We first look at multicast routing in ring networks. In a unidirectional ring network, a multicast tree is simply an arc traversing all the destination nodes from the source node along the ring direction. It is unique. In a bidirectional ring network, there are multiple multicast trees for a given multicast request. Specifically, given a multicast request (s, D), where s is the source node and D is the set of destination nodes, there are $|D| + 1$ ways of constructing the multicast tree on a bidirectional ring. The reason is as follows. In a bidirectional ring, a multicast tree in general consists of two arcs, with one clockwise and the other counterclockwise. Let k denote the number of destination nodes covered by the clockwise arc. The possible values of k are 0, 1, ..., $|D|$, with each value corresponding to a multicast tree. Therefore, for a multicast request with group size $|D|$ the total number of possible multicast trees is $|D| + 1$. In Fig. 12, we show the three ways of constructing the multicast tree for a request with $s = 0$ and $D = \{2, 5\}$. Let $r(i, j, cw)$ and $r(i, j, ccw)$ denote the arc from node i to node j in the clockwise and counterclockwise direction, respectively. The three multicast trees for the request are $T_1 = r(0, 2; ccw)$, $T_2 = r(0, 5; ccw) \cup r(0, 2, cw)$, and $T_3 = r(0, 5; cw)$, as shown in Fig. 12a–c, respectively.

As in the wavelength-routed networks, the shortest path tree (SPT) and minimum spanning tree (MST) heuristics can be used to construct the multicast tree. In the above example, if SPT heuristic is used, the multicast tree T_2 will be constructed, while if MST heuristic is used, the multicast tree T_3 will be constructed. A comparative study of the performance of SPT and MST in WDM ring networks is given in [53].

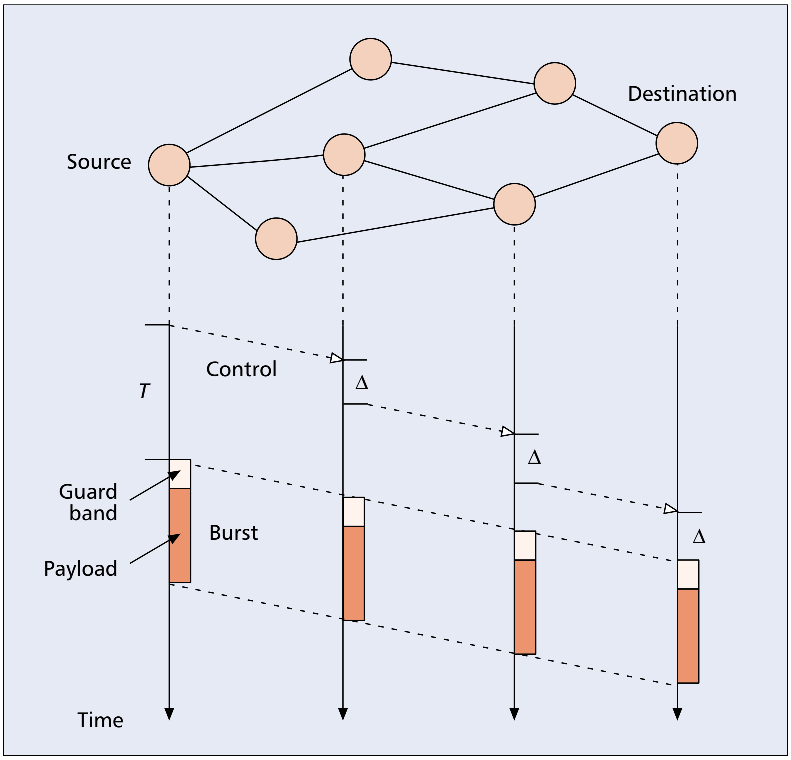
We next discuss the wavelength assignment in ring networks. In a unidirectional ring network, since the multicast tree is a fixed arc, multicasting is performed simply to find a wavelength that is available on all the segments of the arc (assuming no wavelength conversion). If such a wavelength cannot be found, the multicast call is blocked. This is basically the Fixed MC-RWA. In a bidirectional ring network, since there are multiple multicast trees that can be chosen in a bidirectional WDM ring network, the Alternate MC-RWA and Dynamic MC-RWA schemes that have been discussed can be easily extended to fit this situation.

Other Research on Multicasting in Wavelength-Routed WDM Networks

Besides the above discussions, there is other research on MC-RWA that we briefly summarize as follows.

One area of work addresses the bounds on the minimum number of wavelengths required for wide-sense nonblocking (or rearrangeable) multicasting in WDM networks [54–56]. The systems considered assume a *single reception constraint*, which says that each node can be the destination of at most one multicast connection at any given time. In [54] it is further assumed that each node can be a source of at most one multicast connection at any given time. It is shown that the number of wavelengths needed to support multicasting in such networks is $O(\log N)^2$, where N is the number of nodes in the network. The bounds on the number of wavelengths needed for wide-sense nonblocking multicasting and rearrangeable multicasting in some regular networks, such as rings and linear arrays, are derived in [55] and [56], respectively.

Another area of work studies MC-RWA in linear lightwave networks (LLNs) [57–59]. LLNs use a special switching unit called a linear divider-combiner (LDC), as opposed to the wavelength-routing switches used in general wavelength-routed networks.

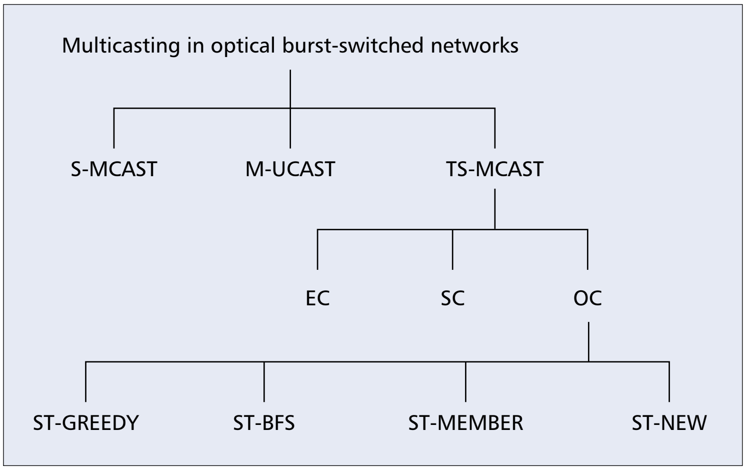


■ **Figure 13.** *An optical burst-switched network.*

Multicasting in Optical Burst-Switched WDM Networks

Optical burst-switched (OBS) networks differ from wavelength-routed networks primarily in the following ways. First, a control packet is sent before the transmission of a data burst to set up a connection and reserve the resources accordingly, and the burst is sent without waiting for the acknowledgment of the connection establishment (which actually does not exist). In other words, only a one-way reservation is made. Second, the resources (e.g., wavelengths on the links) reserved for a burst are released as soon as they are used by the burst. In other words, the wavelength on a link reserved for a burst can be used for other bursts as soon as the burst passes through the link, as opposed to being reserved for the duration of the session in wavelength-routed networks. Third, in forming a burst from packets, guard bands (GBs) need to be used in the burst to accommodate possible timing jitters at each intermediate node. An example of the burst-switched operation is shown in Fig. 13, where a burst is transmitted T time after the control packet which is processed at each node in time Δ to reserve appropriate bandwidth and configure the switch.

To multicast in OBS networks, several issues must be considered. First, a multicast tree (or forest) needs to be built for each multicast transmission. This problem has been addressed in [60, 61], and the ideas are essentially the same as those presented in [35], which are for wavelength-routed networks. Sec-



■ **FIGURE 14.** *Multicasting schemes in optical burst-switched WDM networks surveyed in this article.*

ond, a control packet needs to be sent before the transmission of each multicast burst. Therefore, a multicast scheme having a small overhead of control packets is desirable. Third, the GBs waste the bandwidth, therefore we also want the multicast scheme to have a small overhead of GBs. The second and third issues are addressed in [62, 63], and the proposed schemes are summarized as shown in Fig. 14.

Separate Multicasting (S-MCAST): This is a straightforward scheme, in which each multicast group (session) constructs its own multicast tree (employing the control packets) along which the assembled multicast data bursts consisting of the traffic only for that group are delivered.

Multiple Unicasting (M-UCAST): In this scheme, the multicast traffic of a group is delivered to all the destinations through multiple unicasts. Specifically, for each destination node in a multicast group, during the assembly time of a burst a copy of the multicast traffic will be assembled together with the unicast traffic destined to that node (if such traffic exists) into a unicast burst and then sent to the node. This scheme can reduce the overheads of the control packets and the GBs, since the multicast traffic simply gets a "free" ride for the control packets and GBs from the unicast traffic. However, it may result in low bandwidth efficiency because of the duplication of the multicast traffic. The overall performance of this scheme depends on the network conditions.

Tree-Shared Multicasting (TS-MCAST): In this scheme, the multicast traffic of multiple sessions are mixed together to form a burst that is then delivered by a shared multicast tree. In other words, the control packet and the GBs of a burst are shared among multiple multicast sessions. Therefore, this scheme may achieve low overheads of the control packets and the GBs.

S-MCAST and M-UCAST are simple, while the TS-MCAST scheme will be detailed in the following section. The network is modeled as a set of core routers, a set of edge routers, and a set of links connecting them. A multicast session is composed of an edge router that is the source, a set of some other edge routers that are the destinations, and a set of core routers and links that constitute the multicast tree. In TS-MCAST, the set of multicast sessions (H_i) originating from edge router i is partitioned into a number of subsets, each of which is called a multicast sharing class (MSC) and uses a shared tree (ST). The IP packets from the multicast sessions in the same MSC are assembled together to form bursts. The major problem of TS-MCAST is then the partition strategy (or in other words, the tree sharing strategy) of H_i, which will be discussed in the following.

Equal Coverage (EC): Multicast sessions with the same membership (i.e., the same set of member edge routers) are grouped into one MSC. Note that although these multicast sessions have the same source and destinations, they do not necessarily have the same multicast tree. In this case, one of the existing multicast trees is selected as the new ST.

Super Coverage (SC): If the set of member edge routers of a multicast session is a superset of that of another multicast session, these two multicast sessions are grouped into the same MSC. The multicast tree of the larger multicast session is selected as the new ST.

Overlapping Coverage (OC): A number of multicast sessions having a sufficient degree of overlap in the edge routers, core routers, links, or tree sharing gain are grouped into the same MSC.[4] Four algorithms are proposed to build the ST.

ST-GREEDY — This is a greedy algorithm that simply takes the union of (i.e., merges) all the existing multicast trees in the MSC. It is simple, but may output a ST containing redundant links.

ST-BFS — This is a breadth-first search (BFS) algorithm. It starts at the source edge router (the root), checking each adjacent node of the root to see if it is on any existing multicast tree but not yet on the ST. If so, both the node and link leading to it are added to the ST, and the node is added to a queue for further consideration. The process repeats for each node in the queue until the queue is empty. This algorithm can eliminate some redundant links in the greedy algorithm.

ST-MEMBER — This is a member-initiated algorithm, in which an existing multicast tree with the largest number of members is selected as the base of the new ST, and all the other members join the ST by growing back toward the source along the links on the existing trees. This algorithm does not produce redundant links.

ST-NEW — This algorithm simply constructs a new multicast tree for the MSC with all the members included by applying the multicast tree construction algorithm of the multicast session.

In order to evaluate the performance of the different schemes, the average amount of multicast traffic per link has been chosen as the metric. The reason is as follows. By making the amount of multicast traffic injected to the network the same for different multicast schemes, the measured amount of multicast traffic per link represents the amount of bandwidth consumed by the multicast traffic per link, and thus, the larger the amount, the less efficient a multicast scheme is. The performance of these schemes for static multicast sessions and membership is studied in [62]. It is shown that the TS-MCAST schemes always perform better than S-MCAST schemes, while M-UCAST schemes may perform better than S-MCAST schemes when the GB size is large. Among the TS-MCAST schemes, OC performs better than EC and SC. Moreover, among the four ST algorithms for OC, ST-GREEDY performs the worst, while the other three perform almost the same.

Multicast schemes for dynamic sessions and membership are studied in [63]. In that case, the above schemes need to be extended. The basic idea is that the MSCs (and followed by the corresponding STs) are updated or even re-determined if necessary after the change of the sessions and membership. Simulation results show a similar trend as for static sessions and membership, i.e., TS-MCAST yields better performance than S-MCAST and M-UCAST.

[4] *The tree sharing gain is defined as the ratio of the average amount of multicast traffic carried per link without tree sharing to that with tree sharing. It reflects the amount of bandwidth that can be saved by tree sharing.*

Conclusions

Multicast traffic is expected to account for a larger and larger portion of Internet traffic. Hence, there is a need to support multicast in the next-generation WDM-based Internet. In this article, we have surveyed the multicasting issues and approaches for different types of WDM networks, namely, the broadcast-and-select networks, which are typically for WDM LANs/MANs, the wavelength-routed networks, which are essentially circuit-switched WDM WANs, and the emerging optical burst-switched (OBS) networks.

Broadcast-and-select WDM networks can be either single-hop or multihop. For single-hop networks, the major issue is the design of multicast scheduling algorithms (MSAs) for contention resolution. We have discussed a number of MSAs and have shown that tradeoffs are usually necessary when designing an MSA. In particular, scheduling a single multicast transmission to reach all the destination nodes may result in low throughput and long packet delay. Hence, partitioning a multicast transmission into multiple transmissions is usually used to achieve a higher throughput and a shorter delay at the cost of sacrificing the bandwidth efficiency of multicast. Moreover, reservation-based MSAs schedule the best for individual multicast transmissions with high computational complexity and control message overhead, while simple pre-allocation-based MSAs can only be optimized for static traffic patterns. For multihop networks, supporting multicast is not as efficient as in single-hop networks. We have shown how channel sharing can effectively improve multicast performance.

For wavelength-routed WDM networks, the key issue is the multicast routing and wavelength assignment (MC-RWA) problem. We have reviewed various schemes for building a physically realizable multicast tree (or forest) for each multicast request in a sparse splitting network, as well as the schemes for minimizing the blocking probability when multiple multicast requests exist. We have also discussed the logical topology design in wavelength-routed networks.

For OBS WDM networks, the major consideration is reducing the overheads of the control packets and guard bands. We have discussed some multicast schemes that achieve this goal by sharing the control packets and guard bands between unicast traffic and multicast traffic or among multiple multicast sessions.

Although multicasting in WDM networks is currently still in the research stage, it will find real deployments and become indispensable as the Internet evolves to an optical network and multicast applications further increase.

References

[1] B. Mukherjee, *Optical Communication Networks*, McGraw-Hill, 1997.

[2] C. Qiao and M. Yoo, "Optical Burst Switching (OBS): A New Paradigm for An Optical Internet," *J. High Speed Networks*, vol. 8, no. 1, 1999, pp. 69–84.

[3] B. Mukherjee, "WDM-Based Local Lightwave Networks Part I: Single-Hop Systems," *IEEE Network*, vol. 6, no. 5, May 1992, pp. 12–27.

[4] B. Mukherjee, "WDM-Based Local Lightwave Networks Part II: Multihop Systems," *IEEE Network*, vol. 6, no. 7, Jul. 1992, pp. 20–31.

[5] G. N. Rouskas and M. H. Ammar, "Analysis and Optimization of Transmission Schedules for Single-Hop WDM Networks," *IEEE Trans. Net.*, vol. 3, no. 2, 1995, pp. 211–21.

[6] M. S. Borella and B. Mukherjee, "A Reservation-Base Multicasting Protocol for WDM Local Lightwave Networks," *IEEE Int'l. Conf. Commun. (ICC)*, 1995, pp. 1277–81.

[7] L. H. Sahasrabudhe and B. Mukherjee, "Probability Distribution of the Receiver Busy Time in a Multicasting Local Lightwave Network," *IEEE Int'l. Conf. Commun. (ICC)*, 1997, pp. 116–20.

[8] J. P. Jue and B. Mukherjee, "The Advantage of Partitioning Multicast Transmissions in a Single-Hop Optical WDM Network," *IEEE Int'l. Conf. Commun. (ICC)*, 1997, pp. 427–31.

[9] H.-C. Lin and C.-H. Wang, "A Hybrid Multicast Scheduling Algorithm for Single-Hop WDM Networks," *Proc. IEEE INFOCOM '01*, 2001, pp. 169–78.

[10] H.-C. Lin and C.-H. Wang, "Minimizing the Number of Multicast Transmissions in Single-Hop WDM Networks," *Proc. 2000 IEEE Int'l. Conf. Commun. (ICC)*, 2000, pp. 1645–49.

[11] H.-C. Lin, P.-S. Liu, and H. Chu, "A Reservation-Based Multicast Scheduling Algorithm with a Reservation Window for Single-Hop WDM Networks," *Proc. IEEE Int'l. Conf. Networks (ICON)*, 2000, p. 493.

[12] M. Bandai, S. Shiokawa, and I. Sasase, "Performance Analysis of a Multicasting Protocol in WDM-Based Single-Hop Lightwave Networks," *Proc. IEEE GLOBECOM'97*, 1997, pp. 561–65.

[13] S.-T. Sheu and C.-P. Huang, "An Efficient Multicast Protocol for WDM Star-Coupler Networks," *Proc. IEEE Symp. Computers and Commun.*, 1997, pp. 579–83.

[14] T. Kitamura, M. Iizuka, and M. Sakuta, "A New Partition Scheduling Algorithm by Prioritizing the Transmission of Multicast Packets with Less Destination Address Overlap in WDM Single-Hop Networks," *Proc. IEEE GLOBECOM '01*, 2001, pp. 1469–73.

[15] E. Modiano, "Random Algorithms for Scheduling Multicast Traffic in WDM Broadcast-and-Select Networks," *IEEE/ACM Trans. Net.*, vol. 7, no. 3, 1999, pp. 425-434.

[16] G. N. Rouskas and M. H. Ammar, "Multidestination Communication over Tunable-Receiver Single-Hop WDM Networks," *IEEE JSAC*, vol. 15, no. 3, 1997, pp. 501–11.

[17] Z. Ortiz, G. N. Rouskas, and H. G. Perros, "Maximizing Multicast Throughput in WDM Networks with Tuning Latencies using the Virtual Receiver Concept," *European Trans. Telecommun. and Related Technologies*, vol. 11, no. 1, 2000, pp. 63–72.

[18] A. Bianco *et al.*, "Scheduling Algorithms for Multicast Traffic in TDM/WDM Networks with Arbitrary Tuning Latencies," *Proc. IEEE GLOBECOM '01*, 2001, pp. 1551–56.

[19] W.-Y. Tseng and S.-Y. Kuo, "A Combinational Media Access Protocol for Multicast Traffic in Single-Hop WDM LANs," *IEEE GLOBECOM '98*, 1998, pp. 294–99.

[20] W.-Y. Tseng, C.-C. Sue, and S.-Y. Kuo, "Performance Analysis for Unicast and Multicast Traffic in Broadcast-and-Select WDM Networks," *Proc. IEEE Int'l. Symp. Computers and Commun.*, 1999, pp. 72–78.

[21] M. S. Borella and B. Mukherjee, "Limits of Multicasting in a Packet-Switched WDM Single-Hop Local Lightwave Network," *J. High Speed Networks*, vol. 4, no. 2, 1995, pp. 155–67.

[22] E. Modiano and R. Barry, "Design and Analysis of an Asynchronous WDM Local Area Network using a Master/Slave Scheduler," *Proc. IEEE INFOCOM '99*, 1999, pp. 900–07.

[23] O. Gerstel, "On the Future of Wavelength Routing Networks," *IEEE Network*, vol. 10, no. 6, Nov./Dec. 1996, pp. 14–20.

[24] G. N. Rouskas and V. Sivaraman, "Packet Scheduling in Broadcast WDM Networks with Arbitrary Transceiver Tuning Latencies," *IEEE Trans. Net.*, vol. 5, no. 3, 1997, pp. 359–70.

[25] S. Banerjee, V. Jain, and S. Shah, "Regular Multihop Logical Topologies for Lightwave Networks," *IEEE Commun. Surveys*: http://www.comsoc.org/pubs/surveys, First Quarter 1999, pp. 2–18.

[26] F. K. Hwang, D. S. Richards, and P. Winter, *The Steiner Tree Problem*, Elsevier Science Publishers B. V., 1992.

[27] G. B. Brewster and M. S. Borella, "Multicast Routing Algorithms for the WDM Shufflenet Local Optical Network," *Proc. IEEE Int'l. Conf. Commun. (ICC)*, 1997, pp. 111–15.

[28] S. B. Tridandapani and B. Mukherjee, "Channel Sharing in Multi-Hop WDM Lightwave Networks: Realization and Performance of Multicast Traffic," *IEEE JSAC*, vol. 15, no. 3, 1997, pp. 488–500.

[29] L. H. Sahasrabuddhe and B. Mukherjee, "Light-trees: Optical Multicasting for Improved Performance in Wavelength-Routed Networks," *IEEE Commun. Mag.*, vol. 37, no. 2, 1999, pp. 67–73.

[30] M. Ali and J. S. Deogun, "Power-Efficient Design of Multicast Wavelength-Routed Networks," *IEEE JSAC*, vol. 18, no. 10, 2000, pp. 1852–62.

[31] M. Ali and J. S. Deogun, "Cost-Effective Implementation of Multicasting in Wavelength-Routed Networks," *IEEE/OSA Journal of Lightwave Tech.*, vol. 18, no. 12, 2000, pp. 1628–38.

[32] K.-D. Wu, J.-C. Wu, and C.-S. Yang, "Multicast Routing with Power Consideration in Sparse Splitting WDM Networks," *Proc. IEEE Int'l. Conf. Commun. (ICC)*, 2001, pp. 513–17.

[33] R. Malli, X. Zhang, and C. Qiao, "Benefit of Multicasting in All-Optical Networks," *Proc. SPIE*, vol. 3531, 1998, pp. 209–20.

[34] M. Ali and J. S. Deogun, "Allocation of Splitting Nodes in Wavelength-Routed Networks," *J. Photon. Network Commun.*, vol. 2, no. 3, 2000, pp. 247–65.

[35] X. Zhang, J. Wei, and C. Qiao, "Constrained Multicast Routing in WDM Networks with Sparse Light Splitting," *Proc. IEEE INFOCOM '00*, 2000, pp. 1781–90.

[36] N. Sreenath *et al.*, "Virtual Source-Based Multicast Routing in WDM Optical Networks," *Proc. IEEE Int'l. Conf. Networks (ICON)*, 2000, pp. 385–89.

[37] N. Sreenath *et al.*, "Virtual Source-Based Multicast Routing in WDM Networks with Sparse Light Splitting," *Proc. IEEE Workshop on High Perf. Switching and Routing*, 2001, pp. 141-145.

[38] W.-Y. Tseng and S.-Y. Kuo, "All-Optical Multicasting on Wavelength-Routed WDM Networks with Partial Replication," *Proc. 15th Int'l. Conf. Info. Net.*, 2001, pp. 813–18.

[39] S. Yan, M. Ali, and J. Deogun, "Route Optimization of Multicast Sessions in Sparse Light-Splitting Optical Networks," *Proc. IEEE GLOBECOM '01*, 2001, pp. 2134–38.

[40] Y. Sun, J. Gu, and D. H. K. Tsang, "Multicast Routing in All-Optical Wavelength-Routed Networks," *Optical Networks Mag.*, vol. 2, no. 4, 2001, pp. 101–09.

[41] B. Chen and J. Wang, "Constrained Wavelength Assignment for Multicast in WDM Networks," *Proc. 10th Int'l. Conf. Comp. Commun. and Networks*, 2001, pp. 388–94.

[42] W. Liang and H. Shen, "Multicasting and Broadcasting in Large WDM Networks," *Proc. 1st Merged Int'l. Parallel Processing Symp. and Symp. on Parallel and Distributed Processing (IPPS/SPDP)*, 1998, pp. 365–69.

[43] H. Shen, F. Chin, and Y. Pan, "Efficient Fault-Tolerant Routing in Multihop Optical WDM Networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 10, no. 10, 1999, pp. 1012–25.

[44] I. Chlamatac, A. Ganz, and G. Karmi, "Pure Optical Networks for Terabit Communication," *Proc. IEEE INFOCOM '89*, Apr. 1989, pp. 887–96.

[45] G. Sahin and M. Azizoglu, "Multicast Routing and Wavelength Assignment in Wide Area Networks," *Proc. SPIE*, vol. 3531, 1998, pp. 196–208.

[46] J. He, S.-H. G. Chan, and D. H. K. Tsang, "Routing and Wavelength Assignment for WDM Multicast Networks," *Proc. IEEE GLOBECOM '01*, 2001, pp. 1536–40.

[47] X.-H. Jia *et al.*, "Optimization of Wavelength Assignment for QoS Multicast in WDM Networks," *IEEE Trans. Commun.*, vol. 49, no. 2, 2001, pp. 341–50.

[48] R. Ramaswami and K. N. Sivarajan, "Design of Logical Topologies for Wavelength-Routed Optical Networks," *IEEE JSAC*, vol. 14, no. 6, 1996, pp. 840–51.

[49] B. Mukherjee *et al.*, "Some Principles for Designing a Wide-Area WDM Optical Network," *IEEE Trans. Net.*, vol. 4, no. 5, 1996, pp. 684–95.

[50] D. Banerjee and B. Mukherjee, "Wavelength-Routed Optical Networks: Linear Formulation, Resource Budgeting Tradeoffs, and a Reconfiguration Study," *Proc. IEEE INFOCOM '97*, Apr. 1997, pp. 269–76.

[51] M. Mellia *et al.*, "Optimal Design of Logical Topologies in Wavelength-Routed Optical Networks with Multicast Traffic," *Proc. IEEE GLOBECOM '01*, 2001, pp. 1520–25.

[52] S. Aleksic and K. Bengi, "Multicast-Capable Access Nodes for Slotted Photonic Ring Networks," *Proc. European Conf. Optical Commun. (ECOC) 2000*, Sept. 2000, pp. 83–84.

[53] X. Jia *et al.*, "Multicast Routing, Load Balancing, and Wavelength Assignment on Tree of Rings," *IEEE Commun. Letters*, vol. 6, no. 2, 2002, pp. 79–81.

[54] R. K. Pankaj, "Wavelength Requirements for Multicasting in All-Optical Networks," *IEEE/ACM Trans. Net.*, vol. 7, no. 3, 1999, pp. 414–24.

[55] C. Zhou and Y. Yang, "Multicast Communication in a Class of Wide-Sense Nonblocking Optical WDM Networks," *Proc. Int'l. Conf. Comp. Commun. and Networks*, 1998, pp. 321–28.

[56] Y. Wang and Y. Yang, "Multicasting in a Class of Multicast-Capable WDM Networks," *Proc. 9th Int'l. Conf. Comp. Commun. and Networks*, 2000, pp. 184–91.

[57] K. Bala, K. Petropoulos, and T. E. Stern, "Multicasting in a Linear Lightwave Network," *Proc. IEEE INFOCOM '93*, 1993, pp. 1350–58.

[58] S. Jiang and T. E. Stern, "Regular Multicast Multihop Lightwave Network," *Proc. IEEE INFOCOM '95*, 1995, pp. 692–700.

[59] H. Harai, M. Murata, and H. Miyahara, "Multicast Routing Method in Optical Switching Networks," *Electronics and Communications in Japan, Part 1*, vol. 79, no. 8, 1996, pp. 12–23.

[60] C. Qiao *et al.*, "WDM Multicasting in IP over WDM Networks," *Proc. Int'l. Conf. Network Protocols (ICNP)*, 1999, pp. 89–96.

[61] X. Zhang, J. Wei, and C. Qiao, "On Fundamental Issues in IP over WDM Multicast," *Proc. Int'l. Conf. Comp. Commun. and Networks (IC3N)*, 1999, pp. 84–90.

[62] M. Jeong *et al.*, "Efficient Multicast Schemes for Optical Burst-Switched WDM Networks," *Proc. IEEE ICC '00*, 2000, pp. 1289–94.

[63] M. Jeong *et al.*, "Bandwidth-Efficient Dynamic Tree-Shared Multicast in Optical Burst-Switched Networks," *Proc. IEEE ICC'01*, 2001, pp. 630–636.

BIOGRAPHIES

JINGYI HE (eehjy@ust.hk) received the B.Eng. and M.Eng. degrees, both in opto-electronic engineering, from Huazhong University of Science and Technology, P. R. China, in July 1996 and June 1999, respectively. He is currently a Ph.D. candidate in the Department of Electrical and Electronic Engineering at Hong Kong University of Science and Technology.

S.-H. GARY CHAN (gchan@cs.ust.hk) received the PhD in electrical engineering with a minor in business administration from Stanford University, Stanford, CA, in 1999, and the B.S.E. degree (highest honors) in electrical engineering from Princeton University, Princeton, NJ, in 1993. He is currently an assistant professor with the Department of Computer Science in the Hong Kong University of Science and Technology, Hong Kong, and an adjunct researcher with Microsoft Research Asia in Beijing, China. He was a visiting assistant professor in networking at the University of California, Davis, from September 1998 to June 1999. During 1992-93, he was a research intern at the NEC Research Institute, Princeton. His research interests include multimedia networking, high-speed and wireless communications networks, and Internet technologies and peer-to-peer networking. He was a William and Leila fellow at Stanford University during 1993-94. At Princeton he was the recipient of the Charles Ira Young Memorial Tablet and Medal, and the POEM Newport Award of Excellence in 1993. He is a member of Tau Beta Pi, Sigma Xi, and Phi Beta Kappa.

DANNY H. K. TSANG (eetsang@ee.ust.hk) received the B.Sc. degree in mathematics and physics from the University of Winnipeg, Winnipeg, Canada, in 1979, the B.Eng. and M.A.Sc. degrees, both in electrical engineering, from the Technical University of Nova Scotia, Halifax, Canada, in 1982 and 1984, respectively, and the Ph.D. in electrical engineering from the University of Pennsylvania, Philadelphia, PA, in 1989. He joined the Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Canada, in 1989, where he was an assistant professor in the computing science division. Since 1992 he has been with the Department of Electrical and Electronic Engineering at Hong Kong University of Science and Technology. His current research interests include statistical modeling of variable-bit-rate video traffic, queuing analysis of asynchronous transfer mode (ATM) multiplexers, congestion controls in B-ISDN/ATM networks, and wireless ATM. He has served as Technical Program Committee member of INFOCOM since 1994.