A Thesis for the Degree of Ph.D. in Engineering

A study of multicast on WDM optical network and IP network

August 2010

Graduate School of Science and Technology

Keio University

Fung Chun Cheong Alex

To all people who helped me in this work

ACKNOWLEDGEMENTS

This work would not have been succeeded without the help of all the people around me. I would like to use this place to thank all the people who have given valuable support to me during my studies as a PhD student in Keio University.

First of the all, I would like to give my sincere gratitude to *Prof Iwao Sasase* of Department of Information and Computer Science, Keio University. He is the one who accepted me in his lab such that I can proceed to study in the graduate school of Keio University. In these five years, he supervised me all the know-how in conducting research which has surely built up my ability. Also privately he has been closed to the students to share his interest such as wine, travelling and high-end audio equipment. It was surely a fruitful time in these five years. Also I would like to thank *Prof Naoaki Yamanaka* and *Prof Tomoaki Ohtsuki* of the same department, as well as *Prof Hiroyuki Tsuda* of the Department of Electrical and Electronic Engineering of same university for reading and giving valuable advices on my work. Also, *Prof S.H. Gary Chan* of the Department of Computer Science and Engineering of Hong Kong University of Science and Technology was the one who encouraged me to proceed studies in the graduate level. He has given me lots of advices and supports when I was preparing to proceed further studies in Japan.

The Ministry of Education, Culture, Sports, Science and Technology (Monbukagakusho) has been the main financial support in my studies of five years in Keio University. It is financially impossible for me to continue studies in the graduate school by my own. I am proud to be elected by this scholarship and that surely changed my life by giving me a chance to study and live in Japan.

In Sasase lab, I am glad that I was treated well by my fellow lab members. *Mr Maduranga Liyanage* has been sharing my hard time and giving lots of advices both about research and privately. Also *Mr Takamasa Isohara* and *Mr Motoki Shirasu*, who were my senior during the time I entered Sasase lab, have given me a lot of support in adapting the life in Japan as well as research. I would also like to thank *Mr Maung Sann Maw*, *Mr I-Te Lin*, *Mr Junji Asada* for their encouragement and stimulating my work. Also I would like to thank *Mr Hidetoshi Kajikawa*, *Mr Shinji Kano* and *Mr Kentaro Toyoda* for sharing some of my works in the lab. Surely credit must also be given to all the members of Sasase Laboratory, Department of Information and Computer Science, Keio University, for stimulating and encouraging my work.

During the time as a research associate (RA) in the Global Center of Excellence (GCoE) program, many of the fellow RAs have given me support on my works, especially to the members in the Project 3 "Real-world and Real-time Network for Multi-dimensional Processing and Communication". I would like to specially thank *Mr Souihli Oussama*, *Mr Chinnapat Sertthin*, *Dr Mamiko Inamori*, *Dr Kochi Adachi*, *Dr Sho Shimizu*, *Mr Kou Kikuta* and other members in the GCoE program for giving me support and sharing some of my hard time.

Finally and the most importantly, I would like give my greatest appreciate to my beloved father, mother, brother and sisters. To continue studies in the graduate level was a difficult decision, especially I am studying away from my home town. My family members respected my decision to continue my further studies and have been supporting me continuously. Without them I cannot even start my studies, not to mention to succeed in this work.

Fung Chun Cheong Alex
August 2010

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

With the increasing number of applications and devices making use of the network, the demand of high network bandwidth has been increasing sharply in the recent years. Although the advancement of different device technologies helps satisfying the demand, it involves a very high investment of equipments by the network service providers. The method of distributing data can also improve the usage efficiency of available network resources greatly. Multicast is a technique to deliver data from one point to multiple points effectively that the same stream of data is not sent more than once on the same path. Data stream is duplicated at the different branches of the network such that the data stream can be delivered to multiple destinations. The whole set of paths forms a multicast tree and this tree enables the efficient data delivery. However there are problems of how to allocate network resources on different layers of the network, and how to handle the complexity of the multicast delivery model.

Chapter 1 presents an introduction to the multicast data delivery. I will first introduce multicast in the Internet Protocol (IP) layer, including the related protocols which realize multicast data delivery. Next the reliable multicast will be introduced, in which data have to be delivered correctly to the recipients correctly. Although a large application area multicast is multimedia data delivery in which some data loss is tolerable, there are increasing demands for reliability in multicast such as distributed computing. Structure of the multicast tree for reliable multicast and the data recovery techniques are discussed. Then multicast tree in the optical network will be introduced in which wavelength division multiplexing (WDM) technique is employed. In the optical WDM network resources are being allocated to different sessions. I discuss the issues and limitations on the WDM network in order to setup a multicast session.

Chapter 2 presents the proposed reliable multicast protocol using local retransmission and forward error correction (FEC) based on group-aided multicast (GAM) scheme. In

reliable multicast, feedback and recovery traffic limit the performance and scalability of the multicast session. In the proposed scheme, the original GAM is being improved by making use of FEC locally in addition to negative acknowledgement (NACK)/retransmission in its local-group based recovery. Our scheme produces FEC packets and multicasts the packets within the scope of a local group in order to correct uncorrelated errors of the local members in each group of the multicast session, which reduces the need for NACK/retransmission. By using the proposed scheme, redundancy traffic can be localized in each group within a multicast session, and the overall recovery traffic can be reduced.

Chapter 3 explains the proposed scheme for multicast routing and wavelength assignment for dynamic multicast sessions in WDM network using minimum Δ. In this scheme a light-tree for dynamic multicast session for the WDM network is established by choosing the wavelength that leads to a reduction in blocking probabilities by using a parameter Δ. Δ is defined as the overall reduction of connectivity of the nodes in the network caused by a wavelength assignment process when using a particular wavelength, and wavelength resources to the multicast session are being assigned by choosing Δ which leads to smallest reduction in connectivity. Through computer simulation, it is shown that the proposed scheme has lower blocking probabilities when compared with minimum cost scheme under the condition that wavelength conversion is not allowed.

Chapter 4 concludes this dissertation with an overall discussion of the multicast data delivery and techniques discussed throughout the thesis.

CHAPTER 1

INTRODUCTION

1.1 General introduction

Use of the network has become an inseparable part in every day life and business. The Internet traffic growth rate has been increasing rapidly in recent years. The consumer applications and services enabled by the Internet has evolved from traditional low bandwidth contents such as text and image to bandwidth hungry traffic like voice and video. In term of the business, they make use of the speed of the network in order to keep the most up-to-date information to remain competitive. Also with the decreased cost of the network-enabled mobile devices, the population of Internet users is increasing rapidly. While users in the developed countries are demanding for more bandwidth intensive contents, in developing countries there is a huge increase the number of user who have access to the network. For instance, in Japan the total Internet traffic as estimated by the Ministry of Internal Affairs and Communications has been increased from around 500Gbps to 4000Gbps, both uplink and downlink traffic (Figure 1).

In recent years, the mobile devices especially characterized by the recent growth of smart phone type of devices, are having the processing power of a personal computer just a few years ago. Compared to a traditional mobile phone few years ago which is capable for voice call and short message service (SMS), a typical smart phone is equipped a mega-pixel camera and high resolution screen. Also in contrast to the light contents which are specially written for mobile device, like pages written with Wireless Application Protocol (WAP) or i-mode, high bandwidth contents such as full webpages (as seen by normal computer) with which multimedia contents, are being sent in and out of the mobile devices. These devices are usually equipped with camera which is capable to take high resolution photos and video, and users are only required a small amount of operations to upload the mobile contents online. With the raise of these kinds of device, the mobile operators are offering flat rate

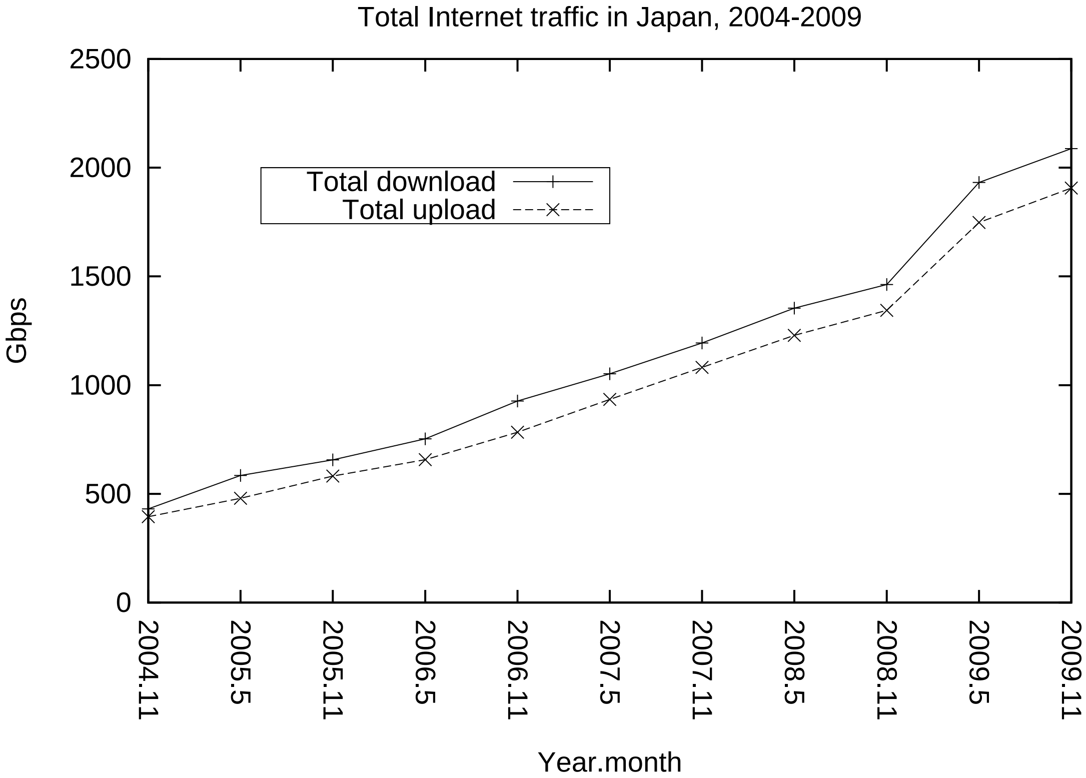


Figure 1: Total Internet traffic in Japan (source: Ministry of Internal Affairs and Communications, Japan, Feb 2010)

service where users can use virtually unlimited amount of data. Also a recent trend that mobile devices are no longer limited to only laptops and mobile phones. As of year 2010, the tablet type devices are gaining popularity. The diversity of the mobile products and services leads to enormous increase of mobile data traffic.

Figure 2 shows the global consumer traffic growth from 2004 to 2009 which features the traffic flowing to the ISPs for the consumers. Multicast traffic has the potential to reduce the overall bandwidth usage of some of the scenarios. For example, when update files for a popular program become available, it can be distributed into different locations of the network, or the mirrors, such that people can download it quickly from different mirrors. It is also important for realizing cloud computing [1] where in cloud computing people do not aware the real location of where the data are stored, and where the computation is being done. For some computational intensive operations, it is needed that data are to be sent and retrieved to different locations of the network in order to distribute the load and utilize

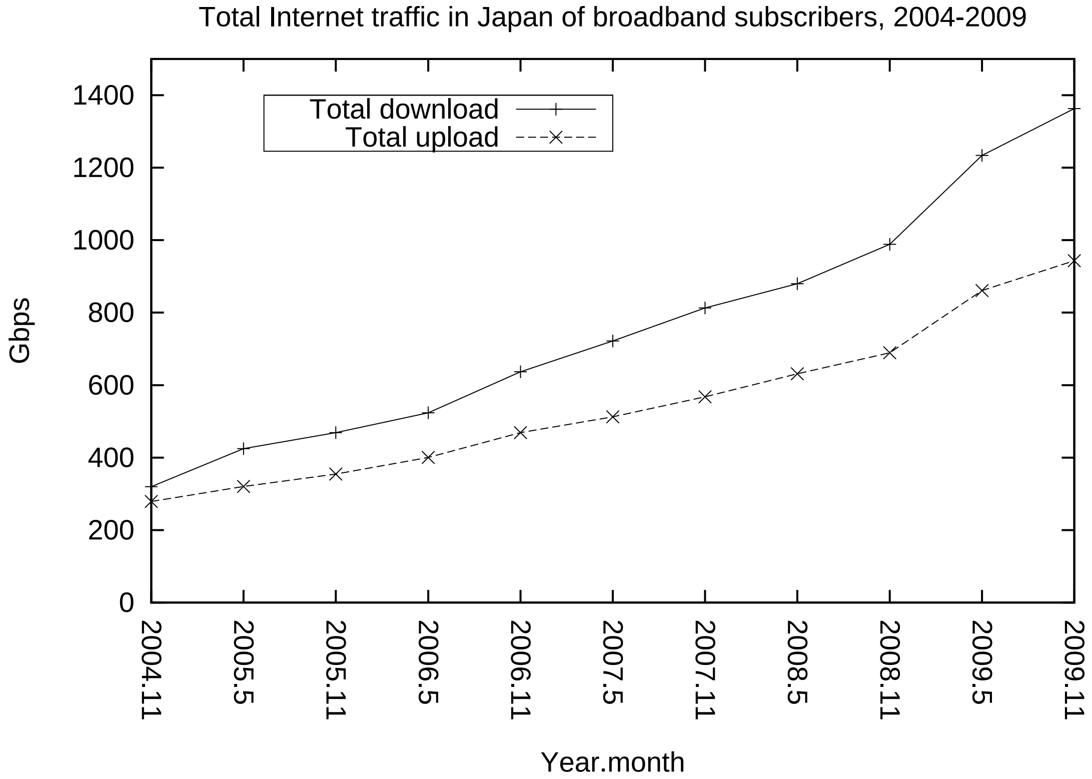


Figure 2: Total Internet traffic of broadband subscribers in Japan (source: Ministry of Internal Affairs and Communications, Japan, Feb 2010)

the computational resources.

Video contents are major consumers of the traffic due to the high bandwidth requirement and increasing popularity featured by the growth of video sharing website, paid video-on-demand services and real-time broadcasting of video. For video contents of some live events such as an important soccer match or debate on president election in some countries, many people consume the same contents at the same time either on their PC or TV set-top boxes. As the video quality is becoming better in order to satisfy the high-definition TV, larger bandwidth will be required. This broadcast creates an additional burden to the network providers.

In client-server model, peer-to-peer model or the broadcast model, when the same content is demanded by group of people, the transfers involve a source and multiple recipients in the process. If the transfer to the recipients are to be occurred at the same time, it would be more network efficient to have the source to send a copy of the data to the network, and let the network to duplicate the copy of the data to the corresponding recipients, where the recipients can be end-consumer of the content, or the mirror sites which are separated geographically. This process is called multicast. And in this work some issues in the multicast are being discussed.

Figure 3 shows the difference among multicast, unicast and broadcast. The network is represented as nodes and edges as shown. In this scenario data packet is being transferred from a source node S to destination nodes D_1, D_2 and D_3 as shown in the figure by means of unicast, broadcast and multicast respectively. In unicast the data packet is being transferred to D_1 and D_2 as two separate copies through link $S - R_1$. Thus it can be seen that R_1 has received the same data packet twice in case of unicast. In case of large set of recipients this can be inefficient as the intermediate node has to handle the same data packet repeatedly.

In broadcast, data packet is being sent by a router to every possible link except the link where the data packet arrived from. In contrast to unicast, data is being sent only once on each link of the network. However the nodes which do not want the data or those who has no nodes want the data in the downstream are going to receive unnecessary data. Because of the flooding nature of broadcasting, data stream usually only travels for a limited range, for example, within the same company network.

In multicast, the intermediate nodes are supposed to duplicate and send data only to relevant downstream nodes to avoid the flooding problem as seen as broadcast. In Figure 3 it can be seen that R_1 duplicates the data stream received from S into two copies and send the data to D_1 and D_2 respectively. It can be seen that data stream only travels on the network for exactly once and D_1, D_2 and D_3 can receive the data they want, and other nodes which have no interest on the data will not get irrelevant transfer.

To realize multicast, it is required that the intermediate nodes setup and remember the "state" of the multicast in order to duplicate and send data packets to only relevant downstream nodes. The "states" of the involving hosts and routers connect together and form a distribution tree. This is referred as a multicast tree, which is a sub-tree on the network topology which connects the source node to the destination nodes. Intermediate nodes or routers in the IP network, even they are not the destination nodes, help maintaining the multicast tree and direct the incoming data into one or more downstream nodes, which can be destination nodes or other intermediate nodes.

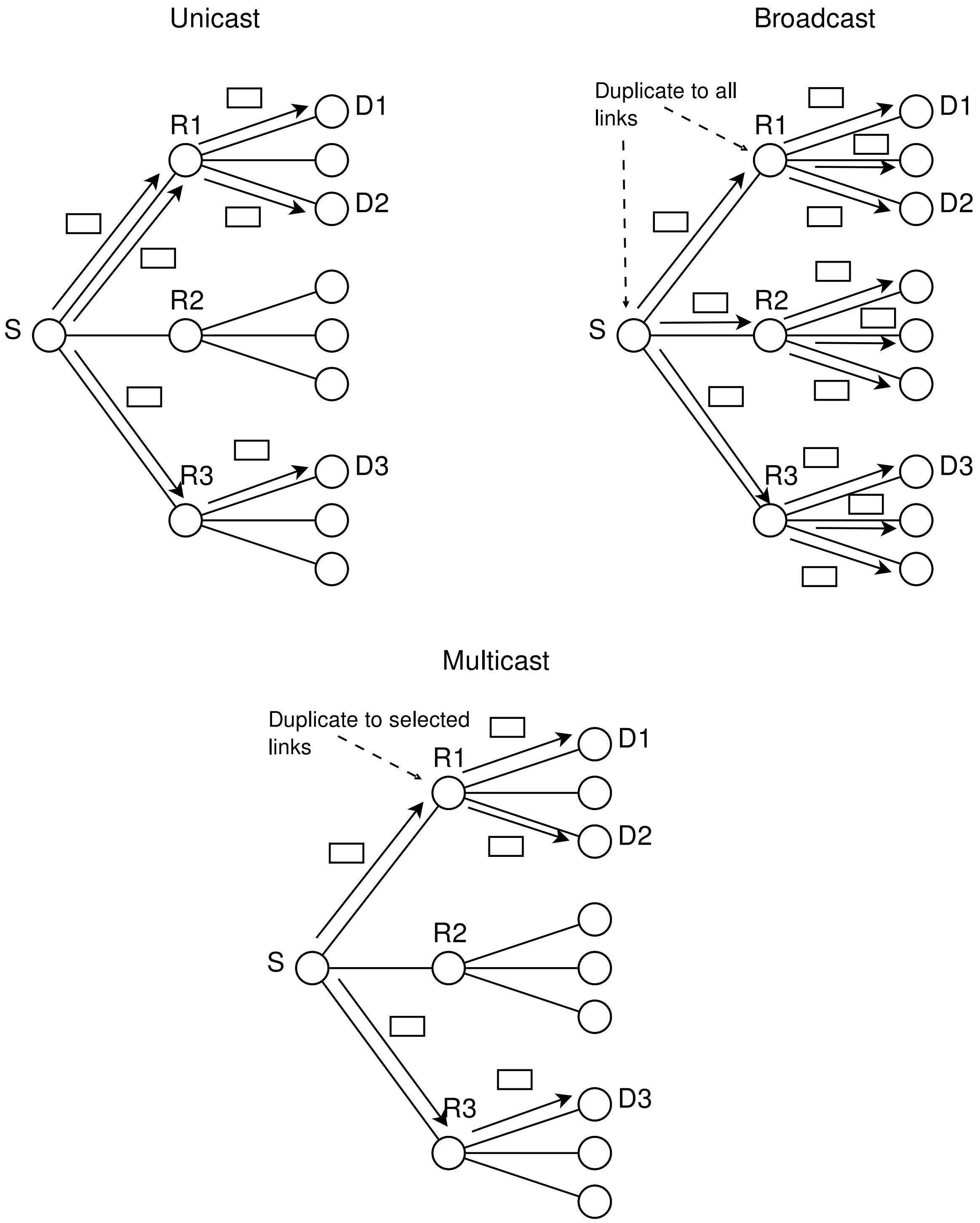


Figure 3: Multicast vs unicast vs broadcast

Multicast can be realized on different layers on the network. On the Internet Protocol

(IP) level (Chapter 1.2) the routers which support multicast protocols can be responsible for setting up the multicast session such that data can be sent to a specific multicast address and valid subscribers of this multicast address can receive the data packets. On an overlay network constructed over peer-to-peer nodes, the underlying network structure is less relevant. The multicast tree is constructed over the overlay nodes. It is not completely network efficient in the IP level as it ignores the underlying network structure, but it is simple to realize as long as the peer-to-peer application supports the signaling such that the end hosts can distribute the data efficiently. This is referred as the application layer multicast [2, 3, 4, 5]. On the other hand, lower layer network can also realize multicast. In wireless ad-hoc network, there is no wire connecting each node and data is being sent a broadcast manner. Multicast is realized in some wireless ad-hoc network routing protocols that in general, when a node hears broadcasted "multicast" message from its neighbor, it determines whether or not it should broadcast according to the current topology of the ad-hoc network. On the other hand, for backbone network using Wavelength Division Multiplexing (WDM), signal are carried with optical carrier of different wavelength in the optical fiber. In the optical node, this optical signal can be split into multiple copies in the optical domain with splitter. This enables multicast where the same optical signal to be transmitted from one optical node to multiple nodes.

1.2 IP multicast and reliability

IP multicast refers to delivering IP datagrams to multiple users in a single transmission. In IP multicast, it is necessary to have multicast capable routers exist in at least some part of the networks. In general, to enable multicast transmission from a source S, first a multicast address G is assigned to a session characterized by (S, G). G is the address which source node sends the data packet to. In order to receive the packet from the sender, receivers need to subscribe to the corresponding multicast address G via the router that the receiver is attached to. When a router determines there is at least one host attaching to it subscribes to the multicast group, it uses protocols like the Protocol Independent Multicast (PIM) protocol to build the distribution tree which connects all the other multicast capable routers.

Multicast tree can be classified into two types in the IP network: "shared tree" and "source tree". The shared tree only the group address is important. Any group members can send data to the multicast address G, and the data packet is being distributed to all other members in the shared tree. In "shared tree" each branch of the multicast tree can support sending data packets in both directions. This is suitable for session like a video conference applications that each participant is sending and receiving data. The other mode, "source tree" is suitable for the distribution of data to a large number of recipients where the data are basically sent from a single source. "Shared tree" is usually built with lowest overall possible cost such that for overall traffic will be minimized, where in "source tree" the distance from the source of the multicast session is usually minimized.

Unlike a unicast protocols like Transmission Control Protocol (TCP), the members in the multicast group can join and leave at any time while the multicast transmission still take place for the other recipients. Therefore the multicast routers need to maintain the multicast tree according to the active members in the session. When member leaves sometime it's necessary for the router to prune the tree if there is no other member downstream, and on the other hand new branch may have to be added when there is new member joining the session.

1.2.1 Reliable IP multicast

The Internet relies on the packet-switched network to provide efficiency of data delivery by allowing packets come from different sources to arrive at different destinations where packets are switched to effective routing methods. However the packet-switched nature of the network is not reliable. Packets can be lost due to the following reasons:

- Incorrect routing states caused by route change
- Corrupted signal in the physical medium, such as lossy wireless environment
- Bursty traffic causing the routers running out of buffer for accepting new packets
- Attackers who cause the network to malfunction

- Large delay on the network that causes the receiver determines data are lost

Some principles and problems in reliable multicast follows will be discussed.

1.2.1.1 Loss Notification

Positive acknowledgement (ACK) is a widely used technique for the sender to learn about the receiving status of a receiver. In ACK, the receiver reports which packet (or group of packets) it has received to the sender. The ACK contains the receiving status of the receiver which is identified by a sequence number. Upon receiving the ACK, the sender can determine if loss has occurred at the receiver's side based on the information contained in the ACK packet, and send out outstanding packets or lost packets to the receiver. ACK packet is sent by the receiver upon receiving a certain number of packets. Positive ACK has the advantage that the sender can make sure the receiver has received the packets correctly, as feedback is provided regularly. The sender and the receiver use round-trip-time (RTT) as an indicator to estimate when the feedback packet should be arriving. If the feedback packet is lost in the network, and the sender/receiver has been waiting a long time (relative to RTT), it will request for the other side to send the packet again to compensate the lost packet.

However when the same approach is used in the case of multicast, every receiver needs to report the receiving status to the sender, and the sender has to keep track of all individual receivers' statuses, which include information like RTT and received sequence number), and handles all the feedback traffic. This causes the ACK implosion problem [6], especially when the group size grows. From Figure 4, even the data are sent by the sender in the form of multicast, feedback packets are sent by receivers in unicast and individual recover has to be done in unicast as well.

On the other hand, using NACK (negative acknowledgement) is much more suitable in reliable multicast. While ACK allows sender to know the receiving status of the receiver, even if all packets are received correctly, in the case of NACK the receivers are responsible of detecting loss and explicitly request for retransmission. If data are received correctly, no NACK will be sent from the receivers (see Figure 5). This has the advantages that the

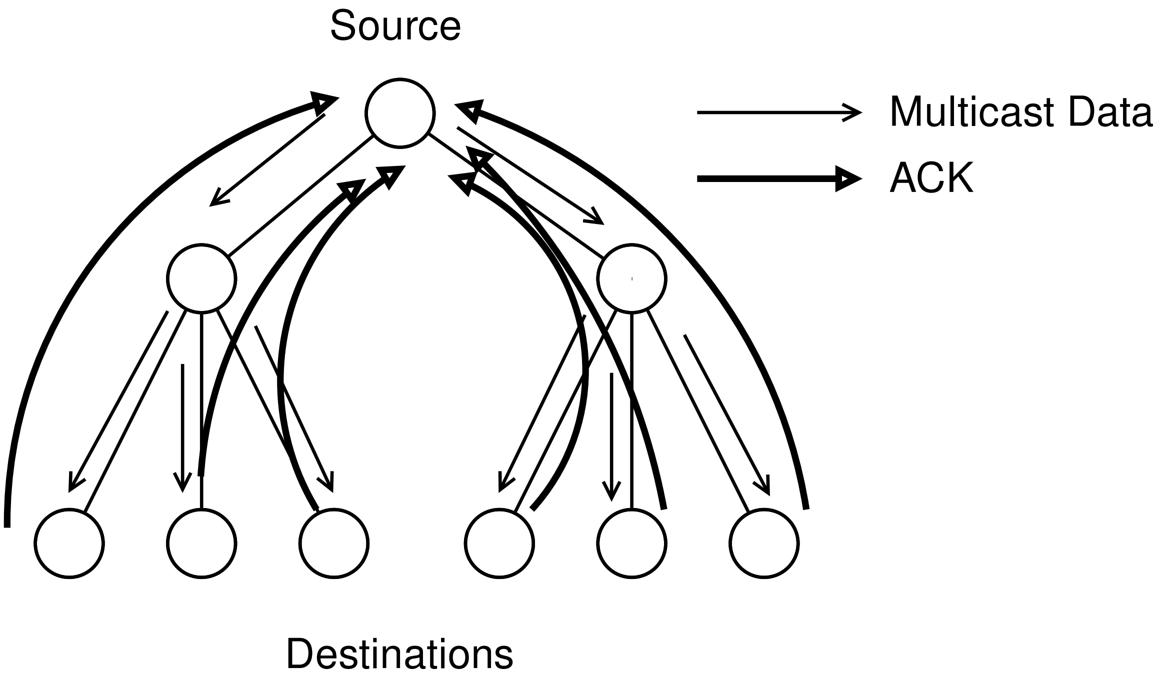


Figure 4: ACK implosion in reliable multicast

request is only sent when needed and the sender no longer need to keep track of receivers' status, but at the same time if the feedback is lost the sender will not be notified and the burden of keeping reliability will fall onto the receiver side. For multicast session, this is especially good when the loss rate is low. However, using NACK can only reduce, but not eliminate the need of feedback traffic when compared with the ACK approach. Especially when the loss rate is high, the NACK feedback traffic can be very large. Therefore the implosion problem is not solved by simply using NACK. Also when there is no feedback from a particular receiver, it is impossible for the sender to determine whether the receiver is receiving the data perfectly, or the receiver (or link to receiver) is down and cannot produce any response. Depending on the reliability requirements of the multicast application, NACK could be inappropriate sometimes.

When sending the NACK, choices between using multicast or unicast have different impacts on the multicast session. If NACK is sent using multicast, other receivers of the multicast session could also listen to the NACK and suppress their own NACK, thus avoid

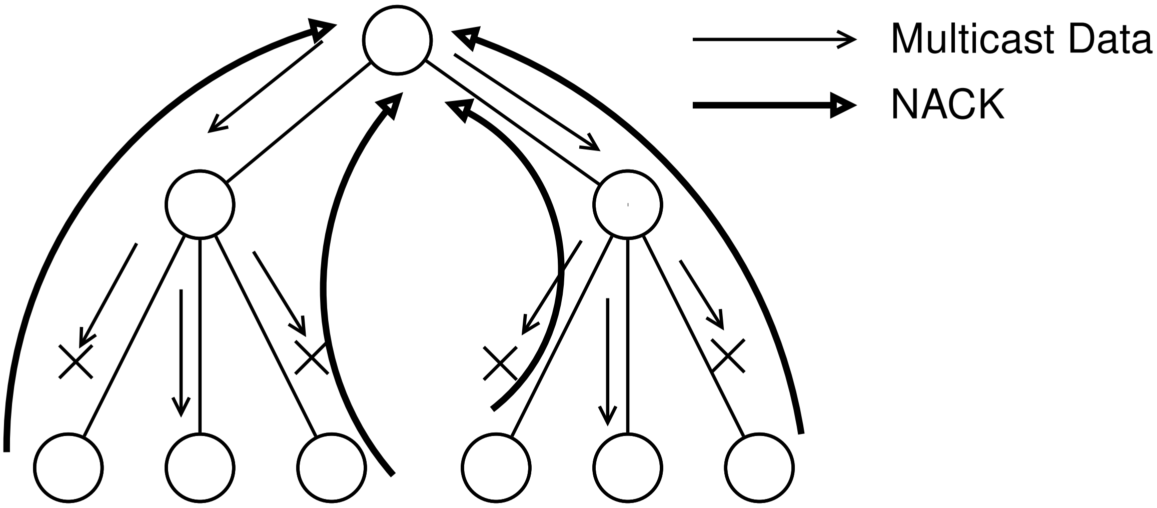


Figure 5: Using NACK as multicast loss feedback

overwhelming the sender of the multicast network. In this case, the receivers should maintain a timer which send out NACK in a probabilistic manner in order to prevent sending same NACK at the same time of other receivers. If the losses suffered by the receivers in the multicast sessions are relatively high, sending NACK by multicast can effectively suppress excess feedback traffic to the sender. On the other hand, when most receivers suffer from relative small loss, multicasting of NACK as a signal of global packet loss is not representative. In this case unicast would be more preferable.

Different methods have been proposed to provide reliable multicast service on the network while maintaining the scalability of the multicast session. The follow sub-sections show some of the important techniques in reliable multicast.

1.2.1.2 Recovery with Resend

When the multicast source receives the ACK/NACK from the receivers, it could send again the data packet(s) to the receiver. This process is referred as Automatic ReQuest (ARQ). In a particular instance within a multicast session, there is chance that many receivers fail to receive a particular packet P, while many other receivers received the packet correctly. If the sender sends the repair packets by unicast, the bandwidth usage on the sender's link will become large when many receivers have lost the packets. If the sender instead sends repair packets by multicast, in this case those receivers who have correctly received the repair packets will receive duplicate packets, and again the bandwidth will be wasted. This problem is known as the repair locality problem. Thus determining whether using unicast or multicast for packet resending is difficult. When packets are rarely lost, unicast resending is naturally a good choice because of its simplicity.

It is important to note that in IP multicast, data delivery path from a particular source is only in one direction. This means that there will be no return path provided in the IP multicast mechanism. An explicit ACK tree is needed in order for receivers to send feedback to the sender. Ideally this return path will be in the opposite direction of the multicast tree, however because of some limitations like asymmetric link, a different path will be used.

1.2.1.3 Recovery with FEC Packets

Sender can append additional redundancy packets which are served as a proactive approach for the clients to repair the loss packet themselves. Packet level FEC recovery is originally designed for time sensitive multicast applications like multimedia streaming. As in the ARQ approach, at least one round trip time is required for the repair packets to come. Additional FEC packets from the sender eliminate this round trip time at the expense of increased bandwidth requirement, and overhead on encoding/decoding FEC packets.

However the FEC approach can also be useful for reliable data multicast. FEC packets are produced from a block of data packets, and it can be used to recover any loss data packets in that block. When different receivers lose packets independently, each of them will observe independent loss. In this case FEC packets can help repair the loss without the need of prior knowledge of which exact packet the client lost. This means in many cases feedback from the receivers are not needed. In the case of block erasure codes [7] if $(n - k)$ FEC packets are produced from k data packets, up to $(n - k)$ packets can be repaired as the sequence number is known in packet level. If more than $(n - k)$ packets are lost in a particular block, the block cannot be repaired with the FEC packets alone. Therefore it is common to use FEC recovery technique with the NACK/resend technique (Hybrid-ARQ) together to provide reliability.

The FEC repairing approach, however, suffers from varying network conditions, as well as heterogeneity of a network. The number of redundancy packets $(n - k)$ per block is determined from some estimations of network condition, and in general this number should be close to the average number of loss of a block. However the sudden changes in network conditions can make the FEC packets insufficient to serve for recovering loss, while in some parts providing excessive recovery traffic. Also in a large multicast session, receivers on different parts of the network can have very different long term loss observation. So it is very difficult for the source to determine the proper redundancy size in advance.

In the past the FEC scheme was not popular because of its computational cost on encoding/decoding redundancy packets. However with the rapid advancement of CPU technologies, this cost is becoming less significant.

1.2.1.4 Local recovery

It is not possible to achieve scalability if the sender is the only node who are going to provide recovery services, whether it is ARQ based or FEC based. In order to be scalable, the processes including processing of feedback messages and providing repair traffic, should be off-loaded to other recipients in the network. In a reliable multicast session, designated receivers are assigned and are responsible for providing local recovery on behalf of the original sender. Figure 6 gives an illustration of the local recovery which helps to reduce the loading of the original source. As multicast topology formed a spanning tree on the network, designated receivers are assigned on the mid-way of the multicast tree from the sender to the receivers. When receivers detect loss, they simply send request (NACK) upstream (left-hand side of Figure 6). This request will be captured by the designated receivers, and these designated receivers will provide local recovery for the receivers, if the requested data are stored in this designated receivers. However there is a chance that the designated receivers do not keep the received data packet because of limited memory, or the designated receivers have not received the data packet from the beginning (right-hand side of Figure 6). In this case the final receivers of the session will report the loss with NACK to the designated receiver. If the designated receiver determine that the lost packets are the same, it can aggregated all the NACKs sent by the receivers and report to the original sender for recovery. In this case, the sender needs to process only one NACK request from the designated receiver and sending out one recovery packet instead of three as shown in the figure. The designated receivers help in distributing the recovery loading in the multicast session.

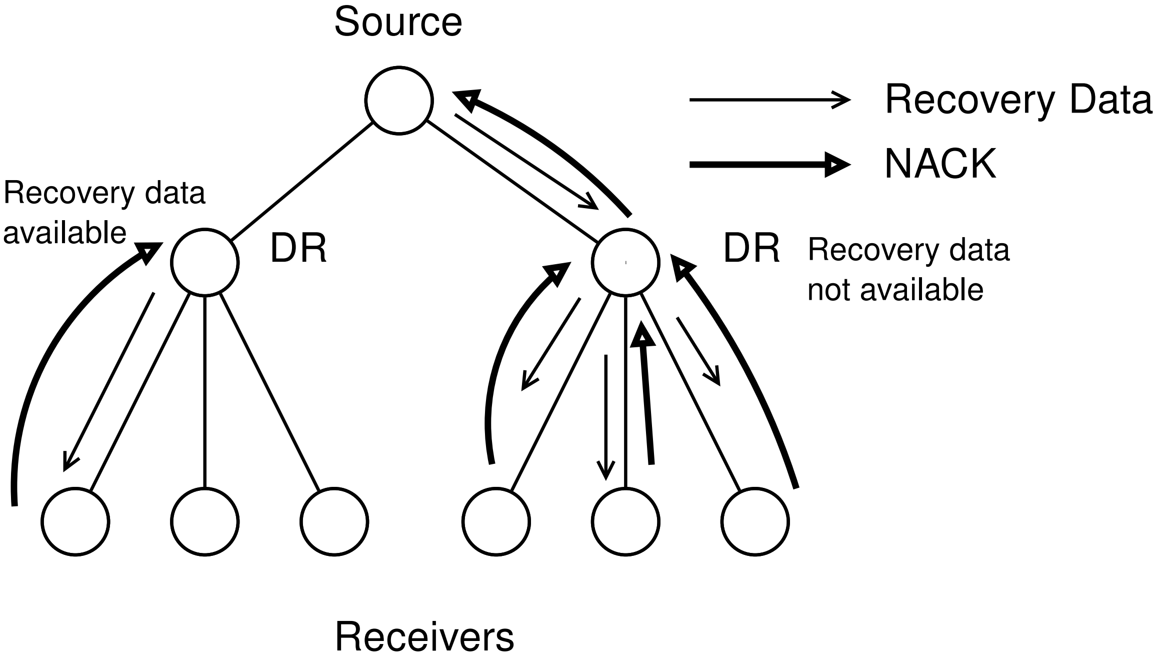


Figure 6: Local recovery

Based on the dynamic nature of a multicast session, the proper assignment of the designated receivers has been a challenging problems. Also in case of a many-to-many (m-to-m) multicast session, it will become more difficult for the proper assignment of the intermediate nodes, as the senders can be located anywhere within a network, and a top-down approach does not work anymore.

1.2.2 Related Works on Reliable Multicast

1.2.2.1 Reliable Multicast data Distribution Protocol

Reliable Multicast data Distribution Protocol (RMDP) [8] suggests how to use error erasure code to provide reliable multicast, Suppose a particular resource (e.g. a file) of length L packets will be sent to the client. The source data are encoded with high redundancy, using an (n, k) code with $n \gg k$, and ideally $k = L$. With the use of erasure code, as long as any

k different packets are received, the receiver can reconstruct the original data from these k packets. The multicast sender sends packets (original data and redundancy) at a variable rate, depending on the number of clients connected. Clients on a fast link could keep up with the transmission rate, and prune themselves from the multicast tree after complete receiving the necessary packets. However some other clients who are on slower links or congested link will suffer from packets loss. These clients need to receive more data from the sender in order to recover the whole resource. If the server finishes transmitting all the data and redundancy packets and the receivers cannot reconstruct the original data (i.e. have received less than k packets) because of high loss rate or late joining, these receivers will initiate a continue request to the server and ask for the data and the redundancy to be sent again from the beginning. Given a sufficient large n in the erasure code, this continue request should be rare.

1.2.2.2 Reliable Multicast Transport Protocol II

Reliable Multicast Transport Protocol II (RMTP II) [9], based on previous RMTP, divides the nodes into clients (senders and receivers) and interior nodes (top node and designated receivers) for distributing recovery processing on the network. In the clients (senders and receivers) multicast data distribution through the data channel (the multicast tree) is take place. On the other hand, reliability control signal is collected and handled in the interior nodes through the control channel, which is part of the infrastructure and is controlled by network administrators.

In the reliable multicast session, data are distributed to the designated receivers in addition to the intended receivers. Tree-based ACK (TRACK) is used by the receivers to report to the designated receivers about their receiving status. The designated receivers aggregate the TRACKs and report to the top node of the hierarchy, such that the top node can make sure that all receivers are alive and receiving the data correctly. In case of data loss, it will be the responsibility for the designated receivers to provide local data recovery. Also, optional use of NACK for reporting error, and optional use of FEC recovery from the sender are supported in this protocol.

In this scheme, explicit network resources (the top node and designated receivers) have to be pre-assigned to provide the recovery service. This requires an expected knowledge of physical distribution of nodes in order to provide best service, and this is in some cases impractical.

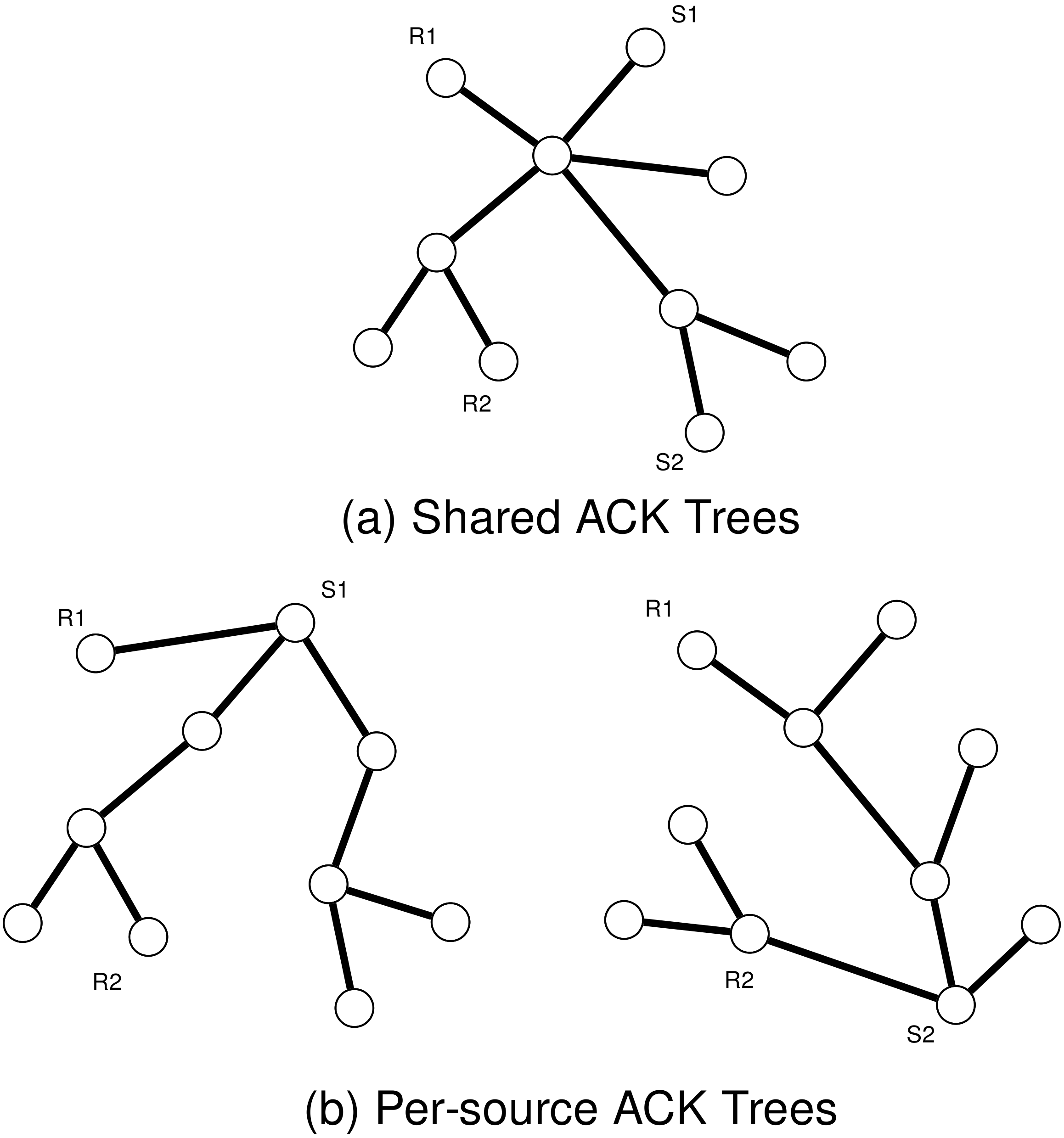
1.2.2.3 Scalable Reliable Multicast

Scalable Reliable Multicast (SRM) [10] suggests way of suppressing NACK traffic, and is designed for many-to-many reliable multicast scenario. SRM guarantees the eventual delivery of all the data to other group members, but data packets can be delivered out of order. This situation is considered to be the minimal requirement in reliable multicast, and ordering can be reconstructed in the higher level, if needed.

As there are more than one sender in the multicast session (all the participants of the session can be senders), tree hierarchy cannot be used as a method of aggregation of traffic. Instead, each receiver keeps packet arrival time T from a particular source as an estimation to the distance from the source. Whenever loss is detected from that particular sender, the receiver will wait for a time cT, where $c > 1$, and multicast a "repair request". This "repair request" is propagated along the multicast topology upstream and downstream (relative to the source). In this case those who are nearer to the source will have a higher chance of getting the required data, and thus, it can capture the repair request and retransmit the loss data downstream. On the other hand, those receivers who are further away (downstream) from the source will have a longer timeout. When they receive the "repair request" from upstream, they know some nodes on the upstream have lost the packets and therefore wait for the recovery to be multicast.

1.2.2.4 Group-aided Multicast

In Group-aided Multicast (GAM) [11] a two level hierarchy, which consists of one core group and number of local groups, is defined for many-to-many reliable multicast. It is observed that the other schemes, ACK tree (feedback path) assignment comes into two extremes: and single shared tree as shown in Figure 7(a)(as in SRM) and per-source logical trees as shown in Figure 7(b)(as in RMTP II). While these two extremes have their own advantages and disadvantages, GAM makes use of both of them and found a balance point between the two.



(a) Shared ACK Trees

(b) Per-source ACK Trees

Figure 7: Shared ACK tree versus per-source ACK tree

Per-source logical trees require all the clients to remember the return path (or upstream path) to each source in order to keep the shortest path from the receivers to the data source. If the number of senders grows, receivers will need to maintain a huge table to remember each source location which in turn limits the scalability of the multicast session. In the shared tree approach the receivers only need to remember one tree in the multicast session. However in this case the it is possible that the route from the receivers to the data source is far from optimal when compared with per-source approach (see the route from R1 to S1, R2 to S2 in Figure 7).

In GAM, as shown in Figure 8, nodes which are close to each other will form local group. In this local group, a shared ACK tree will be used to connect the nodes as they are close to each other and overhead on routing is considered to be small. In each local group one node is elected as a core (or statically assigned manually by the network administrator). Cores from different local groups form one core group. This core group is connected by per-source ACK trees and therefore optimal routing between groups is possible. To each node, data source is either within the local group or in the other group which is connected to the core. Thus routing is kept simple for all local members in the expense on the complexity of the core nodes, while when compare to the pure per-source ACK tree approach the complexity is still significantly reduced.

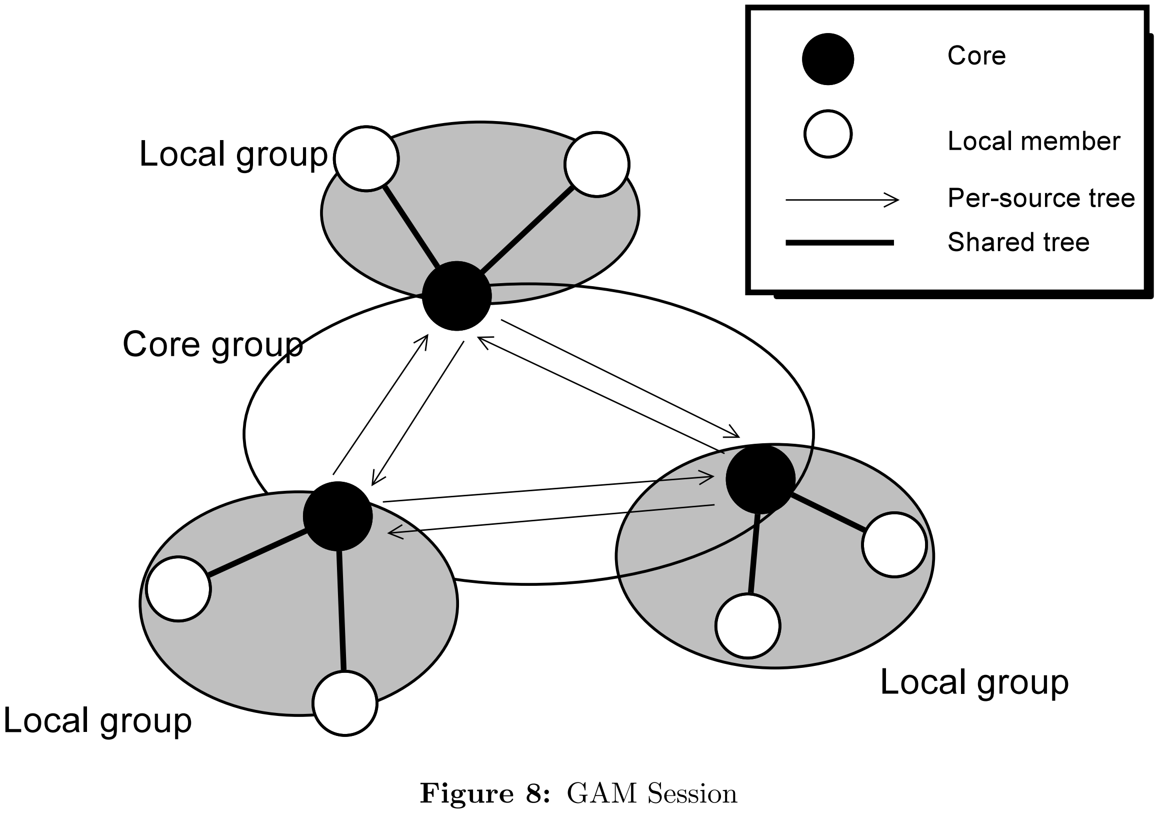


Figure 8: GAM Session

The GAM method employs a two-level simple NACK based scheme. In this scheme when loss is detected (a gap in the received sequence number), local member first sends NACK to the local group root (the core) to see if recovery packet is available. The core retransmits the copy of the loss packet to each receiver with unicast, if it has correctly received the required packets. If loss occurs also at the core, the core sends NACK to the original sender to ask for the repair packet, and after the core has received the repair packet, the packet is sent to the local group members with multicast. However, if the size of local group becomes large, the core will suffer from large NACK traffic and will need to send large number of repair packets to its local group members. This causes NACK implosion problem to the local group core.

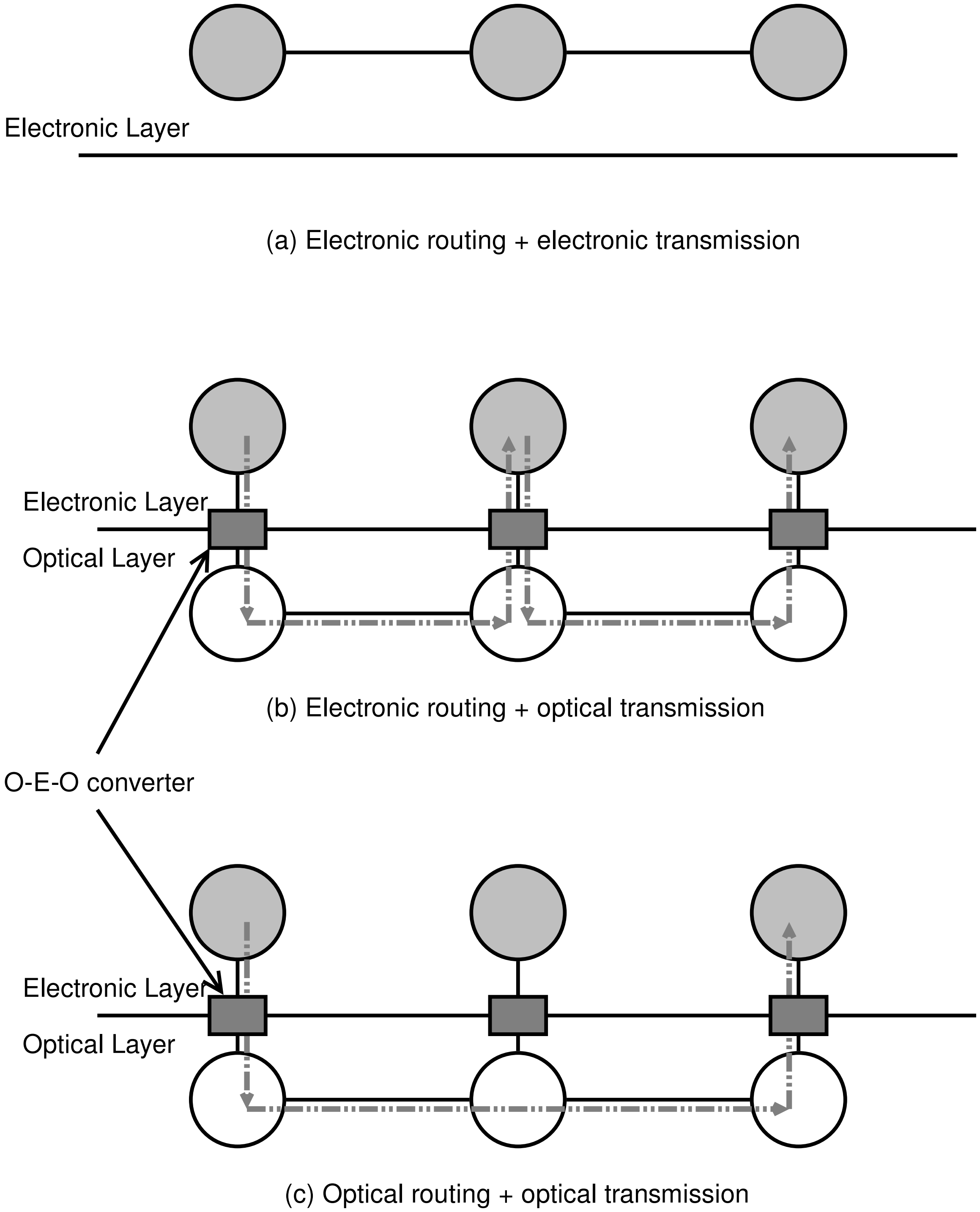
1.3 Multicast in optical domain: light-tree

In this session the nature of the optical network, the principles in realizing multicast in optical network, the challenges from the limitations of optical network for multicast, and some related works in multicast routing and wavelength assignment in the WDM networks will be discussed.

1.3.1 Transition from electrical network to optical network

While IP multicast works on network layer protocol which operates on the topology as seen by all the hosts and routers on the network, making use of the lower layer network properties is a key to provide higher quality service for the multicast data transfer. On the Wavelength Division Multiplexing (WDM) network, the optical multicast can further improve the use of network resources with its multicast capabilities.

Before the use of optical fiber, the network has been based on copper wires (twisted pair or coaxial cables) to transmit signal electronically (Figure 9) (a)). Network equipments and protocols are designed to cope with the electronic signal transferred. For the IP network, a router process each incoming packet it received from the Media Access Control (MAC) layer. First a packet is decoded from the physical interface and stored in the input buffer of the router for processing. When processing a packet, the router read the header information of the packet, where destination address is included in the header. Then the router moves the packet to the output queue of corresponding port according to the look-up result of the



(a) Electronic routing + electronic transmission

(b) Electronic routing + optical transmission

(c) Optical routing + optical transmission

Figure 9: Transition from electronic to optical in routing

routing table. After that the packet is sent on the corresponding MAC layer interface such that the packet can arrive at the other network node.

With the introduction of optical fiber, the connections between network nodes are replaced by the optical fiber to provide much higher bandwidth (Figure 9) (b)). However, the mechanism that a data packet is delivered to the MAC layer of the router and being processed with the same process as described in the previous paragraph remains unchanged. This creates another problem of heavy router load as the optical fiber is delivering data packets in a much higher rate than the conventional copper wire. This requires a large amount of buffer to store the packet, as well as fast enough processing and routing for the packets. This technique is widely used in networks like Synchronous Optical Networks (SONET) and Synchronous Digital Hierarchy (SDH) networks. With this technique, traffic of gigabit per second is feasible. However the gigabit per second data rate is limited by the bottleneck caused by the electronic processing but not the fiber itself, which has potential to carry much higher data rate.

With the advancement of the optical switching technology, it becomes more reasonable to just setup the optical route in the optical node such that the routing can remain in the optical domain (Figure 9) (c)). In contrast to the packet routing approach where the header of the packet is used for routing, in optical network each stream of data is sent on particular wavelength carrier. This approach is called the Wavelength Division Multiplexing (WDM). Each stream carried by the optical wavelength channel can carry traffic of very large bandwidth (100Gbps per channel is available, where 10Gbps and 40Gbps can be commonly found). The high bandwidth is meeting the increasing demand of high-definition video and 3D supports for video. While normal Internet usage usually would not reach this bandwidth, it is possible that different providers can aggregate the traffic into the edge router such that bandwidth of this large traffic is not uncommon in nowadays Internet usage. This process is called grooming, where traffic streams of bandwidth less than a wavelength can share the same wavelength channel for data delivery. With the increasing amount of traffic with the internet population and contents/services provided on the Internet, a huge guaranteed bandwidth is always preferable to provide stable service to different parts of the network.

As an optical node is able to distinguish the wavelength of the light-signal without looking at the contents of the packets carried, the signal of a particular wavelength can be easily routed or directed to the upper layer. Also because when the routing is purely operated in the optical domain, there is no electronic processing of the optical signal until the stream arrives at the end point, and the speed will not be limited by the electronic counterparts of the intermediate nodes of the network. Thus WDM is said to be overcome the bottleneck exists in the electronic routing components. In WDM routing network, optical crossconnect (OXC) are used in the optical switch to route the input port to output port. This will be discussed in further details in next sub-session.

In real-world network, the network resources of different parts are controlled by different providers, however in order to provide inter-connection, the providers have to agree with each other on providing information on the network resource which is open for interconnecting. For instance, Internet Exchange Point (IXP) is the entity where different providers connect different cities in the world. The resource and cost are shared by different providers under mutual peering agreement. Therefore while traffic information within a provider's autonomous domain is not opened, traffic on the IXP resources are shared by the participants of the network.

1.3.2 All Optical Multicasting

For one-to-one unicast where stream of data is transferred from one node to another node. However if the light signal is split into multiple copies to different destinations by passive optical device. After the splitting the signal is no longer a "path" but a "tree", or the light-tree. By using the light-tree approach, multicast can be realized in the optical WDM network. As the routing and multicasting do not require the signal to convert into the electronic domain, data transfer to multiple recipients within only the optical domain become possible. The multicast support in optical network is sometime referred as All Optical Multicasting (AOM). The AOM shares some of the advantages when comparing unicast in optical domain and electronic domain, namely:

1. Signal transparency: The optical routing is independent of signal type, modulation, coding, bit rates and underlying protocols used. This independence is referred as

signal transparency. This allow the optical network to be used by different upper layer protocols. Also the there is no need for the intermediate optical node to understand the protocols or decoding as data are not required to be sent to the upper layer, this means simpler device can be designed.

2. Multicast simplicity: light signal is split by a power splitter in the optical domain, as compared to the copying operations which occur in the electronic domain. This has the advantage of fast splitting of signal into different output of an optical node. Also as only splitting but no electronic copying is required, the node does not need to store the data stream into a buffer, which saves the huge memory requirements and processing power of the electronic counterparts.

3. Reduced network layer complexity: As the multicast operation is moved to the optical layer, routers in the electronic domain do not have to handle the multicast operation. This reduces the load of the routers in the upper layer. Also in the optical layer the optical nodes have less need to convert the data into electronic domain for the upper layer router to conduct the multicast service, and this reduces the workloads of the OXC.

4. Robustness: The multicast operation in electronic domain which includes store-and-forward within all the buffer which creates some processing delay. Splitting light signal in the optical domain using passive device to achieve multicast does not involve the delay. Therefore small delay would be suffered from the system.

1.3.2.1 Optical switch supporting multicast

To support network multicast, splitter have to be included in the optical switch such that light-signal from an input port can be split and sent to multiple output ports. Different structures of an multicast capable switches were proposed. Figure 10 [12] shows an example of how the multicast capable switch can be designed.

In Figure 10 an optical signal arrived at incoming fiber D with wavelength λ_b, which is supposed to be split and direct to two output ports as well as the local drop (upper layer

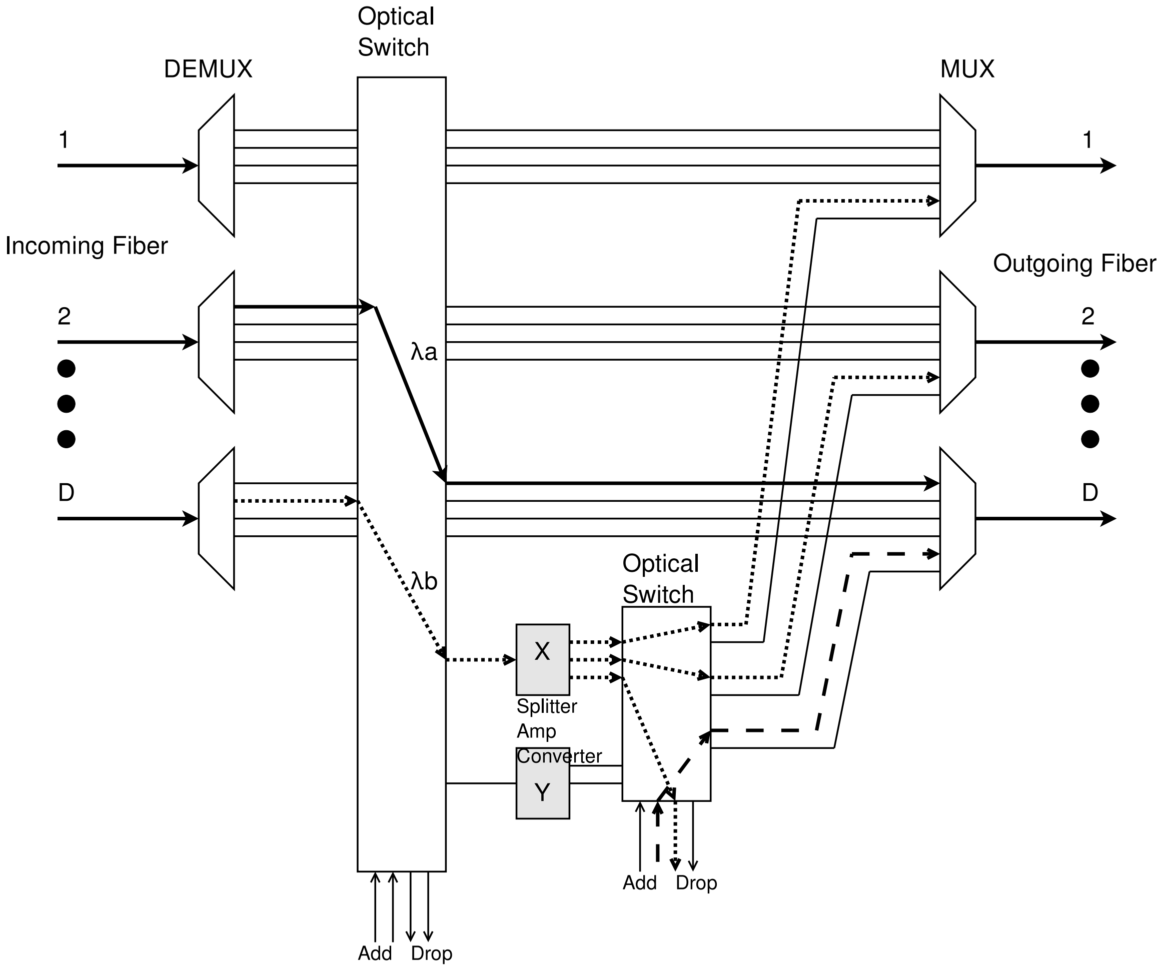


Figure 10: Example structure of a multicast capable node

of this optical switch). The first optical switch in the figure first switch the light signal in to the splitter X. Splitter X splits the signal into three identifcal copies (as well as amplifying the signal) and these three copies are switched into the second optical switch. Here the second optical switch directs the three copies into the two corresponding output ports (and one copy for the local drop). Inside this switch all of the operations are operated in the optical domain. The optical switch in the figure are controlled by some control signal to redirect the data stream into appropriate ports in the OXC. Note that in this figure, another signal coming from port 2 with wavelength λ_a does not branch in this node. The first optical switch just direct the signal into the outgoing port, without passing through the splitter.

1.3.3 Multicast routing and wavelength assignment

Similar to the packet routing network where the routers exchange routing messages and use some algorithms to route the packets to the destination through a certain path, in optical routing, no matter being a light-path or a light-tree, is needed to be found out such that data can be delivered. However different from the packet switching network, in optical WDM network the wavelength channels along the light-path/light-tree have to be reserved by setting up the optical switch, such that the upper layer knows how to encode the data and inject into the optical network such that the data stream can reach the destinations. Because a channel is being reserved in optical domain, quality of service (QoS) can be guaranteed within the light-path/light-tree.

The process of setting up route for a session from a source node to a destination node, as well as assigning wavelength channel to the session is called "Routing and Wavelength Assignment", or RWA in short. If it is a light-tree to be assigned where multiple receivers are involved, the process is called then Multicast RWA, or MC-RWA in short.

The RWA/MC-RWA can be roughly be further divided into two cases: static session and dynamic session. A static session refers to reserving wavelength resources forever, while a dynamic session refers to reserving the wavelength channels for a specific period of time. As static session will always on the network it would be preferable to setup multiple static sessions together such that the overall assignment results will be optimized in cost. However, for dynamic sessions the assignment process should be fast even the cost is not optimal.

The optical network comes with some limitations that have to be considered when running any RWA/MC-RWA algorithms:

1. Wavelength continuity: In the optical domain, wavelength conversion is difficult because of the expensive wavelength converter. Some optical switch has a limited number of wavelength converters while some do not have the wavelength converter inside. In considering routing, it is more appropriate to consider routing using the same wavelength along the whole light-path or the whole light-tree. Due to the number of wavelength conversion allowed, the available wavelength is usually a constraint in

RWA problems.

2. Connection-based network: As the wavelength resources are being reserved, a new session will not be always possible to find a light-path/light-tree to the destination(s). That is, a session request can be rejected if the required wavelength channels are not available.

3. Signal strength requirement: Transmission of optical signal n fibers suffers from noise and interference as in other medium. In order for the signal to be decoded correctly, there is a minimum required signal-to-noise ratio (SNR). The signal strength itself deteriorate other travelling for a long distance. Also in optical multicast with light-tree, for example, if a signal is being split into two copies in a branch. Ideally the signal power is reduced by half, or -3dB in this case. Although passive amplifier can be used to amplify the light signal, the noise component of the signal is also being amplified. So a light-tree can have only limited number of branches, which is dependent on the optical technology. In order to satisfy this constraint, sometime the light-tree will be sub-optimal in term of network cost due to the limitation from the signal strength requirement.

4. Computation cost: To search for a light-tree in the optical network, the current condition of a network is modeled as a graph and algorithm of tree searching is being run on top of the graph. The tree-search algorithm which search for an optimal tree (minimal Steiner tree) is known to be NP-complete. And this problem becomes more complicated when different wavelength channels are considered all together to get the most optimal route as the search has to be repeated for all wavelength channels. It depends on the different network constraints such as wavelength continuity constraint for the most optimal route searching method.

5. Full-acceptance criterion: In IP multicast, a source with address s sends data to a multicast group address G without the need to know the destination nodes as the routers in between will automatically route the packets to the members who joined the multicast session when possible. However in setting up a light-tree for multicast, the

source node and destination nodes are required to be known in advance for resource allocation. If the light-tree fails to deliver data to at least one member of the session, this multicast session is usually blocked, that is, no resource will be allocated in this session. It is possible to setup a multicast tree to accept partial tree, which some of the member are omitted from the light tree. Depends on the application requirements, the omitted member can be served by offering light-path/light-tree of different wavelength channel. But this usually involves either wavelength conversion in optical domain, which is expensive and consideration becomes complicated, or the Optical-Electronic-Optical (O-E-O) conversion is necessary which introduce a large delay.

1.3.4 Formulation of MC-RWA

While there are different requirements to be satisfied in the MC-RWA process, to setup a light-tree, in general some basic constraints of a WDM network are modeled as layers of network graph, where each layer represents different wavelength. In [12], the formulation of the MC-RWA is given as a Mixed Integer Linear Programming (MILP) problem. The general problem statements are given as below for static multicast sessions, and in this formulation wavelength of a session can be converted when passing through the optical node as described in [12]:

1. A physical topology $G_p = (V, E_p)$ consist of a undirected graph in which V represents the set of network nodes and E_p represents the set of links connecting the nodes in V. A network node i is assumed to be equipped with a $D_p(j) \times D_p(j)$ multicast wavelength-routing switch (MWRS), where $D_p(j)$ is called the physical degree of node j equals the number of physical fiber links emanating out of node j.
2. Number of wavelength channels carried by each fiber = W.
3. A group of k multicast session.

The formulation of the problem is given below:

- Notations:
  - s and d refer to source node and destination node respectively in a multicast session.
  - m and n denote end-points of a physical link that might occur in a light-tree.
  - i is used as an index for multicast session number, where $i = 1, 2, \dots, k$. Note that unicast sessions are a special case such that the destination set size is one.
- Given:
  - Number of nodes in the network $= N$
  - Maximum number of wavelengths per fiber $= W$
  - Physical topology P_{mn}, where $P_{mn} = P_{nm} = f$, where $f = 0, 1, 2, \dots$ is the number of fiber connecting node m and n and $m, n = 1, 2, 3, \dots, N$. When $P_{mn} = P_{nm} = 0$, the node m and node n are not connected.
  - Every physical link between nodes m and n is associated with a weight w_{mn}.
  - Capacity of each channel $= C$
  - A group of k multicast sessions S_i for $i = 1, 2, 3, \dots, k$. Each sessions S_i has a source node and a set of destination nodes characterized by $\{s_i, d_{i_1}, d_{i_2}, \dots\}$. The size of a session will be denoted as L_i which is the sum of source and destinations nodes in S_i.
  - Every multicast session is operating at the full capacity, i.e. C.
  - Every node is equipped with wavelength converters capable of converting a wavelength to any other wavelength among W channels.
- Variables:
  - A boolean variable M^i_{mn}, which is equal to one if the link between nodes m and n is occupied by multicast session i or zero otherwise.

- A boolean variable V_p^i, which is equal to one if node p belongs to multicast session i, otherwise $V_p^i = 0$. A node belongs to a session if it is either the source or one of the destination nodes or an intermediate node in the light-tree for the multicast session.
- An integer commodity-flow variable F_{mn}^i. Each session needs one unit of commodity. That is, L_i units of commodity flow out of the source s_i for session i. F_{mn}^i is the number of units of commodity flowing on the link from node m to node n for the session i.

- Optimize:

$$Minimize : \sum_{i=1}^{i=k} \sum_{m,n} w_{mn}.M_{mn}^i \tag{1}$$

This objective functions sum up the cost of each multicast session

- Constraints
  - Tree-creation constraints

$$\forall i, \forall n \neq s_i : \sum_m M_{mn}^i = V_n^i \tag{2}$$

$$\forall i : \sum_m M_{ms_i}^i = 0 \tag{3}$$

$$\forall i, \forall j \in S_i : V_j^i = 1 \tag{4}$$

$$\forall i, \forall m \neq d_{i_j}, j = 1, \ldots, (L_i - 1) : \sum_n M_{mn}^i \geq V_m^i \tag{5}$$

$$\forall i, m : \sum_n M_{mn}^i \leq D_p(m).V_m^i \tag{6}$$

$$\forall m, n : \sum_i M_{mn}^i \leq P_{mn}.W \tag{7}$$

  - Commodity-flow constraints:

$$\forall i, \forall m \notin S_i : \sum_n F_{nm}^i = \sum_n F_{mn}^i \tag{8}$$

$$\forall i, \forall m = s_i : \sum_n F_{s_i n}^i = L_i - 1 \tag{9}$$

$$\forall i, \forall m = s_i : \sum_n F_{ns_i}^i = 0 \tag{10}$$

$$\forall i, \forall m = d_{i_j}, j = 1, \ldots, (L_i - 1) : \sum_n F^i_{nm} = \sum_n F^i_{mn} + 1 \tag{11}$$

$$\forall i, m, n : M^i_{mn} \leq F^i_{mn} \tag{12}$$

$$\forall i, m, n : F^i_{mn} \leq N.M^i_{mn} \tag{13}$$

– Additional constraint:

$$\forall i, m, n : F^i_{mn} \leq (L_i - 1) \tag{14}$$

- Explanation of equations: The equations here are to describe a multicast tree where commodity flow is used to model the multicast session. Equation (2) ensures that every node that belongs to a multicast session (except the source) has one incoming edge. Equation (3) states that the source node has no incoming edge since it is the root of the tree. Equation (4) ensures that every source node and the destination node of a multicast session belongs to the tree. Equation (5) ensures that every node except the destination nodes belonging to the tree has at least one outgoing edge on the tree. Equation (6) ensures that every node with at least one outgoing edge belongs to the tree. Equation (7) restricts the number of light-tree segments between nodes m and n by $P_{mn}.W$ in either direction.

  Equation (8) ensures that intermediate node which is neither source or destination must have its incoming flow and outgoing flow the same. Equation (9) ensures the outgoing flow of the source equals to the number of recipients. On the other hand, equation (10) ensures that there is no incoming flow to the source. Equation (11) states that for destination nodes has its outgoing flow one less than its incoming flow. Equation (12) and equation (13) ensures that link occupied by a session has a positive flow or otherwise no flow. And equation (14) ensures that flow on any link is upper bounded by the number of destinations.

Figure 11 gives a brief illustration of the commodity flow as described in the equations. The figure shows a source node s and three destination nodes $d1$, $d2$ and $d3$. For simplicity we only consider one session in this figure. In this figure the nodes in gray are the where the multicast tree (in bold arrows) spans. The numbers on the link show the flow of on the link. There is an intermediate node I in the figure which does not belong to the multicast session S. As there are one source and three receivers in this session, $L_i = 4$. From the source there is no incoming flow (Equation (10)) and outgoing flow is the number of recipients (Equation (9)). When the flow first arrived at the non-receiver node I, the sum of incoming flow equals to the sum of outgoing flow (Equation (8)). When the flow arrives at the receivers node $d1$, $d2$ and $d3$, the sum of outgoing flow is one less than the incoming flows (Equation (11)). All links spanned by the multicast tree has a positive flow (Equation (12), (13)) and the maximum flow is bounded by the number of destinations in the session (Equation (14)). Note that the "flow" here does not mean the bandwidth required by the session but only represents how many receivers exist downstream.

As stated before there are more constraints which can be introduced by various limitation of the network (such as wavelength continuity constraint). So depends on the requirements and limitations, additional constraints or variables can be added to this model when necessary. In [12] the formulation of MC-RWA in which wavelength conversion is not allowed is also given:

Notations:

- s, d, i, m and n remain the same as the case with wavelength convertor
- c is the index for the wavelength assigned to a multicast session, $c = 1, \ldots, W$

Given:

- The parameters are the same with the case with wavelength convertor, except that wavelength cannot be converter in the optical nodes

Variables:

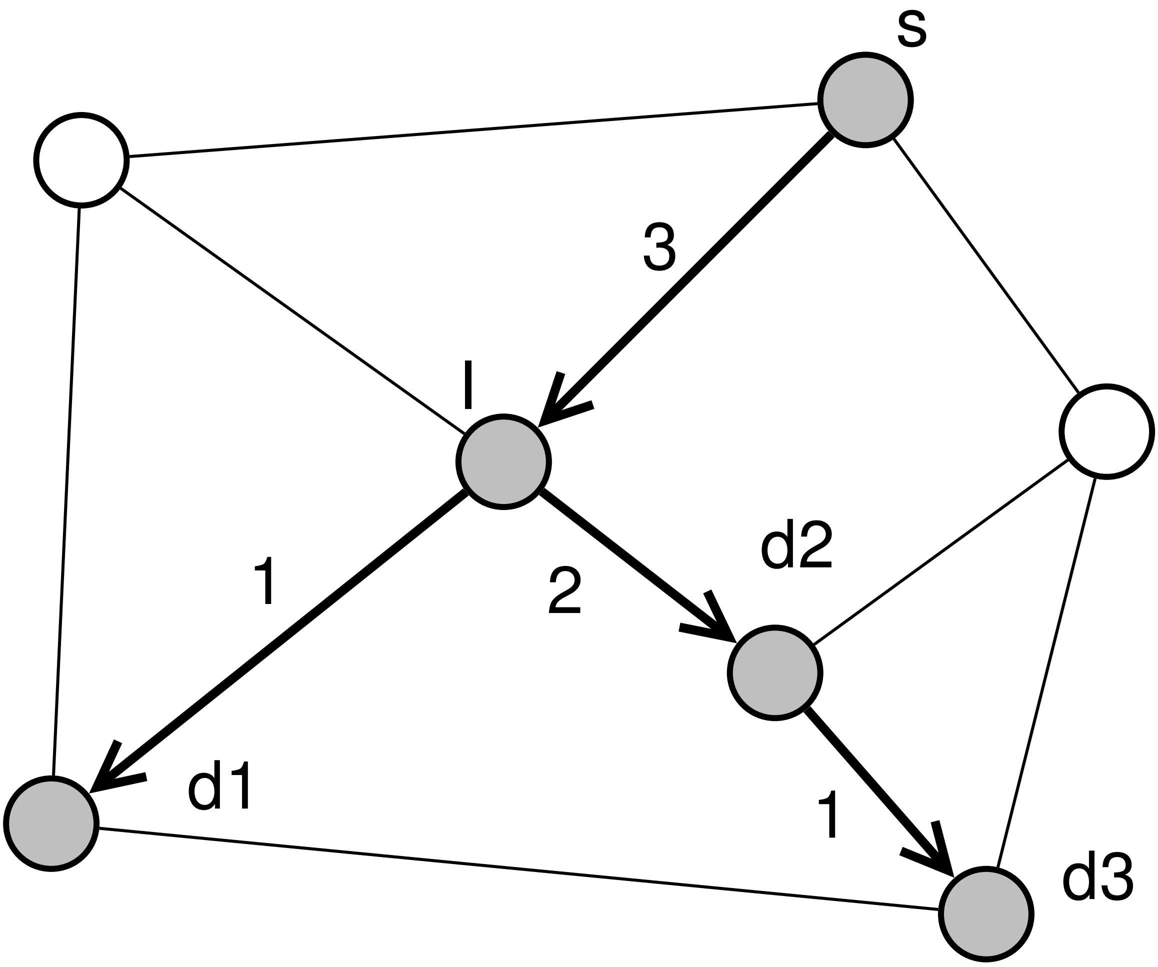


Figure 11: Illustration of commodity flow as used in MC-RWA modelling

- A boolean variable, M_{mn}^{ic}, which is equal to one if the link between nodes m and n is occupied by multicast session i on wavelength c, otherwise M_{mn}^{ic} $= 0$
- Definition of V_p^i and F_{mn}^i remain the same
- A boolean variable, C_c^i which equal to one if multicast session i is on wavelength c or otherwise $C_c^i = 0$.

Optimize:

$$Minimize : \sum_{i=1}^{i=k} \sum_{c=1}^{c=W} \sum_{m,n} w_{mn}.M_{mn}^{ic} \tag{15}$$

Constraints:

- Tree-creation constraints

$$\forall i, \forall n \neq s_i : \sum_{m,c} M^{ic}_{mn} = V^i_n \tag{16}$$

$$\forall i : \sum_{m,c} M^{ic}_{ms_i} = 0 \tag{17}$$

$$\forall i, \forall j \in S_i : V^i_j = 1 \tag{18}$$

$$\forall i, \forall m \neq d_{i_j}, j = 1, \ldots, (L_i - 1) : \sum_{n,c} M^{ic}_{mn} \geq V^i_m \tag{19}$$

$$\forall i, m : \sum_{n} M^{ic}_{mn} \leq D_p(m).V^i_m \tag{20}$$

$$\forall m, n : \sum_{i,c} M^{ic}_{mn} \leq P_{mn}.W \tag{21}$$

$$\forall m, n, c : \sum_{i} M^{ic}_{mn} \leq P_{mn} \tag{22}$$

- Commodity-flow constraints: The first four constraints are the same as Equations (8), (9), (10) and (11) respectively. Equations (12) and (13) are respectively modified as follows:

$$\forall i, m, n : sum_c M^{ic}_{mn} \leq F^i_{mn} \tag{23}$$

$$\forall i, m, n : F^i_{mn} \leq N.M^{ic}_{mn} \tag{24}$$

- Wavelength-related constraints:

$$\forall i : \sum_{c} C^i_c = 1 \tag{25}$$

$$\forall m, n(n > m) \forall i, c : M^{ic}_{mn} + M^{ic}_{nm} \leq C^i_c \tag{26}$$

- Explanations of equations: Equations (22), (25) and (25) are new constraints. The other equations serve with the same purpose in the modelling with the previous model in which wavelength converter is allowed. Equation (22) restricts number of sessions on the same wavelength between a node pair by P_{mn}. Equation (25) ensures a session uses exactly one wavelength. Equation (25) ensures that no link is occupied by a session on the wavelength not chosen by it and all the links occupied by a session are on the same wavelength.

1.3.5 Related Works on Multicast Routing and Wavelength Assignment

1.3.5.1 Traffic grooming in light tree

In [13], the authors first presented mathematical formulations for RWA of multiple light-tree based multicast sessions. They formulated the RWA optimization problem as a mixed integer linear programs (MILPs), in which the optimal solution of the MILPs represents the optimal routing and wavelength assignment on the optical layer. Then the authors further expanded their work for fractional-capacity RWA, which represents grooming of sub-wavelength traffic. Consideration of constraints on light splitting is also examined as this process is different from copying and multicast copies of data in the electronic domain that light splitting is much more limited by the hardware's capability. Finally the authors proposed fast heuristics for establishing a set of multicast sessions under different dynamic scenarios.

In the paper, two kinds of optical network switch which support multicast (i.e. light-tree) are considered. The first kind is an opaque switch. Incoming optical signal is converted into electronic signal first. After adding and dropping data to and from this switch, the electronic signal is converted into optical signal again to its output ports. This kind of switch is very popular as the electronic cross-connects fabrics are a mature technology. Also incoming and outgoing signal which pass through this node are not necessary to be of the same wavelength, as all signals are converted into electronic domain which can be converted into any wavelength. However the store and forward mechanism of this electronic switch performs much slower than the fully transparent switch, where operation is carried on the optical domain.

The transparent switch has built-in optical splitters for separating input signal into multiple copies. The optical signal, after splitting, will has its power weaken and therefore amplifiers are required. Wavelength converters are not necessarily present in this kind of switch, and therefore wavelength continuity constraint should be considered in the MC-RWA problem modelling. Also splitting degree is also limited in this kind of switch.

The authors also suggested a heuristics approach on solving the MC-RWA problem under different situations (switches with or without wavelength conversion, grooming allowed or not). Given a list of multicast connection requests, sequentially assign wavelength to the requests using an approximated minimum steiner tree algorithm according to respective constraints, will allow robust solution for the MC-RWA problem.

In this paper, all the optical switches are assumed to be identical in terms of wavelength conversion capabilities, fan-out (how many light tree branches can be split), and grooming capabilities. As optical switches on the backbone network have different capabilities a further investigation on the heterogeneous network environment is needed. Also in the heuristic approach suggested, the multicast connection requests are considered sequentially according to some order (e.g. ascending order of cost of the multicast request). However as multicast connections are being setup and torn down continuously, consideration of the dynamic nature is also needed in the MC-RWA problem.

1.3.5.2 Hybrid provisioning of low speed unicast/multicast traffic

In [14], dynamic provisioning of low speed traffic into existing connection is proposed. In this paper, the authors suggested that the dynamic multicast grooming problem can be divided into four subproblems (routing, logical topology design, provisioning and grooming). Established sessions acted as logical topology, and the logical topology is combined with the physical topology to form a hybrid approach that allows traffic to be groomed in the same wavelength, which leads to a lower blocking probability when compared with the conventional MC-RWA approach.

The idea of this paper is that assuming there exists a light-tree which is already setup, if newer multicast session is targeting to the same set (or subset) of the destination nodes of the existing session, it can be groomed into the existing session such that minimum wavelength assignment is needed for the new comer. The existing light-tree is acting as the logical topology (Hypergraph Logical Topology, HGLT). If a new multicast connection request of the same source and destination set arrives and there are residual capacity, the new call is simply groomed into the existing session. If the source is outside the logical topology but an existing light-tree is found connecting the targeting destinations set, a light path (physical path) is first setup and connect the requesting source to the original light-tree source. This two-hop hybrid connection allows traffic to be groomed to the original logical topology. Figure 12 shows how a new multicast session B is groomed into an existing session A, where both of them share the same set of destinations D(AB). First by using some methods (out-of-band or in-band) the new session B knows there exists a light tree which is already connected to a set of receivers which session B intends to connect to, and there is residual bandwidth on this light tree which satisfy the new connection. The new session will use the existing light tree for its traffic delivery. As the multicast source is not the same of the original one in this case, a new light path is setup to connect into the multicast source in session A by using conventional unicast RWA approach. In this case this physical connection, with the original logical connection (light tree), forms a hybrid approach for grooming of multicast sessions.

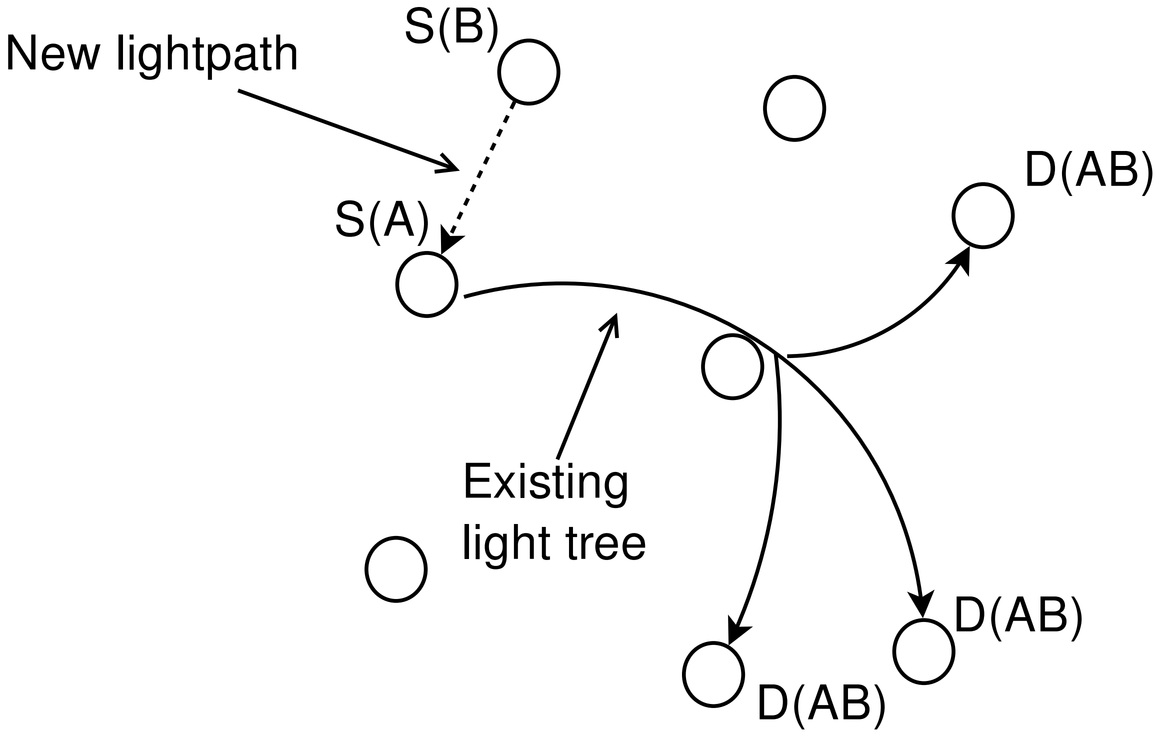


Figure 12: Hybrid approach of grooming into existing light tree

This paper also suggests that unicast traffic can also be groomed into the existing light-tree. However in this case as some destinations on the multicast tree will receive some traffic which is not intended to them, multicast traffic is given a higher priority in the scheme such that the unintended traffic can be kept at a minimum level. By this way, when multicast traffic contributes to a certain level of the total network traffic, different multicast sessions with same destinations set will be able to exploit the under-utilized bandwidth in the wavelength channel.

In this paper, it is assumed that all the optical switches on the optical layer have full splitting capabilities. However as the splitting process is limited by the light splitters and the amplifiers, an out-degree of two or three are quite typical. In additions grooming will not be success if the new multicast request's destination set is slightly different from the existing light tree. Provided a fairly large amount of nodes on the optical layer, the probabilities of the grooming to fail will increase with the number of nodes, wavelength available, and the amount of multicast requests.

1.3.5.3 Dynamic multicast traffic grooming in WDM mesh networks

In [15], the authors further expand the idea in [14] that when searching in existing light trees, even when the destinations set does not match the destinations set of an existing light tree, but close to, new branches can be added in order to satisfy the new multicast request such that the new light tree can accommodate both the existing session and the new coming multicast request. As branch is added to the existing light tree, there is no disturbance of the existing connection that data are still being delivered to the same destination sets. New coming multicast request makes use of most of the unused bandwidth in existing light tree branches and wavelength channel required to be assigned for the new branches can reduced. However as there are different sets of destination on the same light tree, some nodes on the optical layer will be receiving unnecessary traffic from the other groomed multicast sessions. Especially when the correlation between the destinations sets are low, the mismatch between the multicast sessions will become larger. Figure 13 shows the bulky tree formed by grooming two sessions. At the beginning there is a light tree formed from S(a) to three recipients (D(a) or D(ab) in Figure 13), where light branches are split at node B. As S(B) request to form a new multicast session to its destinations (D(B) or D(AB) in figure 2). This process forms a new light tree that includes all the destination nodes of both sessions (D(A), D(B), D(AB)). In the worst case, with grooming of other multicast sessions with relative low correlations of destination nodes, it becomes easier for the new coming sessions to join this light tree, and therefore the more uncorrelated will be the traffic flow on this light tree. This negative feedback mechanism causes more unmatched light tree for different multicast session on the optical layer, especially when the number of wavelength channel increases. Also this bulky multicast tree for large number of branches will cause management problem of the light tree, especially when the number of groomed sessions is large.

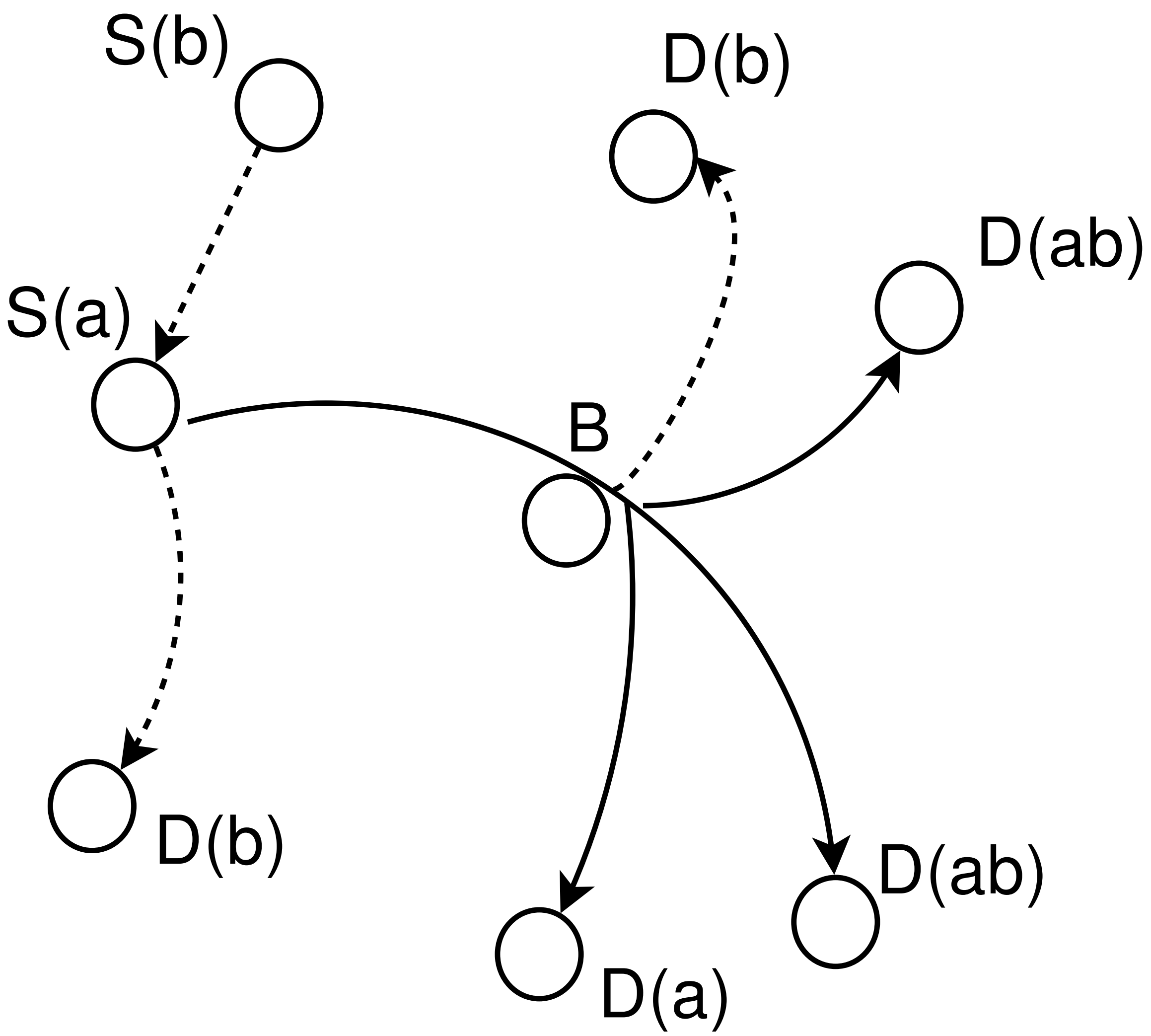


Figure 13: Bulky tree formed by grooming multicast sessions

1.3.5.4 Partial virtual light-tree

In [16], instead of the light-tree approach, the authors suggest that partial virtual light tree (PVLT) or virtual light tree (VLT), which consists number of light paths (and optional light tree), could minimize the total number of fibers needed to support multicast asymmetric traffic. The authors also used ILP formulation to prove that the PVLT/VLT approach, as opposed to the light tree approach, can minimize the number of wavelengths used per fiber. However in this paper the optical nodes are assumed to have full wavelength conversion capabilities. While the traditional opaque optical switches (O-E-O switches) are naturally equipped with such capability, they are operating at a much slower speed compared with transparent switch where all the routing/splitting is done in the optical domain. Furthermore the authors did not consider the possibility of bandwidth grooming, although which is not directly correlated to their proposed scheme.

1.3.5.5 Max-first and re-treeing

In [17], instead of minimizing session blocking probability in which a session is blocked if a multicast session cannot serve at least one user provided the existing network conditions, the authors suggest a MAX-FIRST algorithm which try to minimizing user blocking probabilities which multicast session will always try to setup to serve maximum number of users even if some of them is blocked. In this work the authors also suggested that in dynamic network, re-treeing should be allowed when the network condition is updated. If the algorithm can find another tree which serve a larger number of receivers, re-treeing is done in order to serve more users.

However this method depends on the requirements of the application. If a set of receivers can be missed from the multicast session it is allowable to setup partial tree as described in the paper. However the re-treeing method is to serve another biggest set of receivers, where existing receivers of the session can be missed out after the re-treeing of the method if the new tree can serve more users than the existing tree.

1.4 Positions of study

The position of the studies included in this thesis is summarized in Table 1 and the relationship with some conventional methods is shown in Figure 14.

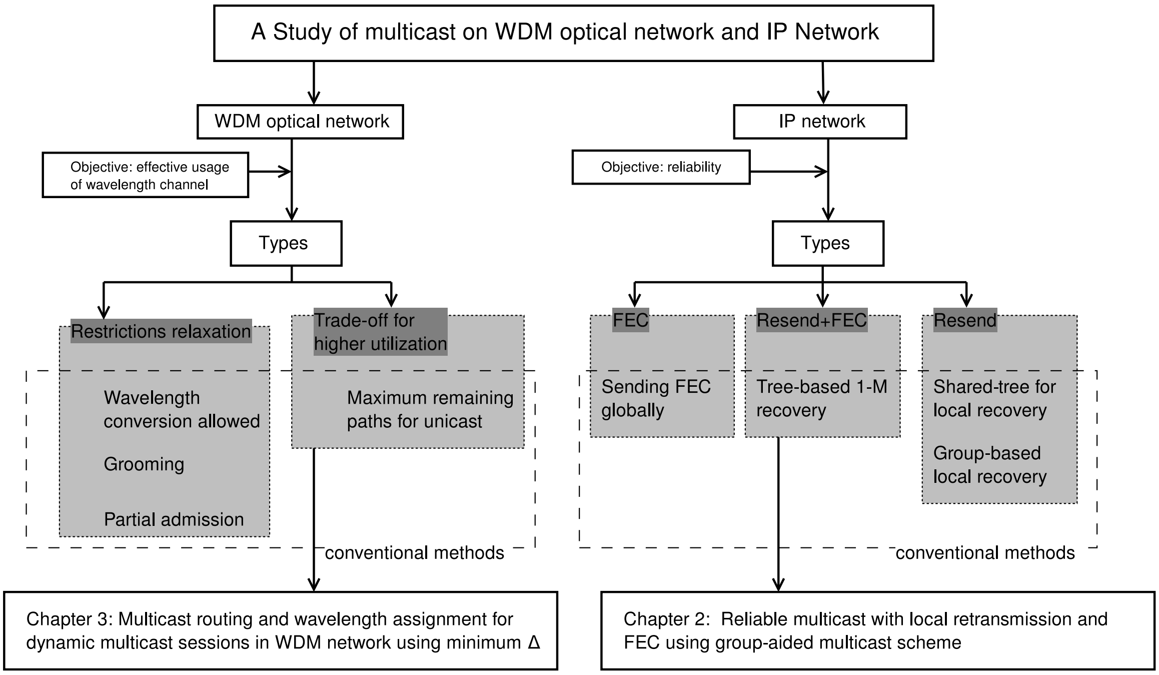

A Study of multicast on WDM optical network and IP Network
WDM optical network
IP network
Objective: effective usage of wavelength channel
Objective: reliability
Types
Types
Restrictions relaxation
Trade-off for higher utilization
FEC
Resend+FEC
Resend
Wavelength conversion allowed
Grooming
Partial admission
Maximum remaining paths for unicast
Sending FEC globally
Tree-based 1-M recovery
Shared-tree for local recovery
Group-based local recovery
conventional methods
conventional methods
Chapter 3: Multicast routing and wavelength assignment for dynamic multicast sessions in WDM network using minimum Δ
Chapter 2: Reliable multicast with local retransmission and FEC using group-aided multicast scheme

Figure 14: The positions of study on this thesis

Multicast provides an efficient way for data to be delivered from sources to destinations by letting the network to duplicate the data stream to different destinations. The Internet itself is a network of heterogeneous networks which include but not limited to wireless network, Ethernet, synchronous optical network and WDM optical backbone network, . Each kind of network has its own issues and limitations. When multicast service is to span across different networks, different challenges come with the property of the network and requirement of the traffic.

In some networks the recipients of the network suffers from packet loss such as low quality physical channels in wireless network. A lots of the applications require reliable data transfer service. Scalability is often the main issue to support multicast with reliability across different networks because of the recovery traffic needed. It is important to distribute the recovery loading away from the sender such that ACK/NACK implosion problem can be avoided. Recovery traffic with NACK/resend packet is simple to implement where FEC recovery method can make use of the multicast nature to reduce overall recovery traffic. The work in Chapter 2 addresses the issue of localizing recovery traffic in lossy networks for many-to-many multicast traffic such that overall recovery traffic can be reduced.

For the relatively error-free backbone optical network data can often be delivered accurately. However in order to provide multicast service in a wide area which travel through the optical WDM backbone network, wavelength resource must be allocated such that data flow is allowed on the backbone network in order to provide high capacity traffic. The nature and limitations of the circuit-switched WDM networks create challenges for effective usage of the WDM network. The conventional methods mostly study the high utilization when a certain of constraints in WDM optical network. In some older studies wavelength assignment schemes for unicast which leads to higher utilization have been proposed. The study in Chapter 3 proposed a parameter for wavelength assignment in multicast which leads to lower blocking probabilities.

Table 1: Problems of existing schemes and the contribution of the proposed schemes

Chapter 2	Topic	Reliable multicast with local retransmission and FEC Using Group-aided Multicast Scheme
	Problem of existing research	Existing schemes either do not have a good structure to distribute the loading across the network, or use simple retransmission method to recover loss packets. There are still room for reducing the required recover traffic
	Proposed method	By grouping the members of the multicast members in an many-to-many multicast session according to their loss statistic, the recover action can be bounded into the local group. With the use of FEC packets to recover uncorrelated loss of the receivers
	Effect of proposed scheme	The recovery traffic as observed in the proposed scheme is less than the conventional scheme
Chapter 3	Topic	MC-RWA for dynamic multicast sessions in WDM network using minimum Δ
	Problem of existing research	The existing schemes assign wavelength usually with some simple methods such as in order of wavelength index or randomly. This could lead to a situation that the wavelength channels are ineffectively assigned and causing high blocking probabilities
	Proposed method	A new parameter Δ is proposed as an indicator to choose the wavelength during MC-RWA process. This Δ is calculated based on the current network states such that by choosing Δ with minimal value the blocking probabilities can be reduced
	Effect of proposed scheme	The proposed scheme can achieve lower blocking probabilities of dynamic multicast sessions. Also the average size of these session is observed to be large in the proposed scheme, which means that higher overall throughput is achieved in the proposed scheme.

CHAPTER 2

RELIABLE MULTICAST WITH LOCAL RETRANSMISSION AND FEC USING GROUP-AIDED MULTICAST SCHEME

2.1 Introduction

Multicast is a network efficient way of delivering copies of data to multiple specific recipients, group members, across the public network. Only one copy of the data travels across each multicast link, and the network duplicates the data to the members as needed. However, IP multicast does not provide any reliable delivery mechanism. With the evolution of network based applications, data exchanges between servers and clients, or peer-to-peer, have become much more frequent. For instance, stock quotes exchange, active news/media feeding, multi-player online gaming, software patching servers, can make use of reliable multicast to provide service with more efficient use of network resources. Reliability on the best-effort IP network can be achieved in two ways: resending same copy of data upon loss detected, or by using transport layer forward error correcting code (FEC) to recover error. This is very similar to the FEC used in lower layer, but FEC packet approach is operating on the packet level instead of bit level. In the case of unicast, like TCP, resending data on receiver's demand is a natural choice as it is simple to implement. In unicast it makes no practical advantage to use transport layer FEC to recover receiver's error, and this can be easily implemented on the upper layer if needed.

However, in the case of reliable multicast, use of transport layer FEC can improve the recovery performance. When compared with the unicast case, multiple recipients suffer from independent loss. A single FEC packet, however, can recover different errors on different recipients. If the FEC packets are sent pro-actively, loss can be protected, and thus, recipients can reproduce the desired data more promptly, as they do not need to wait for at least one round-trip-time (RTT) for the process of requesting repair packets. Different FEC based schemes were proposed mostly on an end-to-end basis [18, 9]. In these schemes the original sender produces FEC packets, and multicasts to all receivers. The heterogeneity of the different parts of the network involved in a multicast session is probably the biggest problem to determine the appropriate parameters for producing the FEC packets (e.g. size of redundant traffic). When the sender multicasts the redundant traffic to all members of the multicast session, members who suffer from relatively low packet loss will receive an excessive amount of FEC packets, while for some receivers who suffer from high packet loss, FEC packets will fail to protect the loss in many cases.

In reliable multicast careless implementation of a recovery mechanism might cause problems to both senders and receivers. ACK implosion problem, where many recipients request resending of a particular copy of data, causes a huge amount of traffic to the sender's link. On the other hand, when the sender multicasts repair packet to the group, it causes the repair locality problem, where receivers who did not suffer from loss receive a duplicated copy of the same data. These two problems, together with different management problems associated with the multicast group, limit the scalability of reliable multicast session.

Scalability in a 1-to-many (1-to-m) model can be achieved by hierarchical structure of the multicast tree [8]. Intermediate routers or designated receivers are responsible for aggregating ACK/NACK from their child nodes. Thus, the ACK traffic being sent to the senders will be bounded by the number of immediate child nodes. Also these routers or designated receivers can provide recovery service locally to their child nodes. In many-to-many (m-to-m) cases, tree hierarchy becomes difficult as there is more than one sender and they may be located anywhere on the multicast tree. To solve the m-to-m multicast load distribution problem, the Scalable Reliable Multicast (SRM) [10] was proposed in which the nodes lie on a common shared ACK tree (feedback path to senders). Nodes listen to the "repair request" from the neighbors, provide recovery if data packet is available, or otherwise suppress their own "repair request". Later the Group-aided Multicast (GAM) [11] was proposed and is shown to be effectively providing reliable multicast traffic for m-to-m multicast scenario by using a two-level hierarchy (which consists of one core group and a number of local groups). The two-level hierarchy is formed by per-source ACK trees between the cores in the core group, and one shared ACK tree within each local group. Per-source ACK trees in the core group ensure the shortest distance between cores, and shared ACK tree in the local group ensures the small maintenance cost. In the original GAM scheme, it is shown that this two-level hierarchy is able to achieve effective recovery in terms of low recovery latency and lower maintenance overhead for a many-to-many reliable multicast. Each local group in GAM is effectively regarded as another m-to-m multicast session, in which the core of the group is responsible for handling all the recovery on behalf of the data sources outside the local group. However, each local member suffers from uncorrelated loss on different branches of the multicast tree, and sends NACK to the core for the repair traffic individually. If the number of members in the local group grows large, the core has to handle large numbers of NACK packets, as well as resending large numbers of recovery packets. This causes an implosion problem to the local core.

In this work, I further extend the idea of GAM by using transport layer FEC locally in addition to NACK/retransmission in order to reduce the feedback and overall recovery traffic. Provided the current processing power of the network nodes, FEC packets can be produced as soon as the required data packets are received. By providing FEC recovery pro-actively, the need of NACK/retransmission for uncorrelated error recovery reduced at the core of a group. Also using localized FEC recovery within a group allows more optimized recovery traffic when compared to the sender based FEC, which uses the same set of FEC parameters across the whole network. Consider the recovery traffic in relation to group size, heterogeneous loss conditions across the multicast session, and also effect of under/over sending of FEC packets. Through simulation it is shown that for a group where uncorrelated losses occur at different hosts, the proposed scheme can reduce the recovery traffic, especially when the group size is large. Also, optimized recovery traffic is localized to each group according to the observed loss.

In section 2.2, in order to improve the scalability for reliable multicast, a hybrid local recovery scheme based on Group-aided Multicast is proposed. In section 2.3 the simulation analysis is presented in order to show the effectiveness of the proposed scheme. Finally this work will be concluded in section 2.4.

2.2 Proposed Scheme

2.2.1 Protocol overview

In the proposed scheme, a hybrid local recovery scheme (retransmission + FEC) is used with the GAM hierarchy in order to provide a reliable many-to-many multicast service with a reduced amount of recovery traffic. Packet recovery is achieved first by FEC packets from the core. Individual local members only send NACK to the core to trigger resending if FEC fails to recover the packets. The core of each group adjusts the amount of the redundancy traffic according to the loss rate as observed in the local group, as shown in Figure 15. In order to produce FEC packets, the core has to make sure that it has received a block of k data packets in advance. If any loss occurs in the core, it will immediately ask for retransmission from the original sender. After the core has received the k packets from a specific sender, it produces additional $(n - k)$ FEC packets based on a linear systematic code [7]. In systematic code, suppose the original packets are $X_1, X_2 \ldots X_k$, the generator matrix G of size k by n will produce a sequence of packets $Y_1, Y_2 \ldots Y_k, Y_{k+1} \ldots Y_n$, where for any i between 1 and k, $X_i = Y_i$. Therefore, the core, which is also a receiver of the multicast session, needs only to produce and send the additional $Y_{k+1} \ldots Y_n$ packets to its local members. After the FEC packets $Y_{k+1} \ldots Y_n$ are produced, the core sends them to the local members through multicast. When the local members receive the FEC packets, they try to recover the packet loss with the FEC packets. The systematic code allows local members to recover lost packets as long as a total of k packets, either original data packets or redundancy packets, are received.

If any local member judges that the packet loss it has suffered cannot be recovered with all the other correctly received packets (data and FEC), it will send NACK to the core to trigger retransmission. Note that the local member will only ask for the data packet but not the FEC packets. This eliminates the need for the core to store the FEC packet it has generated.

Feedback (NACK) are sent upstream with ACK-tree, which can be a separate path from

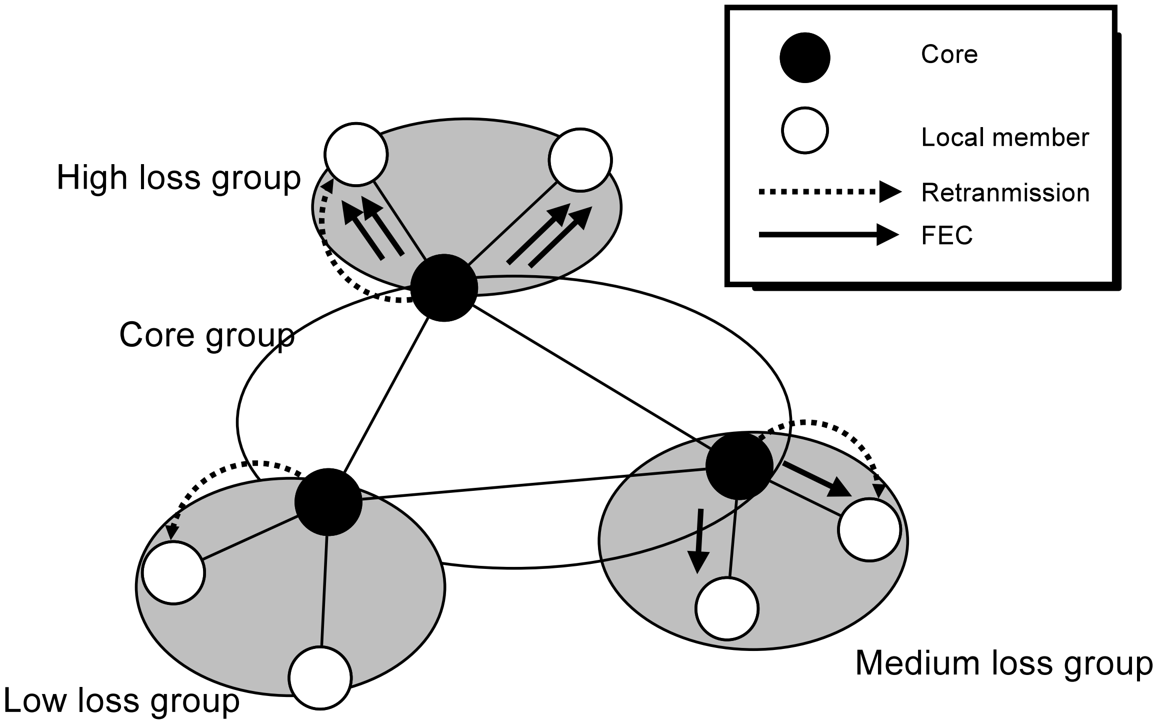


Figure 15: The proposed scheme: core increases redundancy traffic locally when the loss rate is high

the multicast tree. Recovery mechanism in the proposed scheme works on the transport layer, and feedback among the cores are sent with packet switching IP network. For example, if the packet is to be traveled on the WDM core network, the service provider only has to make sure there is channel to deliver the packet, but no special supervisory channel should be needed as the feedbacks should be handled in the upper layer.

With the multicast of FEC packets within the local group, even the local group members suffer from uncorrelated packet loss and it is possible for them to repair their own different losses with the same set of FEC packets. In the perspective of the core, one set of FEC packets to the local members can correct a large number of uncorrelated errors. This greatly helps decreasing the number of unicast NACK and unicast retransmission.

For core-to-core recovery, it is done by unicast with NACK/retransmission as suggested in the original group-aided multicast scheme. As each group is formed by nodes sharing similar loss characteristic (for example, nodes in the same wireless LAN), the FEC packets could become useful. However for edge nodes across the network (e.g. across the MAN), the packet loss characteristics in different parts of the network are relatively uncorrelated. For example, if the multicast session spans across multiple ISPs, some ISPs may have inadequate equipments to handle traffic and loss which may occur at that part of the network. In that case sending FEC packets with multicast does not effectively cover the packet loss, and thus the NACK/retransmission way will be used for core-core recovery.

An example data exchange scenario in a local group is shown in Figure 16. For simplicity, suppose a systematic code (9,8) is used in this scenario (i.e. one redundancy packet for every eight data packets), and the delivering and recovering two blocks of data (sixteen packets) is shown:

- At point (1), a data packet is lost before reaching the core.
- At point (2a), the core detects the loss (sequence number 82) as it finds a gap in the sequence number. The core will immediately send a NACK request to the multicast source. At point (2b), the local member detects the same loss. However it will suppress the NACK request and expect recovery from FEC at a later time.
- At point (3), the data source receives the NACK from the core, and resends the lost packet.
- At point (4), another data packet is lost within the local group, although it was delivered to the core.
- At point (5), by receiving the repair packet (sequence number 82) from the data source, the core multicasts the packet which is lost locally within the local group (this is the same as the original GAM, as loss at core usually indicates loss at local members)
- Right before point (6), the core receives all of the required packets (eight packets, from sequence number 80 to 87). The core produces FEC packet and multicasts to the local members (in this case, one packet).
- At point (7), by receiving the FEC packet, the local member can recover the loss.

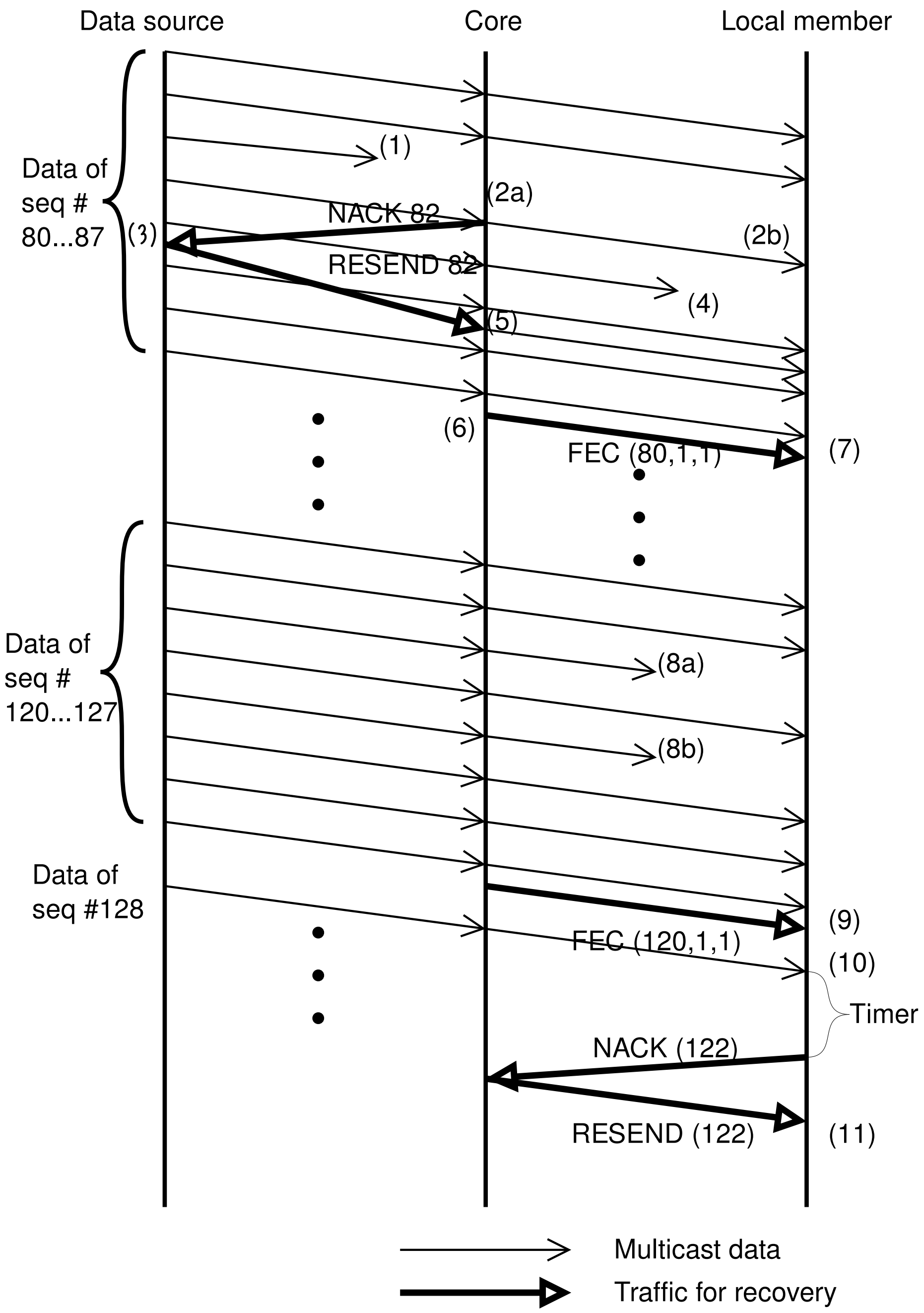


Figure 16: An example scenario of data exchange in the proposed scheme

In this case, the whole block of packets can be recovered without asking the core for resending.

- At point (8a) and point (8b), two multicast packets (sequence number 122 and 124) are lost. The local member does nothing at this moment.

- At point (9), the FEC packet for the block with sequence number from 120 to 127 arrived at the local member. However in this time the local member cannot recover the two lost packets from one FEC packet.
- Upon receiving the recovery packet at point (11), the corresponding block can be reconstructed, and is thus considered to have been correctly received.

2.2.2 Protocol details

In this sub-section, the detailed operation on how to recover packet loss is explained. Data recovery methods and control messages are discussed. Data recovery involves both the error correcting and retransmission of lost packets. Control messages which include a loss-query, loss-reply and redundancy announcement, are also discussed.

2.2.2.1 Data recovery

In the beginning phase of the multicast session, when any node joins the local group, it will receive control message from the core which contains the systematic code parameters (n, k). During the multicast session, each local member maintains a receive status bitmap, which is based on the sequence number for each data source, and is shown in Figure 17.

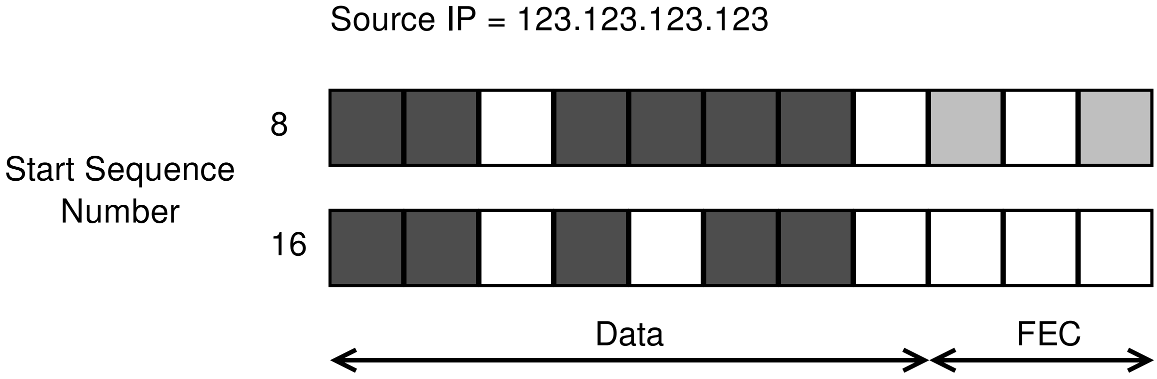


Figure 17: An example receive status bitmap

For a particular data source (identified by IP address), the receive status bitmap shows the receive blocks which have not been completely received. Each receive block contains a start sequence number S and a bitmap of length n (data + redundancy size of the systematic code). And thus, this receive block is used to show the receive status of data packets with sequence number (S, S+1...S+k-1), and also FEC packets for the block started with S, which are generated by the core and will be explained in the next paragraph. Upon receiving a new packet of sequence number i, the local member should either fill in to the existing bitmap or create a new receive block, which depends on the current receive status.

The core produces and multicasts FEC packets when it has received all the data packets for a particular block (S to S+k-1). Each FEC packet contains the start sequence number S, the redundancy size $(n-k)$ used, and its FEC index. Three parameters are combined and they uniquely identify a FEC packet. As mentioned in the previous section, the FEC packets should use a systematic code. This is because data packets are sent from some senders to the whole multicast session, while the additional FEC packets are sent by the cores of each group. After the FEC packets are generated, they are then multicasted from the core to the local group members. As the core is responsible for producing the FEC packet, whenever it detects a gap in the packets it has received, it should immediately send NACK to the original data source for recovery packet.

After filling in the bitmap (of data or FEC), the local members should check whether the receive block is "completed" or not. The "completed" means k or more packets, either data or FEC, are correctly received. The "completed" receive block will be removed from the receive status bitmap as this block is considered to be correctly received.

An example of receive status bitmap is shown in Figure 17. Systematic code of (11,8) is used in this case, in which 8 packets are for data and 3 packets are for FEC. If a packet from 123.123.123.123 with sequence number 25 arrives, due to it not fitting in any existing receive blocks, the local member will create a new receive block of start sequence number 24 with the second bit set to 1 (packet of sequence number 24 is considered to be delayed or lost). On the other hand, when a packet of sequence number 10 arrives (this can be retransmission or an out of order packet), the local member fills in the 3rd bit of the receive block of start sequence number 8. In this case the local member finds that six data packets and two FEC packets are correctly received. As it can reconstruct the two lost data packets (13 and 15) with the FEC packets, this block is considered to be complete, and will be removed from the receive bitmap.

The next question is when should the local members send NACK to the core for retransmission. In this scheme, NACK is sent when there are more than one receive blocks in the receive status bitmap. Gaps inside a receive block indicate delayed or lost packet. However, as FEC packets are expected to correct some errors, a NACK request is first suppressed. When the number of receive blocks grows as more packets are received, all packets of the previous block have been sent from the data source. At this moment, the local member expects the core to have received the complete block and produce FEC packets. A timer starts and NACK request will be sent to the core after this timer expires and the local member is still not able to recover its error. The timer duration will be set based on some statistic on the time of previous FEC packets arrival time with respect to packets of the same receive block (FEC packets with same start sequence number of the receive block). The timer is needed for the processing time to produce the FEC packet, traveling time of the packets from the core, and also round trip time between the core and the original sender, as the core can also suffer from loss. When the timer expires, the local member picks the lowest sequence number, which it has not received correctly, and sends NACK with this sequence number. In this case, the core simply retransmits the requested copy of the packet with unicast. After repair packet arrives, the local member carries out the bitmap filling procedure and sees if the receive block can be completed. If not, the local member repeats the same procedure again with the next missing sequence number until the block can be completed.

2.2.2.2 Control messages

Control messages are exchanged between core and local members to serve for three main purposes. First, core sends loss query to the local group members to estimate the average loss in their receive blocks. Then, the local members provide loss report to the core, and thus the core can decide the proper redundancy size. Finally, the core announces to the local group if it has decided the redundancy size should be changed.

The core periodically multicasts loss query message to the local members, which includes a value of block size k which the core is going to generate the FEC packets from. Upon receiving of the loss query message, the local members will base on their receive history and reply an estimated loss value per receive block of size k to the core. This value can be short-term average packet loss in receive blocks (e.g. the number of data packet loss in last three receive blocks) such that this loss value can accurately reflect the changing network conditions.

After gathering the loss reply messages from the local members, the core calculates the average number of loss of the local members, and compares the result with the redundancy size $(n - k)$ used currently. In case they are different, the core will announce a new redundancy size to the local members. Then the core and the local members will use the new redundancy size starting from next receive block. This allows the core to adjust to the most appropriate redundancy size, when the underlying network conditions have changed (e.g. newly joined local member or congested router).

The average number of loss for the local members is chosen as the redundancy size. This is based on the assumption that each packet losses suffered by the local members are independent to each other, such that number of packet loss in each receive block is a binomial random variable, parameterized by the data block size k and the perceived drop probability p. For a binomial random variable, it can be shown that $P[loss = l]$ is maximum at l_{max}. This indicates that if the average number of loss is chosen, which is given by kp, as the redundancy size, a reasonably large number of lost packets that have occurred in the local members can be recovered. In a more aggressive approach, the average number of $loss + 1$ can be chosen in order to further protect loss at the local members, as $P[Loss = l_{max} + 1]$ has also a relatively high value. Although a further increase of redundancy size can decrease the number of NACK (and thus the number of retransmission packets), this decrease may not be worth the increase of FEC packets sent by the core.

In case of degrade of network conditions, the core might receive and increase of NACK packets as reported from the local members. In this case, the core can update the number of FEC packets sent according to the updated information from the receivers (which can be piggybacked in the NACK packet as the information size is supposed to be small). When the network later recovers from the bad network conditions, the periodic query can detect feedback from the receivers such that it is going to reduce the FEC packets sent to fit the network characteristics.

Regarding a change of redundancy size, there are chances that the core and some local members are using different redundancy sizes, due to loss of announcement packet from the core, or out of order data packets to the local members. In this case, the local member will treat the received FEC packet as an announcement, and increases/decreases the created receive block's redundancy size corresponding to the FEC packet, and uses the new FEC to create a new receive block.

In the operation of the proposed scheme, different control messages carry only a small amount of information, like the number of FEC packets produced per block, or average number of loss of the local members. These messages can be included in the header of the NACK message which local members send to the core, or in the header of the FEC packets which the core multicasts to the local members. However the proposed scheme does not limit the implementation of control messages between core and local members.

2.3 Simulation Results

2.3.1 Simulation model

The hybrid scheme and the conventional NACK only GAM are implemented and their performances are compared using the OMNet++ discrete event simulation system[19]. The topology used is shown in Figure 18. This topology consists of 12 groups and total of 84 non-core recipients, and all groups are connected by four backbone routers. In this setup, a many-to-many multicast session is simulated and the recovery traffic on the core link is observed. Each link in the simulation model represents a logical path, and suffers from independent loss.

The traffic consists of two types: one type is the global traffic which includes data (to all recipients), NACK/retransmission (between cores and senders). The other type is within a local group, which includes local NACK/retransmission (between cores and local members), FEC data (from core to local members) and control traffic (between core and local members). However as the control messages carry only small amounts of information, the messages is piggybacked into NACK (from local members to core) and FEC packets (multicasted from core to local members) such that the overhead caused by the control messages does not need to be considered.

The group size (number of local non-core members) of each group is fixed and ranged from 3-10 local members, as shown in Table 2. For simplicity only the 12 cores act as senders in this m-to-m multicast session. The parameters used in the simulation are shown in Table 3. The simulation for the conventional NACK only GAM scheme, and the proposed scheme with FEC as local recovery are run. It can be observed the recovery traffic sent by the cores within different groups. Backbone router links (the four routers in the middle) have a loss rate of 0.01, backbone router to core links have a loss rate of 0.05. and packet loss rate of core to non-core member links is set to 0.1, 0.2 and 0.3 in different scenarios, which will be discussed later. It is also assumed that all local members suffer from uncorrelated loss to each other, in which the uncorrelated loss occurs at the logical links.

In real world scenario wireless communication can suffer from high loss rate as suggested in the simulation parameters. For instance, in [20, 21] two studies are conducted on experimenting packet loss for video multicast over IEEE 802.11b and IEEE 802.11g network. In both papers two experiments were conducted. In one experiment the terminals are put in a variable distance from the access point (10-100 meters) in an outdoor environment, while in the other is an indoor environment with different obstacles of a typical office environment. It is found that in IEEE 802.11b network, terminal suffers from PER up to around 30% when the it is more than 60m from the access point in an outdoor environment, or there are a few obstacles between the terminal and the access point. For IEEE 802.11g network, PER of 30% is observed when the terminal is around 30m from the access point outdoor.

The bit error rate (BER) is a lower layer quantity used in lower layer transmission (for example, wireless LAN or 3G network). The simulation model does not consider the lower layer quantity. When the error occurs under bad BER condition such that the lower layer fails to deliver packet to the upper layer, packet error occurs and that is what the proposed scheme concerns about. However if the BER is not very bad such that the packet can be reconstructed with lower layer mechanism and transferred properly to the upper layer, the protocol does not consider it as a loss condition.

In the simulation, all the links are assumed to have a bandwidth of 100Mbps such that bandwidth issue is not under consideration in the proposed model. In reliable multicast session, receivers which always have insufficient bandwidth are excluded from the session. These receivers should be taken care later using some out-of-band method (e.g. request required data later through unicast). However, occasional insufficient bandwidth can always occur, in which case the loss is modeled as burst loss, and the simulation results will be presented in section 2.3.5.

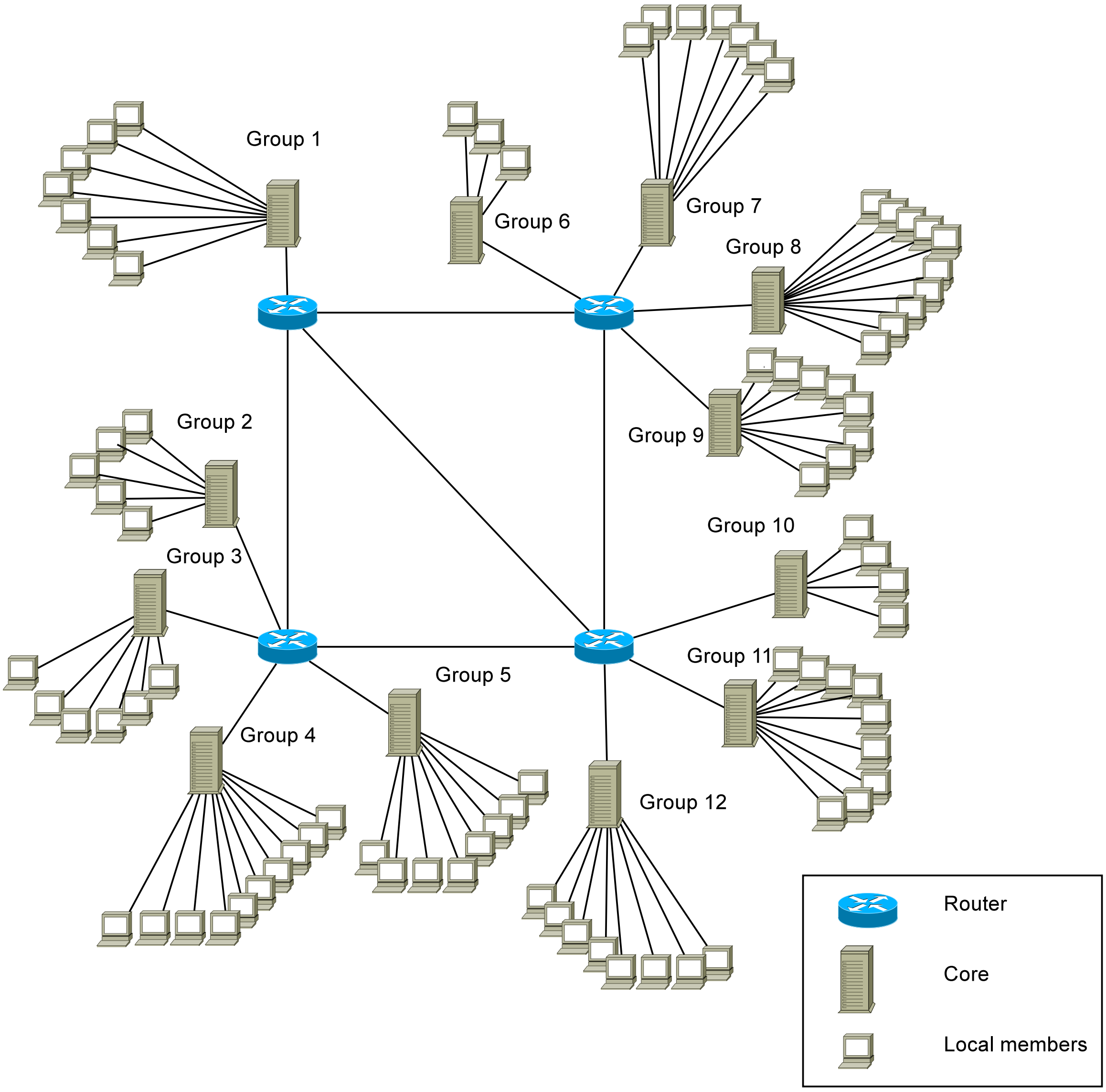


Figure 18: Network topology in simulation

Table 2: Group ID and group size (number of local members)

Group ID	1	2	3	4	5	6	7	8	9	10	11	12
Group Size (number of local members)	7	5	6	10	8	3	7	10	8	4	9	8

Table 3: Simulation parameters

Parameters	Values
Simulation duration	1 hour (except burst loss model in section 2.3.5)
Data sending frequency	Uniform from 0.5s to 1.5s (except burst loss model in section 2.3.5)
Bandwidth of all links	100Mbps
Receive block size (data)	8 packets
Data packet size	1 kB
Backbone router to backbone router loss rate	Random packet drop with probability 0.01
Backbone router to core loss rate	Random packet drop with probability 0.05
Core to non-core members loss rate	Random packet drop with probability 0.1 to 0.3 (see each sub section), or burst loss (see section 2.3.5)
FEC code	Systematic code
FEC data block size	8 packets
FEC redundancy size	variable
Number of under/over sending of FEC packets per block	-1 to 3 (relative to observed loss)

2.3.2 Relationship between group size and the effectiveness of the proposed scheme

In this simulation the loss from the core to the local members is set to be constant for all groups to see the effect of group size and reduction of recovery traffic as compared to the conventional scheme. In Figure 19, the amounts of repair traffic from different cores are shown, where the x-axis is arranged from group of lowest group size (group 6) to group of highest group size (group 8). All the core to non-core member links have a loss rate of 0.1. When the group size is only 3, it can be seen that the core produces about same amount of recovery traffic in both the conventional and proposed schemes. However, the increase rate for repair traffic with the number of members is higher in the conventional scheme. This is because in the proposed scheme, local members first suppress sending NACK and wait for FEC packets to see if they can correct the missing packets using the FEC packets. This helps in reducing the NACK traffic, and therefore the number of retransmission packets. Moreover, suppose two data packets are lost and one FEC packet is received in a particular receive block, the local member only needs to request for one data packet from the core, as it can correct another missing packet with the received FEC packet.

In the simulation a low packet sending rate is used. Referring to Table 3, data packets are sent every second on average. Also, in this many-to-many multicast scenario, the 12 cores are also acting as the senders. As the simulation time is one hour a total of approximate $12 \times 60 \times 60 = 43200$ data packets are sent in this case (approximate because the packets are not sent at regular interval).

For instance in group 11 (group size: 9), when comparing 43200 packets to the recovery traffic in both conventional (47,000 packets) and proposed scheme (31,000 packets), they represent and amount of about 108% (conventional) and 71.7% (proposed) recovery traffic with respect to the total received data packets in the multicast session. The acceptable amount of recovery traffic mainly depends on the capability of the core and the bandwidth of its link, and this depends on the actual network conditions.

2.3.3 Groups with different packet loss rates

In this simulation three groups of the same number of local members are observed to see how the cores adjust the FEC parameters according to the observed loss within their groups. Figure 20 shows the repair traffic from different cores, where their local members suffer from different packets loss rates. It is observed that cores of group 5, 9 and 12 as in Figure 18, where each group has eight local members. The loss rates of the router to non-core links in group 5, 9, and 12 are set to 0.1, 0.2 and 0.3, respectively, such that it can be observed within a multicast session how heterogeneous network conditions affect each other. In all cases the proposed scheme produces less recovery traffic than the conventional scheme.

In the case of 0.3 loss rate on the network links (group 12), it can be seen that a relatively large amount of FEC packets are sent from the core compared with FEC packets sent by the

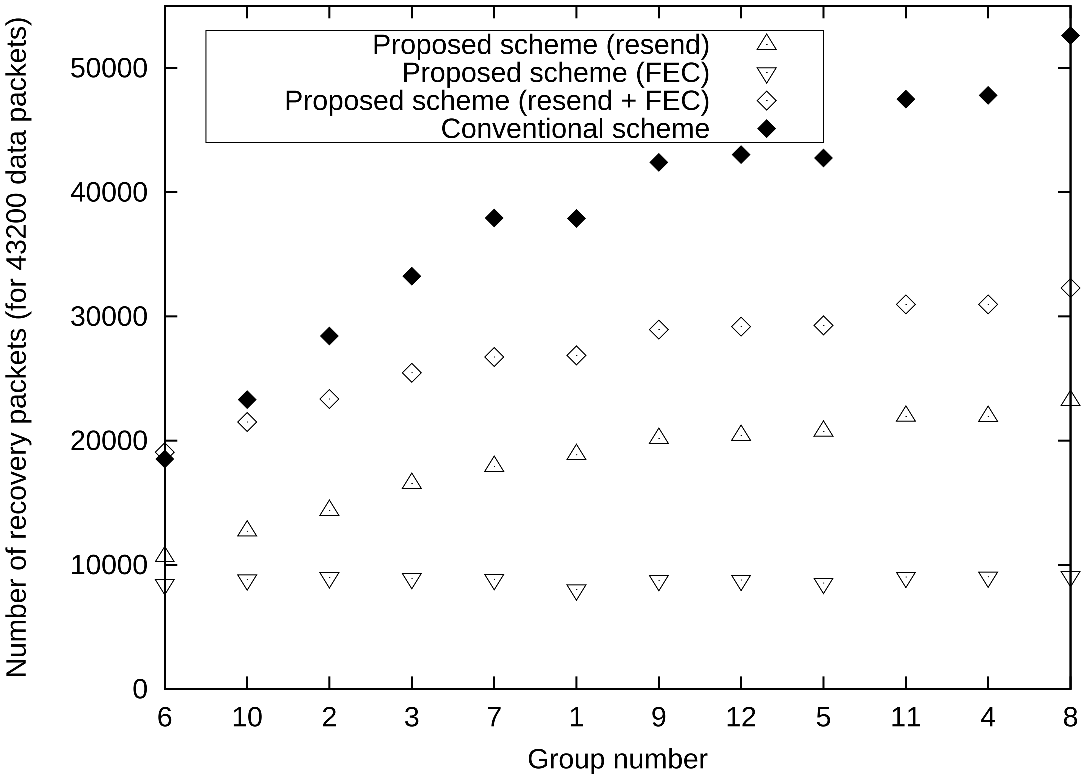


Figure 19: Repair traffic from core (packet loss rate=0.1)

cores of the lower loss groups (group 5 and group 9). This shows that by using a local basis to produce redundancy traffic, the core can control the appropriate amount of FEC traffic to send, according to the loss characteristic of the local group. However when comparing group 5 and group 9 (with 0.1 and 0.2 packet loss rate on the logical links respectively) the number of FEC packets does not differ by a large amount. Recall that the proposed scheme determines the number of FEC packets for each block depending on the average loss observed by the local members. For an 8-packet receive block used in the simulation, a 0.1 packet loss rate means on average 0.8 packets are lost for each block, while a loss rate of 0.2 means an average of 1.6 packets lost for each block. Small fluctuation of this average will lead to the core to produce one FEC packet per block, instead of two. Although a higher receive block size can be used to fine-tune for different error rates, the drawback is that it delays the local members to trigger resending from the core in case of loss.

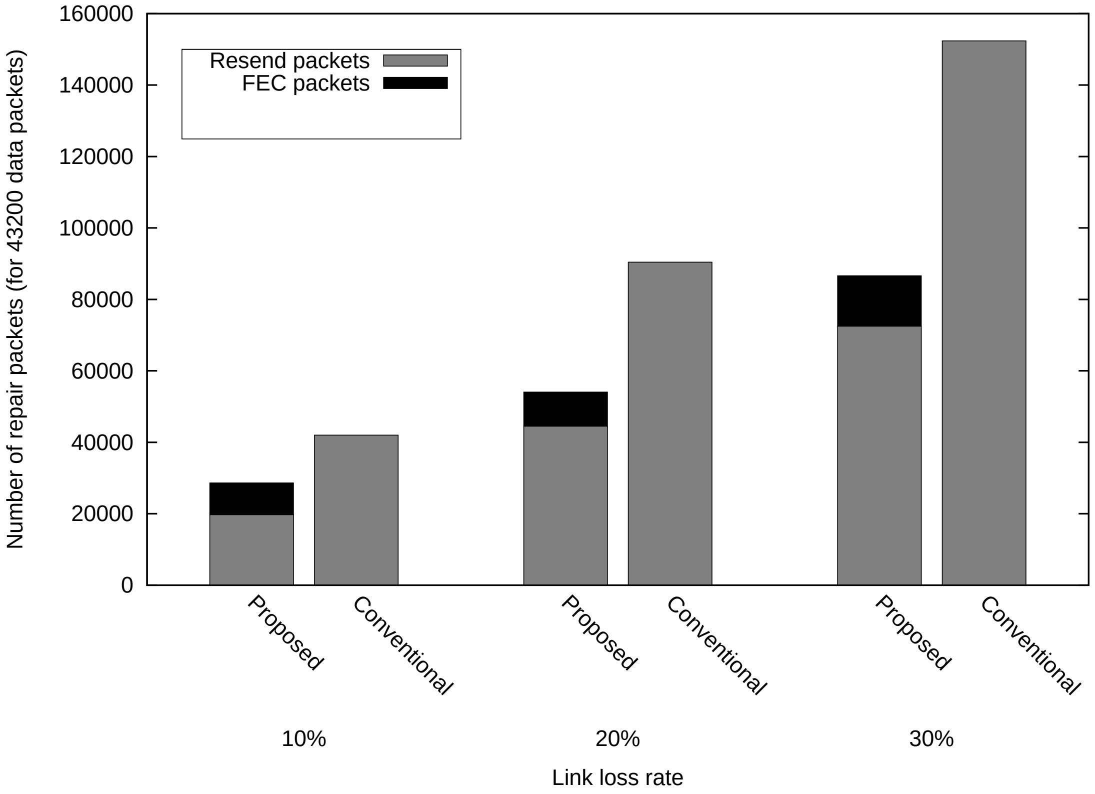


Figure 20: Recovery traffic at different packet loss rates (Group size = 8)

2.3.4 Under/over sending of FEC packets from core

In this simulation it is tested if under/over sending FEC packets can further reduce the recovery traffic. In section 2.2.2.2, the average number of loss is chosen as the redundancy size $n - k$ (or average number of loss + 1 for more aggressive approach). Next the question of what would happen if the core under/over sends FEC packets is discussed. Three groups of different sizes (Group 6: size 3; Group 1: size 7; Group 4: size 10) are observed and the number of FEC packets sent is adjusted. Let R be the average number of loss as observed from the feedback of local members, the redundancy size is set from $R - 1$ to $R + 3$ and observe the repair traffic from the core. Figure 21 shows the repair traffic from the three groups at a loss rate of 20%. It can be seen that the overall repair traffic does not decrease with the over sending of FEC packets.

From previous simulation setups the increase in number of local members would benefit from FEC approach, but an aggressive approach does not bring much benefit because the

FEC traffic increases. Depending on the actual loss rate and number of members in the local group, choosing R or $R+1$ as the redundancy size is good enough to decrease the overall repair traffic from the core.

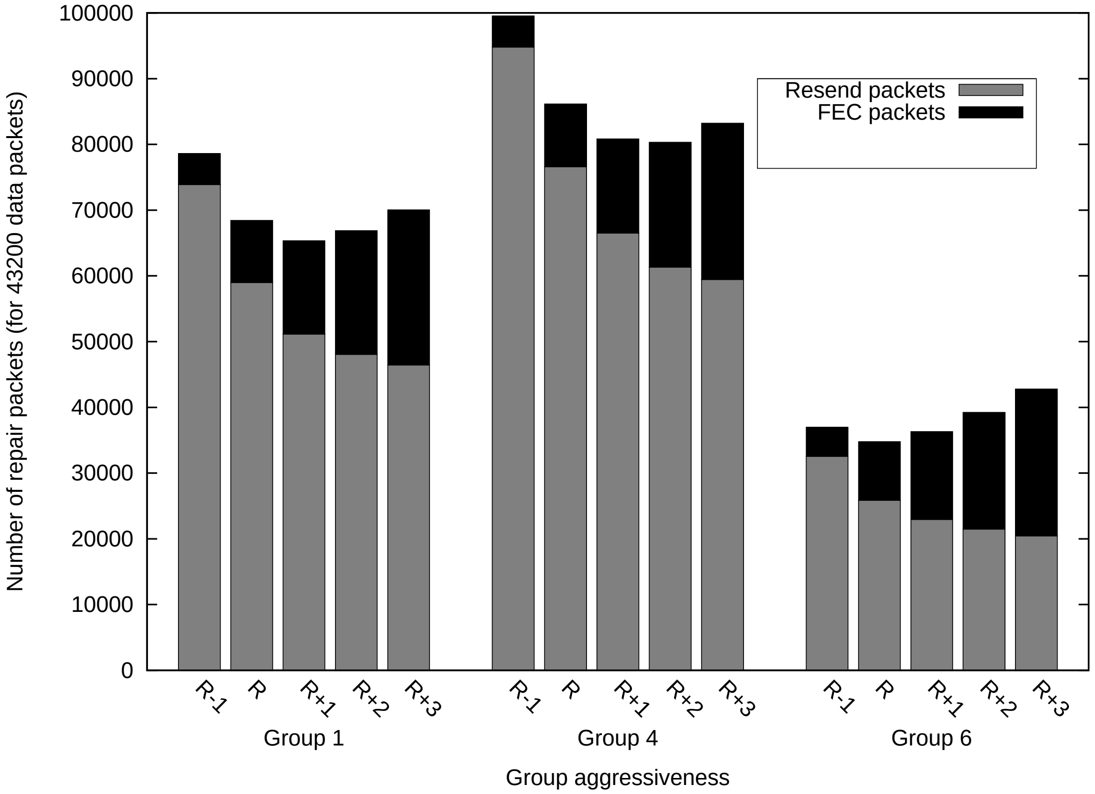


Figure 21: Aggressiveness and repair traffic (packet loss rate=20%)

2.3.5 Bursty traffic loss

In this simulation in order to observe the performance of the proposed scheme under burst loss conditions, the simulation parameters are adjusted as shown in Table 4. In this setup a higher data rate is set such that the packet loss occurs in burst. In the setup local members suffer burst loss with lengths of mean 36 and normal packet reception with lengths of mean 60. This is because as there are 12 senders in the multicast session, each 8-packet data block 3-packet long burst loss can be observed statistically. Due to the nature of geometric random variable, burst loss of a wide range is observed during the simulation.

The results of the simulation is shown in Figure 22. When compared with the conventional scheme, the overall recovery traffic can still be reduced in the burst loss scenario.

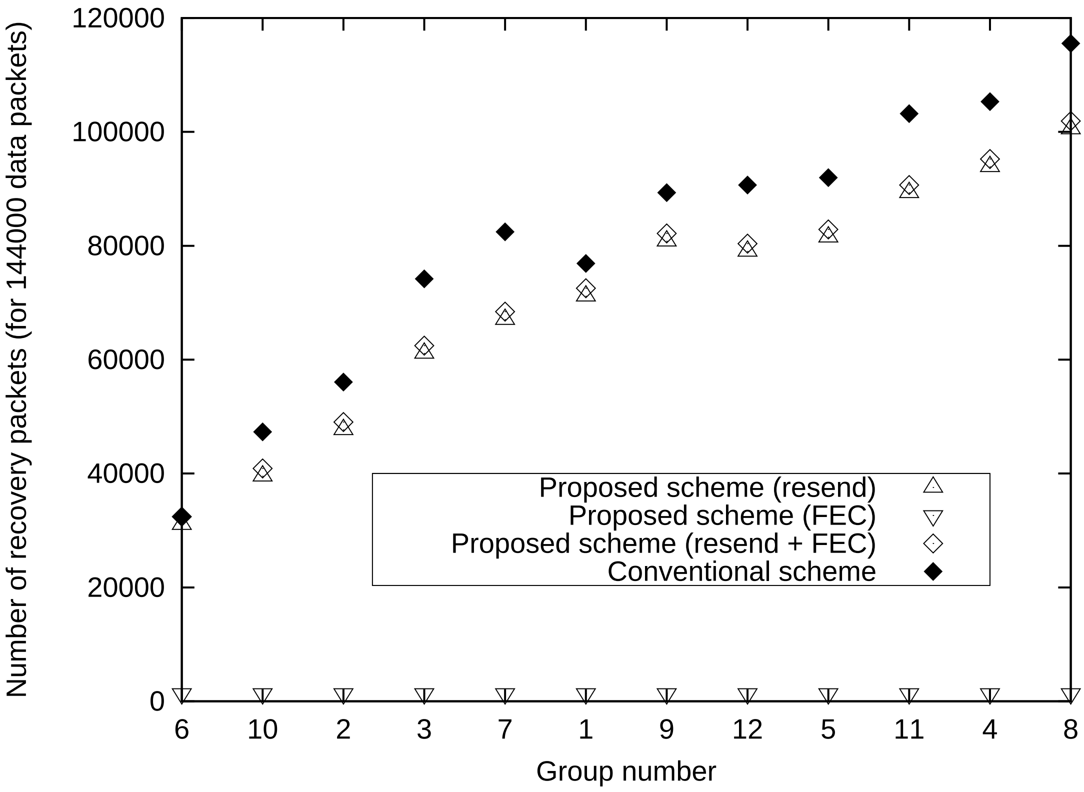


Figure 22: Repair traffic from core (burst loss model)

Table 4: Simulation parameters for burst loss scenario

Parameters	Values
Simulation duration	50s
Data sending frequency	exponential RV with mean 0.05s
Average local members' burst loss length	Geometric RV with mean 36
Average local members' normal reception length	Geometric RV with mean 60

As in section 2.3.2, it is found that the number of FEC packets sent from the core is not affected by the number of local members in the group. However, differing ratios between the FEC traffic and resend traffic is observed in the case when loss occurs in burst. In the simulation, a total of around 12000 data packets are sent from all data sources. Each core sends out around 1300 FEC packets for error protection. It means that on average less than one FEC packet is produced for each block, or an average of one packet loss is reported from local members.

In the proposed implementation the local members report their loss based on the loss occurred in the recent data blocks. In the burst loss model the local members fail to receive data during the burst, thus they cannot detect the loss promptly as no new packets are received. However, in this state when they calculate their loss, they will use the state before the burst occurred, for which the loss is probably lower than the actual loss which the local members suffer. This caused the observed loss lower than the actual situation, and resulting less FEC packets produced from the core.

2.4 Conclusion

A hybrid local recovery scheme for reliable multicast using both NACK and FEC is proposed, in which redundancy traffic is produced according to each local group's own perceived error rate, in order to provide fine-grained control of recovery traffic. Using computer simulations it is shown that the proposed scheme reduces recovery traffic sent by the core when compared with the conventional NACK only recovery scheme.

CHAPTER 3

MULTICAST ROUTING AND WAVELENGTH ASSIGNMENT FOR DYNAMIC MULTICAST SESSIONS IN WDM NETWORK USING MINIMUM Δ

3.1 Introduction

Wavelength Division Multiplex (WDM) is a technique to transmit multiple independent data streams by a single fiber. In WDM, each independent data stream is transmitted on a particular wavelength. Optical routers are configured to switch the traffic streams to different outputs. These traffic streams are preferred to be switched all the way with the same wavelength across the edges they are transfered. If the same wavelength for the stream is not available along the path of the stream, conversion of wavelength is possible, but it has to be done by a wavelength converter, or the optical signal has to be converted into an electric signal and re-encoded into optical signal of different wavelength (O-E-O conversion). However, wavelength converters are still expensive and they are usually not available on every node of the WDM network. Also O-E-O conversion gives the flexibility of signal regeneration and wavelength conversion at the cost of large delay during the conversion.

In contrast to multicast routing protocols in the IP network, multicast in the WDM layer can be done by splitting the incoming signal into multiple copies in the optical switches. This approach is referred as the light-tree approach [22, 23]. As long as the signal strength is strong enough, the light stream can be split multiple times such that the stream can be delivered to multiple clients. This splitting process is very fast compared to the multicast routing in the electronic domain. It has the advantage of being protocol independent. With the advancement of optical switch technology and optical fiber, the light-tree approach is raising research interest in recent years.

There are some previous works suggested for establishing light-tree on WDM network.

In [22], the MultiCast Routing and Wavelength Assignment (MC-RWA) for static multicast sessions is modeled as a Mixed Integer Programming (MIP) problem, where sessions occupy the network forever. Minimization of the objective function of the MIP problem leads to the minimum cost trees for all of the static multicast sessions. While this is a natural choice, the minimal Steiner tree is ideal for low cost delivery, the WDM optical network is connection oriented, and unlike packet-switched network, once wavelength resource is allocated, other request to the same wavelength resource will be blocked. In [13], the MIP model formulation with no wavelength converters for static sessions is described. In this MIP model formulation, if one cannot find a solution for the objective function, it means that no session can be established under the MIP constraints, and therefore, the request is blocked.

Sessions, either unicast or multicast, can be static or dynamic. Unlike static sessions, for dynamic sessions network usages are not known at the very beginning. They request network resources at a certain point of time. If network resources are allocated to the dynamic sessions, these resources are occupied for a certain period of time. After a session has finished using the resources, the resources are released and will be available again for future sessions. However at the situation that wavelength channels are heavily occupied such that the resources are not available to satisfy a session request, this request is said to be "blocked", and the blocking probability P_b is given in Equation (27):

$$P_b = \frac{\text{Number of blocked sessions}}{\text{Total number of sessions requested}} \tag{27}$$

In [16, 24, 25, 26, 27], schemes which utilize wavelength conversion to achieve lower blocking probabilities are introduced. With the use of wavelength conversion, the constraint that single wavelength must be used for the whole multicast tree is removed, and therefore the multicast tree can consist of different wavelengths on different parts of the network. With the use of wavelength conversion combined with Steiner tree algorithm, one can find the set of Steiner trees which results in lowest link cost. However, wavelength conversion is expensive due to wavelength converters and inefficient due to O-E-O conversion in WDM network.

For unicast cases, [28] gives a comparison of well-studied approaches where wavelength conversion is not allowed. In [29, 30] the MAX-SUM method is suggested to maximize the remaining path capacities after light-path establishment. Later on the Relative Capacity Loss (RCL) method [31] based on MAX-SUM method is suggested. Both methods attempt to choose a wavelength on a predefined set of routes, which are prepared before wavelength assignment, such that the remaining wavelength channels on the network are serving best for future sessions.

The problem of applying the concept of MAX-SUM/RCL directly to multicast scenario is that it is impractical to predefine multicast tree candidates for the network. In unicast scenario for a network with n nodes, the number of combinations of source-destination pair is $n(n-1)$. However, for multicast session, an arbitrary source s is chosen, and the subset of remaining $n-1$ nodes can be the receiver nodes. The number of combinations of source-destination sets is given by $n\sum_{i=1}^{n-1}\binom{n-1}{i} = n(2^{n-1}-1)$. Because of the large combinations, defining all possible multicast trees and applying the concept of RCL or MAX-SUM are impractical.

In this work, using a new parameter Δ as an indicator in MC-RWA in order to lower the blocking probabilities of dynamic multicast sessions is proposed, where Δ is defined as the sum of decrease in the number of nodes that each node can connect to caused by a particular wavelength assignment. It is different from the MAX-SUM or RCL that Δ does not rely on a pre-defined set of paths of all the source-destination pairs. The number of nodes that each node can connect to after a certain wavelength assignment directly affects the blocking probabilities for setting up a multicast session. Given a network condition characterized not only by the topology, but also the wavelength usage information, searching for a set of light-trees is focused, where each tree consists of a single wavelength, and choose the light-tree for dynamic multicast session on the WDM network. The Δ for each single wavelength light-tree is calculated and the tree with lowest value of Δ is chosen such that blocking probabilities can be reduced. With lower blocking probabilities more sessions can use the network at the same time. Consider dynamic sessions in some scenarios, like a video broadcast of a football match, network resources only need to be reserved for a specific period of time. The scheme aims to exploit the WDM network capabilities for establishing light-tree without wavelength conversion. It is shown in the simulation results that the proposed scheme can achieve lower blocking probabilities at slightly higher tree cost, and this means more multicast sessions can be setup in the network and thus higher utilization of the network.

3.2 Proposed Scheme

In this section the use of proposed parameter Δ in MC-RWA for reducing blocking probabilities for dynamic multicast sessions at the cost of slightly higher tree cost is described. Δ takes minimum cost trees at different wavelength as parameters, and it reflects sum of "loss of connectivity" of all the nodes in a network caused by a particular wavelength assignment. Δ acts as an indicator of impact to future dynamic sessions by choosing a light-tree at particular wavelength. It can be calculated with the current wavelength usage information. In the proposed scheme, light-tree is chosen based on values of Δ of different wavelengths. First the problem being dealt with is defined, and then the operation of proposed scheme and the reason why it is expected to reduce blocking probabilities for dynamic multicast sessions are explained.

3.2.1 Multicast tree searching for dynamic multicast session

During operation of the WDM network, the network topology $G(V, E)$, where $G(V, E)$ is a directed graph containing a set of nodes V, and E is the set of edges connecting the nodes in V. Each link in E contains W channels, the wavelength channels are being assigned and released based on the multicast sessions. A dynamic multicast session i is characterized by $S(s_i, D_i, t_{start(i)}, t_{end(i)})$ where s_i represents the source node, D_i represents the set of destination nodes, $t_{start(i)}$ and $t_{end(i)}$ represent the start time of the session and end time of the session, respectively, and $s_i \in V$, $s_i \notin D_i$ and $D_i \subset V$.

Let the wavelength resource usage information from node m to node n on wavelength λ at time t be $U(m, n, \lambda, t)$, where $m, n \in V$ and $\lambda \in 1, 2, ..., W$. At time t, $U(m, n, \lambda, t) = 1$ if the wavelength channel from m to n on wavelength λ is occupied or $U(m, n, \lambda, t) = 0$ otherwise.

For multicast session $S(s_i, D_i, t_{start(i)}, t_{end(i)})$, it is aimed to find out the multicast tree $T_i(\lambda)$ (the set of wavelength channels on the wavelength λ) and the wavelength assignment given $U(m, n, \lambda, t_{start(i)} - 1)$, where $t_{start(i)} - 1$ is one time unit before $t_{start(i)}$. $T_i(\lambda)$ can only be assigned on available wavelength channel. Assume the network has no knowledge about other future sessions. Let $M(m, n, i, \lambda, t)$ be the wavelength channel usage of multicast session i of the edge connecting from node m to node n on wavelength λ at time t. If the dynamic multicast session i uses the wavelength channel of edge (m, n) on wavelength λ , $M(m, n, i, \lambda, t)$ equals 1 when $t_{start(i)} \leq t < t_{end(i)}$, or otherwise 0. During the wavelength assignment for dynamic multicast session, a constraint as in Equation (28) exists, which means that wavelength resource must be assigned to an empty channel:

$$M(m, n, i, \lambda, t_{start(i)}) + U(m, n, \lambda, t_{start(i)} - 1) \leq 1 \tag{28}$$

Other than this constraint, the other constraints of searching for a multicast tree are the same as the static multicast sessions scenario [13]. As wavelength conversion is not allowed, the MIP problem can search the light-tree independently for each wavelength. After this step, a set of light-trees $\mathbb{T}_i$ of different wavelengths which is found based on the wavelength utilization conditions and the requesting session information. If the set of light-trees $\mathbb{T}_i$ is an empty set, it means that under the network conditions there is no wavelength channel which satisfies the requesting multicast session i, and therefore, this multicast session will be "blocked". The definition of blocking probability P_b for multicast scenario is the same as unicast scenario, and is given in Equation (27).

3.2.2 Minimum Δ routing wavelength assignment scheme

Given the wavelength channel usage condition, a node can reach to a number of other nodes with wavelength λ. Let $R_{before}(n, \lambda, t)$ be the set of nodes which node n can reach using wavelength λ at time t, and $R_{after}(n, \lambda, t, T_i(\lambda))$ be the set of nodes which node n can reach after wavelength assignment on wavelength λ using the tree $T_i(\lambda)$. It should be note that $R_{before}(n, \lambda, t)$ and $R_{after}(n, \lambda, t, T_i(\lambda))$ only depend on the network conditions at time t, but not any pre-defined paths set as MAX-SUM or RCL do. Define $d(n, \lambda, t, T_i(\lambda))$ in

Equation (29), which represents the decrease of number of nodes which node n can reach using wavelength λ if the multicast tree $T_i(\lambda)$ is assigned for multicast session i:

$$\begin{aligned} d(n, \lambda, t, T_i(\lambda)) =& |R_{before}(n, \lambda, t)| \\ & - |R_{after}(n, \lambda, t, T(\lambda))| \end{aligned} \tag{29}$$

In the conventional schemes, the main objective is to minimize the cost of the multicast tree. Therefore the wavelength is chosen such that the paths which the multicast tree spans sum up with the minimal cost. In the proposed scheme it is aimed to reduce the blocking probability for future multicast sessions. Therefore for all the nodes, if a tree that sum of the $d(n, \lambda, t, T_i(\lambda))$ for all nodes is minimum is chosen, then the lower blocking probabilities can be expected as overall the nodes are expected to reach the highest number of other nodes among the selection of wavelength. This is because $d(n, \lambda, t, T_i(\lambda))$ is defined as the decrease in number of nodes that a node can connect to, and minimal total value of $d(n, \lambda, t, T_i(\lambda))$ of all the nodes means nodes on the network remain connecting to a maximal number of nodes. Thus the objective is to choose the value of λ such that (30) is minimum. If there are multiple λ such that (30) achieves the minimum value, λ is chosen such that the tree cost is minimum.

$$\Delta(\lambda, t, T_i(\lambda)) = \sum_{\forall n \in V} d(n, \lambda, t, T_i(\lambda)) \tag{30}$$

While the proposed scheme and MAX-SUM are in nature having the same objective to assign wavelength such that the other sessions follow are more likely to have wavelength channel assigned to them, the evaluations of the path capacity loss as in MAX-SUM and the Δ in the proposed scheme are different. Consider the two example assignment scenarios in Figure 23, the arrows show the available channels of a particular wavelength. The arrows in dash-line ((6→7) in (a) and (0→1) in (b)) show two independent cases of wavelength assignment for a session. The differences on how the MAX-SUM method and the minimum Δ evaluate the network state are compared in these two scenarios.

In MAX-SUM method, the path capacity loss depends on the pre-defined set of paths P. According to the work of the MAX-SUM method in [30], P can contain all the possible paths of every source-destination pair. Consider the path capacity loss for the node 0 for the wavelength assignments (a) and (b) in Figure 23. In assignment (a), there are total of 5 possible paths from node 0 to node 7, namely (0→1→6→7), (0→2→6→7), (0→3→6→7), (0→4→6→7) and (0→5→6→7). Since P refers to the set of all possible paths of every source-destination pair, these 5 paths are inside set P. After assignment of scenario (a) where (6→7) is reserved, all these 5 paths are affected by this assignment, so the path loss is 5 here for the pair (0,7) for scenario (a). There is also a path loss of 1 for each source-destination pair in (1,7), (2,7), (3,7), (4,7), (5,7) and (6,7). This is because there is only one path from source node 1-6 to destination node 7, and all these paths are affected by the assignment of link (6→7). So the total path loss of all source-destination pairs will be 11 by assigning the link (6→7) as in scenario (a).

Similarly in scenario (b), (0→1) is to be assigned to a session. Consider all source-destination pairs, it can be easily seen that the only source-destination pairs (0,1), (0,6) and (0,7) will be affected by assigning (0→1) to a session. The three affected paths (0→1), (0→1→6) and (0→1→6→7) cause the total path capacity loss to be 3 in scenario (b).

In the scheme,Δ is calculated by the searching the reachable nodes set before and after the wavelength assignment. In Figure 23, consider the reachable nodes of all the nodes before and after the assignments. In (a) all the nodes 0-6 cannot reach node 7, thus the Δ is 7 in this case. In (b) only node 1 becomes unreachable by node 0, and therefore the Δ is 1 in (b).

The advantage of the minimum Δ method over MAX-SUM method is that minimum Δ method is designed for dynamic multicast session. The path capacity loss as described in the MAX-SUM method is less accurate to describe the feasibility of setting up multicast sessions, especially when the predefined path set is a sub-set of the all possible source-destination paths. The minimum Δ simply searches the network for reachable destinations of each source which does not depend on predefined parameters and shows directly how possible each source is to reach destinations. For example in the scenarios as shown in Figure 23,

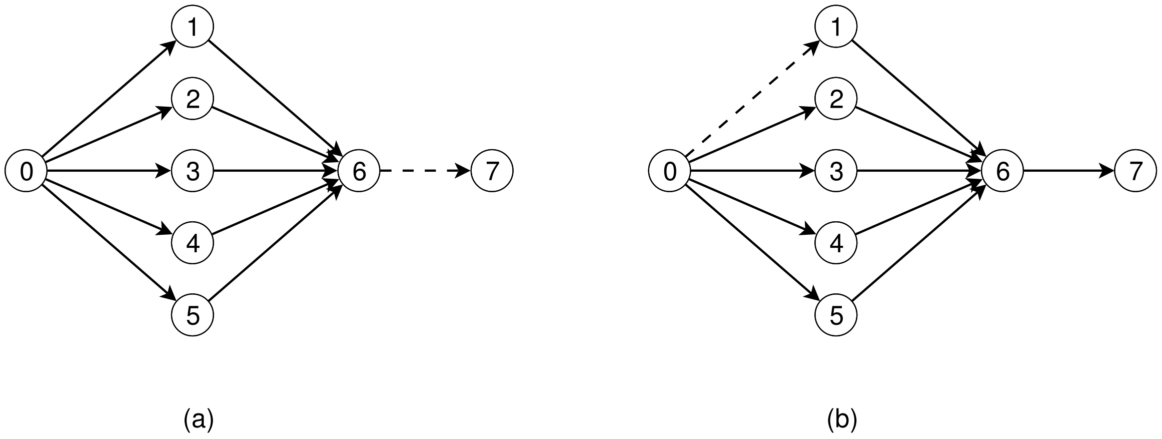


Figure 23: Two example assignment scenarios on a particular wavelength channel

assume all nodes are equally probable to be chosen as a source node or as a destination node, the blocking probabilities can be seen to be 7 times higher in (a) compared to (b) as in (a) any session with source node 0-6 with node 7 included in destination nodes will be blocked, compared to in (b) only the session with source node 0 with node 1 included in destination nodes will be blocked. And this agrees with the values of Δ (7 in scenario (a) and 1 in scenario (b)) in terms of impact when compared to the path capacity loss (11 in scenario (a) and 3 in scenario (b)).

3.2.3 Example of wavelength assignment

Figure 24 shows the available wavelength channel of two different wavelengths λ_1 and λ_2 at a certain instance. A multicast session with source at node 1 and destinations of node 5 and node 6 is requesting wavelength resource from the network. First multicast trees of different λ are found from the network using MIP, and let both the multicast trees found be $T(\lambda_1)$ and $T(\lambda_2)$ be the two links (1,5) and (1,6) as shown in Figure 24 for λ_1 and λ_2. It is easy to see that the value of $|R_{before}(n, \lambda, t)|$ and $|R_{after}(n, \lambda, t, T(\lambda))|$ for all nodes remains the same except for node 1 in the network by taking away the links (1,5) and (1,6) for both $\lambda = \lambda_1$ or λ_2. So consider only the value of $|R_{before}(1, \lambda_1, t)|$ and $|R_{before}(1, \lambda_2, t)|$ for $\lambda = \lambda_1$ and λ_2. Before assignment it can be seen that both $|R_{before}(1, \lambda_1, t)|$ and $|R_{before}(1, \lambda_2, t)|$ have a value of 5 as node 1 can connect to all other nodes in the network using either λ_1 and λ_2. However it can seen that if links (1,5) and (1,6) are assigned on λ_1, $|R_{after}(1, \lambda_1, t, T(\lambda_1))|$ becomes 0 as there are no channel comes from node 1 on λ_1. On the other hand, if the multicast tree is assigned on λ_2 the value of $|R_{after}(1, \lambda_2, t, T(\lambda_2))|$ still has a value of 3. Thus by the proposed scheme λ_2 for the multicast session is assigned, as by this assignment the network only loses connectivity to 2 nodes instead of 5 if λ_1 is used.

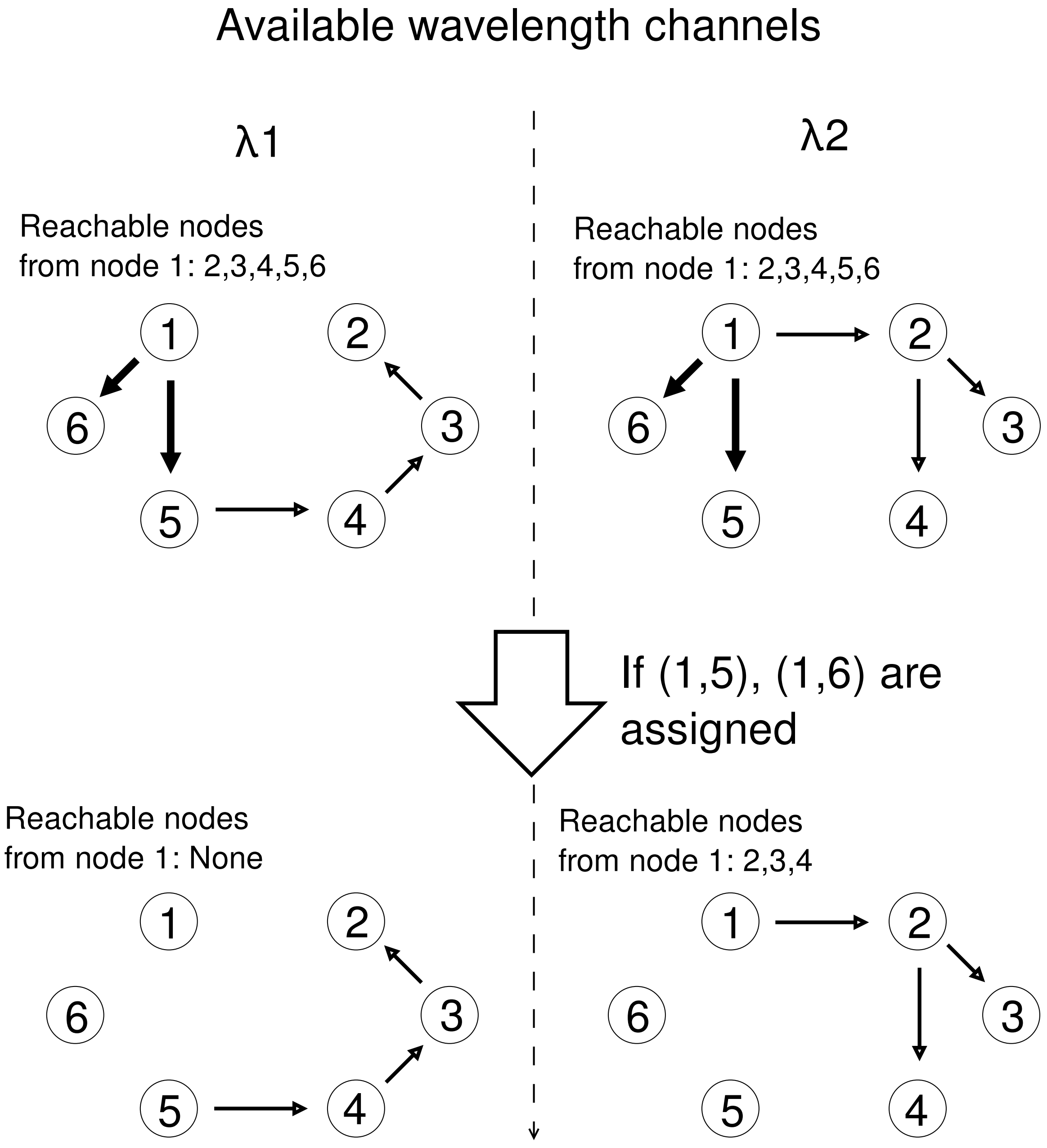


Figure 24: Example scenario

3.2.4 Computational complexity

In the proposed scheme routing and wavelength assignment are performed separately. Routing is performed on each wavelength independently for each wavelength based on the Steiner tree algorithm using MIP. Thus the set of trees $\mathbb{T}$ found from the MIP are minimum cost single-wavelength trees. Given this set of trees, the Δ values are compared and the wavelength such that Equation (30) is minimized is assigned. Compared with the conventional scheme, it involves an additional process of searching for the values of Δ. The value of Δ can be found by a standard node searching algorithm. For example, standard Breadth-First Search (BFS) is known to have a computational complexity of $O(|V| + |E|)$, where $|V|$,$|E|$ are the number of nodes and number of edges of the network, respectively. This process has to be repeated for each node and for each wavelength channels, and thus, the computational complexity is bounded to $O(W|V|(|V| + |E|))$, where W is the number of wavelength channels in the fiber. Note that this is the worst case scenario since when searching for the connectivity of all nodes, information can be shared and lots of the searching steps are saved. The Steiner tree algorithm used to search the multicast tree is known to be NP-complete, thus it would not be causing scalability problem although the proposed method increases computational complexity.

3.3 Simulation Analysis

The proposed scheme with a generic scheme that chooses minimal Steiner tree for each wavelength assignment are compared. The OMNet++ [19] simulator is used to simulate the network. The conventional scheme is called “minCost” and the proposed scheme is called “minΔ”. The topology used is shown in Figure 25. The numbers on the edges indicate the edge cost in the figure. The simulation parameters are given in Table 5. The simulation scenario is as follows: At the beginning there is no multicast sessions established on the network. Therefore all the wavelength channels of all the links are available for assignment. Next different dynamic sessions start requesting for wavelength resource at random time to establish the session, except for Section 3.3.6 where static sessions are assigned in the beginning. Assume the multicast session request is sent to the control plane early enough such that it allows the control plane to finish the MC-RWA process before the session starting time. For these sessions each node is equally probable to be the source node, and receivers set of random size (under uniform distribution), which is formed by any nodes other than the source node. The dynamic sessions also specify clearly the end time of the sessions. When a successful dynamic session finishes using the wavelength channel ("end time" is reached), the wavelength channels it used are released and will be available to be used by future sessions.

Table 5: Simulation parameters

Parameters	Values
Network topology	NSFNET (Figure 25) , 15 nodes total
Number of static sessions	0 (except Section 3.3.6)
Arrival interval of dynamic sessions on each node	Erlang distribution, mean 3600s
Duration of each dynamic session (except Section 3.3.5)	Exponential distribution, mean 900s-18000s
Duration of each dynamic session (Section 3.3.5)	Exponential distribution, mean 9000s (8 wavelength case), 18000s (16 wavelength case)
Offered traffic from each node (except Section 3.3.5)	0.25 - 5.00 Erlang
Offered traffic from each node (3.3.5)	2.5 Erlang (8 wavelength case), 5 Erlang (16 wavelength case)
Size of receiver nodes set of multicast session	Uniform distributed from 1 to 14
Number of wavelength per fiber	8,16
Number of fiber per edge	1

When a multicast session requests for wavelength channels, both minCost and minΔ first search for the light-tree (routing) using MIP for each wavelength, and the cost of the light-tree if the light-tree is found. The minCost method can immediately determine the wavelength of the light-tree to be assigned to the session by choosing the tree with lowest cost, but the minΔ method searches for the wavelength with the lowest value of Δ. If the MC-RWA process succeeds, wavelength channels on the networks are being assigned to the sessions. As time advances, wavelength channels are being occupied by different sessions, and there is possibility that new coming session might fail to get wavelength resources assigned under the wavelength channel condition. It is recorded as "blocked" and no wavelength channel will be assigned. The blocking probability which is calculated using Equation (27) is compared.

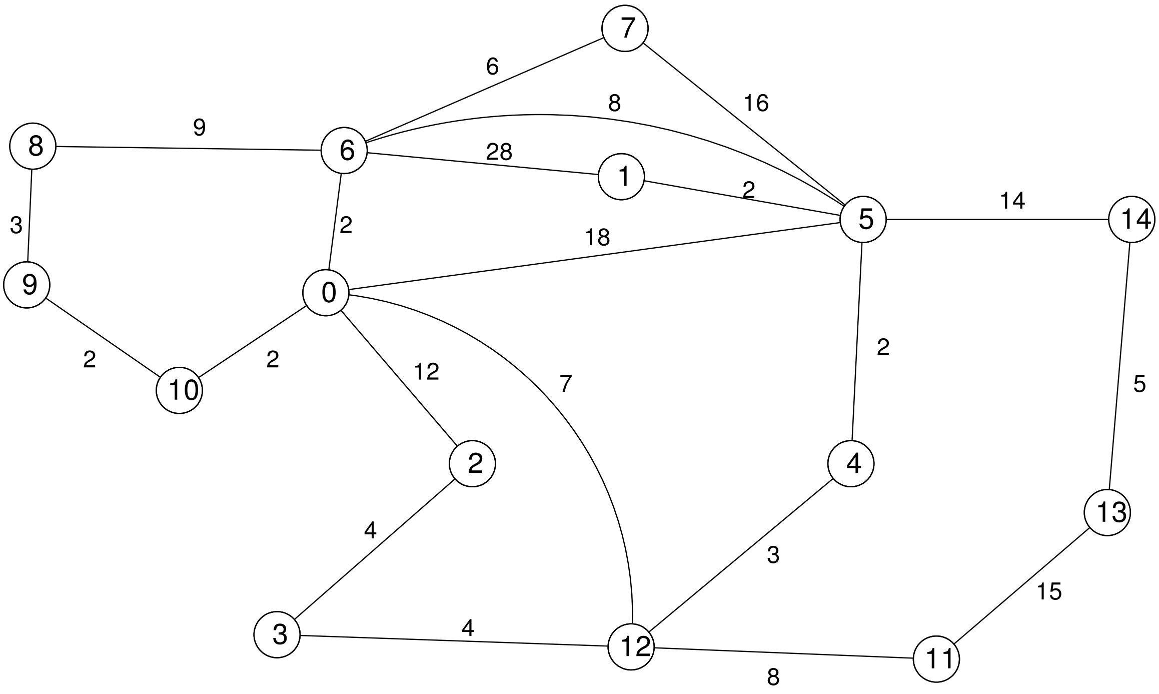


Figure 25: Topology used in simulation

During the process of MC-RWA for a particular multicast session in both minCost scheme and minΔ scheme, the network has no knowledge about the future request. Thus the network has to determine the route and the wavelength for the multicast session based on the current network conditions, i.e. the current wavelength usage of the network. It is assumed the network control plane has global knowledge of wavelength channel usages.

3.3.1 Blocking probabilities

Figure 26 shows the blocking probabilities of the proposed scheme and the minCost scheme. For both schemes, the offered load per node on the network is varied. The traffic load is measured in Erlang, which is given by the product of average request arrival interval and the requested session duration (see Table 5). The larger traffic load leads to higher blocking probability. When compared the minΔ scheme with minCost scheme, it can be seen from the results that the proposed scheme achieves lower blocking probabilities. During the process of multicast tree searching, both minCost and minΔ search for the minimal Steiner tree across the available wavelength. When a set of trees is found, the minCost scheme picks up the tree from the set which has the lowest cost. $d(n, \lambda, t, T_i(\lambda))$ for a particular node n for a wavelength assignment represents the loss of connectivity of node n to a set of other nodes by the assignment. When there are multiple wavelengths available to choose, in the proposed scheme the wavelength is chosen such that the sum of $d(n, \lambda, t, T_i(\lambda))$ values for all nodes is minimum. In the other words, overall loss of connectivity for all nodes is minimum, and this favors the chance of successful future multicast sessions establishment. Therefore, the proposed scheme can decrease the blocking probabilities compared with the conventional scheme.

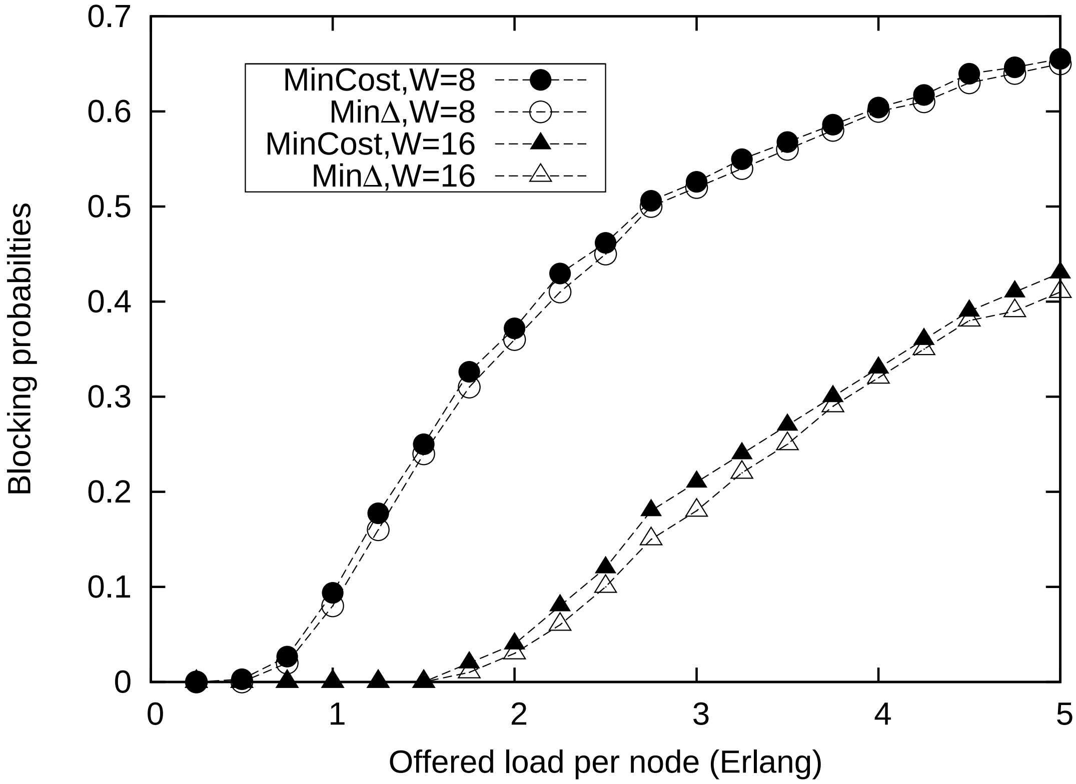


Figure 26: Blocking probabilities vs offered load per node

3.3.2 Variation of wavelength

Refer to Figure 26 again, the blocking probabilities is compared when the available network resource changes. For instance blocking probabilities is compared when number of wavelength = 8 with offered load per node = 1.5, and the case when number of wavelength =

16 with offered load per node = 3. When the network resource and offered load are both doubled at the same time, both minCost and minΔ schemes achieve lower blocking probabilities (minCost: 0.25 → 0.21, minΔ: 0.24 → 0.18). The decrease in blocking probabilities in both scheme can be explained by the fact that more wavelength channels are offered by the diversity of unused wavelength channels. In the minΔ scheme the decrease of blocking probabilities is larger than the minCost scheme. As in the proposed scheme wavelength with reducing blocking probabilities is chosen as the main criteria, diversity of wavelength channels naturally improves the performance of the proposed scheme.

3.3.3 Receivers set size of successful multicast sessions

Figure 27 shows the receivers set size of multicast sessions which are successfully established. When the offered traffic is low and there are nearly no blocking for the sessions, it can be seen that the average receivers set size of the multicast sessions is 7.5. This agrees with the simulation parameters as shown in Table 5 that as the requested session's receivers set size is uniformly distributed from 1-14, the average size of the sessions should be 7.5 when no blocking occurs. When the traffic load from each node increases, the average receivers set size of the multicast sessions decreases for both minΔ and minCost scheme. This is because the more the wavelength channels are utilized, the more difficult for sessions with large receivers set size to be established. However in the proposed scheme, it can seen that a slightly higher average receivers set size can be achieved in the minΔ scheme. This is because the minΔ scheme focuses on the loss of connectivity, which is measured by decrease in number of nodes that the nodes in the network can connect to. $\Delta(\lambda, t, T_i(\lambda))$ in equation (30) reflects the summary of the loss of connectivity, and thus the minΔ scheme has advantage on the receivers set size. Consider that the minΔ scheme has a lower blocking probabilities, this means that the minΔ scheme allows higher throughput than the minCost scheme as more sessions can be established and on average more receivers can be served by each session.

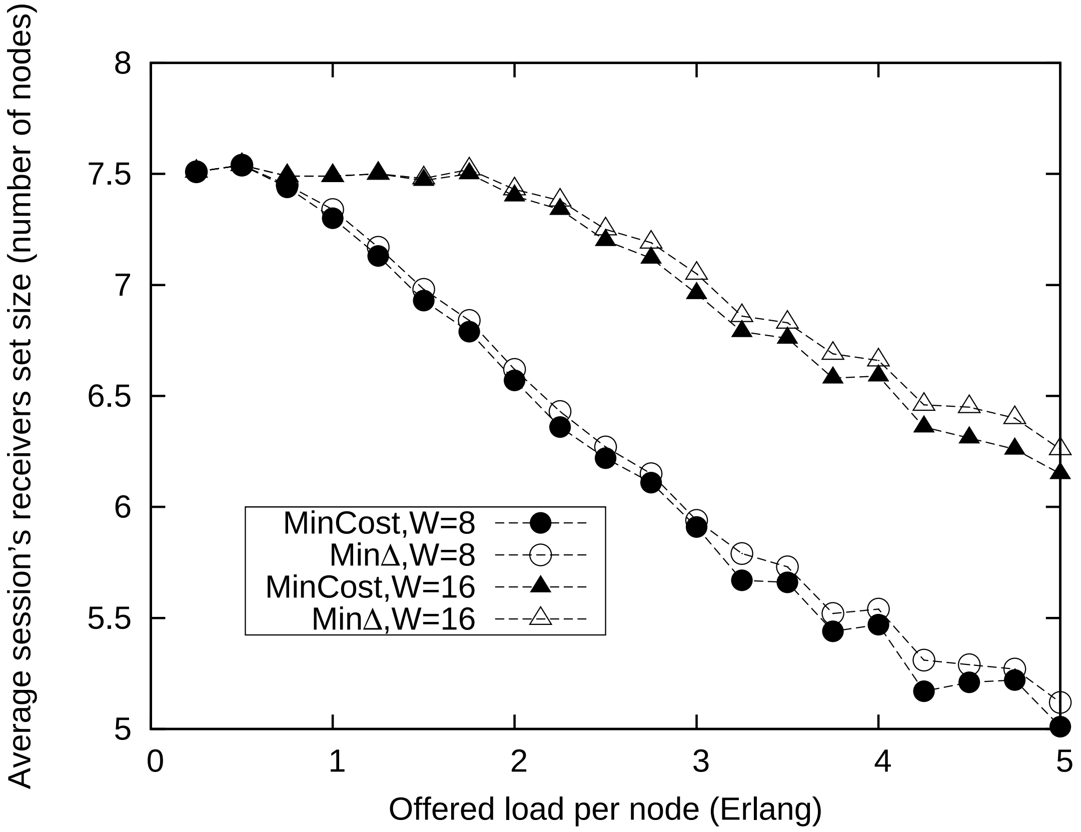


Figure 27: Average receivers set size vs offered load per node

3.3.4 Tree cost

Figure 28 shows the tree cost of the proposed scheme and the minCost scheme. In Section 3.3.1 it was mentioned that the proposed scheme does not choose the tree with minimum cost. Therefore, the proposed scheme will lead to a higher cost compared with the conventional scheme as shown in Figure 28 when the network load is not too heavy. It can be seen that both minCost and minΔ have their average costs decreased slightly when more sessions are put in the simulation (the network is more overloaded). This is because when network is heavily utilized, sessions with small receivers set size are more probable to be established. They tend to span smaller amount of links, and thus, the average cost becomes lower. Our scheme in general has a higher cost than the minCost scheme as expected. In Section 3.3.1 it is shown that the proposed scheme has a lower blocking probability. In order word, the proposed scheme is able to establish more sessions under the same load. The additional sessions are expected to be smaller when the network is congested, and these increased small sessions contributes to the lower average cost in the proposed scheme under the heavy load.

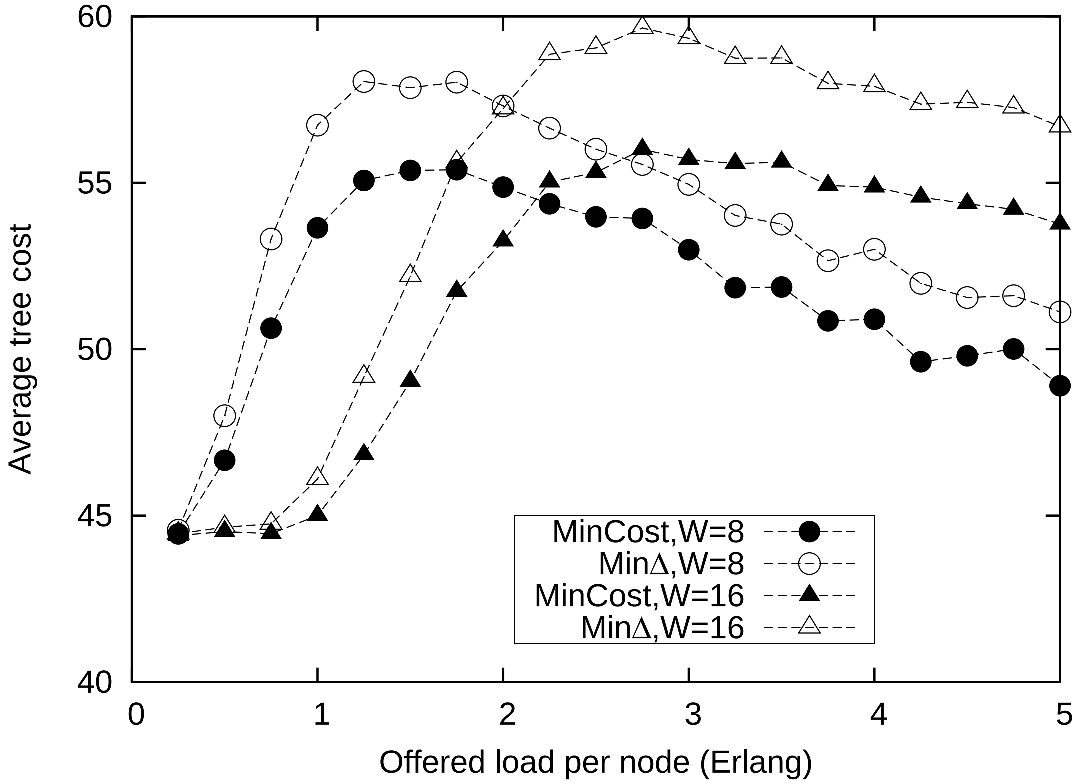


Figure 28: Average tree cost vs offered load per node

3.3.5 Mixing with unicast sessions

In this sub-section a case that unicast traffic streams coexist with the multicast traffic streams in the network is presented. In contrast to Section 3.3.1, the offered traffic per node is fixed to 2.5 and 5 Erlang respectively for the network with 8 and 16 wavelength channels in their edges so that the offered load to the network are in proportion. The ratio of unicast sessions to the multicast sessions is varied to see how both the conventional scheme and proposed scheme perform. The shortest path algorithm is used for routing of the unicast sessions on each available wavelength, and assign the wavelength by using minCost and minΔ as in the previous simulation analysis. That is, unicast is viewed as a special case of multicast, where the number of destination nodes is one, in the wavelength assignment process. The ratio of unicast sessions is varied from 0% to 100% in the simulation.

The results are shown in Figure 29. As unicast sessions only need to establish a path from the source to the destination instead of a tree in the case of multicast, unicast sessions consume less wavelength resources than multicast sessions, thus the blocking probabilities decrease with the increase of percentage of unicast sessions in both schemes. When the number of wavelength channels of each edge is set to 8, it can be seen that a small reduction of the blocking probabilities with the minΔ scheme compared with the minCost scheme. However if the number of wavelength is 16, a larger reduction of the blocking probabilities by using the proposed scheme is observed. This can be explained by the similar argument in Section 3.3.2 that larger number of wavelength provides better diversity for the scheme to choose. As the proposed scheme is designed with the consideration of least reduction of connectivity of all the nodes, wavelength assignment for unicast sessions using the proposed scheme can also lead to lower blocking probabilities of the requesting sessions.

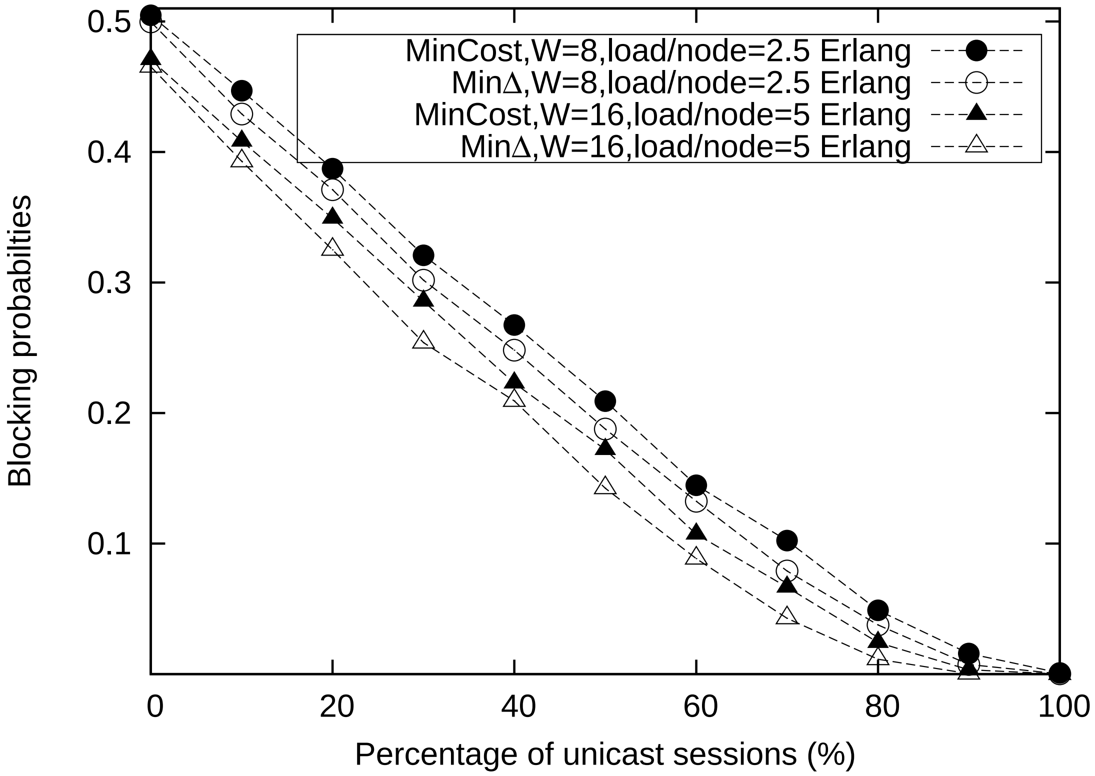


Figure 29: Blocking probabilities vs offered load per node when mixing with unicast sessions

3.3.6 Mixing with static sessions

Static sessions occupy the network resources all the time and never release wavelength channels like dynamic sessions do. Because of the permanent occupation of the network resources, network cost is put at the first priority when conducting MC-RWA for static sessions. On the other hand, the proposed scheme is designed with the consideration on the overall connectivity loss caused by dynamic sessions, and the aim is to reduce blocking probabilities. So the effect when the static sessions coexist with the dynamic sessions is studied. The simulation for two cases are run that each edge on the network has 8 or 16 wavelength channels. For the 8-wavelength case each node in (1,3,5,7,9,11,13) establishes one static session, or a total of 7 static sessions always occupy the network. For the 16-wavelength case each node in the network establishes one static session, or a total of 15 static sessions are assigned. These static sessions are assigned using the minCost approach, where the sessions' destinations are random. MinΔ approach cannot be used here because all the static sessions are considered altogether to achieve lower overall cost [13]. Then similar to Section 3.3.1, dynamic multicast sessions request and release network resources at different time, and the blocking probabilities are observed.

Figure 30 shows the simulation results of the blocking probabilities of the dynamic sessions. With the presence of cost-optimized static sessions, the proposed scheme can still achieve a lower blocking probabilities than the minCost scheme. With the presence of static sessions, the wavelength channels are permanently occupied and some links are more congested because of their lower cost. Similar to pure dynamic cases as described in 3.3.1, the proposed scheme chooses the wavelength channel which leads to overall least connectivity loss and thus gives better chance for future session to be established. And therefore the proposed scheme achieve better blocking probabilities even with the presence of static sessions. The results is less obvious for the case that there are only 8 wavelength channels of which 7 nodes establish multicast sessions. This is because the network is already occupied by the static sessions and choices of available wavelength channels are very limited. Therefore this leads to the improvement on blocking probabilities to be very limited.

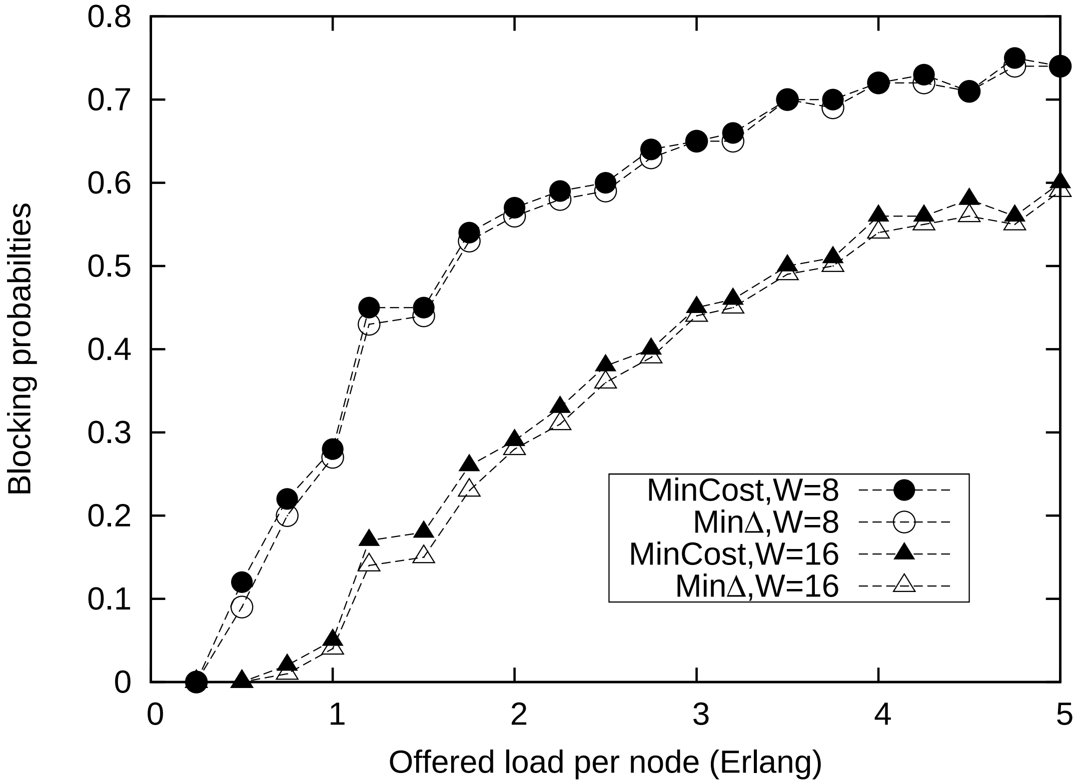


Figure 30: Blocking probabilities of dynamic sessions when mixing with static sessions

3.4 Conclusion

A scheme of MC-RWA to establish light-tree for dynamic multicast session for the WDM network by using the newly defined parameter Δ to choose the wavelength has been proposed to reduce blocking probabilities. As Δ reflects the overall connectivity loss of the network nodes, choosing the wavelength where Δ is minimum can lead to a higher chance of connection establishment in the future. The simulation results show that higher number of multicast sessions can be achieved with proper wavelength channel assignment with the tradeoff of increasing network cost for the sessions.

CHAPTER 4

CONCLUSION

In this dissertation multicast in both IP network and also optical WDM network were studied.

Chapter 1 presented an introduction to the multicast data delivery. Multicast in the IP layer was introduced, including the related protocols which realize multicast data delivery. Next the reliable multicast was introduced, in which data has to be delivered correctly to the recipients correctly. Although a large application area multicast is multimedia data delivery in which some data loss is tolerable, there are increasing demands for reliability in multicast such as distributed computing. Structure of the multicast tree for reliable multicast and the data recovery techniques were discussed. Then multicast tree in the optical network was introduced in which WDM technique is employed. In the optical WDM network resources are being allocated to different sessions. The issues and limitations on the WDM network in order to setup a multicast session was discussed.

Chapter 2 presented the proposed reliable multicast protocol using local retranmission and FEC based on group-aided multicast scheme. In reliable multicast, feedback and recovery traffic limit the performance and scalability of the multicast session. In the proposed scheme, the original GAM was improved by making use of FEC locally in addition to NACK/ retransmission in its local-group based recovery. The proposed scheme produces FEC packets and multicasts the packets within the scope of a local group in order to correct uncorrelated errors of the local members in each group of the multicast session, which reduces the need for NACK/retransmission. By using the proposed scheme, redundancy traffic can be localized in each group within a multicast session, and the overall recovery traffic can be reduced. In this chapter recovery mechanism can be used in the local network.

It is also important to effectively utilize the core network in order for multicast to be realized in a wide area. Chapter 3 explained the proposed scheme for multicast routing and wavelength assignment for dynamic multicast sessions in WDM network using minimum Δ. In this scheme a light-tree for dynamic multicast session for the WDM network is established by choosing the wavelength that leads to a reduction in blocking probabilities by using a parameter Δ. Δ is defined as the overall reduction of connectivity of the nodes in the network caused by a wavelength assignment process when using a particular wavelength, and wavelength resources to the multicast session are assigned by choosing the Δ which leads to smallest reduction in connectivity. Through computer simulation, it was shown that the proposed scheme has lower blocking probabilities when compared with minimum cost scheme under the condition that wavelength conversion is not allowed.

As a concluding remarks, this dissertation reviewed some issues and solutions in applying multicast for reliable data delivery using multicast, as well as a new parameter in MC-RWA in order to reduce the blocking probabilities.

APPENDIX A

PUBLICATIONS BY THE AUTHOR

A.1 Journal papers

- Alex Fung, Iwao Sasase, “Reliable multicast with local retransmission and FEC using group-aided multicast scheme”, IEICE Trans. on Communications, Vol. E92-B, No.3, pp.811-818, March 2009
- Alex Fung, Iwao Sasase, “Multicast routing and wavelength assignment for dynamic mmulticast sessions in WDM network using minimum Δ”, IEICE Trans. on Communications, to be published

A.2 International Conferences

- Alex Fung, Iwao Sasase, “Hybrid local recovery scheme for reliable multicast using group-aided multicast scheme”, 2007 IEEE Wireless Communications and Networking Conference, Hong Kong, pp.3478 - 3482 , March 2007
- Alex Fung, Iwao Sasase, “Survey on Light-Tree Based Multicast Data Delivery on WDM Networks”, 1st Gent Univ. and Keio Univ. G-COE joint workshop for future network, Belgium, March 2008
- Alex Fung, Iwao Sasase, “Wavelength assignment scheme for dynamic multicast sessions in optical WDM network”, The First Joint Taiwan National Chiao Tung University and Japan Keio University Wireless Workshop, Taiwan, December 2008
- Alex Fung, Iwao Sasase, “Multicast Routing and Wavelength Assignment for Dynamic Multicast Sessions in WDM Network Using Minimum Delta”, The First University of Sydney-Keio University Joint Workshop on Information Technology, Sydney, February 2010

A.3 Domestic Conferences

- Alex Fung, Iwao Sasase, “Hybrid local recovery scheme for reliable multicast using group-aided multicast scheme”, IEICE Domestic conference on Network System, NS2006-67, pp.93-96, July 2006

REFERENCES

[1] B. Rimal, E. Choi, and I. Lumb, “A taxonomy and survey of cloud computing systems,” in *INC, IMS and IDC, 2009. NCM ’09. Fifth International Joint Conference on*, pp. 44 –51, 25-27 2009.

[2] G. Popescu and Z. Liu, “Stateless application-level multicast for dynamic group communication,” in *Distributed Simulation and Real-Time Applications, 2004. DS-RT 2004. Eighth IEEE International Symposium on*, pp. 20 – 28, 21-23 2004.

[3] J. Zhu, T. Tsuchiya, and K. Koyanagi, “Peer-to-peer collaborative application-level multicast,” in *E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services, 2007. CEC/EEE 2007. The 9th IEEE International Conference on*, pp. 228 –238, 23-26 2007.

[4] P. RenJie, S. JunDe, and H. L. iong, “Application level multicast in hierarchical topology,” in *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, vol. 2, pp. 805 – 808 vol.2, 4-7 2003.

[5] K.-I. Kim, D.-H. Choi, and S.-H. Kim, “Deployment issues for application level multicast,” in *Communications, 2003. APCC 2003. The 9th Asia-Pacific Conference on*, vol. 3, pp. 1087 – 1091 Vol.3, 21-24 2003.

[6] M. Handley, S. Floyd, B. Whetten, R. Kermode, L. Vicisano, and M. Luby, “The reliable multicast design space for bulk data transfer,” *Internet Request For Comments*, vol. 2887.

[7] L. Rizzo, “Effective erasure codes for reliable computer communication protocols,” *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 2, p. 36, 1997.

[8] L. Rizzo and L. Vicisano, “A reliable multicast data distribution protocol based on software FEC techniques,” in *High-Performance Communication Systems, 1997. (HPCS ’97) The Fourth IEEE Workshop on*, pp. 116 –125, 1997.

[9] B. Whetten and G. Taskale, “An overview of reliable multicast transport protocol II,” *Network, IEEE*, vol. 14, pp. 37 –47, jan/feb 2000.

[10] S. Floyd, V. Jacobson, C.-G. Liu, S. McCanne, and L. Zhang, “A reliable multicast framework for light-weight sessions and application level framing,” *Networking, IEEE/ACM Transactions on*, vol. 5, pp. 784 –803, dec 1997.

[11] W. Yoon, D. Lee, H. Y. Youn, S. Lee, and S. J. Koh, “A combined group/tree approach for scalable many-to-many reliable multicast,” in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp. 1336 – 1345 vol.3, 2002.

[12] B. Mukherjee, *Optical WDM networks.* Springer-Verlag New York Inc, 2006.

[13] N. Singhal, L. Sahasrabuddhe, and B. Mukherjee, “Optimal multicasting of multiple light-trees of different bandwidth granularities in a WDM mesh network with sparse splitting capabilities,” *Networking, IEEE/ACM Transactions on*, vol. 14, no. 5, pp. 1104–1117, Oct. 2006.

[14] A. Khalil, A. Hadjiantonis, C. M. Assi, A. Shami, G. Ellinas, and M. A. Ali, “Dynamic provisioning of low-speed unicast/multicast traffic demands in mesh-based WDM optical networks,” *Lightwave Technology, Journal of*, vol. 24, no. 2, pp. 681–693, Feb. 2006.

[15] L. Liao, L. Li, and S. Wang, “Dynamic multicast traffic grooming in WDM mesh networks,” *Next Generation Internet Design and Engineering, 2006. NGI ’06. 2006 2nd Conference on*, pp. 366–370, 3-5 April 2006.

[16] W. Wattanavarakul, S. Segkhoonthod, and L. Wuttisittikulkij, “Design of multicast routing and wavelength assignment in multifiber WDM mesh networks for asymmetric traffics,” *TENCON 2005 2005 IEEE Region 10*, pp. 1–6, Nov. 2005.

[17] J. He, S.-H. G. Chan, and D. H. Tsang, “Routing and wavelength assignment for WDM multicast networks,” *Global Telecommunications Conference, 2001. GLOBECOM ’01. IEEE*, vol. 3, pp. 1536–1540 vol.3, 2001.

[18] A. Araar and H. Khali, “A simulation study of a dynamic multicast using FEC,” in *Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on*, vol. 2, pp. 523 – 526 Vol.2, 14-17 2003.

[19] A. Varga, “OMNeT++ discrete event simulator.” `http://www.omnetpp.org/`, 2010.

[20] O. Alay, T. Korakis, Y. Wang, and S. Panwar, “An experimental study of packet loss and forward error correction in video multicast over IEEE 802.11 b network,” in *Proceedings of IEEE CCNC*, 2009.

[21] O. Alay, T. Korakis, Y. Wang, and S. Panwar, “Is physical layer error correction sufficient for video multicast over IEEE 802.11g networks?,” in *Consumer Communications and Networking Conference, 2009. CCNC 2009. 6th IEEE*, pp. 1 –5, 10-13 2009.

[22] L. H. Sahasrabuddhe and B. Mukherjee, “Light trees: optical multicasting for improved performance in wavelength routed networks,” *Communications Magazine, IEEE*, vol. 37, no. 2, pp. 67–73, Feb 1999.

[23] G. Rouskas, “Optical layer multicast: rationale, building blocks, and challenges,” *Network, IEEE*, vol. 17, pp. 60–65, Jan/Feb 2003.

[24] Y. Yang, J. Wang, and C. Qiao, “Nonblocking WDM multicast switching networks,” *Parallel and Distributed Systems, IEEE Transactions on*, vol. 11, pp. 1274–1287, Dec 2000.

[25] W. Yao, G. Sahin, M. Li, and B. Ramamurthy, “Analysis of multi-hop traffic grooming in WDM mesh networks,” *Broadband Networks, 2005 2nd International Conference on*, pp. 165–174 Vol. 1, Oct. 2005.

[26] Z. Zhang and Y. Yang, "On-line optimal wavelength assignment in WDM networks with shared wavelength converter pool," *Networking, IEEE/ACM Transactions on*, vol. 15, pp. 234–245, Feb. 2007.

[27] X. Zhang, J. Y. Wei, and C. Qiao, "Constrained multicast routing in WDM networks with sparse light splitting," *Lightwave Technology, Journal of*, vol. 18, no. 12, pp. 1917–1927, Dec 2000.

[28] H. Zang, J. P. Jue, and B. Mukherjee, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks," *Optical Networks Magazine*, vol. 1, pp. 47–60, Jan 2000.

[29] R. Barry and S. Subramaniam, "The MAX SUM wavelength assignment algorithm for WDM ring networks," in *Optical Fiber Communication. OFC 97., Conference on*, pp. 121–122, Feb 1997.

[30] S. Subramaniam and R. Barry, "Wavelength assignment in fixed routing wdm networks," in *Communications, 1997. ICC 97 Montreal, 'Towards the Knowledge Millennium'. 1997 IEEE International Conference on*, vol. 1, pp. 406 –410 vol.1, 8-12 1997.

[31] X. Zhang and C. Qiao, "Wavelength assignment for dynamic traffic in multi-fiber WDM networks," in *Computer Communications and Networks, 1998. Proceedings. 7th International Conference on*, pp. 479–485, Oct 1998.

[32] L. H. Sahasrabuddhe and B. Mukherjee, "Light trees: optical multicasting for improved performance in wavelength routed networks," *Communications Magazine, IEEE*, vol. 37, no. 2, pp. 67–73, Feb 1999.

[33] N. K. Singhal and B. Mukherjee, "Architectures and algorithm for multicasting in WDM optical mesh networks using opaque and transparent optical cross-connects," *Optical Fiber Communication Conference and Exhibit, 2001. OFC 2001*, vol. 2, pp. TuG8–1–TuG8–3 vol.2, 2001.

[34] J. Wang, X. Qi, and B. Chen, "Wavelength assignment for multicast in all-optical WDM networks with splitting constraints," *Networking, IEEE/ACM Transactions on*, vol. 14, pp. 169–182, Feb. 2006.

[35] E. Modiano, "Traffic grooming in WDM networks," *Communications Magazine, IEEE*, vol. 39, pp. 124–129, Jul 2001.

[36] B. Chen, G. Rouskas, and R. Dutta, "On hierarchical traffic grooming in WDM networks," *Networking, IEEE/ACM Transactions on*, vol. 16, pp. 1226–1238, Oct. 2008.

[37] R. Ul-Mustafa and A. Kamal, "Design and provisioning of WDM networks with multicast traffic grooming," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 4, pp. –53, 2006.

[38] V. V. M. Nhat, A. Obaid, and P. Poirier, "Application of game theory in traffic grooming," *Wireless and Optical Communications Networks, 2005. WOCN 2005. Second IFIP International Conference on*, pp. 383–387, March 2005.

[39] R. Shenai and K. Sivalingam, "Analysis of ip grooming approaches in optical WDM mesh networks," *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, vol. 4, pp. 5 pp.–, Nov.-2 Dec. 2005.

[40] O. Awwad, A. Al-Fuqaha, and M. Guizani, "Genetic approach for traffic grooming, routing, and wavelength assignment in WDM optical networks with sparse grooming resources," *Communications, 2006. ICC '06. IEEE International Conference on*, vol. 6, pp. 2447–2452, June 2006.

[41] J. Zhou and X. Yuan, "A study of dynamic routing and wavelength assignment with imprecise network state information," *Parallel Processing Workshops, 2002. Proceedings. International Conference on*, pp. 207–213, 2002.

[42] X.-H. Jia, D.-Z. Du, X.-D. Hu, M.-K. Lee, and J. Gu, "Optimization of wavelength assignment for qos multicast in WDM networks," *Communications, IEEE Transactions on*, vol. 49, pp. 341–350, Feb 2001.

[43] R. Libeskind-Hadas and R. Melhem, "Multicast routing and wavelength assignment in multihop optical networks," *Networking, IEEE/ACM Transactions on*, vol. 10, pp. 621–629, Oct 2002.

[44] S. Sankaranarayanan and S. Subramaniam, "Comprehensive performance modeling and analysis of multicasting in optical networks," *Selected Areas in Communications, IEEE Journal on*, vol. 21, pp. 1399–1413, Nov. 2003.

[45] B. Chen and J. Wang, "Efficient routing and wavelength assignment for multicast in WDM networks," *Selected Areas in Communications, IEEE Journal on*, vol. 20, pp. 97–109, Jan 2002.

[46] T. Makabe and T. Takenaka, "WDM multicast tree construction algorithms for minimizing blocking probability under a delay constraint," in *Computer Communications and Networks, 2008. ICCCN '08. Proceedings of 17th International Conference on*, pp. 1–6, Aug. 2008.

[47] X. Qin and Y. Yang, "Blocking probability in WDM multicast switching networks with limited wavelength conversion," in *Network Computing and Applications, 2003. NCA 2003. Second IEEE International Symposium on*, pp. 322–329, April 2003.

[48] M. Yajnik, J. Kurose, and D. Towsley, "Packet loss correlation in the MBone multicast network," in *Global Telecommunications Conference, 1996. GLOBECOM '96. 'Communications: The Key to Global Prosperity*, pp. 94 –99, 18-22 1996.

[49] S.-W. Tan, G. Waters, and J. Crawford, "A multiple shared trees approach for application layer multicasting," in *Communications, 2004 IEEE International Conference on*, vol. 3, pp. 1456 – 1460 Vol.3, 20-24 2004.

[50] B. Levine, D. Lavo, *et al.*, "The case for reliable concurrent multicasting using shared ack trees," in *Proceedings of the fourth ACM international conference on Multimedia*, p. 376, ACM, 1997.

[51] K. Obraczka, "Multicast transport protocols: a survey and taxonomy," *Communications Magazine, IEEE*, vol. 36, pp. 94 –102, jan 1998.

[52] M. Lacher, J. Nonnenmacher, and E. Biersack, “Performance comparison of centralized versus distributed error recovery for reliable multicast,” *Networking, IEEE/ACM Transactions on*, vol. 8, pp. 224 –238, apr 2000.

[53] D. Dujovne and T. Turletti, “Multicast in 802.11 WLANs: an experimental study,” in *Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*, p. 138, ACM, 2006.

[54] C. Index, “Global Mobile Data Traffic Forecast Update, 2009–2014,” *Cisco Systems*, 2010.