A Two – Phase Distributed Reconfiguration Algorithm for Node Failures in IP-Over-WDM Networks

N.Sreenath, Balaji Palanisamy
Department of Computer Science and Engineering,
Pondicherry Engineering College, Pondicherry-605104, India.
{nsreen@yahoo.com,balajipec@gmail.com}

Abstract

In Wavelength Routed Wavelength Division Multiplexing (WR-WDM) networks, a set of lightpaths (all-optical communication paths) defines the virtual topology. Internet Protocol (IP) makes use of virtual topology to route its traffic in the optical form. In the event of a node failure, the virtual topology may be partitioned so that a set of nodes may not be able to communicate with another set of nodes. The failure may also affect the objective function value leading to decreased network performance. In such a case, it may be indispensable that the connectivity in the virtual topology is restored and the network performance is improved in terms of objective function value. Hence, it may be necessary to alter the existing virtual topology. Altering of a virtual topology is called as virtual topology reconfiguration. In this paper, we propose a two-phase distributed algorithm for virtual topology reconfiguration in the event of a node failure. The algorithm restores the failed connectivity during phase I. Phase II of the algorithm improves the network performance by bringing down the objective function value of the network.

Key Words: WDM Optical Network, Node Failures, Reconfiguration of Virtual topology.

1.Introduction

With the Internet Protocol (IP) playing a dominant role in wide area networking technology and advancements in wavelength-routed wavelength division multiplexing (WDM) technology to provide huge bandwidth, the IP-over-WDM technology becomes the right choice for next generation Internet networks [1]. In these networks, messages are carried from one node to another node using *lightpaths*. A *lightpath* is an all-optical path established between two nodes in the network by the allocation of same wavelength on all links of the path. In IP-over-WDM networks, lightpaths are established between IP routers. The problem of establishing lightpaths statically (*a priori*) is referred to as static lightpath establishment (SLE) problem. IP-over-WDM networks make use of these lightpaths to transfer data. This set of pre-established lightpaths is called as *virtual topology.* Virtual topology is a graph with nodes as the routers in physical network (IP-over-WDM network) and edges corresponding to the lightpaths between them. A node pair may not have a lightpath due to non-availability of wavelengths between them or transmitters at the source or receivers at the destination. Traffic between a node pair *sans* a lightpath is routed using multiple lightpaths (logical hops). A virtual topology (which is a graph) is said to be *connected*, if every node in the virtual topology can communicate either using single lightpath (single hop) or using multiple lightpaths (multihop) to every other node. A virtual topology may be designed with an objective of minimizing certain metric (objective function) such as maximum congestion on a lightpath [3], average packet delay [4], or average weighted hop count [5].

Virtual topology needs to be altered in various situations. This procedure is referred to as *reconfiguration.* The first situation, which calls for reconfiguration, is when the traffic demand changes at higher layers. Here, a set of lightpaths needs to be changed to reflect the new traffic. The second situation for reconfiguring the virtual topology arises when the network elements such as links and nodes fail. Due to a node failure, even if the physical topology remains connected, the virtual topology may become disconnected. It means that physically the network may be connected, but the virtual topology may be partitioned into a number of components. In such a case, the virtual topology may be reconfigured to provide connectivity among the nodes in the network. This process could succeed only if the number of transceivers at every node is greater than one. Hence in our work, we assume that more than one transceiver exists in every node. Also we consider a 2-Connected physical topology which remains connected even after the removal of a node.

0-7695-2699-3/06/$20.00 (c) IEEE

2.Related Work

In [5-7], different algorithms for reconfiguration due to traffic changes were proposed. These algorithms try to minimize the objective function value in the event of changes in traffic but they do not consider the restoration of connectivity in the virtual topology, if connectivity is lost due to the node failure. Moreover, these are centralized algorithms which involve huge control traffic to and from the central node. Also, the central node in these algorithms is a single point of failure. In [8], the authors have proposed ILP formulations for survivable routing of virtual topology. In this, the virtual topology is designed in such a way that the removal of a node does not disconnect the virtual topology. However, for a given physical topology, it is possible to find survivable routing only for a restricted set of logical topologies. Also, the added survivability comes only at the expense of additional network resources. In [9], centralized algorithm for virtual topology reconfiguration due to node failures has been proposed. This algorithm tries to establish the connectivity of the virtual topology after a node failure. Although centralized algorithms are well-suited for country wide networks, they do not scale well for world wide networks like the Internet. Distributed networks like IP over WDM networks demand a fully distributed process for reconfiguration. The major thrust behind the current work is to provide a distributed reactive solution to the problem of virtual topology reconfiguration after node failure.

The rest of the paper is organized as follows. Section 3 describes the problem. Section 4 explains the distributed reconfiguration methodology. Section 5 discusses the performance study of our method. Section 6 concludes the paper.

3. Problem Description

We assume the physical topology is 2-connected and hence it will remain connected even after a node failure. However, the virtual topology may not be 2-connected and hence a node failure can partition the virtual topology into two or more components. Since there is no connectivity existing between the divided components, a node in one component cannot communicate with nodes in other components. We also assume that every node has T tunable transmitters and T tunable receivers and a fiber carries W wavelengths. The physical link between two nodes is assumed to be bi-directional, in the sense that if there exists a link between a node pair (s,d), there also exists a link between the node pair (d,s). We use *average weighted hop count* (H_{avg}) [5] as the objective function which is defined as the average number of hops on the virtual topology traversed by an unit traffic.

Menger's Theorem: *The edge connectivity of a graph G is k if and only if every pair of vertices in G is joined by k or more edge-disjoint paths, and at least one pair of vertices is joined by exactly k edge disjoint paths.*

Let us consider a physical network G with edge connectivity k. It means that the smallest cut-set of the network is k. Since we consider a 2-connected physical network (failure of at least 2 nodes disconnects the network), failure of a single node does not disconnect the network. It means that any cut-set contains the links that are corresponding to (incident on) more than one node. Hence, failure of a single node does not affect all the links of a cut-set. Assume a cut-set c that contains k links. Failure of a node s causes failure of l links where $l < k$. These l links are incident on node s. Now let us consider a virtual topology VT that is constructed on G. VT is not guaranteed to be 2-connected, since the number of transceivers available in a node may be low. In such a case, if a node fails, some lightpaths may fail and hence the virtual topology may get disconnected. It is because, these failed lightpaths are using the failed physical links that are incident on the failed node. Our reconfiguration process tries to route the traffic carried by the failed lightpaths on links (k-l) that are not failed but in the cut-set c.

Lemma 1 *If G is a simple connected graph (network) and v is a vertex (node) with degree d(v), then the maximum number of one-way disconnected components formed due to the failure of node v is d(v) x W, where W is the number of wavelengths carried by a physical link.*

Proof If node v fails, then all d(v) links connected to it also fail. If a link carries W wavelengths, then in the worst case, d(v) x W lightpaths may fail. Assume that these lightpaths belong to different single element cut-sets of virtual topology. Then, the virtual topology gets divided into d(v) x W one-way disconnected components. As a corollary to this, we can state that the maximum number of two-way disconnected components due to the failure of node v is $\frac{d(v) \times W}{2}$.

4. Reconfiguration methodology

The various control messages used by our protocol are sent through the control channel. The control channel is established by duplicating the physical topology on the virtual topology using the wavelength W_0. Since physical topology is 2-connected and the control channel is a duplication of the physical topology, it remains connected even after the node failure. Every node is assumed to know the following details: All lightpaths emerging from it, all lightpaths incident on it, load on each lightpath associated with that node and route of each lightpath. Each node also knows the details of the set of lightpaths associated with its neighboring nodes. A lightpath is considered for deletion only when its deletion does not affect the connectivity of the network. Such lightpaths are referred to as *non-critical lightpaths*. The lightpaths whose deletion disconnects the virtual topology are referred to as *critical lightpaths*. Criticality of a lightpath is verified by flooding a control packet on the virtual topology. If the source of the lightpath receives an acknowledgement from the other end node of the lightpath, it concludes that the lightpath is non-critical. Our reconfiguration process has two phases. Phase I restores the connectivity in the virtual topology and phase II improves the network performance in terms of objective function value. The sequence of events that takes place during phases I and II of the reconfiguration process are discussed below:

4.1 Phase I

The failure of a node say node v is first identified by the physical neighbors of node v. The failure of the node affects all lightpaths associated with it. Once the failure is detected, a node adjacent to node v prepares a message containing the set of sources and destinations of the lightpaths routed through the failed node and sends it to the source and destination of each failed lightpath using the control channel.

Assume $\{s_1,s_2,...s_n\}$ be the various sources and $\{d_1,d_2,....d_n\}$ be the various destinations of the lightpaths routed through the failed node v. Upon receiving the set of sources and destinations of the failed lightpaths, a node say t (can be a source or destination of the failed lightpaths) sends a message to every other source and destination of the failed lightpaths on the virtual topology. If a node say r receives this message, it sends an acknowledgement back on the control channel. From this acknowledgement, node t comes to know that there exists one-way connectivity from the component containing node t to the component containing node r. Node t also understands that node r is present in its own component as node r is reachable from node t. Thus, every node in the set $\{s_1,s_2,...s_n\}$ U $\{d_1,d_2,....d_n\}$ finds all the other nodes that it can communicate. If a node finds that it is able to communicate to all nodes in the set $\{ s_1,s_2,...s_n\}$ U $\{d_1,d_2,....d_n\}$, then it assumes that the connectivity of the virtual topology is not lost due to the node failure. In such a case Phase I terminates and Phase II is trigerred to improve the network performance by bringing the objective function value down.

If it is found that the virtual topology is disconnected, then a co-ordinator is elected in every component. Within a component, the node having the least IP address among the set of nodes in the set $\{s_1,s_2,...s_n\}$ U $\{d_1,d_2,....d_n\}$ is elected as the co-ordinator for that component. The co-ordinator node in each component sends a control message using the control channel to all the nodes in the set $\{s_1,s_2,...s_n\}$U $\{d_1,d_2,....d_n\}$ to which the co-ordinator does not have connectivity. The nodes on receiving this message reply with the IP address of their co-ordinator node in their component. From this, the co-ordinator in every component is informed about the co-ordinators of other components and hence the total number of logically divided components. Every component prepares the list of co-ordinators to which there is no connectivity.

The co-ordinator of each component broadcasts a *Req_Estab_lp* message on the virtual topology for establishing connectivity to every non-reachable component. This message requests the establishment of a set of new lightpaths that provide connectivity to the set of non-reachable components. Every *Req_Estab_lp* message consists of the list of co-ordinators to which the establishment of connectivity is requested. A node (say k) on receiving the *Req_Establ_lp* message, looks for a physical link connecting k and a node say q which is belonging to a component whose co-ordinator is mentioned in the received *Req_Estab_lp* message. If such a link exists, then node k checks for the availability of a transmitter with it. If it is not available, it checks whether a transmitter can be made free by deleting some other non-critical lightpaths. If so, it sends *Req_Bid_Resource* message to node q using the control channel. Node q checks for the availability of receiver in a similar way as node k did. Similarly, availability of wavelength on link kq is also verified. Once the information regarding the availability of resources is available, node k computes the weight of the lightpath that it wants to establish. The weight indicates a measure of the disturbance to the network due to deleting some other lightpaths inorder to provide the required resources. Here, lightpaths carrying less traffic or lightpaths whose hop count on the physical topology is less are given less weight. A weight value of infinity indicates that the resource cannot be granted.

Once the weight of the lightpath is computed, the same may be informed to the co-ordinator using *Bid_estab_lp* message using the control channel.

The co-ordinator after getting the *Bid_Estab_lp* messages from various nodes, finds the node which sent the lowest weight in its *Bid_Estab_lp* message. The co-ordinator then sends a *Estab_lp* message to the node which sent the least weight in its *Bid_Estab_lp* message. This node establishes the connectivity by deleting the least weighted lightpaths whose resources are necessary for the establishment of the new lightpath. After the establishment of the new lightpath, this node sends an *Estab_lp_Ack* message to acknowledge the lightpath establishment. If a co-ordinator C_1 establishes connectivity with another co-ordinator C_2, it informs to all the other co-ordinators through the control channel. The other co-ordinators on receiving this message, will try to establish connectivity to either of the component containing C_1 and the component containing C_2 as there is connectivity from C_1 to C_2. Thus, all the co-odrinators view the component containing C1 and the component containing C_2 as a single component containing $\{C_1,C_2\}$. This process continues till all the divided components become connected and all the components get coalesced into a single component containing all the co-ordinators. This ensures that the connectivity between the various components is restored.

4.2 Phase II

To minimise the objective function value, every node runs Phase II of reconfiguration after Phase I gets terminated. Every node is assumed to know the traffic from that node to various other destinations, and watermark levels (W_{cong}, W_h and W_l) for every destination. A lightpath is said to be overloaded or congested if it carries load greater than W_{Cong}. It is said to be underloaded if the load carried by it is less than W_l. In addition to this, every node keeps a set of candidate routes for each destination. If a node finds that the traffic to a particular destination is greater than the watermark level W_h or if a lightpath to a destination is congested (carrying load greater than W_{cong}), it tries to establish a new lightpath to that destination node. If there exists a lightpath to that destination, the new lighpath shares the traffic of the previously existing lightpath. This avoids any lightpath getting overloaded. In turn, it improves the objective function value and thereby the network performance. If there exists no lightpath to that destination previously, then the traffic to the destination is carried by the newly established lightpath.

Algorithm

Assume C is the capacity of a lightpath, SD is a source destination pair emerging from the node. SSDset gives the set of SD pairs emerging from the node S. Traffic(SD) gives the traffic between the nodes S and D and Load(SD) gives the load of the lightpath between nodes S and D. W_l, W_h *and* W_{cong} *are the watermark levels.*

```
For all SDs in the Sdset do
{
If ( $(Traffic(SD)>W_h)$ or $(Load(SD)>W_{cong})$ )
{
 Send control messages [using distributed
 reconfiguration protocol] along a candidate route R to
 Dest(SD) to find a wavelength for the new lightpath.

If (a wavelength is found)
{
 Send control messages [using distributed
 reconfiguration protocol] along the route R, to setup
 the lightpath. Route the traffic through the newly
 established Lightpath.
}
}
}
```

To establish a lightpath, every node makes use of the following protocol which is similar to the backward reservation protocol for dynamically establishing a lightpath. The protocol reserves free wavelengths as in backward reservation protocol. But in addition to that, it deletes some underloaded (<WL) lightpaths, whose resources are necessary for the establishment of the new lightpath. Since the lightpaths with low load are selected for deletion, it may not affect the objective function value significantly. The resources released after the deletion of a lightpath are used for establishing a new lightpath which will carry higher traffic. Hence, the overall objective function value of the network may be minimized. An underloaded lightpath is deleted only when it is a non-critical one.

In this protocol, all wavelengths (free wavelengths and wavelengths that can be made available by deleting underloaded lightpaths) for establishing the new lightpath are collected during the forward transversal of the control message along the candidate route. This set of wavelengths is referred to as *candidate wavelengths.* Inorder to establish a new lightpath, the source verifies if it has a free transmitter or if it can get a transmitter by deleting some underloaded lightpath. It sends a control message across the candidate route to collect all candidate wavelengths for establishing the new lightpath. The candidate wavelengths are collected but not reserved during the forward traversal. Once this message reaches the destination, the destination node knows the set of candidate wavelengths. It verifies if it has a free receiver or a receiver can be made free by deleting some underloaded lightpath. It prepares another control message containing a subset of candidate wavelengths. This message traverses backward from the destination to the source node reserving potential wavelengths that are selected on the links. When it reaches the source node, one wavelength is selected and the other wavelengths are released. The source sends an *estab_lp* message to establish the lightpath. Since our protocol tries to establish new lightpaths only for those source-destination pairs with high traffic, the number of changes made to the virtual topology is considerably less and hence, it does not cause much traffic disruption in the network during the reconfiguration process.

5.Discussion of the results

In this section, the performance of the proposed methodology is studied. Extensive simulation experiments are conducted on NSFNET T1 backbone network. To design a virtual topology, wavelength continuity constraint is enforced, and it is assumed that only one lightpath may exist between a node pair, and the lightpath is chosen on the shortest path between source-destination pairs. We use the model given in [5] to simulate the values of traffic matrix. Traffic is measured between all pairs of nodes and is given as an N x N matrix, say Traffic, where $\text{Traffic}_{i,j}$ represents the average traffic from node i to node j. It is to be noted that $\text{Traffic}_{i,j}$ may not be the same as $\text{Traffic}_{j,i.}$ The objective of the reconfiguration process is to provide connectivity amd improve the objective function value with minimum number of changes N_{ch} and minimum message overhead. The parameters with respect to which these metrics can be studied are, the number of transceivers T and the number of wavelengths W. In our experiments, we observed the values for performance metrics such as Average weighted hop count H_{avg}, Number of messages sent and number of changes N_{ch} made to the virtual topology. These observations are made while fixing one of the parameters. The failures of all the fourteen nodes of NSFNET are observed by varying the above specified parameters. It is observed that the increase in number of resources causes increase in the number of messages that are to be sent. It is because of the existence of more number of lightpaths in a topology designed with more resources. Also, the number of messages is less for connected cases than that of disconnected cases. The cases where the virtual topology remains connected after the removal of the failed node are referred to as connected cases.

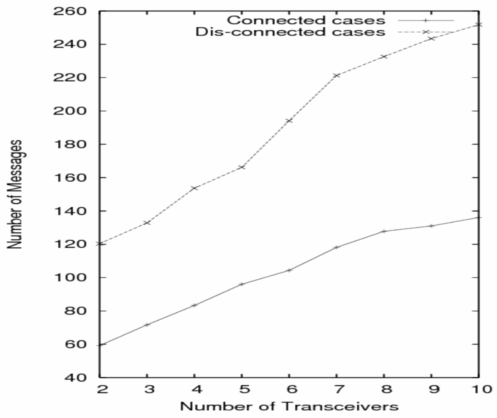

Fig 1. Effect of number of transceivers on number of messages when W=5

Figure 1 shows the variation in number of messages sent with respect to the number of transceivers T (transmitters and receivers) used. Here, the number of wavelengths W is taken as 5. Figure 2 shows the variation in the number of messages sent with respect to the number of wavelengths carried by each fiber. Here the number of transceivers T is taken as 5.

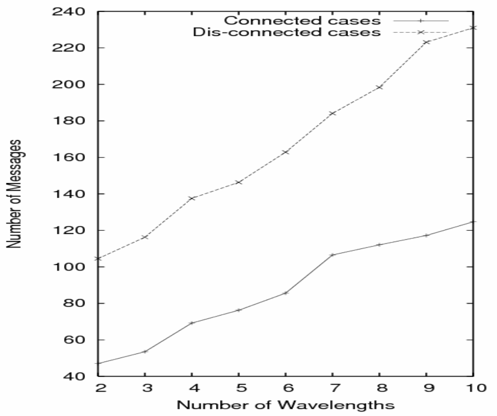

Fig 2. Effect of number of wavelengths on number of messages when T=5

Figure 3 shows the effect of transceivers on the objective function H_{avg} of the network after the reconfiguration process. Here, we considered the cases where the node failure disconnects the virtual topology. Figure 4 shows the effect of transceivers on the objective function H_{avg} in the cases where the connectivity is not affected due to the node failure. In both figures 3 and 4, the number of wavelengths is taken as 5. The results show that the value of H_{avg}

obtained after reconfiguration is almost equal to that of a newly designed virtual topology. Table 1 indicates the number of disconnected topologies N_{disc}, average number of components formed Avg_{comp} and the number of changes N_{ch} that are to be carried out under various resource conditions.

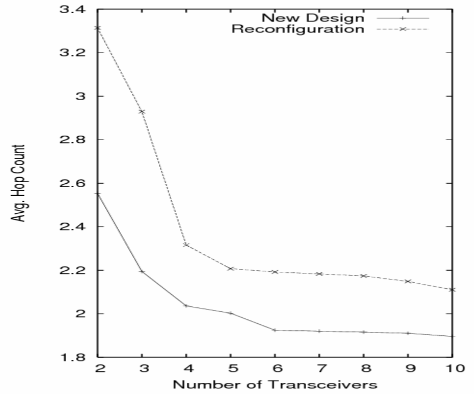

Fig 3. Effect of number of tranceivers on H_{avg} in disconnected cases when W=5

Table 1. Reconfiguration results when T=5

W	T	N_{disc}	Avg_{comp}	N_{ch}
2	5	2	2	2.3
3	5	2	2	3.7
4	5	6	2.333	4.2
5	5	5	2.4	5.1
6	5	4	2.5	7.8
7	5	3	2	8.167
8	5	2	2	6.7
9	5	1	2	6.3
10	5	0	0	5

6.Conclusions

In IP-over-WDM networks, the IP traffic may make use of a set of lightpaths that are already established (virtual topology). A virtual topology is said to be *connected*, if every node in the virtual topology can communicate either using single lightpath (single hop) or using multiple lightpaths (multihop) with every other node. Failure of a node may cause disconnection of the virtual topology, even if the physical network remains connected. In such a case, some lightpaths may be established to achieve a connected virtual topology. This is possible only if every node has at least two transceivers. In this paper, we made an attempt to device distributed protocols that provide connectivity in the event of a node failure. Once the virtual topology connectivity is achieved, the reconfiguration procedure is further applied to improve the network performance. Extensive simulations are performed on NSFNET to support our results empirically.

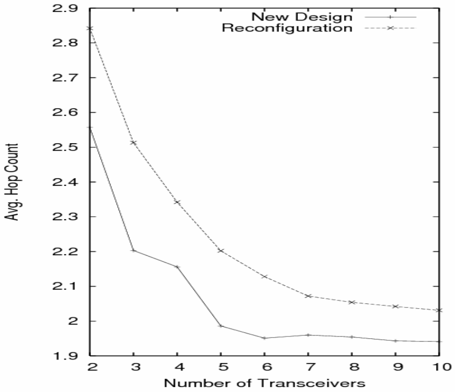

Fig 4. Effect of number of tranceivers on H_{avg} in connected cases when W=5

References

[1] J.Y. Wei, “Advances in the Management and Control of the Optical Internet", *Journal of Selected Areas in Communications*, vol. 20, no.4, pp. 768-785, May 2002.

[2] R. Dutta and G. N. Rouskas, “A Survey of Virtual Topology Design Algorithms for Wavelength Routed Optical Networks,” *Optical Networks Magazine*, vol. 1, no.1, pp. 73-89, January 2000.

[3] R. Ramaswami and K. N. Sivarajan, “Design of Logical Topologies for Wavelength Routed Optical Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 5, pp. 840-851, June 1996.

[4] B. Mukherjee, S. Ramamurthy, D. Banerjee, and A. Mukherjee, “Some Principles for Designing a Wide- Area Optical Network,” *IEEE/ACM Transactions on Networking*, vol. 4, no. 5, pp. 684-696, October 1996.

[5]D. Banerjee and B. Mukherjee, “Wavelength-Routed Optical Networks: Linear Formulation, Resource Budget Trade-offs and a Reconfiguration Study,” *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 598-607, October 2000.

[6] B. Ramamurthy and A. Ramakrishnan, “Virtual Topology Reconfiguration of Wavelength-routed optical WDM Networks,” *Proceedings of Globecom 2000*, November 2000.

[7] N. Sreenath and C. Siva Ram Murthy, “On-line Reconfiguration of Virtual Topologies in Wavelengthrouted WDM Networks,” *Journal of High Speed Networks*, vol. 12, no.3-4, pp. 141-169, December 2002.

[8] E.Modiano and A. Narula-Tam, “Survivable Lightpath Routing: A New Approach to Design of WDM-Based Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 20, no.4, May 2002.

[9] N.Sreenath and Prakash Ramaswamy, “Virtual Topology Reconfiguration for Node Failures in Wavelength Routed Optical Networks,” in the Proceeding of ICNS 05.