GMPLS FAULT MANAGEMENT AND ITS IMPACT ON SERVICE RESILIENCE DIFFERENTIATION

M. Brunner, C. Hullo
NEC Europe Ltd., Adenauerplatz 6, D-69115 Heidelberg, Germany, brunner@ccrle.nec.de
EANTC AG, Sprembergerstr. 6, D-12047 Berlin, Germany, hullo@eantc.de

Abstract: Generalized Multi-Protocol Label Switching (GMPLS) is currently under standardization. It basically reuses the MPLS control plane (IP routing and signaling) for various technologies such as fiber switching, DWDM, SONET, and packet MPLS. Since GMPLS runs in core networks, fault management is of major concern. However, fast fault recovery and backup capacity assignments are very expensive and not all customers need this or are willing to pay for it. Therefore, we propose in this paper to use several protection and bandwidth-sharing schemes on the same network in order to provide differentiated services in the resilience space. This means an operator can offer and provide several customized services. The service management system implementing the schemes is built on top of a GMPLS network management system developed in our Lab.

Key words: Optical Networks, Fault Management, Resilience, and Service Management

1. INTRODUCTION

Fault management aims at helping network operators automatically detecting and recovering from faults before human beings, and so before customers notice them. By localizing faults when they happen and improved speed of repair, they might save paying refunds against broken SLAs (Service Level Agreement).

Recently, a technology called Generalized Multi Protocol Label Switching (GMPLS) has showed up and is currently under standardization at the IETF [1]. The primary goal of GMPLS is reusing the MPLS [2] control plane, namely IP routing protocols and path setup protocols for different kinds of networks such as SONET, DWDM, and packet MPLS. Since these technologies operate mainly in backbone networks, fault management is of primary interest, as it ensures the reliability of data delivery.

On the other hand, fault management might be expensive in terms of allocating resources for failure cases and on mechanisms (hardware or software) for fast detection and restoration. So fault management has to respond to the trade-off between making the network robust and avoiding too much resources to be allocated for not directly paid use, but for fail-over cases.

Another key feature of GMPLS is the fast and dynamic provisioning of optical or packet paths using an IP/MPLS control plane. This implies that also the fault management issue needs to be concerned about the low provisioning times and the dynamic behavior of the network.

On the other hand various customers have various requirements and demands on the service availability. So it does not make sense to install only one mechanism for all customers. That's where differentiation of the service in terms of resilience is used.

In the literature, different possible schemes for fast restoration and protection switching for various technologies have been published ([3], [4], and [5]). However, none of them propose the approach of differentiation of services based on different protection and bandwidth sharing schemes in a dynamic service creation environment. In this paper, we propose to integrate different schemes into a single network, in order to offer various availability parameters for customers. The availability needs to be specified on the service request, and it must be provided by different means in the network.

The only work the authors are aware of proposing a similar scheme is [13]. However, they have looked into IP and MPLS restoration only. They do not consider the GMPLS case, where different technologies are controlled and managed by the same management system. Additionally, they have not implemented their scheme for real world hardware as we did. And we tried to apply the different random generated topologies, where they had one topology of their test network. However, they have been more extensive in simulating various algorithms without mentioning the results in the publication, whereas we have chosen only one algorithm.

2. GENERALIZED MULTI-PROTOCOL LABEL SWITCHING (GMPLS)

GMPLS is, as its name suggests, a generalization of MPLS. GMPLS first generalizes the control plane such that it is not only used for packet switched networks, but also for optical switched, TDM-based, and physical networks. This requires that the control-plane and data-plane are no longer only logically separated but might also be physically separated. Second, GMPLS extends the notion of a label in order to support multiple switching layers. For instance a label might be an optical wavelength number.

The major goal of GMPLS is to control optical backbones as flexibly and dynamically as IP backbones today. GMPLS intend to reuse existing technologies by combining MPLS control technology namely RSVP-TE and CR-LDP [6] with operational experience of IP routing with some extensions towards GMPLS in order to provide a general network solution.

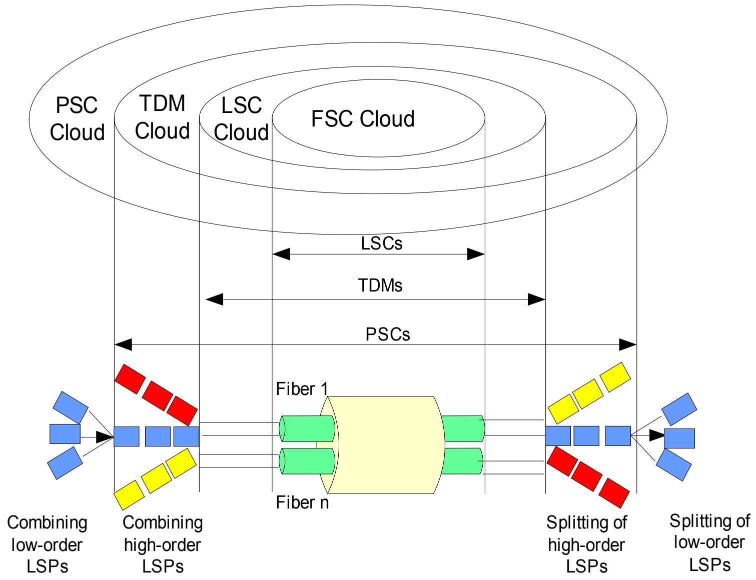

Figure 1. GMPLS Architecture

The strength of GMPLS is to also provide a uniform semantic for network management, operations, and control in hybrid networks. Additionally, GMPLS provides a tool for real-time service provisioning of different types of channels as well as easy and cheap equipment operation and management. Finally, MPLS allows performing traffic engineering on packet-based network, and its generalization GMPLS to perform this on optical networks as well.

The following are the most often-heard interface types to be supported with GMPLS: (1) Packet Switch Capable (PSC) interfaces (e.g. IP, MPLS, and Ethernet). (2) Time-Division Multiplex Capable (TDM) interfaces (e.g. SONET/SDH Cross-Connect). (3) Lambda Switch Capable (LSC) interfaces (e.g. Photonic Cross-Connect (PXC) or Optical Cross-Connect (OXC)). (4) Fiber-Switch Capable (FSC) interfaces.

As a consequence of the various interfaces, GMPLS establishes a link hierarchy (Figure 1). And as MPLS Label Switched Path (LSP) can be nested, optical channel trails have discrete bandwidth granularity in units of individual wavelength capacity.

A generalized label contains enough information to allow the receiving node to program its cross connect (switching hardware), regardless of the type of this cross connect. Since the nodes sending and receiving the new forms of label know what kinds of link they are using, the generalized label does not contain a type field, instead the nodes are expected to know from context what type of label to expect.

3. FAULT MANAGEMENT IN GMPLS

Fault management includes detection, localization, and recovery of/from failures. The detection of the fault by continuous or periodic checking is the very first step in order to take provisions to repair it. Various mechanisms allow for immediate localization of the fault, others need more work to find out where the fault really is, by analyzing or testing. Fault notification includes notifying an entity to perform recovery and possibly raises alarms in the operations center. Or an entity that performs service management is notified about the inability to provide the network part of a service (e.g. alarms).

3.1 Fault Detection

For fault detection, various approaches exist. Those mainly used for plain MPLS, where some of them also work in more general cases. For MPLS networks that can be classified into three groups: the first one using IP capabilities to detect MPLS defects (ICMP extensions, GTTP, LSP Ping), the second one using the MPLS layer only to detect MPLS defects (Y.1711), the third one operating only on a hop-by-hop basis (LMP, data plane encoding). Note that the first two groups work end-to-end of an LSP.

ICMP Extensions [7] use ICMP messages to convey control information to source hosts. Extensions to ICMP allow an LSR to append MPLS stack information to ICMP messages. Generic Tunnel Trace (GTTP, [8]) supports enhanced tunnel-tracing applications that network operators use to trace paths through an IP or MPLS network's forwarding plane. LSP PING [9] verifies the availability of a connection (Label Switched Path). ITU-T Recommendation Y.1711 [10] provides mechanisms for user-plane fault management, by defining OAM packets sent over an LSP on a predefined label. GMPLS introduces a new protocol called the Link Management Protocol (LMP, [11]). It runs between adjacent nodes and is responsible for establishing control channel connectivity as well as failure detection. LMP also verifies connectivity between channels. The detection of optical data-channel failure is measured by detecting Loss Of Light (LOL). LOL propagates downstream along the connections path and therefore all downstream nodes may detect the failure.

3.2 Fault Localization

The fault localization of the control-link is done simultaneously with the fault detection in many cases, since it is applied locally. E.g., the localization of a data-link failure might be achieved by the Link Management Protocol's (LMP) fault localization procedure that sends LMP Channel-Status messages between adjacent nodes over a control channel maintained separately from the data-bearing channels. In other cases, it is pretty difficult to localize the fault, since only end-to-end fault detection mechanisms are involved.

3.3 Fault Notification

Depending on where the faults are detected, there is an upstream notification needed or at least a management system alarm needs to be raised.

RSVP specifies that errors be notified to upstream node using PathErr messages and to downstream nodes using ResvErr messages. Moreover, a Notify message (that contains the affected LSP and failed resource) has been added to RSVP-TE for GMPLS to provide a mechanism for informing non-adjacent nodes of LSP-related failures. These Notify messages do not replace existing RSVP error messages as they differ from them in that they can be targeted to any node other than the immediate upstream or downstream neighbor.

3.4 Fault Recovery

The purpose of fault recovery is to trigger for corrective actions when failures occur. The goal is that this must happen as fast as possible. However, the recovery speed has an impact on the potential service downtime and on the resources allocated within the network. That's where service differentiation makes a lot of sense, because also prices will be differentiated.

Several differentiators are possible including the following:

Recovery is done via the control plane or the management plane, where control plane recovery is faster, but needs configuration to do the expected thing.

Secondary paths are pre-established (protection) or established on demand in case of a failure (restoration).

The failure might be addressed either at the transit node where the failure is detected, or at the endpoints. The transit node handling is faster, since it does not need a signaling action taking place in order to notify the end nodes of a path. However, it might mean that several secondary paths are established from several transit nodes, if protection switching is used.

Different bandwidth sharing schemes are possible. No sharing at all, share several secondary paths for the same primary LSP, share bandwidth for several primary paths.

Secondary, pre-established paths are carrying traffic all the time (1+1 protection) or it will be switched over in case of failure (1:1 protection). The fist choice is faster or even no service interruption at all.

In the following we constrain ourselves to only part of the problem.

4. GMPLS SERVICE MANAGEMENT

GMPLS Service Management is applied to administrative boundaries such as the user-to-network or the network-to-network interface. In general, two ways shown in Figure 2 are envisioned to request a GMPLS-based service. The managed way is over any communication means the service request is received by the service management system (in our case we use a web interface). The second way, mainly envisioned by the IETF allows using signaling messages (RSVP-TE or CR-LSP) for requesting an end-to-end service.

Since, none of the standardization bodies use resilience as a service parameter so far, and since we provide other services such a Gigabit Ethernet over MPLS as well, we favor the managed way, but will work on the signaled way in the future.

Another issue for fault management in GMPLS is that the control plane is applied to different underlying technologies. Therefore not all of the possible protection schemes are possibly implemented. This means the managed approach allows for better targeting the service requested knowing the service able to be provided.

Finally, there is the issue of hierarchy, where GMPLS is applied to several networking technologies within the same domain and potentially IP runs over a GMPLS network. In this case, one needs to decide on what layer what protection scheme is used. Otherwise all protection schemes will be implemented at all layers, which is an overhead not needed. On the other hand, the layer performing fault recovery must be chosen based on protection and restoration capabilities and operators policy.

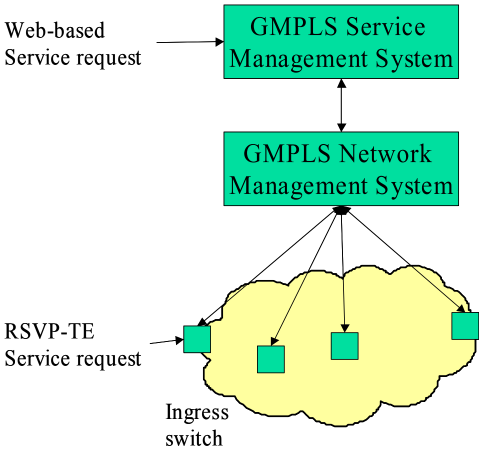
Web-based
Service request
GMPLS Service Management System
GMPLS Network Management System
RSVP-TE
Service request
Ingress
switch

Figure 2. General System Overview

Figure 2 also shows the interaction between the service management and the network management system and the interaction between the network management system and the GMPLS switches. The functionality available on the network management level is in our case topology discovery, basic configuration management, monitoring, and viewing capabilities. For setting up a service the network management system gets the GMPLS LSP (the primary and secondary paths) and configures it into the network using different means. It mainly needs to manage label space or it triggers RSVP-TE to setup the LSP.

4.1 Service Definition

The service definition for bandwidth and resilience might include several parameters depending on what is possible in the network and what a customer requires, and what technology is used. In our service definition the following parameters are used:

Source and Destination Address of the service, where we currently use IP addresses, but GMPLS in general allows for unnumbered links, where an interface is a number on a switch.

Service Type specifies what type of service a customer requests. At the moment we can support packet MPLS (packet over anything), lambda paths (DWDM), Gigabit Ethernet over MPLS (a specific application).

Bandwidth, specifically, for packet MPLS cases we have a high granularity, for other technologies the granularity is given by the technology, e.g. SONET granularity.

Maximum Service Interruption Time specifies the time a customer is willing to accept a service outage. Which means the service is not available at all during that time.

Minimum Bandwidth in failure cases denotes the bandwidth, which needs to be available when a failure occurs. This translates into the bandwidth allocated for secondary paths, where the traffic is switched to in case of a failure, or the bandwidth signaled for in fast rerouting.

4.2 Application to an Optical Label Switch

In the following we are getting more specific towards our implementation. For that reason some explanations of the underlying hardware restrictions and constraints are given. See [12] for a more extensive description.

The main functionality of the Optical Label Switch (OLS) is that it runs GMPLS for optical paths and for MPLS packet paths. We implement the overlay model only, which means that the links seen on the packet layer are optical paths. Both path setups are triggered by the management system and use RSVP-TE with GMPLS extension in order to setup paths in the network.

For fault recovery we have a constraint that on the optical level path setup, update, and switching to a secondary path are very slow. On the other hand, on MPLS level switching to a secondary path is basically instantaneous after the failure has been detected. Failure detection is possible on the sending side of an optical path. This means on the MPLS level each hop of an MPLS LSP can switch to a secondary path.

Naturally, any failure raises an alarm in the GMPLS network management station. In case of the optical path having troubles, the management system can decide on what secondary path the traffic might be mapped. So here we use on-demand, managed secondary path setup scheme only. Due to a limitation of the number of optical interfaces per OLS, it is in many cases not possible to even setup a secondary optical path.

On the packet MPLS level however, we are able to use the fast protection switching capability of the OLS. However, this means we need to pre-compute and pre-setup secondary paths in advance during service provisioning time.

4.3 Protection Schemes Used

In the following, we describe the protection schemes evaluated. However note again that most of them are well known from the literature [13], [6], [5]. As described in the previous chapter the schemes and the evaluation take into account some of the constraints of the OLS specific features. So we assume only pre-established paths, where the establishment is different and the failure notification mechanism is different.

For protection schemes we use end-to-end, link local, and local-to-egress, (see Figure 3 - Figure 5).

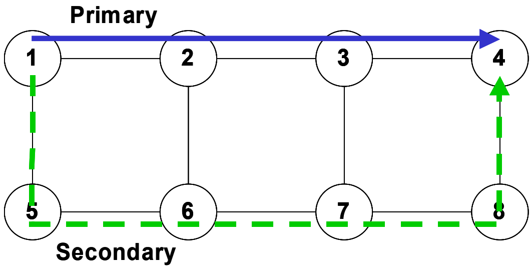

Figure 3. End-to-end Protection Scheme

End-to-end means one secondary path is setup from the ingress switch to the egress switch. The benefit of this scheme is the number of secondary LSPs used for backup. Namely there is only one secondary used per primary and the bandwidth allocated for the secondary is pretty low depending on sharing schemes. However, the worst-case notification time is bigger then in the following schemes. Additionally, a signaling protocol is used to signal failures upstream such that the traffic can be switched to the secondary path.

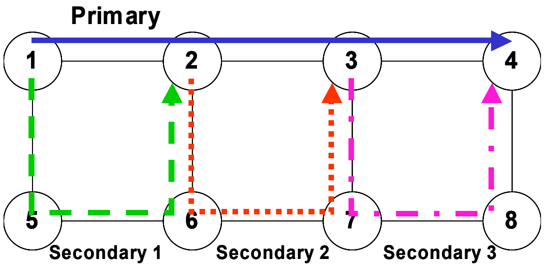

Figure 4. Link Local Protection Scheme

Link local requires the setup of secondary paths from the ingress and from each intermediate switch to the following switch on the primary path (in Figure 4 from 1 to 2, from 2 to 3, from 3 to 4). The benefit lies in the fast reaction time in case of failures. Basically only link failure detection is needed and no upstream signaling is required. However, the large number of secondary is a problem, as well as the bandwidth allocated for the secondary LSPs is a problem, since in most cases they do not share the same path for the same primary LSP.

Local-to-egress means, from each intermediate switch a secondary path is setup to the egress switch (in Figure 5 from 1,2,3 to 4). In this scheme, we have the same fast recovery as in the Link-Local Scheme, but we have better sharing capabilities.

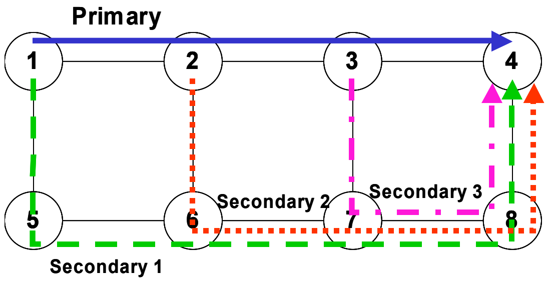

Figure 5. Local-to-Egress Scheme

For certain topologies and a small number of nodes one can run linear programming based optimization and achieve about the same sharing ratio as the end-to-end protection scheme [14]. One of the drawbacks of this scheme is, that it is very difficult to implement it in the control plane. It is difficult to change the signaling protocols to work such that the protection scheme is setup. In our case, this is not a problem, because we mainly rely on managed networks and only setup the LSP with pre-computed secondary paths.

4.4 Sharing Schemes

For bandwidth sharing schemes we use no sharing, per path sharing, max bandwidth sharing, percent sharing per path, and percent sharing global.

No sharing means there is no bandwidth sharing at all. Also in link-local or local-to-egress protection schemes, all the bandwidth is allocated on the secondary paths. Specifically, in local-to-egress, this is a bad solution. In Figure 5 it would mean to allocate three times the bandwidth of the primary LSP on the link from 7 to 8.

Per path sharing is useful for local-to-egress and link local protection schemes, where several secondary paths are used for the same primary path. In Figure 5 it would mean to allocate only once the bandwidth of the primary LSP on the link from 7 to 8.

Max bandwidth sharing allocated only the maximum bandwidth of all secondary paths crossing through that link. This does not give any guarantee on the bandwidth, but the hope is that in cases where only one primary path is affected it still has enough backup capacity. So the backup capacity is a function of the number of failed LSPs using a particular link as backup.

The percent sharing schemes allocate a certain percentage of the primary path to the secondary path. This is mainly used for service specification, where a customer is willing to reduce his bandwidth demand for a certain time in failure cases.

Percent sharing per-LSP is an optimization such that in protection cases, where several secondary paths are allocated and that they share the bandwidth. This is basically a combination of per-LSP and percent sharing.

5. EVALUATION

In the following, we give a numerical evaluation of the different schemes with a set of topologies. Since we do not have enough Optical Label Switches, we need to simulate the schemes, however we have implemented some of the schemes also for the real system. We use a homegrown Java-based simulation tool for that task. Some of the simulation results are qualitatively known in advance, but if it comes to charging for a service, we need to get some quantitative measures of the different schemes and their behavior.

The routing algorithm used is constraint shortest path first (CSPF), which has known problems but is very easy to implement. All the numbers are averaged numbers over 100 simulation runs and averaged over all links in the network. The stopping criterion is defined as 10 consecutive failed service setups. The link bandwidth of all topologies is 2.4 Gbit/s and the degree of the topology is 4. We have chosen the degree of 4, because of hardware limitations on the OLS.

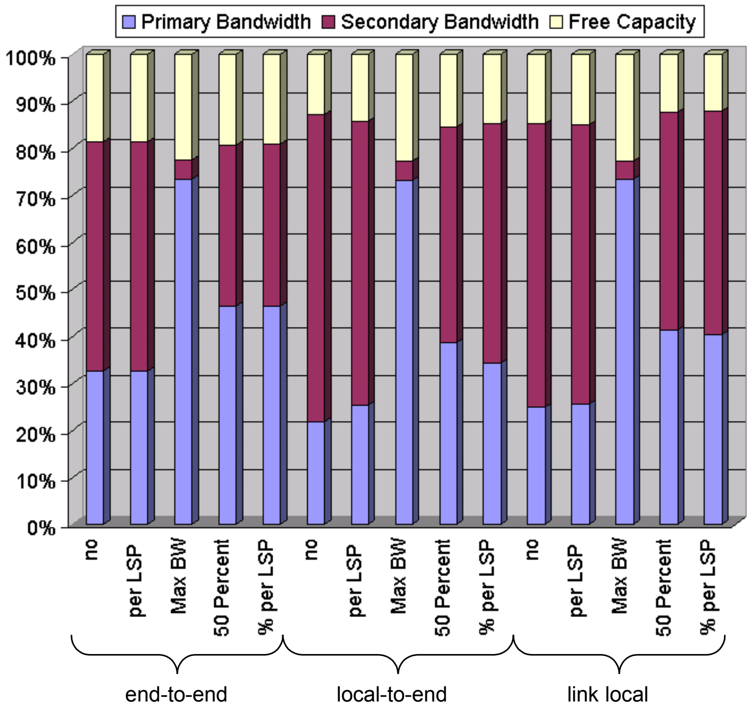

Figure 6. Topology 16 Nodes, Degree 4, Ring-Mesh

5.1 Experiment 1

In the first experiment, service requests are chosen with random originating node, terminating node, and bandwidth in the range of 10-100 Mbit/s. Results of one topology with 16 nodes are shown in Figure 6. The topology is basically a ring with additional links between each second and third node in the ring.

Figure 6 also shows the effect of sharing schemes. For end-to-end protection it does not matter whether we have no or per LSP sharing, because there is only one secondary path anyway. Also in the other protection schemes it does not matter too much. Not surprisingly we see that the Max BW sharing scheme performs best in terms of secondary path bandwidth consumption. However, no guarantees can be given that these paths get the requested bandwidth in failure cases. In case of sharing based on 50% of the primary path the primary to secondary bandwidth ratio is much better compared to other schemes.

In summary, the Max BW sharing scheme does perform by far the best, with the least guarantee. So all service requests with very small bandwidth guarantee in failure cases will use this scheme. Most likely this service type also does not need fast protection switching. So this class of service will use end-to-end with Max BW allocation for secondary paths.

5.2 Experiment 2

As seen before per LSP sharing naturally makes no sense for end-to-end protection schemes. Therefore we excluded it and all Max BW sharing schemes in the following simulations. In this experiment we study the effect of topology on the results of the scheme. We use three different topologies, the random, the ring-mesh, and a mesh, all of them again with degree 4. The difference from the ring-mesh to the mesh is that in ring mesh the links are nearer to the ring, where in the meshed topology links run across the complete network. Figure 7 shows the result of simulating various schemes in various topologies. The values shown are the ratio between total capacities allocated for primary paths divided by the total capacity allocated for secondary paths.

We definitely see that end-to-end protection is performing better compared to local-to-egress or link local. However that scheme needs longer detection and notification time. In case of link failure detection, the head end of an LSP must be notified. Or in case of end-to-end connectivity check the failure detection takes more time. So we use this scheme for service requests with moderate service interruption time allowed, because the capacity cost is less than with the other schemes.

Furthermore, it can be seen that for link local and local-to-egress the per-LSP sharing scheme (with or without percentage) performs well compared to without sharing. Since per LSP sharing does not influence the service performance but is using the capacity more efficiently we choose this one.

The other important observation is that depending on the topology local-to-egress and link-local performs different. Since in more meshed topologies such as mesh 16 and mesh 32 in Figure 7, the local-to-egress scheme performs better, we have chosen that one for our implementation. However, there is more work needed to exactly figure out the impact of topological issue influencing what part of the problem.

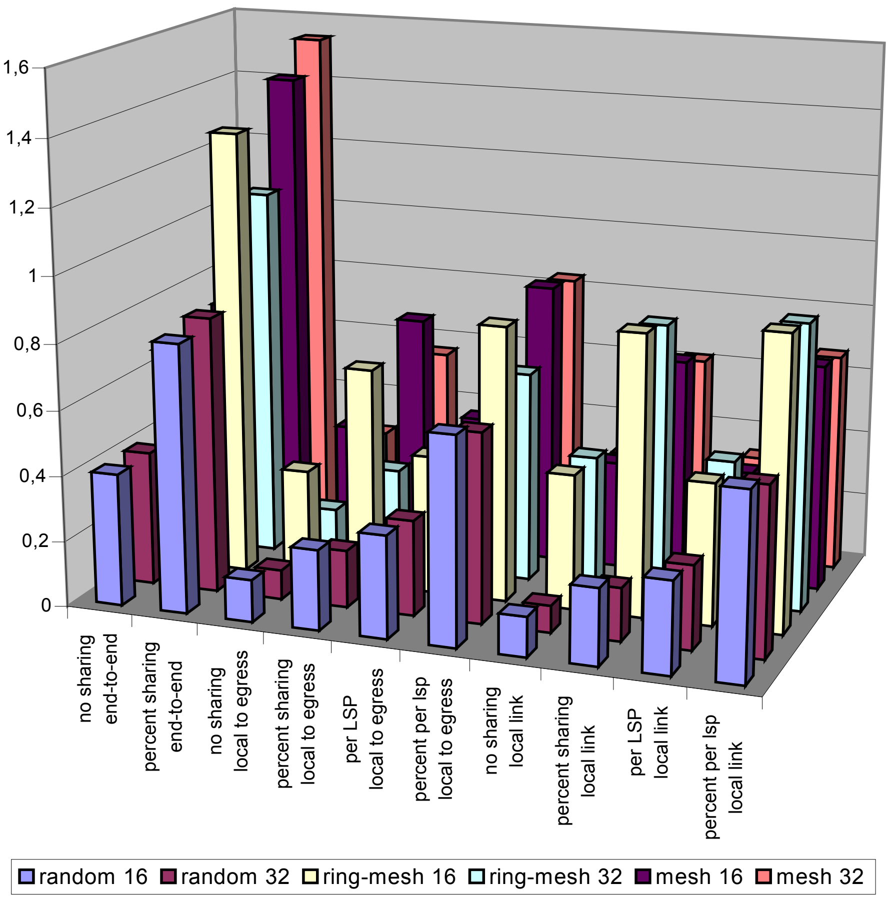

Figure 7. Ratio Primary to Secondary Capacity Used (Y-axis)

Comparing per-LSP and percent per-LSP sharing, we see that percent per-LSP sharing performs better. But naturally secondary paths might not have full capacity. They only get guaranteed X% of the primary capacity. However, taking into account that only a small set of failures occurs, it still might get full capacity, but without any guarantee.

5.3 Experiment 3

So far we have simulated only one scheme per network. In the following we simulate various schemes running at the same time on the network. This is the simulation of different service types running on one network. With the above observations we choose end-to-end and local-to-egress protection scheme, and per-LSP, percent sharing, and percent sharing per-LSP.

Additionally, we have broadened the topology scope and added a 64-node topology of each type, and added two new random generated types of topologies know as the Waxman topology [15] and the Barabasi and Albert topology model [16].

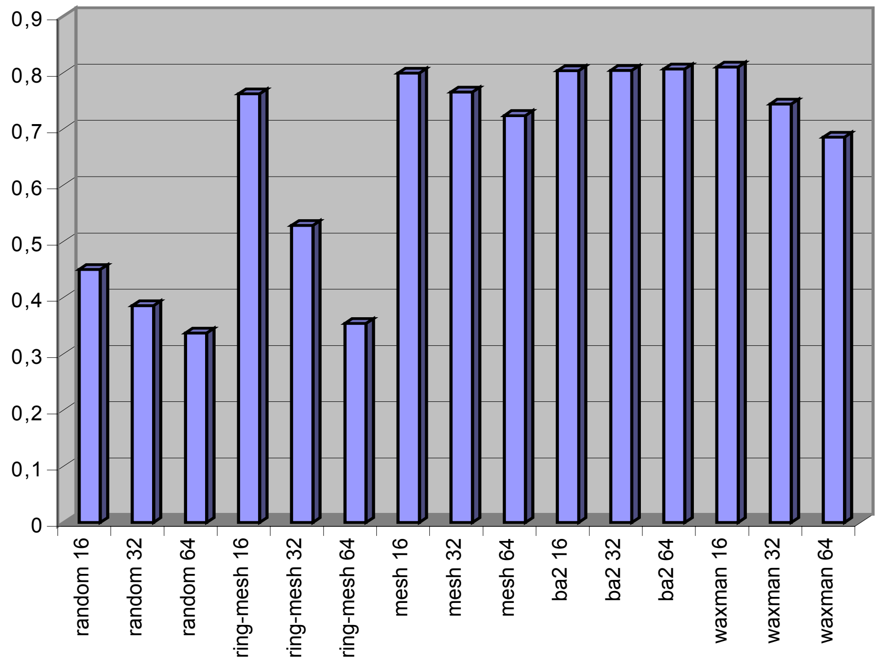

Figure 8. Primary/Secondary BW Ratio (Y-axis) for Services Mixed on the Network

Figure 8 shows results, which basically are similar for all the topologies. The bad numbers for the random topologies again stem from many unavailable secondary paths (no two distinct to reach the nodes). The remarkable result here is in the numbers for the ring-mesh 32 and 64 topology. Here we get worse result. This stems from the average hop count of primary paths allocated in the network. The average hop count for the "ring-mesh 32" is 4.31, whereas it is 2.33 for "ring-mesh 16", 2.04 for "mesh 16", and 3.16 for "mesh 32". And it is even worse for ring-mesh 64, where the average hop count is 8.25. Note that the bigger the average hop count the more secondary paths are allocated for local-to-egress protection scheme, which impacts the overall allocation of mixed protection schemes.

6. CONCLUSION

The idea of differentiating GMPLS-based services proposed in this paper is a reasonable way of efficiently use the bandwidth of a network and still provide the customers the service they want. We mainly differentiate the service based on resilience parameters. We implemented the most appropriate schemes as an add-on for a GMPLS network management system built for an Optical Label Switch (OLS). However, we used simulation as a method to numerically evaluate the schemes first to figure out the benefits and drawbacks.

So far we have not considered the hierarchical nature of GMPLS in the simulations. For the less general GMPLS service management system implementation, we included also the hierarchical issue. Basically, we kept the two hierarchies pretty independent, but changed the algorithm for the secondary path calculation such that not only the underlying 'virtual' topology is used, but also the physical topology beneath is taken into account for secondary path calculations.

The simulation regarding the mixture of different services on the same network needs much more attention and work to detect the influencing parameters. Additionally, we have not considered a dynamic system, where customers come and go. One of the problems is the estimation of the call arrival process and the service type distribution.

REFERENCES

[1] Mannie et al., "GMPLS Architecture", IETF Internet Draft, work in progress, draft-ietf-ccamp-gmpls-architecture-03.txt, August 2002.

[2] Davie and Rekhter, "MPLS Technology and Applications", Morgan Kaufmann Publishers, 2000.

[3] Banerjee et al., "GMPLS: An Overview of Routing and Management Enhancements", IEEE Communications Magazine, January 2001.

[4] G. Li et al., "Control Plane Design for Reliable Optical Networks", IEEE Communications Magazine, Feb 2002.

[5] Sharma and Hellstrand (Editors), "Framework for MPLS-based Recovery", IETF Draft, draft-ietf-mpls-recovery-frmwrk-08.txt, work in progress, draft-ietf-mpls-recovery-frmwrk-08.txt, October 2002.

[6] Banerjee et al., "GMPLS: An Overview of Signaling Enhancements and Recovery Techniques", IEEE Communications Magazine, July 2001.

[7] Bonica et al., "ICMP Extensions for Multi Protocol Label Switching", Internet Draft, draft-bonica-icmp-mpls-02.txt, work in progress, Nov. 2000.

[8] Bonica et al., "Generic Tunnel Tracing Protocol (GTTP) Specification", Internet Draft, draft-bonica-tunproto-01.txt, work in progress, July 2001.

[9] Pan et al., "Detecting Data Plane Liveliness in RSVP-TE", Internet Draft, draft-pan-lsp-ping-02.txt, work in progress, 2002.

[10] ITU-T Draft Recommendation Y.1711, "OAM mechanism for MPLS networks", work in progress, 2002.

[11] Lang (Editor), "Link Management Protocol (LMP)", Internet Draft, draft-ietf-ccamp-lmp-07.txt, November 2002.

[12] M. Arai et al., "High Performance Network with Merged Optical and IP", IEEE High-Performance Switching and Routing Workshop (HPSR'02), Tokyo, May 2002.

[13] A. Autenrieth, A. Kirstaedter, "Engineering End-to-End IP Resilience Using Resilience-Differentiated QoS", IEEE Communications Magazine, Vol 40(1), January 2002.

[14] H. Saito, M. Yoshida, "An optimal recovery LSP assignment scheme for MPLS fast reroute", 10th International Telecommunication Network Strategy and Planning Symposium (Networks 2002), Munich, Germany, June 2002.

[15] B. Waxman, "Routing of Multipoint Connections", IEEE Journal of Selected Areas of Communications (JSAC), December 1988.

[16] A.L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks", Science, 286:509–512, October 1999.