Reprinted from Globecom 2000

Fault Restoration and Spare Capacity Allocation with QoS Constraints for MPLS Networks

Tae H. Oh, Thomas M. Chen
Dept. of Electrical Engineering
SMU
PO Box 750338
Dallas, TX 75275

Jeffery L. Kennington
Dept. of Computer Science Engineering
SMU
PO Box 750122
Dallas, TX 75275

***Abstract*— This paper investigates distributed fault restoration techniques for multiprotocol label switching (MPLS) to automatically reroute label switched paths in the event of link or router failures while maintaining quality of service (QoS) requirements. Protocols for path and partial path restoration are evaluated. A backup route selection algorithm based on optimization of equivalent bandwidth is formulated and demonstrated for an example network.**

I. INTRODUCTION

The IETF is establishing common agreements on a framework for multiprotocol label switching (MPLS) to integrate various implementations of IP switching that use an ATM-like "label swapping" technique [1,2]. Although better scalability and faster packet forwarding are the most apparent motivations for MPLS, the approach is also attractive for the support of traffic engineering and quality of service (QoS). QoS guarantees can be supported when MPLS is used in conjunction with constraint-based routing and resource reservation. QoS assurances are important for guaranteed services with hard bound requirements on packet loss, packet delay, and delay variation (jitter) [3]. Guaranteed services also need assurances about reliability in terms of continuity of services maintained by fast rerouting around network failures [4]. As a connection-oriented technique, MPLS is potentially more vulnerable to faults than IP.

This paper investigates distributed fault restoration protocols for MPLS that quickly reroute connections in the event of failures while maintaining QoS requirements. For fault restoration, we assume that a set of QoS mechanisms already operate in the network, such as QoS or constraint-based routing and a signaling protocol for reserving resources [5-9]. Different strategies for fast rerouting are described and evaluated in Section 2. Section 3 presents a integer linear program based on equivalent bandwidth that can be used to select optimal backup paths. Section 4 presents results for an example network.

II. DISTRIBUTED FAULT RESTORATION TECHNIQUES

Traditional fault restoration depends on a centralized network manager to detect faults and manually reroute traffic around effected areas. This process is relatively slow, not scalable to large networks, and vulnerable to failure. Fault restoration functions can be distributed among the LSRs (label switched routers) to automatically detect and carry out rerouting without the need for manual intervention. Distributed fault recovery can be faster and more scalable than centralized fault recovery, at the expense of more complexity in the LSRs. LSRs must be automated with protocols to carry out these steps: fault detection, backup route selection, fast rerouting, and re-optimization after fault restoration.

In distributed approaches, backup routes may be either selected dynamically at the time of a fault or pre-established before a fault occurs. Dynamic searching usually involves flooding messages after fault detection to discover all possible backup routes, and a selection of one of them according to some criterion. This approach has several advantages: no topology information is necessary; the condition of backup routes are discovered when they are needed; and no additional work is necessary in the absence of failures. Unfortunately, the flooding mechanism can involve many messages and substantial delays to discover the best route.

Dynamic searching is unnecessary if a dynamic link-state routing protocol such as OSPF (open shortest path first) is being used to distribute link-state information globally. In that case, each LSR will be able to store complete network topology and status information, and therefore make a consistent selection of optimal backup routes. Compared to dynamic searching, the number of messages and delay can both be reduced by pre-establishing a backup route before it is needed, for example, at the same time that the active path is established. The approach is to prepare the backup route as much as possible before it is needed, so that minimum work must be done when a fault occurs. At least the backup route is selected and the routing tables are configured with the backup route information (including label assignments). Only a single message is then needed to activate the backup route. Although the backup route is pre-established, bandwidth does not necessarily have to be reserved along the backup route. Reserved backup resources will ensure that failed FECs can be certainly restored, but may unnecessarily restrict the admissible traffic load. For efficiency, we will study only distributed approaches where the backup route is pre-established but backup bandwidth is not reserved. Hence, there will be a possibility that the pre-selected backup path will be inadequate at the time when it is needed, in which case alternative paths will have to be found.

0-7803-6454-6/00/$10.00 (C) 2000

A. Partial Path Restoration

To restore around any possible upstream link or node failures, partial path restoration attempts to find an alternative route from the downstream LSR detecting the fault to the ingress LSR at the edge of the MPLS domain, as shown in Figure 1. Although one rerouted FEC is shown, all traffic flows from the failed link must be rerouted at the same time and may be distributed to different backup routes for traffic balance. For restoration speed and efficiency, the backup route is pre-established at the same time as the active route, but bandwidth along the backup route is not reserved in order to make it available for active traffic. The backup route is selected according to the path selection algorithm described later, and the routing tables at the LSRs along the backup route are configured with the appropriate routing table entries. Hence, when traffic is rerouted to the backup route, only changes in bandwidth allocations but not routing table changes are needed at the LSRs.

The restoration process for each effected connection consists of the following steps:

Step 1: Restoration request: After the LSR on the downstream end of the failed link detects the fault, it must check the pre-selected backup route for adequate resources by sending a Restoration Request (RR) message in the backward direction along the backup route. The RR message is derived from the Label Mapping message, a label assignment message used in constraint-based routing, with additional fields to identify the effected connection, labels, requested QoS (packet loss ratio, packet delay bound), traffic characteristics, failed link, and downstream/upstream LSRs.

Each LSR calculates the requested equivalent capacity based on its buffer resources, requested QoS, and specified traffic characteristics. Equivalent capacity is an approach to analytically derive the precise needed bandwidth to support a specific QoS for a given traffic flow [10]. If the equivalent capacity is available, the LSR makes the resource allocation change and propagates the RR message backward to the next LSR in the backup route. If insufficient resources are available, the LSR will terminate the RR message and return a Release message in the forward direction to the LSR on the downstream end of the failed link. The Release message contains the information from the RR message, as well as the identity of the rejecting LSR and reason for rejection. The Release message will clear the resources allocated already along the backup route. When the downstream LSR receives the Release message, it will proceed to Step 2.

Step 2: Restoration reattempt (if a restoration attempt fails): If the LSR on the downstream side of the failed link receives a Release message, it will reattempt to check for adequate resources on another candidate backup route. Assuming the use of a link-state routing protocol such as extended OSPF, the LSR has complete information about network topology and status, and there is no need for a dynamic search by flooding. The LSR will exclude the pre-selected backup path from its set of feasible alternative paths, and perform the path selection algorithm on the other feasible alternative paths to select another backup route. Step 1 is reattempted on the second alternative path. Because this second alternative path has not been pre-established, the RR message is now source routed with the alternative path specified by the originating LSR. Source routing will relieve the LSRs from having to make routing decisions for the RR message. Also, the RR message will have to carry out label assignments, routing table updates, and bandwidth allocation changes at each LSR.

If the RR message is successful along the candidate route, the restoration process proceeds to Step 3. If the RR message is rejected by any LSR, a Release message is returned as before and Step 2 is reattempted on the remaining feasible alternative paths. When the set of feasible alternative paths are exhausted without success, the failed connection cannot be restored with its QoS guarantees. If reliable services are critical, the network provider must ensure that sufficient backup resources will be properly provisioned in the network such that an adequate backup route will always be found even under high traffic load.

Step 3: Reroute to backup path: If the RR message is successfully propagated to the ingress LSR, the ingress LSR will change its routing table to reroute traffic from the failed link to the backup route.

For this entire restoration process, we are interested in the degree of restorability of a failed connection measured by the probability that a connection will be restored within M attempts, P_M, assuming that there are only M feasible backup paths. Restorability may be expected to increase with larger M, which is a design parameter for the network provider. Another performance metric of interest is the restoration time measured in the mean time to complete the three restoration steps, $E(T_r)$, and mean number of attempts, $E(m)$.

We assume that an RR message has a random probability of rejection q by an LSR, and acceptance or rejection of an RR message is independent between consecutive LSRs although it may be correlated in actuality. Also, the probability will vary between the first attempt and consecutive attempts; the probability of success should presumably be higher on the first attempt if the path selection algorithm is correct, but the analysis here will make the simplifying assumption that each the probability q is the same on each attempt. The probability of rejection q will be a complicated function of many factors such as active traffic load, available bandwidth and buffers, requested QoS, and traffic characteristics.

Additionally, we assume that the M feasible backup paths will all be N hops where the value of N will depend on the location of the fault. If the fault occurs near the ingress LSR, the backup route may be short and restoration may be fast. If the fault occurs near the egress LSR, the restoration may be slow. For analysis, we assume that a typical connection (called label switched path or LSP) consists of H hops and a fault is equally likely with probability $1/H$ to be located at any hop. In the Internet, LSPs could be quite lengthy and H could be much more than the value in N in link restoration. If the fault occurs on the nth hop, feasible backup routes will all consist of n hops. The length of backup routes may range between 1 and H with uniform probability.

An important metric of performance is the probability of successful restoration on the first attempt:

$$P_1 = \sum_{h=1}^{H} \tfrac{1}{H}(1-q)^h = \frac{1-q-(1-q)^{H+1}}{Hq} \qquad (1)$$

Under the assumptions, each attempt has the same probability of success, and the probability of eventual restoration within M attempts is clearly

$$P_M = 1 - (1 - P_1)^M \quad (2)$$

The restoration time T_r for a successful attempt is the sum of store-and-forwarding delays and processing delays at each LSR (propagation delays are ignored for simplicity but can be added easily). RR messages may be expected to be assigned high priority, so queueing delays are assumed to be minimal compared to a fixed time t_p representing packet processing (including acceptance/rejection decisions) and packet forwarding. Because the length of backup routes is uniformly distributed between 1 and H, the mean delay in a successful restoration attempt will be $t_p \sum_{h=1}^{H} h/H$, which is approximately $t_p H/2$ for large H.

On any unsuccessful attempt, the RR message will be rejected at a random LSR and additional delays will be incurred by the return of the Release message. The probability of being rejected at the ith LSR along the backup route is $q(1-q)^{i-1}$. If the time to store, process, and forward Release messages at each LSR is a fixed time t_r and the RR message is rejected at the ith LSR, then the total delay involved in an unsuccessful attempt will be $i(t_r + t_p)$ including the time to forward the RR message and return the Release message. If the backup route is h hops, the conditional mean delay is

$$E(T_{rel} \mid h) = (t_r + t_p)[1 - (1-q)^h (1 + hq)]/q \quad (3)$$

Unconditioning on the probability of h hops, the mean delay involved in each unsucccessful attempt is

$$\begin{aligned} E(T_{rel}) &= \sum_{h=1}^{H} \frac{1}{H} E(T_{rel} \mid h) \\ &= \frac{t_r + t_p}{Hq^2} [Hq(1 + (1-q)^{H+1}) \\ &\quad - 2(1-q)(1-(1-q)^H)] \end{aligned} \quad (4)$$

The probability of exactly m unsuccessful attempts before a successful restoration is $P_1(1-P_1)^m$, $0 \le m < M$, and the mean number of unsuccessful attempts is therefore

$$\begin{aligned} E(m) &= \sum_{m=1}^{M-1} mP_1(1-P_1)^m \\ &= \left[1 - (1-P_1)^{M-1}(1 + (M-1)P_1)\right]/P_1 \end{aligned} \quad (5)$$

Combining (4) and (5), the total mean restoration time including failed attempts is

$$E(T_r) = t_p \sum_{h=1}^{H} h/H + E(m)E(T_{rel}) \quad (6)$$

B. Path Restoration

In partial path restoration, the segment of the original active LSP downstream from the failure is uneffected by the rerouting. For more flexibility, it might be desired to reroute the entire failed LSP to another path between the ingress and egress LSRs, as shown in Figure 2. In the path restoration approach, when an LSP fails, the LSR detecting the fault will send a Fault Notification message downstream to the egress LSR. A backup route is pre-established between the ingress LSR and egress LSR. When the egress LSR receives the Fault Notification message, it will send a RR message in the backward direction along the backup route to check for bandwidth and reserve resources. If the RR message reaches the ingress LSR successfully, the LSR will switch the indicated FEC to the backup path. If any LSR along the backup route rejects the RR message, a Release message will be sent back to the egress LSR which will try another feasible backup path.

Any ingress-to-egress alternative path may be chosen, so path restoration has more flexibility than link restoration or partial path restoration. However, the restoration time may be significantly more because the Fault Notification message must be sent to the egress LSR. If the fault occurs near the ingress LSR, the Fault Notification message would have to traverse most of the LSP. If the fault occurs near the egress LSR, then the transit delay for the Fault Notification message may not be substantial.

As before, we assume that a typical LSP consists of H hops and a fault is equally likely with probability $1/H$ to be located at any hop. If the fault occurs on the nth hop, the Fault Notification message will travel H-n hops to the egress LSR. The transit delay for a Fault Notification message is assumed to be linearly proportional to the number of hops, so a message traveling i hops will experience it_d delay where t_d is a delay constant depending on the characteristics of LSPs. The mean delay experienced by Fault Notification messages will be $t_d \sum_{h=1}^{H} h/H$. This delay must be added with times for unsuccessful restoration attempts and a successful attempt.

First, we will need to find the restorability in terms of the probability of eventual restoration. We assume that backup routes will all have H hops, and there are M feasible backup paths. The probability of successful restoration on the first attempt is clearly

$$P_1 = (1-q)^H \quad (7)$$

Again, each attempt has the same probability of success, and restorability in terms of the probability of eventual restoration within M attempts is (2) where P_1 is now (7).

In a successful restoration attempt, a RR message will travel H hops with a total delay of Ht_p. On any unsuccessful attempt, the RR message will be rejected at the ith LSR along the backup route with probability $q(1-q)^{i-1}$. The total delay involved in that unsuccessful attempt will be $i(t_r + t_p)$. The mean total delay involved in each unsuccessful attempt is therefore

$$\begin{aligned} E(T_{rel}) &= \sum_{i=1}^{H} i(t_r + t_p) q(1-q)^{i-1} \\ &= (t_r + t_p)\left[1 - (1-q)^H (1 + Hq)\right]/q \end{aligned} \quad (8)$$

The probability of exactly m unsuccessful attempts before a successful restoration is $P_1(1-P_1)^m$, $0 \le m < M$, and the mean number of unsuccessful attempts is given by (5) where P_1 is now (7). Combining (5) and (8), the total mean restoration time including Fault Notification message and failed attempts is

$$E(T_r) = t_d \sum_{h=1}^{H} h / H + Ht_p + E(m)E(T_{rel}) \quad (9)$$

where $E(T_{rel})$ is given by (8).

III. BACKUP ROUTES WITH QOS CONSTRAINTS

Searching for an optimal backup LSP is not a simple task because of the time-varying rates of packet flows. The bandwidth requirements of a variable bit-rate packet flow is not straightforward to calculate. More bandwidth will result in better QoS, while less bandwidth will allow more traffic to be restored using a given amount of backup resources. The QoS will also depend on the amount of buffer space at each LSR and the traffic characteristics of other packet flows. The problem is made more difficult by the requirement to reroute all failed FECs simultaneously. Backup routes must be selected optimally to make the most use of a fixed set of available resources.

To overcome the first difficulty, we use the equivalent bandwidth concept to reduce the requirements of each FEC solely into terms of bandwidth. Equivalent bandwidth is a reversal of queueing analysis: given traffic characteristics and QoS requirements, the minimum required resources are calculated. More specifically, we used the equivalent bandwidth formulation that finds the required bandwidth given the source traffic peak rate P, mean burst period b, utilization factor ρ, buffer size B, and target packet loss ratio ε [10]. The delay requirement is considered later as an additional constraint. For convenience, the equivalent bandwidth is rewritten here as

$$c = P \frac{a - 1 + \sqrt{(a-1)^2 + 4\rho a}}{2a} \quad (10)$$

where $a = -\frac{b}{B}(1-\rho)\ln \varepsilon$. We note that the equivalent bandwidth for an FEC will vary on different links because of the dependence on an LSR's buffer space. This fact complicates the usual capacity optimization problem encountered in connection-oriented networks, where a fixed amount of bandwidth is required end-to-end for each flow.

Using the notion of equivalent bandwidth to translate the overall requirements of each FEC into bandwidth terms, an arc-path model can be developed for optimization to simultaneously route all backup LSPs satisfying their QoS requirements while efficiently utilizing available spare capacity [11]. Let G=[N,E] denote a network with node set N={1,2,…,n} and link set $E = \{e_1, e_2, \ldots, e_m\}$ of ordered pairs of nodes. An ordered pair of nodes is $e_m = (i, j)$ with $i \neq j$ and $i, j \in N$. A path in [N,E] with origin i_1 and destination i_{p+1} is a sequence $\{i_1, e_{m1}, i_2, e_{m2}, \ldots, i_p\}$. Let o_r and d_r denote the origin and destination for a demand $r \in \{1,2,\ldots,g\}$.

The delay of a path is calculated as the sum of maximum queueing delays, i.e., buffer size divided by link transmission rate. Let D denote the maximum delay permitted for any backup LSP, and let P_r denote the set of backup LSPs whose delay is less than or equal to D. Let $Q_{re} \subset P_r$ denote the subset of paths in P_r containing link e. Let y_e denote the actual capacity of link e and F_e denote the set of demands affected by the failure of link e. For each $p \in P = \bigcup_{r=1,2,\ldots,g} P_r$, let the constant c_{ip} equal the equivalent bandwidth in (10) if $i \in p$, and $c_{ip} = 0$ otherwise. Let the binary variable $g_{ep} = 1$ if path $p \in P$ is a backup LSP when e fails and $g_{ep} = 0$ otherwise. When link e fails, we wish to determine backup LSPs for each demand in F_e such that the total equivalent capacity for each link in these LSPs is less than or equal to the actual capacity. Mathematically, we seek a set of binary variables, g_{ep}, such that

$$\sum_{p \in P} c_{ip} g_{ep} \le y_i \quad \forall i \in E \quad (11)$$

and

$$\sum_{p \in P_r \setminus Q_{re}} g_{ep} = 1 \quad \forall r \in F_e \quad (12)$$

Since our goal is to minimize the total equivalent capacity needed for restoration, the optimization objective can be written as

$$\sum_{i \in E} \sum_{p \in P} c_{ip} g_{ep}. \quad (13)$$

Because of the computational difficulty, we rely on the CPLEX [12] optimization software for solution.

IV. AN EXAMPLE NETWORK SIMULATION

To demonstrate the spare capacity allocation procedure using the arc-path based optimization model for guaranteed service, the small example MPLS network shown in Figure 3 was solved. Table A describes the working LSPs, which include routing information and traffic characteristics for each working LSP. The link capacity and delay for traveling from one LSR to the next are shown in Table B. The buffer capacity for every LSR in the network is given in Table C.

Using this known information, the arc path based optimization model was derived. The arc path model includes all feasible backup LSPs for all working LSPs in the network using the equivalent bandwidtch concept. Choosing an optimal backup LSP for each working LSP was accomplished using the CPLEX optimization system. For each failure e, (11)-(13) was solved using CPLEX 6.0 to obtain the optimal backup LSPs shown in Table D. For the QoS restrictions given (packet loss ratio and total delay), the paths found have the absolute minimum total equivalent capacity.

Due to the structure of the mathematical model, each link failure requires solving a simple discrete optimization problem having $m + g$ constraints and $|P|$ binary variables. Realistic problems having 30 nodes and 45 edges are tractable using the latest version of CPLEX. In previous work involving QoS, searching for backup routes can be very complex, but our strategy exploits equivalent bandwidth and well-known optimization techniques to reduce the complexity of the problem.

V. Conclusions

The described techniques for partial path restoration and path restoration attempt to minimize restoration time by distributing protocol functions to LSRs and pre-establishing backup routes. The techniques have trade-offs in flexibility, restorability, and restoration time.

Spare capacity allocation is usually a difficult problem, and even more difficult with additional QoS constraints, but we reduce the complexity using equivalent bandwidth and an arc path based optimization model.

References

[1] B. Davie, P. Doolan, and Y. Rekhter, *Switching in IP Networks*, Morgan Kaufmann, 1998.

[2] A. Viswanathan, N. Feldman, Z. Wang and R. Callon, "Evolution of multi-protocol label switching," *IEEE Commun. Mag*., vol. 36, pp. 165-173, May 1998.

[3] K. Van Der Wal, M. Mandjes, and H. Bastiaanse, "Delay performance analysis of the new Internet services with guaranteed QoS," *Proc. IEEE*, vol. 85, pp. 1947-1957, Dec. 1997.

[4] T. Chen and T. Oh, "Reliable services in MPLS," *IEEE Commun. Mag*., vol. 37, pp. 58-62, Dec. 1999.

[5] E. Crawley, et al., "A framework for QoS-based routing in the Internet," Internet RFC 2386, Aug. 1998.

[6] R. Guerin, A. Orda, and D. Williams, "QoS routing mechanisms and OSPF extensions," GLOBECOM'97, pp. 1903-1908, Nov. 3-8, 1997.

[7] B. Jamoussi, ed., "Constraint-based LSP setup using LDP," Internet draft draft-ietf-mpls-cr-ldp-02, Aug. 1999.

[8] D. Awaduche, et al., "Extensions to RSVP for LSP tunnels," Internet draft draft-ietf-mpls-rsvp-lsp-tunnel-00, Nov. 1998.

[9] B. Davie, et al., "Use of label switching with RSVP," Internet draft draft-ietf-mpls-rsvp-00, Mar. 1998.

[10] R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE J. Selected Areas in Commun*., vol. 9, pp. 968 –981, Sep. 1991.

[11] M. MacGregor and W. Grover, "Optimized k-shortest paths algorithm of facility restoration," *IEEE Trans. Commun*., pp. Sept. 1992.

[12] *Using the CPLEX Callable Library*, CPLEX Optimization, Inc., Incline village, NV, 1994.

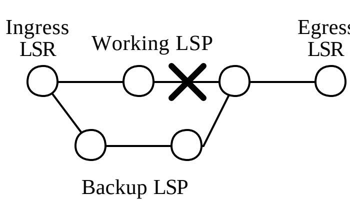

Fig. 1. Partial path restoration

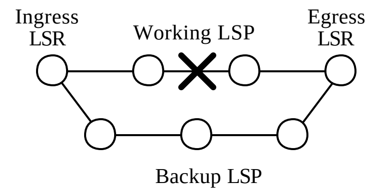

Fig. 2. Path restoration

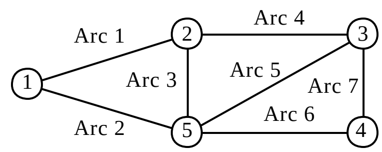

Fig. 3. Simulated example network.

TABLE A. WORKING LSPS AND TRAFFIC FOR EXAMPLE

Work LSP #	LSP	Peak rate	Avg rate	Burst size	Loss prob	Total delay
1	1-5-3-4	6 Mb/s	3 Mb/s	6 Mb	10^{-6}	5
2	2-3-4-5	7 Mb/s	5 Mb/s	6 Mb	10^{-6}	5
3	1-5-4-3	9 Mb/s	2 Mb/s	7 Mb	10^{-6}	5

TABLE B. LINK CAPACITIES AND DELAYS

Link	Capacity	Delay
Arc 1 (1-2)	15 Mb/s	1
Arc 2 (1-5)	17 Mb/s	2
Arc 3 (2-5)	20 Mb/s	1
Arc 4 (2-3)	22 Mb/s	2
Arc 5 (3-5)	19 Mb/s	1
Arc 6 (4-5)	18 Mb/s	1
Arc 7 (3-4)	18 Mb/s	2

TABLE C. BUFFER CAPACITY AT EACH LSR

LSR	Buffer size
1	6 Mb
2	10 Mb
3	5 Mb
4	4 Mb
5	7 Mb

TABLE D. BACKUP LSP FOR ANY LINK FAILURE

Failed link	Affected working LSP #	Optimal backup LSP	Total capacity	End-to-end delay
Arc 2	1 (1-5-3-4)	1-2-3-4	16.6 Mb/s	5
	3 (1-5-4-3)	1-2-3	16.4 Mb/s	3
Arc 4	2 (2-3-4-5)	2-5	6.3 Mb/s	1
Arc 5	1 (1-5-3-4)	1-2-3-4	16.6 Mb/s	5
Arc 6	2 (2-3-4-5)	2-5	6.3 Mb/s	1
	3 (1-5-4-3)	1-2-3	16.4 Mb/s	3
Arc 7	1 (1-5-3-4)	1-2-5-4	16.4 Mb/s	3
	2 (2-3-4-5)	2-5	6.3 Mb/s	1
	3 (1-5-4-3)	1-2-3	16.4 Mb/s	3

0-7803-6454-6/00/$10.00 (C) 2000