An Alternative Path Fast Rerouting in MPLS[1]

Xavier Masip-Bruin, Sergio Sánchez-López,
Josep Solé-Pareta, Jordi Domingo-Pascual

Departament d'Arquitectura de Computadors, Universitat Politècnica de Catalunya
C/ Jordi Girona, 1-3 – 08034 Barcelona, Spain
{xmasip, sergio, pareta, jordid} @ac.upc.es

Abstract. Some different models have been proposed in order to obtain QoS from a network. One of these is related to *Traffic Engineering* (TE). Its goal is achieved based on network performance control. This is carried out by always selecting the best path in order to optimize the network resources. *Multiprotocol Label Switching* (MPLS) and *Explicit Routing* are mechanisms used to perform the TE. Moreover, it is possible to utilize alternative paths in MPLS so that the traffic can be rerouted when a node or a link failure occurs. Therefore, if QoS is desired, a fast rerouting method must be implemented. In this paper a method for selecting the alternative path and a solution for fast rerouting between the primary and the alternative paths is proposed.

1 Introduction

Originally, the Internet only offered a simple Quality of Service (QoS), based on best-effort delivery of data point-to-point. At the moment, applications in real time such as remote video, multimedia conferences or virtual reality, are not performed under this network definition, due to both the delay variable in the queuing process and the problem of congestion. Before these applications are used, the network has to be modified to support QoS end-to-end.

Optimizing the network performance to provide the network with a guaranteed QoS, will be the best modification. In order to achieve this, TE [1] is applied. Improving the utilization of network resources and providing features for quick recovery when a node or a link fails, are the goals of TE. The routing mechanisms are very important whenever the QoS is a vital network feature. Traditionally, obtaining an effective TE in an IP domain is very difficult. This is due to the limitations of the routing mechanisms used as OSPF or IS-IS. The remainder of this document is organized as follows. In section 2, the MPLS and a routing protocol are described. In section 3, the need of rerouting and a fast rerouting are analyzed. Moreover, different approaches and a new suggestion are described in section 4. Finally, in section 5 the conclusions and future works are presented.

[1] This work was supported by the Spanish Research Council (CICYT) under contracts 2FD1997-2234-C03-02 (MIRA) and TEL99-1117-C03-03 (SABA2)

2 MPLS and Constraint-Based Routing

In order to obtain both better scalability and faster packet forwarding performance the MPLS has appeared. The MPLS [2,3] is an advanced forwarding scheme, which allows streams from any particular ingress node *(Label Switching Router*, LSR, in an MPLS domain) to any particular egress node, to be individually identified with a simple label. Therefore, MPLS provides a straightforward mechanism to manage the traffic associated with each ingress node to egress node pair and obviously enhances the source routing. In addition, the path between an ingress node and an egress node is called *Label Switched Path* (LSP).

A label is a short, fixed-length, locally significant identifier that is used to identify a *Forwarding Equivalence Class* (FEC). An FEC is a group of IP packets, which are forwarded in the same manner. The label, which is put on a particular packet, represents the FEC to which that packet is assigned. The routing method used is known as *Constraint-based Routing* (CR) [4], and obviously the constraint is the explicit route. To perform this, the CR not only processes the network topology, but also the attributes of paths and links. This constraint allows the network to be optimized, since all the traffic are controlled.

The network and the MPLS need a mechanism that allows the label distribution and any other negotiation between LSRs. This mechanism is *the Label Distribution Protocol* (LDP) [5]. Currently, a number of different label distribution protocols are being proposed. Existing protocols have been extended so that label distribution can be piggybacked on them, e.g. MPLS-RSVP [6]. Moreover, new protocols have also been defined for the explicit purpose of distributing labels, i.e. MPLS-LDP [5], MPLS-CR-LDP [4]. In this paper the RSVP [7] option is analyzed. Document [6] proposes several additional objects, which extend RSVP allowing the establishment of explicitly routed label switched paths using RSVP as a signaling protocol.

3 Rerouting and Fast Rerouting

In critical MPLS networks the rerouting option is needed. With this feature the traffic flow through the network is guaranteed even though the initial or primary path fails. Moreover, the network congestion will be a very important parameter if QoS is desired. Therefore, the rerouting does not be necessary only to restore the path when a network failure appears but also when congestion exists.

Once the rerouting problem has been presented, two different actions can be analyzed [8]: fast rerouting and optimized rerouting. The former attempts to minimize the time needed to sort out the flow disruption occasioned by an outage. As a consequence, a non optimal path can result from the quick restoration of a disrupted flow. This new path may not be compliant with the QoS required. The latter, optimized rerouting, deals with the mechanism needed to compute a more optimal path in order to obtain a path compliant with the QoS required. In this way the QoS will be guaranteed.

Actually, the network will have two or more feasible paths to send the traffic from an ingress node to an egress node. We can call these paths as the primary path and the

alternative path. It is possible to say that the alternative path protects the primary path. This alternative path can be established either after the failure is detected, or simultaneously with the first LSP establishment. In the first case, when a failure appears, the network needs a significant time to both set up an alternative path and reroute the traffic to this. But there are some applications with important restrictions over the packet loss, where this behavior is not desirable. Therefore, in order to reduce this time it is necessary to previously set up the alternative path. However, some questions are proposed: where should this alternative path be allocated? and what should its length be?. Generating an alternative path between every node in the network and the egress node could be an option to answer these questions. In this way all the nodes are protected nodes, since it is possible to reach the egress node from any node using both the primary (or protected) path and the alternative path. We can say that the protected path is protected by a set of protected nodes. Nevertheless, a major and complex computation along with extensive signaling will be necessary to protect every node along an LSP. In this document, a new method for setting alternative paths is defined. In order to correctly use the alternative path, some new mechanisms must be added. We can summarize these as:

- A mechanism must be added to allow the ingress node of an MPLS domain, to collect information about the network, before applying the routing algorithm.
- When this information is known by the ingress node a routing algorithm is needed. Now, two paths should be established, a primary path and an alternative path. These routes are calculated according to the network topology, available resources, traffic type and any other policy which could be defined.
- A mechanism to allow the node to reroute the traffic between both paths is needed.

A problem exists when it is necessary to reroute the traffic between two paths. This process must be performed minimizing the time. In [9] a new mechanism to allow fast rerouting is described. Basically, having an alternative path to the protected path in every ingress node is desired. Thus, when a link or node failure on the protected path is produced, the ingress node must both detect the failure and reroute the traffic through the alternative path. In this way, only the ingress node should know the different paths between edge routers, and as a consequence, the information that the intermediate nodes should store is minimized. However, when the failure is produced in a node which is close to the egress node, two undesirable situations occur: firstly, the traffic that is still travelling along the path must reach the failure point and then, it must be returned to the ingress node using the same path but in reverse direction. In addition, even though a failure was produced in the protected LSP, the ingress node will continue sending the traffic along it. When the ingress node detects traffic flow in reverse direction along the protected path, it stops sending traffic to this path and reroutes the traffic to the alternative path. Another mechanism is needed in order to perform the label allocation and distribution along the reverse path.

Finally, the needed extensions to apply this method over RSVP are presented in [10]. Several new objects are added to the PATH and RESV messages to allow the network to correctly perform.

4 A solution for Fast Rerouting

In this section a new solution for fast rerouting is presented. This solution depends on the alternative path computation. In fact, this computation is based on the ability to select which nodes should be protected along the path. This ability is, basically, the solution suggested in order to calculate an alternative path (different of an end-to-end path) so that the overload can be minimized. As a consequence, the primary path will be made up of both protected and non protected nodes. In this way, the alternative path is not a simple path between ingress node an egress node, but a set of alternative paths protecting the selected nodes. Moreover, an ability to select the protected nodes should be added to the ingress node so that this node can take the decision. This decision is taken based on the network topology and the available network resources, so a mechanism to collect and refresh this information must exist.

To obtain an exhaustive control of the network a routing algorithm, which allows explicit routes to be created, is needed. If the RSVP is used as an LDP a new object, the *Explicit Routing Object* (ERO)[4] with the intermediate node addresses, is added to the PATH message in order to determine the LSP explicitly. Moreover, a new object to request a label for the flow that is to be sent is added as well. The ERO is unique and will be made up of a set of subobjects with the addresses that the message PATH must follow. Thus, when alternative routes are added to the path, the ingress node has to perform the following:

- selecting a path between edge nodes within the MPLS domain,
- finding out which nodes along the LSP have a major probability to be non compliant with the QoS required, and marking those nodes as protected,
- finding an alternative path for the protected nodes.

4.1 A proposed solution

Since an ERO is used to describe the route, the route information is piggybacked in the PATH message. Therefore, a new field to indicate whether the node is protected or not, is added to the ERO subobjects. This field, named *Protect* (P), consists of one bit which is set when the node is protected. In Figure 1 a subobject format of the ERO (with IPv4 address), with the Protect field added, is shown. When the routing algorithm computes the explicit route, it decides which nodes are protected and which are not. How the routes are computed is not an objective of this paper.

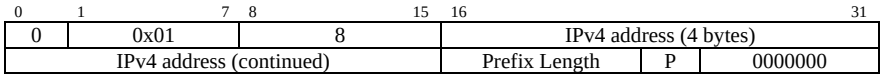

Fig. 1. IPv4 suobject of ERO

A solution to reduce the overhead, is to only make alternative paths between protected nodes. In this way, the information in each protected node is reduced, major network control is obtained and the number of used labels is reduced. This can be performed adding special PATH messages, in order to carry out the label allocation between the protected nodes. It is the routing algorithm, depending on both the

network topology and the network available resources, which indicates to the protected nodes which are the edge nodes of the alternative paths. Determining the nodes which can present inconsistencies is a routing algorithm task. Therefore, the routing algorithm is a major component to be perfectly defined. The following items must be designed in order to implement the proposed solution:

- The subobjects format of the ERO of the PATH message must again be modified. A new field named *Alternative Path Address* (APA) is added. It is a variable length field, always a multiple of 32 bytes. This field is only useful when the node is protected. In this case, it provides the *n* addresses that make up the alternative explicit route, from this node to the next protected node, that has been established by the routing algorithm.
- A new mechanism to allow the node to trigger an alternative path setup when the node detects that it is protected.
- A new message AL_PATH to perform the label request through the route given by the APA field between protected nodes.
- A new mechanism in order to handle the labels in the alternative paths.

The format of an IPv4 subobject is shown in Figure 2. Two different options are described. In (a) the node which receives the ERO (@IPv4-i) is a non protected node, and in (b) it is protected. In this case, two IPv4 addresses are added in the APA field. These addresses indicate the explicit route of the alternative path between this protected node and the node addressed by @IPv4-2.

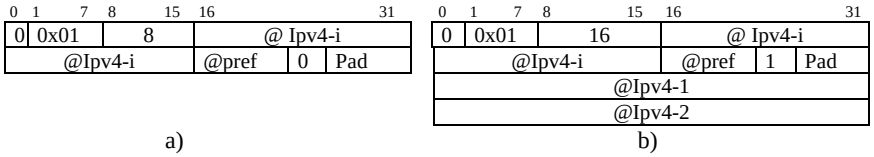

Fig. 2. IPv4 subobject of ERO, for a non protected node (a), and a protected node (b)

A restriction exists over this implementation. The alternative path between two protected nodes must be made up of the same node type as the node which is receiving the message. In this way, knowing the alternative path type address is very easy. Furthermore, with this restriction adding the address type in the APA field is avoided. Therefore, APA field size is minimized.

With the suggested solution, the performance is as follows (Figure 3): when a PATH message is received by a node, protected or non protected, a new PATH message is generated by this node. This message is sent to the next hop. The address of this next hop, is obtained from the next subobject of the ERO. In this new message the LAB_REQ object is sent, in order to perform the label request by the flow that is going to be sent. In addition to this, if the node is a protected node, it generates a new message AL_PATH. This message is forwarded through the route given by the APA field, which is located in the subobject where this node is defined in the ERO. Two objects are contained in this message:

- AL_LAB_REQ: is used to perform the label request in the alternative path.
- AL_ERO: is used to describe the explicit route of the alternative path. It is filled with the addresses provided by the APA field.

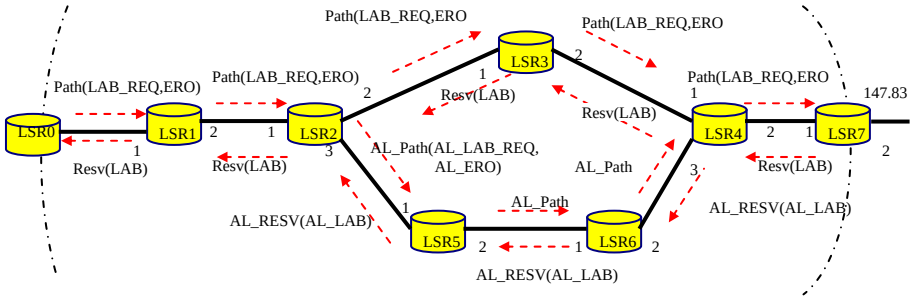

Fig. 3. A network topology

In Figure 3, the message flow is presented over a network topology. Obviously, it is supposed that initially the routing algorithm has chosen the suitable paths. Two messages, PATH and AL_PATH are generated in the protected node identified as LSR 2. In Table 1 (a), the ERO content of the message PATH when is reaching the LSR1 is described. In Table 1 (b) the content of the AL_ERO of the AL_PATH message is presented. In this example, although the LSR4 is not a protected node, it has been chosen as a alternative end path.

Table 1. ERO content (a) and AL_ERO object of AL_PATH message (b)

0	0x01	8	@IPv4-1
@IPv4-1		@ pref	0 0000000
0	0x01	20	@ IPv4-2
@ IPv4-2		@ pref	1 0000000
@ IPv4-5			
@ IPv4-6			
@ IPv4-4			
0	0x01	8	@ IPv4-3
@ IPv4-3		@ pref	0 0000000
0	0x01	8	@ IPv4-4
@ IPv4-4		@ pref	0 0000000
0	0x01	8	@ IPv4-7
@ IPv4-7		@ pref	0 0000000

a)

0	0x01	8	@ IPv4-5
@ IPv4-5		@ pref	0 *3*
0	0x01	8	@ IPv4-6
@ IPv4-6		@ pref	0 *2*
0	0x01	8	@ IPv4-4
@ IPv4-4		@ pref	0 *1*

b)

The routing algorithm takes this decision because either LSR4 is close to the egress node or it is compliant with the QoS constraints required by the flow to which this path has been established. The label allocation for the alternative path will be carried out on the node where the alternative end path is detected. In order to perform this label allocation, a node should be able of detecting if it is the end of the path. Adding one new field in the routing table describing the alternative path end could be one way to achieve it. But it imply the addition of a new field, which is not desirable. Using the seven padding remainder bits of the AL_ERO format as a hop counter is the solution proposed. Therefore, the number of nodes which make up the alternative path is added in this field by the protected node which generates this path. This field will be decreased in each hop, and the node which generates a zero value will make the label allocation. This field is shown in Table 1 (b).

Nevertheless, even a last problem exists. In Figure 3, the LSR2 must know that the alternative path is made up of three nodes. This is necessary because the information added in the hop counter field by this node must be correctly added. But, the number of nodes of the alternative path is easily determined from the length of the subobject corresponding to the protected node. In the example we can see that the subobject length is 20 bytes. If the 8 bytes used by a non protected node are subtracted from this 20 bytes, the 12 bytes represents three IPv4 addresses length. Adding a new field in the APA field to determine the alternative path end is avoided with this solution.

It is also important to analyze the alternative path behavior when there are other LSPs sharing the same resources. Generally, to carry out resource reservation for alternative paths is not desirable. Basically, an alternative path is used only when there is traffic present on that path. When there is no traffic, other LSPs sharing the links, can use all the system resources. Thus, when an alternative path should be used, it is necessary to evaluate the priorities of both others LSPs using these resources and the alternative path. Thus, it is possible to preempt resources for a protected high priority LSP without performing explicit reservation for alternative paths. A new mechanism should be added to the LDP to request the LSP priority in order to avoid resource reservation. In summary, the advantages of this method of fast rerouting are:

- Packet reception disorder is not produced.
- The mechanisms, on the ingress node, to detect an LSP failure and to decide on which interface reroute the traffic are not necessary.
- A faster label allocation and distribution is allowed on the alternative paths at the same time that is performed on the protected LSP, adding a minor overhead.

4.3 Routing tables modifications

Applying the mechanism mentioned above over any network topology, will cause immediate changes on the routing tables structure. However, it is a straightforward difference, there are more than one possible path to connect two possible edge routers.

Different priorities must be allocated in each protected node to allow the selection between protected path (P) and alternative path (AL). The RSVP must be running in order to generate the new routes for the alternative paths in the routing table, according to the new objects added in this paper to perform the label request. The routing tables of the nodes of the Figure 3, with the new field (Type), are in Table 2.

Table 2. Routing tables

LSR	Int In	Label In	Destinat.	Int out	Label Out	*Type*
0	1	-------	147.83	2	0.10	*P*
1	1	0.10	147.83	2	0.50	*P*
2	1	0.50	147.83	2	0.40	*P*
	1	0.50	147.83	3	0.60	*AL*
3	1	0.40	147.83	2	0.30	*P*
4	1	0.30	147.83	2	0.20	*P*
	3	0.80	147.83	2	0.20	*AL*
5	1	0.60	147.83	2	0.70	*AL*
6	1	0.70	147.83	2	0.80	*AL*
7	1	0.20	147.83	2	----------	*P*

5 Conclusions and future works

This paper suggests a mechanism to perform a fast rerouting between different paths. We propose a method in order to avoid the need of holding up an alternative path structure end-to-end. This method is based on the added ability for selecting which nodes must be protected. Thus, a "protected path" made up of several protected nodes exists. To achieve this objective, a new field is added to the routing tables. This is then used by each node to select the most suitable path to forward an incoming packet. Moreover, the RSVP protocol is used as LDP and a minor modification to the PATH message format is suggested. In this way, without interfering with the label allocation and distribution in the protected path, label allocation and distribution by an alternative path are both performed, at the same time. Since the alternative paths are usually generated between protected nodes, the routing algorithm will determine this path easily. When the traffic must be rerouted between two paths, an LSP failure detection mechanism in the ingress node is not necessary. The traffic will be rerouted to an alternative path by the node that has detected an LSP failure. Therefore, the rerouting time is only the time spent to index the labels table. Thus, the goal proposed is achieved with only a minor overhead, which is due to the added modifications to the RSVP messages and to the new field added in the routing table. Finally, the egress node will receive the packets in the same order as sent by the ingress node. In summary, a trade-off exists between the added overhead and the amount of nodes to be protected, i.e. the network reliability. Future works proposed are:

- A mechanism to select which nodes must be protected.
- The routing algorithm, using QoS parameters to take routing decisions.
- A mechanism to indicate the alternative path priority.

References

[1]Awduche,D.O., et al. "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

[2]Rosen,E., et al. "Multiprotocol Label Switching Architecture", Internet draft< draft-ietf-mpls-arch-05.txt>, April 1999.

[3]Xiao,X., et al.,"Traffic Engineering with MPLS in the Internet",IEEE Network Magazine, March 2000

[4]Jamoussi,B. "Constraint-Based LSP Setup using LDP", Internet draft<draft-ietf-mpls-cr-ldp-03.txt>, September 1999.

[5]Anderson,L.,et al. "LDP Specification",Internet draft<draft-ietf-mpls-ldp-05.txt>,June 1999.

[6]Awduche,D.O., et al. "RSVP-TE: Extensions to RSVP for LSP Tunnels", Internet draft<draft-ietf-mpls-rsvp-lsp-tunnel-05.txt>, February 2000.

[7]Braden,R et al., "Resource Reservation Protocol (RSVP)-Version 1, Functional Specification", RFC 2205, September 1997.

[8]Swallow,G., "MPLS Advantages for Traffic Engineering", IEEE Communications Magazine, December 99.

[9]Haskin,D. et al., "A Method for Setting an Alternative Label Switched Paths to Handle Fast Reroute", Internet draft<draft-haskin-mpls-fast-reroute-02.txt>, December 1999.

[10]Krishnan, R. et al., "Extensions to RSVP to Handle Establishment of Alternate Label Switched Paths for Fast Reroute", Internet draft<draft-krishnan-mpls-reroute-rsvpext-01.txt>, June 1999.