Network Working Group
Internet Draft
Expiration Date: Jan 2005

Albert J. Tian
Redback Networks
George Apostolopoulos
ICS-FORTH

July 2004

Source Routed MPLS LSP using Domain Wide Label

draft-tian-mpls-lsp-source-route-01.txt

1. Status of this Memo

2. Abstract

One advantage that MPLS provides is that it allows traffic to be directed through an explicit path that deviates from IP routing. Such ability is widely used in traffic-engineering and fast-reroute. Currently signaling protocols such as RSVP is needed to establish and maintain such an explicit Label Switched Path. When there are a large number of such signaled LSPs in the network, the aggregated signaling and maintenance overhead can be significant.

In this paper, we propose a way to establish a source routed explicit path with zero signaling overhead. The scheme uses a stack of labels

and requires domain wide LDP FEC to label bindings.

3. Introduction

On merging capable LSRs, LDP builds merging LSP trees rooted at the egress of the FEC. LDP allocated labels usually are only of local significance. In other words, an FEC can bind to different labels on different links in a network. By doing so, each LSR can achieve conflict free label allocation without any coordination.

But in some cases, a domain wide FEC to label binding may be desirable. In a domain wide FEC to label binding, a given label is always bound to the same FEC on all links in the network, if a binding for the given label exists. We call such a label a Domain Wide Label(DWL).

Consider the following example where FEC-d corresponds to a loopback interface address d on LSR-D. In traditional FEC to label binding, FEC-d can bind to different labels on different links:

```
           label 30  label 20  label 10
 FEC-d : A ------- B ------- C ------- D
```

In domain wide label binding, FEC-D binds to the same label 10 on all links:

```
           label 10  label 10  label 10
 FEC-d : A ------- B ------- C ------- D
```

4. Terminologies

Domain Wide Label (DWL): A label is said to be a Domain Wide Label if the FECs that map to that label are always the same on all links in a MPLS domain.

Local Label: A label is said to be a local label if multiple distinctive FECs can map to that label on different links in a MPLS domain.

5. Source Routed LSP

DWL allows an efficient way to support source routing in an LDP enabled network using a stack of labels.

5.1. An example

For example, in the following network there are 6 LSRs A through F. Each LSR has a loopback interface with a domain wide label allocated for it. Assuming LDP is running on all the LSRs and LDP can be enhanced to distribute such domain wide label bindings throughout the MPLS domain. The domain wide labels still have the same semantics as other LDP labels, just that the same label here always maps to the same FEC on all LSRs in the MPLS domain. Later in this document, we will give out the details on the LDP enhancements.

```
         +-------> 60/30/P ---> 30/P -------->+
         ^                                    |
         |      D-----------E-----------F     |
         |      | 40        50       60 |     v
    50/60/30/P  |                       |     P
                | 10        20       30 |
                A-----------B-----------C
```

Fig. 1

The domain wide label allocation on all LSRs are as follows:

LSR	Label	Loopback-Interface-prefix/FEC
A	10	a
B	20	b
C	30	c
D	40	d
E	50	e
F	60	f

In this example, all LSRs would use label 10 to deliver packets to address a on LSR A, and use label 30 to deliver packets to address c on LSR C, and so on and so forth.

Lets say that normally LSR A would use label 30 to deliver packets to C along path A-B-C. So FEC c to label 30 mapping must be installed on all LSRs on the path. Here we assume this can be achieved by the enhanced LDP.

Now if LSR A wants to use an alternative path A-D-E-F-C to deliver packets to C, it can push an additional label stack 40/50/60/30 on to

the packet and forward the packet according to label FIB. Here 40 is the top (outer most) label. If PHP is enabled, the top label 40 will be popped on LSR A and the packet will be forwarded to LSR D according to the action associated with label 40. When the packet arrives on LSR D, the top label 50 will determine the nexthop to be LSR E with action pop. Similar procedures will happen on LSR E and F, eventually deliver the payload P to LSR C.

If PHP is disabled on these LSRs, the labels will not be popped at the penultimate hop resulting in one extra label on the label stack in the packet.

Similarly, LSR A can use stack 50/30 to specify a Loosely Source Routed Path A-E-C. In this case top label will deliver the packet to LSR E, and the next label 30 will deliver the packet to LSR C.

5.2. Benefits of Source Routed LSP

There are several advantages of such source routed LSPs.

5.2.1. Zero signaling and maintenance overhead

Since these LSPs are source routed, there is no signaling overhead and no maintenance overhead. Also only the headend of the LSP needs to maintain the state related to the LSP, other LSRs on the LSP are not even aware of the existence of such LSPs.

This makes source routed LSPs perfect for establishing bypass LSPs for fast-reroute. In such applications, numerous bypass LSPs need to be created and maintained yet only to be used very infrequently when some link or node fails in a network.

5.2.2. Zero signaling delay

Also because the LSPs are source routed, they can be used immediately after the stack of labels are determined. This allows LSPs to be adjusted on the fly without any interruption. In other words, make-before-break is inherit in the design.

5.3. Other Benefits of DWL

5.3.1. DWL and LDP node protection

In applications such as LDP node protection as described in [SHEN00], an LSR needs to learn labels allocated by the next-nexthop LSR for a given FEC. Without DWL, protocol extensions as outlined in [SHEN00] will be needed to propagate that information. In a DWL enabled LDP network, the protocol extensions described in [SHEN00] will not be needed since the next-nexthop label for a FEC will be the same as the label allocated on the local box if that label is a DWL.

5.3.2. DWL can help in troubleshooting

DWL makes the network easier to troubleshoot. Since each FECs using DWL bind to the same label on all the hops, packets with such a label can be correlated to the FEC easily.

6. Strictly Source Routed Segments over High Cost Links

Using a stack of DWLs, one can construct a Loosely Source Routed Path(LSRP), with each DWL representing a loose segment on the path.

In most LDP enabled networks, at direct link between two LSRs is the shortest path between the two according to routing. In such a network, a DWL for a directly connected neighbor will deliver packets over one or more of the directly connect links to that neighbor. In this case, strict segments in an explicit path can be implemented using DWLs.

However, in some cases if all direct links between two adjacent LSRs have been configured with higher link costs than the shortest indirect paths, then these direct links will not be used by IP routing except for packets whose destination address is the interface address on the far end of the high cost link.

```
1-------------Z-------------1
|                           |
|                           |
|                           | 1.1.1.1/32 loopback label 100
X-------------3-------------Y
 10.1.1.1/24     10.1.1.2/24 interface address label 10
```

Fig. 2

In the example given in Fig. 2, the costs of the links among three LSRs X, Y and Z are marked on the links. Label 100 is a DWL for FEC 1.1.1.1/32 whose egress is a loopback interface address on LSR Y. Even when there is a direct link between X and Y, MPLS packets arriving on X with top label 100 will still be forwarded to LSR Z, since the path X-Z-Y has a better metric. The only traffic that will be sent over the direct link between X and Y is traffic from X with destination 10.1.1.2, and vise versa, due to the fact that most of the implementations prefer directly connected interface route over any other route types.

In order to guarantee a Strictly Source Routed Segment between X and Y in this scenario, a new Longest Match Address FEC element (LM-Address FEC element) is introduced that uses longest prefix match instead of exact match to find its matching route. The LM-Address FEC element is defined in as follows:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | LM-Address(4) |     Address Family            | Host Addr Len |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   |                     Host Addr                                 |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Address Family
   Two octet quantity containing a value from ADDRESS FAMILY NUMBERS in [RFC1700] that encodes the address family for the address prefix in the Prefix field.

Host Addr Len
   Length of the Host address in octets.

Host Addr
   An address encoded according to the Address Family field.

The LM-Address FEC element is essentially the same as the Host Address FEC element, except that it has a different FEC element type 0x04.

To solve the strict hop over high cost link problem, a DWL needs to be allocated on Y and bound to LM-Address FEC element 10.1.1.2. When the binding is advertised to X, X performs a longest prefix match in its routing table and finds the route 10.1.1.0/24. A LSP will be created with link X-Y as the outgoing interface.

To specify a strict hop over a high cost link in an explicit path, the interface address 10.1.1.2/32 needs to be used.

7. LDP extensions for DWL

7.1. Reserve a pool of labels for DWL

A pool of labels need to be set aside on all LSRs in the domain to be used as DWLs. Local labels MUST not be allocated from this pool, otherwise we can not guarantee that the same label always maps to the same FEC. After the pool of labels are reserved, LSRs can then allocate domain wide labels from the pool.

Implementations MUST allow user to configure the DWL label range. All LSRs in a domain MUST agree on the range of labels reserved for DWL to avoid allocating local labels from the DWL pool.

Since most existing implementations allocate local labels from near the lower end of the label space, label ranges near the higher end of the label space is usually more suitable for DWLs.

7.2. Allocating DWL

In most cases, each LSR is allocated a unique DWL from the DWL pool for its loopback interface address FEC. This FEC to DWL binding will be propagated throughout the MPLS domain using LDP.

This allocation can be achieved in several ways:
  a) manually allocate via configuration, or
  b) automatically allocate through a centralized server, or
  c) algorithmically derived from something else.

Method a) is the most strait forward. In this case the operator needs to make sure there are no conflicts in DWL allocations. Since each LSR only needs one or in some cases two DWLs, this should not be a big burden for operators.

7.3. Advertising and Detecting DWL

An LSR can determine if a label is a DWL by checking if it falls within the DWL range. Hence DWL can be advertised using the existing Generic Label TLV.

7.4. Extensions to Label Mapping Procedures

When a LABEL MAPPING message is received with a DWL in the label TLV, the receiving LSR SHOULD try to allocate the same label for the FEC. If the received DWL is already allocated to a different FEC, a local label SHOULD be allocated for the FEC, and a NOTIFICATION message with non-fatal status code SHOULD be sent to the advertising router.

The value of the status code is TBD.

7.5. PHP

Since DWL label values need to be communicated to adjacent LSRs so that they can be further propagated upstream, implicit-null label can not be used to signal PHP operation. One solution is to infer PHP from ADDRESS messages.

For FEC with /32 Prefix FEC elements, or Host Address FEC elements, or LM-Address FEC elements, if all the addresses in the FEC are among the addresses in the ADDRESS messages from the advertising LSR, and the advertising LSR is not a targeted neighbor, then PHP is assumed for the LSP unless otherwise instructed by local policy.

8. Construct Source Routed LSP using DWL

Assuming an explicitly routed path is specified by a list of IP prefixes, each of which represents either a loose hop or a strict hop. Given such a path, we can construct a Source Routed LSP using the following algorithm:

a) Set the current node to be the last node in the path, and set the label stack to be empty.

b) For the IP prefix of the current node, find an FEC element that exactly matches the IP prefix. Host Address FEC elements and LM-Address FEC elements are considered of having prefix length 32. Then find the DWL that is bound to that FEC element. Push the DWL onto the stack.

If such DWL can not be found, abort the algorithm with an error.

c) If the current node is the first node in the path, terminate the algorithm.
Otherwise set the current node to its predecessor in the path and goto step a).

The resulting label stack represents a source routed LSP, and can be used to forward packets from the starting point to the last hop following the desired path.

9. Other Considerations

9.1. Interface Label Space

A LSR support interface label space needs to reserve the same DWL label range on all interfaces, and needs to allocate the same label for an FEC out of the reserved DWL label range.

9.2. ATM and Frame Lable Encoding

For a network using ATM label TLV or Frame Relay Label TLV, a seperate DWL label range must be defined for each different label encodings. Still DWLs can be advertised using the same ATM Label TLV or the Frame Relay Label TLV.

9.3. Verification of Source Routed Paths

Standard tools such as MPLS ping and MPLS traceroute can be used to verify a source routed path is functioning as expected.

9.4. Compatibility Considerations

The solution proposed in this document is compatible with existing LDP specification. LSRs that do not understand DWL will not get the benefit of DWL, but basic LDP connectivity should remain intact.

9.5. Loop Prevention

One very nice attribute of the source routed LSP is that as long as each hop is loop free, the path will also be loop free.

It is possible to construct a LSP that visits the same LSR twice by including the same DWL twice in the stack, but no infinite loop will be created.

It is recommended for LSRs that support DWL to copy TTL values from the outer label to inner label when a label is popped, to avoid delayed TTL expiration.

10. Security Considerations

This document proposed a new and efficient way to implement source routing. All known security concerns related to source routing may also be concerns here.

Please note that an attacker can use a stack of labels to perform source routing today if label bindings are known on all routers on the path. With this proposal, if label bindings on one router is known to the attacker, then source routing can be utilized by the attacker.

The main security concerns related to source routing include the following scenarios where source routing may be abused:

- to bypass administrative control
- to make a malicious packet appear as if it had come from a trusted system
- to reach otherwise unreachable part of the network such as private address space
- to collect information about a network

The concerns can be eliminated by not accepting DWLs from outside the trusted domain. This can be achieved by doing the following:

1) Do not accept labeled packets from outside the trusted domain.

   If labeled packets must be accepted from outside, then do not accept DWLs from outside the trusted domain. Since the DWL range is known, this policy can be achieved by label based filtering at the entrance points of the trusted domain to block packets with any DWLs in the label stack.

2) Do not accept labeled packets arriving from tunnels (such as GRE or IP-in-IP, etc). This can be achieved by disabling protocol ID MPLS at tunnel next protocol ID demux point.

   If MPLS over tunnel must be supported, then do not accept labeled packets from tunnels originated from outside the trusted domain.

   If labeled packet must be accepted from tunnels originated from outside the trusted domain, then do not accept labeled packets with DWLs from these tunnels.

One difference between MPLS source route and IP source route option is that the IP source route option is designed to specify the path for both the request in the forwarding direction and the response in

the reverse direction, while MPLS source route can only specify the path in the forward direction. Therefore the security risk for MPLS source route is lower than the IP source route option.

11. IANA Considerations

A new LM-Address FEC element TLV is defined in this document with FEC element type 0x04. This LDP extension requires this FEC element type to be allocated by IANA.

A new status code for LDP NOTIFICATION message to notify the conflicts in DWLs needs to be defined and allcoated by IANA.

12. Acknowledgments

The authors would like to thank Naiming Shen, Enke Chen and Acee Lindem for their comments.

13. References

[RFC3036] L. Andersson, P. Doolan, N. Feldman, A. Fredette, and B. Thomas, "LDP Specification", RFC 3036, January 2001.

[SHEN00] N. Shen, E. Chen, A. Tian, "Discovering LDP Next-Nexthop Labels", draft-shen-mpls-ldp-nnhop-label-01.txt, work in progress.

[SHEN01] N. Shen and P. Pan, "Nexthop Fast ReRoute for IP and MPLS", draft-shen-nhop-fastreroute-01.txt, work in progress.

14. Author Information

Albert Jining Tian
Redback Networks, Inc.
300 Holger Way
San Jose, CA 95134
Email: tian@redback.com

George Apostolopoulos
ICS-FORTH, Institue of Computer Science,
Foundation of Research and Technology Hellas
Email: georgeap@ics.forth.gr

15. Full Copyright Statement