A methodology for monitoring LSP availability in MPLS networks

A Brooks,† L Sacks‡
† BTexact Technologies, ‡ University College London
{a.brooks@ee.ucl.ac.uk}

Abstract: Multi Protocol Label Switching (MPLS) is an evolving technology that can enable new service offerings such as Virtual Private Networks (VPNs) to be delivered over an IP infrastructure. However, there is presently no mechanism for confirming availability of the Label Switched Paths (LSPs), or data-plane within such a VPN. This paper discusses some of the ways in which customer connectivity within a MPLS VPN could be monitored.

1. Introduction

Multi-service networks contain mechanisms that allow network operators to detect and collect information about connectivity fault conditions. For example, SDH networks use bit interleaved parity to identify bit errors, and ATM networks employ loopback cells to ensure integrity of the virtual circuits. However, there is presently no way of confirming connectivity at the MPLS layer.

A requirement for MPLS monitoring tools has risen from network operators need to determine Label Switched Path (LSP) availability within their networks. This information alerts network operators to faults as early as possible, and therefore helps to ensure that Service Level Agreements (SLAs) are met. More detailed information about service providers requirements for MPLS Operations and Maintenance (OAM) can be found in [1].

This paper is concerned with monitoring the health of the Label Switched Path (LSP) data-plane within a BGP MPLS Virtual Private Network (VPN) platform [2]. The scale of network being considered is assumed to have over 100 Provider Edge (PE) routers. It is expected that the Label Distribution Protocol (LDP) is used to establish a full mesh of LSPs between all the PE routers. Use of the Resource Reservation Protocol (RSVP) as an alternative method for path establishment has not been considered in this paper.

2. Background

The precursor of MPLS was IP Switching. In the mid 1990s, companies such as Ipsilon and Toshiba realised there was the potential to speed up the process of forwarding IP packets within a router by identifying flows of layer 3 traffic, and then switching the remainder of these traffic flows through the equipment at layer 2. The information about the mapping between layer 3 and 2 could then be shared with neighbouring nodes, so that a layer 3 IP address lookup was only required once.

As the IP packet forwarding capacity of routers has increased so considerably over the last few years, the original emphasis of label switching has changed. The concept is no longer required to speed up packet throughput, but it offers other advantages. One of these is to provide VPN services over an IP infrastructure, as depicted in Figure 1.

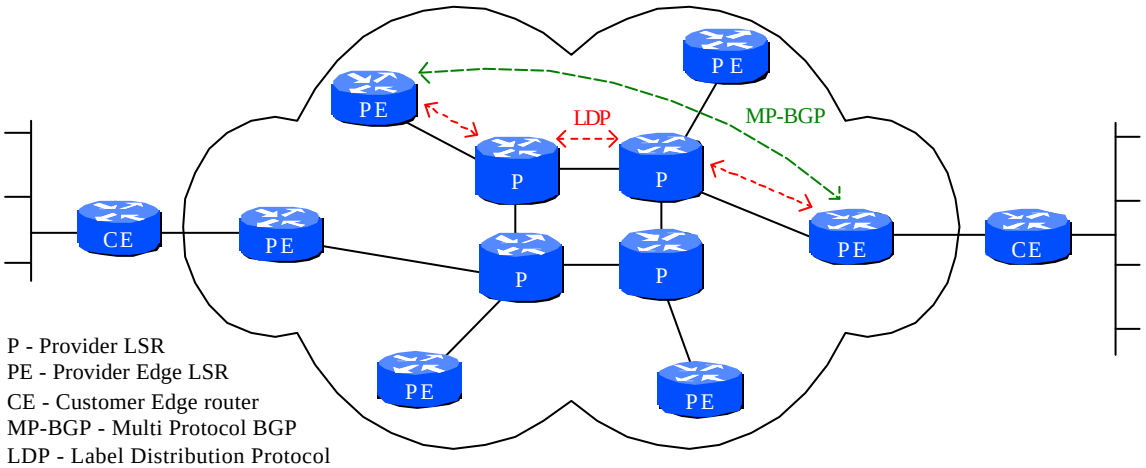

Figure 1 – MPLS VPN network topology

The network provider would own and operate the core label switch routers (LSRs) and the edge LSRs. An MPLS label stack is used for packet forwarding. When the IP packet enters the network at the PE router it is pre-pended with two labels. The outer label is allocated by the Label Distribution Protocol (LDP) and represents the BGP next hop, whereas the inner label is allocated by BGP and represents that VPN's routing table, or the outgoing interface on the PE router. From here onwards, the packet is switched through the network by simply swapping the outer label at each P node. The PE maintains an individual routing and forwarding table for each attached VPN, called a VRF. This routing table is populated by routes from the connected CE device, and relevant routing information received from other PE routers with customers belonging to the same VPN. The Border Gateway Protocol with multi-protocol extensions (MP-BGP) is used for distributing reachability information about customer VPN routes between PEs. Within the core of the network, LDP is used to establish LSPs between the P and PE devices.

3. Fault conditions

Within the MPLS VPN network, there are a number of faults that a monitoring tool should be able to detect and alert the operator to. The most important issues are outlined below:

1. MPLS configuration errors on PE routers. This could happen when a new PE device is added into the network, but global commands relating to the whole router or interface specific commands are not enabled. In this case, although IGP neighbour adjacencies may well be formed, obviously it will be impossible to sent traffic over LSPs.
2. Software failures on routers. MPLS code is not yet as mature as other router operating system components. It is therefore desirable for an error condition to be captured whereby the router appears to be operating correctly, and again, IGP routing protocol sessions may establish, but LDP itself does not run properly. It could be that LDP control plane sessions are established between routers, but that labels are not assigned correctly to address prefixes.
3. BGP VPN label allocation verification. This is a much harder situation to monitor, but ideally the network operator would like to be able to confirm that changes to the network topology, and hence the IP routing tables will be accurately reflected into the label forwarding tables within the routers.
4. Load sharing. In Figure 2, LDP sessions might not establish from the PE router to all the P interfaces. For example, sessions might be established to A and B, but not C or D. The P routers will advertise the same label to an IP prefix out of both interfaces, since labels are allocated on a per-router not per-interface basis. The load balancing mechanism on the PE might cause a stream of identical automated ping echoes within a LSP to always go via A or B, but never C or D. Thus the lack of path establishment to C and D would never be noticed, and traffic will not be load balanced.

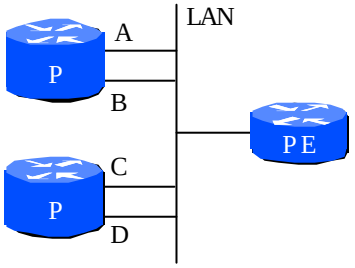

Figure 2 - Simple network topology illustrating load sharing problem

Ideally any tool should be able to detect problems with LSPs regardless of whether they have been set up by LDP or TDP (Tag Distribution Protocol). This will ensure that a monitoring tool would work on networks employing either of the label distribution techniques.

4. Methodology

Several methods for resolving this problem were debated, and these are given below.

4.1. Method 1

Ideally an OAM solution that is engineered into the router operating software is the best way forward. However, a standardised solution obviously requires a standards body such as the IETF to agree on a definition. A draft submitted to the IETF [3] by Harrison in December 2001 presented such a mechanism. This draft defined a connection verification "CV" packet, which would be transmitted periodically from source to sink to determine any LSP mis-routing defects as well as link and node failure. It also defined a forward defect indicator "FDI" and backward defect indicator "BDI" which would carry the defect type and location to the near end and far end respectively.

This particular draft has been dropped by the IETF. However, drafts proposing an MPLS equivalent of ping and traceroute are being discussed [4]. Even if these features are adopted by router manufacturers, services providers will still required a tool to actually make use of these basic facilities, and to automate large scale network monitoring.

4.2. Method 2

This method assumes the monitoring tool must insert test traffic into the data-plane (i.e. the LSPs), but that an OAM solution has not yet been defined by a standards body. This means that ordinary, IPv4 test packets must be injected at the LSP ingress and recovered at the egress.

It is very difficult to make such an approach scalable. If there are more than 100 PE routers in a network, and they have a full mesh of LSPs between them, this equates to approximately 10,000 unidirectional LSPs that must be monitored. There is also the question of where would such a test packet originate from. It could either be created by the PE router itself, or by an external device, such as another router or a type of server. The preferred approach here would be to use an external device. An undesirable amount of interference with the PE routers would be required to generate test probes, and an independent solution would be preferred anyway.

If an external server is used, there is still the quandary about how to get the test packets into each LSP. Two or three servers could be placed centrally, and the appropriate test packets (e.g. ping echo) could be tunnelled through to the PE routers. This might be done using IP-IP, GRE or PPP tunnels. However, the ping reply still needs to be returned to the server. It can't be returned through an LSP in the opposite direction, because if loss does occur it won't be known which of the LSPs was at fault. Therefore the reply should go straight back to the server from the target PE through the ordinary IP plane, or perhaps another tunnel. This solution still has one major failing though - if a problem is identified, there is no way of ensuring that this is due to the MPLS network and not the tunnels. If a catastrophic network event happens, which prevents the server from communicating with a significant number of the PE routers, then it will be impossible to determine the state of the core network, when the reality of the situation is that the core network has probably been severely compromised too.

4.3. Method 3

This method does not actually rely on sending probe traffic through the data plane. Instead, P and PE routers within the network would be interrogated via SNMP on a periodic basis about their LSP connections, and their LDP connection status.

The LSP connection information can be obtained from the Label Switching Router (LSR) Management Information Base (MIB) within each node. By using the cross-connect and other tables within the MIB, incoming labels and interfaces can be mapped to egress labels and interfaces. The IPv4 destination prefix can also be found. The relation between these tables for a Cisco router is described in [5].

The cross connect index identifies a group of entries which all forward frames with the same incoming outer label towards the same prefix. For example, on one of the routers within a network the index "1658385716" might identify all those incoming LSPs using label number 208. Hence all these frames would be routed out of the same egress port towards the same prefix address. Therefore, to create a unique reference for every LSP within the network, just using the router loopback address and the cross connect entry of each LSP will not be sufficient - the associated interface index would also be required.

A database could then be created consisting of all the LSPs between all the routers within the network. Links within this database could be created to show end-to-end connections between PE routers. Each database entry would comprise the following parameters:

- Unique network LSP identifier
- Starting router
- Starting interface
- Terminating router
- Terminating interface
- Destination prefix
- Upstream unique network LSP identifier
- Downstream unique network LSP identifier

Routers within the network could then be polled periodically for LSP status update information, which would go into the database. Any changes within the database would show LSP connectivity issues. Although this may seem like a lot of router polling, each router would only be polled once every minute, and the rate would be controlled so that the SNMP engine load does not get too high. The number of entries within the LSP cross

connect table could be used to dynamically bias the monitoring of some routers more highly than others. Obviously there would be maximum and minimum limits on this to maintain deterministic monitoring.

In addition to the above, it is also important to monitor the TCP connection status of LDP between routers in the network. This would be done through a mixture of standard and private MIBs. For example, the 'iso.org.dod.internet.mgmt.mib-2.tcp.tcpConnEntry' table tells us what state all the TCP connections are in, and between what addresses and port numbers they exist as described in [6]. Also, the 'TcpRetransSegs' entry could be used to indicate which end of a TCP link was trying to re-establish a connection, and which was not. Details about TCP session connection time, which reveals when the session was last reset, can be found from private MIBs, such as 'ciscoMgmt.ciscoTcpMIB.ciscoTcpMIBObjects.ciscoTcpConnTable.ciscoTcpConnEntry' for Cisco devices [7]. Routers could be interrogated for LDP session information more frequently that than for LSP information, since the amount of data to be retrieved is far less.

5. Conclusions and future work

The only realistic method for providing LSP monitoring in the near term is Method 3. Method 1 relies on waiting for a standards body to agree on a solution, and then for the software vendors to implement it. Method 2 requires an excessive amount of probe devices, or uses an unrealistic topology for tunnel probe traffic to and from LSP ingress and egress points. Method 3 on the other hand can be implemented with a small number of servers (2 or 3) and use the existing operating software on the routers.

However, Method 3 has negative points. Fundamentally, the approach does not involve actually putting any probe traffic into the data plane. As a consequence of this, it relies on all the information being taken from the router being accurate, i.e. it assumes that the MIBs are correct. Also, the idea is dependent on seeing changes within the database. Therefore, if a new PE router is added, or load balancing links are added, it must be ensured that the correct number of new LSPs on the correct number of links are indicated by the monitoring tool, and hence go into the database. Otherwise, when these components suffer failure it could go unnoticed.

There are still many aspects of Method 3 to be investigated. The main part of this is determining how the collected data is interpreted, and what event sequences are significant and need to be flagged to network operators. The search algorithms which network operators will use to extract event history from the database also need to be defined. It has not yet been decided whether detecting certain changes within the LSPs mean that additional probing should be done as a result of this. Furthermore, the frequency that the routers should be polled for LSP and LDP session information needs to be established. Once these issues have been tackled, a concept demonstration could be set up with routers within the laboratory. Future, longer term work could also involve the inner label in the stack, and checking label assignment to customer address prefixes within a VPN.

6. References

1. Requirements for OAM in MPLS Networks, N. Harrison, December 2001
   draft-harrison-mpls-oam-req-01.txt
2. BGP/MPLS VPNs, E. Rosen, Y. Rekhter, Cisco Systems Inc, March 1999
   http://www.ietf.org/rfc/rfc2547.txt
3. OAM Functionality for MPLS Networks, N. Harrison, February 2001
   draft-harrison-mpls-oam-00.txt
4. Detecting Data Plane Liveliness in MPLS, K. Kompella, Juniper Networks, March 2002
   draft-ietf-mpls-lsp-ping-00.txt
5. Cisco MPLS Label Switching Router MIB, Cisco Systems
   http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120limit/120st/120st14/lsrmib.htm
6. SNMPv2 MIB for the TCP using SMIv2, K. McCloghrie, Cisco Systems, November 1996
   http://www.ietf.org/rfc/rfc2012.txt
7. Network Management System, MIBs, and MIB User Quick Reference, Cisco Systems, May 2001
   http://www.cisco.com/univercd/cc/td/doc/product/software/ios103/mib_doc/80516.htm#xtocid2791789