LDP Failure Detection and Recovery

Luyuan Fang, AT&T
Alia Atlas, Avici Systems
Fabio Chiussi, Invento Networks
Kireeti Kompella, Juniper Networks
George Swallow, Cisco Systems

ABSTRACT

In the last few years, multiprotocol label switching has been successfully deployed by the majority of service providers worldwide. The Label Distribution Protocol is used in many MPLS networks for distributing labels to establish the label switched paths. This article focuses on LDP failures, namely failures that may occur in LDP while the underlining Interior Gateway Protocol of choice, and the physical connections are operating normally. Since LDP and MPLS in general do not have intrinsic means for detecting these failures, new mechanisms have to be introduced to handle them. Furthermore, the IGP may remain unaware of the LDP failure, and continue to direct traffic to the faulty path. To resolve this situation, coupling between LDP and the routing protocol may have to be introduced. In this article, we discuss all aspects related to handling LDP failures: discovery, location, notification, recovery, and prevention.

INTRODUCTION

In the last few years, Multi-Protocol Label Switching (MPLS) has been successfully deployed in the backbone networks of most service providers worldwide. MPLS is based on dividing the forwarding space into forwarding equivalence classes (FECs), and establishing corresponding label switched paths (LSPs) from sources to destinations in the network. For each FEC, labels are exchanged between adjacent nodes in the network and used to forward packets along the LSPs. A new protocol called the Label Distribution Protocol (LDP) has been explicitly defined for distributing labels [1].

MPLS does not assume LDP as the only label distribution protocol, and other choices are possible. For example, the label distribution function can be piggybacked on existing protocols, such as Resource Reservation Protocol (RSVP) or Border Gateway Protocol (BGP), and extensions of these protocols have been defined for this purpose. Nevertheless, the use of LDP is widespread. The protocol is supported by most vendors of MPLS equipment and used by most service providers. As such, the issues related to the use of LDP are of great practical relevance, and possible failures due to the use of this protocol are a major reliability concern in MPLS networks.

The introduction of LDP as a separate protocol from the routing protocol brings a new set of issues related to the possible failures that may occur when LDP is used. In fact, LDP and MPLS in general do not have intrinsic means for detecting and locating failures in the LSPs, and new mechanisms have to be introduced. Furthermore, the LSPs are formed using information from the underlying Interior Gateway Protocol (IGP) to select the next hop. Yet the routing protocol may not be aware of a failure that may be happening in LDP, so traffic may continue to be forwarded to the location of the fault, unless coupling between LDP and IGP is created to make the routing protocol aware of the problem.

In this article our focus is on this type of failure to which we refer as *LDP failures*, to stress that their occurrence is closely tied to using LDP as the label distribution protocol of choice. Indeed, some of these failures would be simpler to handle if a different protocol were used for label distribution, such as RSVP with traffic engineering (RSVP-TE) or BGP, since MPLS could rely on the intrinsic failure recovery mechanisms of that protocol.

As for any type of failure, four aspects are relevant in LDP failures: detection, location, notification, and recovery. New mechanisms have to be provided in MPLS to detect that an LSP is faulty and locate the failure. To verify the status of an LSP, the mechanism that has been receiving the most attention in the industry is *LSP Ping* [2]. A second useful mechanism to understand if an LSP is up and running, and locate a failure in case of a problem, is *LSP Trace Route*. Mechanisms for automatic LSP validation and self-testing have also been recently proposed [3]. Other mechanisms also exist, and some have been used in proprietary implementations. Some of these mechanisms are inspired by similar mechanisms used in asynchronous transfer mode (ATM) to verify the status of a virtual circuit. For example, operations, administration, and maintenance (OA&M) packets may be periodically inserted

0163-6804/04/$20.00

When LDP fails while all lower-level protocols and physical connectivity are operational, IGP may remain unaware of the failure in LDP. In this situation, IGP continues to make routing decisions regardless of whether an LDP session is established or has failed, and whether or not the LDP labels are correct.

at the source of the LSP and used to verify the continuity of the LSP (connectivity verification). Those packets may be handled at the destination, which is then in charge of communicating the occurrence of a fault back to the source, or simply looped back, leaving all control to the source. The connectionless nature of LDP, with lack of clear correlation between forward and reverse paths, introduces a number of complications in these OA&M schemes (the details are outside the scope of this article).

Once a fault is detected, further procedures may have to be used to determine its location. The knowledge of the exact location is important, especially when local fast recovery procedures are used to redirect the traffic to a backup LSP. If the source is in charge of restoration, the redirection of the LSP can occur even before the fault is located, by choosing a disjoint path. In general, the knowledge of the fault location is necessary in order to repair the cause of the failure and restore normal operation.

After a fault is detected, it needs to be notified. Two aspects, in practice often closely intertwined, are relevant: the label switching router (LSR) in charge of recovery needs to be notified, and the routing protocol needs to be notified so that it can properly react to redirect traffic away from the fault. Once the failure has been notified, it has to be recovered. Rerouting the traffic may involve manual intervention or be fully automatic.

A final interesting aspect pertains to the means of preventing or at least reducing the likelihood of occurrence of LDP failures, in particular procedures to minimize operator errors and guarantee correctness of protocol implementation.

LDP Failure Types and Failure Scenarios

Types of LDP failures

As mentioned above, in this article we are primarily interested in failures that occur in LDP while the underlying routing protocol and physical layer are operating normally, since these are the failures that cannot be detected with existing mechanisms. Failures in LDP that occur simultaneously to failures in other layers, such as a link going down and interrupting operation, are not of primary interest in this article, since they can be handled by existing mechanisms for failure recovery.

When LDP fails while all lower-level protocols and physical connectivity are operational, IGP may remain unaware of the failure in LDP. In this situation, IGP continues to make routing decisions regardless of whether an LDP session is established or has failed, and whether or not the LDP labels are correct. MPLS forwarding continues to use the result of IGP calculation to forward the packets, and keeps directing traffic to the location of the fault. For example, labeled packets may continue to be forwarded through a broken LSP that has one or more LDP failures along its path, and dropped at the location of the fault (black-holed). Large packet loss may result if the situation persists for long periods of time. Manually shutting down the failed link is a drastic measure to resolve the situation and force an IGP reroute, but should be considered as the last option, and we are obviously interested in more intelligent ways to cope with the problem.

The possible types of LDP failures are summarized in Table 1.

LDP Failure Scenarios and Their Impact to the Networks

The impact of certain LDP failures depends on whether the network is configured to label switch all data traffic (it may have IP routing capability for control or management traffic only) or has dual capability of label switching and IP routing the traffic.

If the former case, any type of LDP failure listed above affects the packets since all data traffic depends exclusively on correct labels to be switched to the next hop.

The situation is different in the latter case. Let us first consider a common scenario where each router first tries to label switch traffic if it can find the right label binding for the outgoing interface, and IP routes the packets if the right label binding is not found. In this network, the impact of LDP failures depends on the number of labels carried by the packets and the type of failure. With packets carrying a *single* label, that is, no inner label or virtual private network (VPN) label, a router first attempts to label switch the packets. If there is no label binding available due to a failure (e.g., a protocol or operator error, or a race condition has occurred), the router simply reverts to IP routing, and packets are routed to their destination without being dropped. On the other hand, if the LDP failure is label corruption or a stale label, the router finds an outgoing (but incorrect) label binding, and the packet is label switched and eventually dropped or misrouted.

In this same network, with packets carrying two or more labels as is the case for VPN traffic, the router does not know the inner labels, so it depends on having correct outer labels for the packet to be forwarded. As a consequence, these labeled packets are dropped or misrouted with any kind of LDP failure, including those failures where no valid label binding is found, which do not affect packets with single labels.

Another common configuration of a network with dual capability is one where the network is capable of either LDP label switching or IP routing different types of traffic. This situation may occur, for example, in the transition period of converting a pure IP core to a pure MPLS core. A possible way to achieve this behavior is to configure LDP filtering on all routers to allow only certain types of traffic (e.g., VPN traffic within a specific address space) to be label switched, and all other traffic to be IP routed. In this case, the portion of traffic that is label switched is impacted by the LDP failures, while the IP traffic is not impacted and is routed normally. This is a simple way to reduce the total impact of LDP failures, while MPLS LDP technologies are being developed and made more mature and robust.

Type	Description	Notes and examples
LDP session failed	LDP session is not up or has been terminated incorrectly.	*Possible causes:* MPLS control plane going down on the interface, or LDP Hellos or KeepAlive not received from the LDP neighbors.
Operator's errors	Operator's configuration mistakes. Carriers can avoid or minimize this type of error through more automated provisioning and by strictly following well defined operation procedures.	Not a complicated technical issue, but nonetheless a major contributor of network outages. *Example:* when a new interface or a new link is added to the network, the operator mistakenly turns off LDP on certain interfaces or does not turn on LDP where label switching needs to be supported.
Stale labels	A label becomes stale when it is not updated and is no longer current (LDP session is up). Labeled packets may be dropped or misrouted due to the obsolete label.	Cannot be simply detected by checking if the LDP session is up or down.
Protocol errors	Software bugs in the LDP protocol implementation. May result in failures in either the control plane or forwarding plane.	May not manifest under simple network topologies and applications, yet may surface under more complicated network conditions (e.g., LDP protocol implementation that shows errors under stress of large scaling requirements).
Race conditions with IGP	LDP protocol functioning correctly, but the timing of when information becomes available from the IGP may cause failures.	*Example:* New link is added to the network. IGP computation starts to use the new interface once both ends of the link have exchanged the necessary link-state advertisements (LSAs). If traffic is routed to a new interface before LDP has finished session establishment or all label bindinginformation has been received, MPLS cannot forward the traffic, and discards the labeled traffic across the new interface.
Label corruption	Labels may become corrupted due to a hardware or software problem in the LSR. LDP session properly established, all LDP neighbor adjacencies appear normal, label bindings correctly exchanged between the LSRs and stored in the database.	LDP is *not* the direct cause of this type of error. Can be hard to detect since LDP sessions are up, and problem is hidden in the internals of the router. Correctness of packet delivery needs to be tested in the data plane. *Examples:* corruption of the label information base (LIB) or forwarding information base (FIB), or errors in the handling of the labels.

Table 1. *Types of LDP failures.*

In case of LDP failures occurring at the time an LSP is established, the exact location of where the labeled packets are dropped or misrouted may depend on whether *ordered control* or *independent control* is used to establish the LSPs [1, 4]. With independent control, when LDP has failed at a particular LSR along the LSP, the packets will be dropped (or misrouted) at the failing point. With ordered control, an LSP passing through a failing LSR is never established at the ingress, and the labeled packets may be dropped at the ingress of the LSP (clearly, if the failure occurs during the life of the LSP after it has been established, packet drop or misrouting occurs at the location of the failure).

Failure Detection and Isolation

As described earlier, there are many types of LDP failures. At some level, it doesn't matter what the cause of the failure is: detecting the error is paramount. However, from the point of view of diagnosing the error and fixing it, both immediately and also in the long term, it is vital to localize the exact cause and nature of the failure. Note that while reports from a node's control and data planes are very useful in diagnosing a problem, they are not of much use in either detecting or isolating failures.

When discussing failure detection, there are two key issues: correctness and speed. LSP ping tests end-to-end connectivity and focuses on correctness: minimizing false positives and false negatives by ensuring that there is enough information in each LSP ping packet, the processing of which implies a detection time on the order of seconds. However, ongoing work in the Internet Engineering Task Force (IETF) to combine bidirectional forwarding detection (BFD) with LSP ping [5] is expected to yield subsecond detection times. LSR self-test tests one hop of an LSP and is more amenable to fast processing, so it is conceivable that an LSR can achieve subsecond detection time for its LSPs.

In deciding among the various options (LSP ping, LSP ping+BFD, LSR self-test, or other mechanisms not discussed here), a service provider must first decide an acceptable outage time; this must then be broken down into detection time and recovery time. Acceptable outage times depend on the service type and service level agreements; for example, for IP connectivity, an outage time of several seconds may be acceptable, but for voice over IP, the outage time typically needs to be under 200 ms.

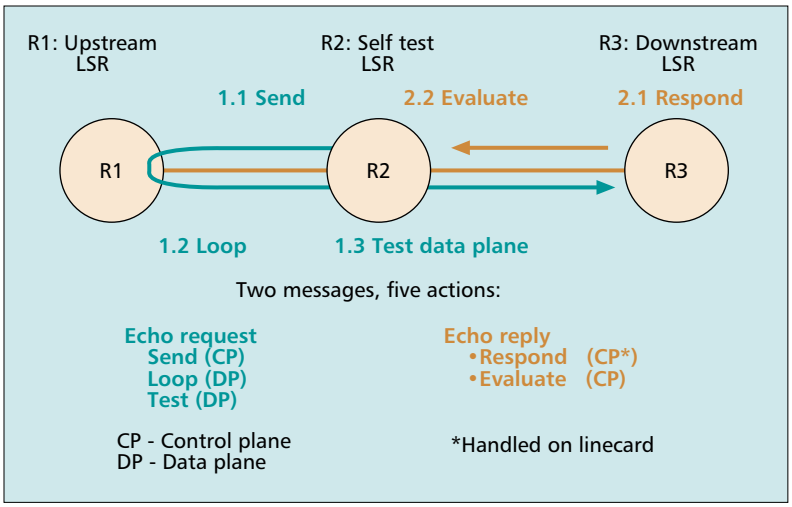

Figure 1. *LSR self test operation: 1.1) Self test LSR forms Echo Request and sends it upstream. 1.2) It is looped back though the data plane, 1.3) and intercepted at the downstream LSR. 2.1) The downstream LSR sends Echo Reply with results. 2.2) Self test LSR evaluates the result.*

LSP Ping and Traceroute

The tools *ping* and *traceroute* have long been used in IP networks to detect and localize failures in IP routing. Ping is used for connectivity verification, and traceroute is used to isolate (to within a node or two) the failure. The tools we describe here are the MPLS analogs of these tools. LSP ping and traceroute share a common paradigm with their IP cousins, in particular the notion that pings *must* follow the same data path and forwarding decisions as would a "normal" packet headed for the same destination. (It should be noted that LSP ping and traceroute are general mechanisms that work for many protocols that distribute labels, such as BGP and RSVP-TE, not just LDP.)

There is one major difference between IP ping and LSP ping: in IP ping, to know whether a given IP address is working, one directly pings the address under investigation, as the address is unique globally (or networkwide). In LSP ping [2], however, to know if a given LSP (for a given FEC) is working, one cannot ping the FEC directly; nor can one ping a label, as the label value only has significance local to a node. To complicate matters further, there may be many valid egress LSRs for a given FEC. There is also a minor difference: while IP pings go all the way to the designated destination, LSP pings must stop where the LSP terminates.

LSP ping addresses all of the above problems. An LSP ping has enough context (FEC) in the packet to ensure that a receiver can tell if the ping arrived at the correct destination, despite the fact that labels only have local significance. This context information also takes into account that LSPs may merge; that is, an LSR may have multiple incoming labels map to the same outgoing label. Furthermore, LSP ping has a solution of sorts for the multipath problem (which IP ping doesn't solve): if there are multiple parallel paths between the source and destination, how can one tell if all of them are working? Both label merging and multipath are common with LDP LSPs.

Having identified that there is an LDP failure, one would like to localize it to within a node or a link, and diagnose the cause of the failure. The tool used here is LSP traceroute, which again borrows from its IP cousin. LSP traceroute packets are sent with MPLS time-to-live (TTL) set to 1, 2, 3, …, until an egress LSR is reached, an error is reported, or no reply is received. An LSP traceroute packet is an LSP ping packet with more information; essentially, each intermediate node has enough information to verify that it is a valid transit node, and to check both the control plane and data plane of the previous hop. The node then returns information about valid next hops. Thus, one can localize the failure to one of two nodes, or the link between them, and generally one has some idea of the nature of the failure.

Details on the operation of LSP ping and traceroute, as well as packet formats, are given in [2] and omitted here for brevity.

Detection by the Router Itself: LSR Self Test

The first line of defense is built into LDP itself. LDP uses TCP, so a catastrophic error by a neighbor will be noticed as a TCP failure. In order to detect a problem limited to a neighbor's LDP process, KeepAlive messages are exchanged on a periodic basis between LDP neighbors when no other messages are sent on the LDP session. If no messages are received for a configured interval, the session is declared dead and restart mechanisms are invoked.

LSR self test provides a means for an LSR to verify both its control and data planes [3]. The control plane test uses LSP ping; the data plane test is a simple extension of LSP ping. It is designed to be a lightweight mechanism that can be used with fairly high frequency. While the data plane test relies on assistance from both its upstream and downstream neighbors, the downstream neighbor is required to do minimal processing, and the upstream neighbor only needs to do normal packet forwarding. LSR self test operation is illustrated in Fig. 1.

An LSR performing self test sends test packets through its data plane. Packets are looped through an upstream neighbor and intercepted by a downstream neighbor. The first function is achieved via a loopback label; the second by proper setting of TTL values. Neighboring routers exchange loopback labels via LDP.

For the control plane test, a normal LSP ping message is sent. This causes the downstream neighbor to invoke the detailed processing described above for traceroute.

To enable high-frequency testing, it is very desirable to minimize the processing required and keep it local to the LSR invoking the self-test. To minimize the processing required in the downstream LSR, a new LSP ping message, the *MPLS Data Plane Verification Request*, is defined. When an LSR receives this message, it simply replies with a message that reports the identity of the downstream router, the interface on which it received the packet, and the label stack on the packet. By using a different message type, the downstream LSR can quickly recognize that the message requires minimal processing.

For the data plane test, an LSR that invokes an MPLS Data Plane Verification Request on an interface prepends the packet with the incoming label stack and the upstream LSR loopback label, and sets the TTL values to ensure that the ping will be intercepted by a downstream neighbor (in the usual case, TTL = 3 for the loopback label and TTL = 2 for the next label in the stack). It then forwards the packet to the upstream LSR. When the upstream LSR receives the packet, it simply pops off the loopback label and forwards the packet back to the invoking LSR, which forwards the packet normally. The TTL expires at the next LSR, causing it to intercept and reply to the message. When the invoking LSR receives the reply, it compares the result with what its control plane expects.

Both LSP ping and LSR self test are near completion in the IETF.

LDP Failure Notification

Once the failure has been detected, the network elements in charge of its recovery need to be notified. In addition, the routing protocol is still unaware of the failure in LDP and needs to be notified so that traffic can be diverted away from the location of the fault. These two aspects may be closely combined or handled separately.

Notification through NMS

A first possibility is to have the network management system (NMS) solely in charge of the notification. In this case, the node that detects the failure simply notifies the NMS, and the NMS then takes care of orchestrating the recovery by informing all the affected nodes and, if necessary, initiating an IGP recomputation. Traps can be used for the purpose of triggering NMS notification.

Notification through IGP

If the recovery procedures need to be initiated without NMS intervention, or in parallel to NMS notification, additional mechanisms have to be implemented.

If the node that has detected the failure is not the node in charge of starting the restoration, a notification mechanism needs to be used. A convenient way to notify the entire network of the failure is to artificially increase the IGP cost associated with that link. The cause of the increase, namely the LDP failure, needs to be appropriately remembered in a database so that the cost can be restored to the original value once that LDP failure is rectified. With Open Shortest Path First (OSPF), the frequency of the Hellos can be increased in order to reduce the notification time. Of course, increasing the potential frequency of flooding may have other adverse consequences on the network, and the method requires careful finetuning.

The higher cost of the link triggers an IGP route recomputation, which in turn produces a redirection of the LSP to a different forwarding path. This is a simple way to notify the entire network of the failure and at the same time correct it by informing the routing protocol. A disadvantage is that both MPLS and IP traffic are affected, due to the routing change that has been triggered. If the failure is in LDP only, but IP is still operating normally, this is an overly conservative remedy. More important, this method is relatively slow, since it relies on IGP convergence and subsequent LDP convergence, and is therefore only appropriate when fast restoration is not required.

Source Notification

When fast restoration is in place, more rapid ways to notify the failure need to be used. With LSP ping, the source of the LSP is made aware of the fault. If a restoration LSP is prepared at the source with a disjoint path from the primary LSP, the source can immediately switch to the secondary LSP without waiting for the fault to be located.

In this scenario the disjoint paths are typically computed using some variations of constraint-based routing algorithms that are able to compute multiple paths. Hence, MPLS may not need to trigger an IGP recomputation since it may already have access to alternate paths. By leaving the IGP untouched, IP traffic may still be successfully routed through the location of the LDP fault, with only the MPLS traffic redirected, so the impact on network capacity is minimized.

Upstream Node Notification

With local fast restoration, a node that detects a failure may need to notify its upstream neighbor, so restoration can be started. Many options are available to achieve this task, and a complete review is beyond the space limits of this article. A simple, albeit draconian, solution to take care of the problem is to turn off the link to the upstream neighbor so that node can initiate the fast restoration procedure. More refined methods involve the insertion of appropriate messages either in-band using the reverse LSPs, or out-of-band. These mechanisms are often implementation-dependent and present considerable interoperability challenges.

LDP Failure Recovery

Given that an LDP failure has occurred and been detected, the next step is recovering from the fault. This can be done manually by an operator, or automatically by the protocol software. Some may view the latter warily, as automatic recovery from software errors may itself be subject to bugs, leading to network instability, but it is the ultimate objective for achieving a self-healing network. The trade-offs for manual vs. automatic recovery are outage time and system stability.

Manual Correction

Once an error has been diagnosed, the root cause has to be identified before corrective action can be taken. If the cause is a configuration error (e.g., an interface not configured for LDP), the corrective action is straightforward. If the error is in LDP protocol operation, resetting LDP might help. As mentioned above, the issue is also to make sure that LDP traffic is rerouted to avoid the link over which LDP failed, by forcing an IGP recomputation.

Once the failure has been detected, the network elements in charge for its recovery need to be notified. In addition, the routing protocol is still unaware of the failure in LDP and needs to be notified so traffic can be diverted away from the location of the fault.

There is no mechanism in LDP to indicate that all label bindings have been exchanged. If new label bindings are frequently being added and removed, it may not be possible to determine whether the process has been completed.

Automatic Correction

If detection and fault isolation is automated, then correction should also be automated. That way, no operator intervention (which can be expensive and slow) is required to rectify the problem. Note that automatic correction need not actually fix the root cause; it need only get traffic flowing again, even if it is on a suboptimal path. Again, the mechanisms include raising IGP metrics to move traffic off the broken interface, or disabling the broken interface. Another avenue is protocol changes whereby a node can inform another that LDP on a specific link is faulty, and the two nodes might cooperatively fix or alleviate the problem.

LDP Convergence

The failure or appearance of one or more links in the network causes LDP to have to change its forwarding paths. As with IP, the change to the new forwarding paths is done independently on each router in the network. The convergence of LDP can be considered for an entire network or for a single router. In this article we consider the *LDP convergence time* to be the duration of the period from when LDP traffic entering a particular router is lost due to the topology change until that LDP traffic is no longer lost. The traffic loss will stop when traffic to the affected FECs is redirected to a new next hop on which the necessary label bindings are known and the appropriate forwarding state has been installed.

Before traffic for a particular FEC can be forwarded without loss using a forwarding state created by LDP, three steps must be completed. First, new primary next hop(s) for that FEC must be determined based on IGP recomputation. Second, LDP must receive the necessary label binding from the primary neighbor(s). Third, LDP must install an out-segment specifying the received label, connect the appropriate in-segment to that out-segment, and connect the appropriate IP prefixes to that out-segment.

Using best practices to minimize the IGP SPF computation time also improves LDP traffic convergence time. In addition, if LDP is using both downstream unsolicited and liberal label retention mode, LDP will generally have the necessary label binding already locally stored at the time of failure; if LDP does not have a label binding for the primary next hop for the FEC, the LSR can act as an LDP egress for that FEC so that some traffic can be sent through. As discussed above, this has the disadvantage that only traffic with a single label (and with destinations known by the router) can be forwarded.

If fast reroute is used, the LDP convergence time can be significantly reduced by precomputing and preinstalling appropriate alternate next hops for each FEC. LDP local protection [6, 7] will limit the local traffic loss to the time to switch from the broken primary next hop to the alternate next hop. Such a switch has been demonstrated to occur within tens of milliseconds.

When a new interface is configured, the IGP will start using it and reporting it as a primary next hop to LDP. If LDP has not yet established a session to the interface's neighbor and learned all necessary label bindings for the FECs the IGP directs across that new interface, when the new interface is enabled and used LDP traffic will be lost.

To avoid this LDP traffic loss, the IGP should not use the new link until LDP has learned the necessary label bindings. However, if the new link provides the only path to a destination, the IGP should be able to use the new link, at least for IP traffic to that destination. To allow such use, the new link must be advertised into the IGP for use even before the LDP label bindings exchange is completed. This can be accomplished by setting the cost on all links from that router where the LDP label binding exchange has not completed to the maximum IGP cost. Once that exchange is complete, the links can be advertised with their configured cost.

There is no mechanism in LDP to indicate that all label bindings have been exchanged. If new label bindings are frequently added and removed, it may not be possible to determine whether the process has been completed. There are three methods for making the determination that an LDP session has exchanged all label bindings:

- Assume that the LDP label binding exchange is completed a configurable interval after the LDP session was established. Although simple, this does not provide a guarantee of completion. Since there is no requirement to rapidly bring a new link into use, a conservative value for the interval can be used. This method works with a talkative LDP session, where label bindings continue to be frequently added and removed.
- Configure the number of label bindings to learn. This number may be different on different interfaces on the same router. A network failure that reduces the number of label bindings learned means the required number is not reached.
- Use an implicit or explicit "end of LIB" marker. An implicit marker might be receiving two KeepAlive messages without any label bindings in between. A new explicit marker could be sent by LDP when the LSR believes it has sent all label bindings. This does not handle talkative LDP sessions; the LSR would need to determine when sufficient label bindings have been sent, while additional ones continue to be sent.

LDP Failure Prevention

The prevention of LDP failures means to eliminate or minimize the causes of LDP failures described above. This includes efforts by providers on the operation side and vendors on the implementation side.

A first issue is automation of network provisioning. Many operator mistakes can be prevented by eliminating or reducing human intervention. This requires service providers to develop reliable and sophisticated fully automated operation support systems (OSSs) to support all backbone and customer provisioning needs. The development of such systems is very costly, but is key to reducing human errors caused by manual provisioning. In addition, operators must have a set of well defined operation procedures

and well trained staff who understand the process and are able to follow these procedures without mistakes. As the complexity of their network increases, providers have to evolve their operation practices to higher standards than ever before.

For LDP, it is necessary to manually configure LDP on each interface. This is an "opt-in" configuration model. In this model, to reduce the impact of the operator's error in not configuring LDP on the interface, it is necessary to detect that LDP should have been configured and have the IGP take the appropriate action of costing out the new interface. A better way to reduce the likelihood of an operator's error occurring is an "opt-out" configuration model. In this model, a feature such as LDP and IGP coordination is enabled on the LSR; this feature applies to every interface on the LSR unless the operator explicitly configures an interface to not participate. Thus, if the operator forgets to configure anything about LDP, it will be detected due to the lack of an LDP session and the coordination between LDP and the IGP being active.

Another important aspect is for vendors to minimize protocol or equipment implementation errors by using validation methods for their implementations and providing robust reliability features in their routers.

CONCLUSIONS

Because of the popularity of LDP as a label distribution protocol, LDP failures have great practical relevance and constitute a major reliability issue in MPLS networks. Means to combat them are receiving great attention in the industry and in the standards bodies.

In this article we have addressed all the relevant aspects of LDP failures. Our overall objective is to bring attention to this new type of failure and describe the various technical advances in handling it. This is important in order to improve the availability/reliability of MPLS networks, and avoid or minimize service interruptions.

REFERENCES

[1] L. Andersson *et al.*, "LDP Specification," IETF RFC 3036, Jan. 2001.

[2] K. Kompella and G. Swallow, "Detecting MPLS Data Plane Failures," draft-ietf-mpls-lsp-ping-06.txt, July 2004, work in progress.

[3] G. Swallow, K. Kompella, and D. Tappan, "Label Switching Router Self-Test," draft-ietf-mpls-lsr-self-test-02.txt, Feb. 2004.

[4] B. Davie and Y. Rekhter, *MPLS Technology and Applications*, Academic Press, 2000.

[5] R. Aggarwal *et al.*, "BFD for MPLS LSPs," draft-ietf-bfd-mpls-00.txt, July 2004.

[6] A. Atlas *et al.*, "Loop-Free Alternates for IP/LDP Local Protection," draft-atlas-ip-local-protect-loopfree-00.txt, June 2004.

[7] A. Atlas *et al.*, "U-turn Alternates for IP/LDP Local Protection," draft-atlas-ip-local-protect-uturn-00.txt, June 2004.

BIOGRAPHIES

LUYUAN FANG (luyuanfang@att.com) is a principal technical staff member in the IP and MPLS Network Architecture design team at AT&T Laboratories. She is a lead architect for AT&T MPLS IP VPN network design and deployment, and plays a key role in the technical realization of MPLS IP VPN service features in the AT&T IP network. She is also involved in evaluating forward-looking network technologies and equipment capabilities to support emerging services. Prior to her current position, she worked on the design and deployment of AT&T International Frame Replay Services. Prior to joining AT&T, she held various research positions in artificial intelligence and data networking in Telstra, Nortel, and Racal Datacom. Her current interests are MPLS, BGP/MPLS VPNs, IP routing in MPLS networks, VPLS, traffic engineering, and generalized MPLS (GMPLS). She is active in the IETF, where she has co-authored several Internet drafts and RFCs in the MPLS, L3VPN, Traffic Engineering, and Inter-Domain Routing Working Groups. She received her B.S. in physics from Jiangxi University in China, her M.S. in computer science from Brigham Young University, and her Ph.D. in computer science at Flinders University of South Australia.

ALIA ATLAS is a consulting software engineer for Avici Systems. Her current interests include MPLS, QoS, and protection mechanisms for IP and MPLS. She is an active contributor in the IETF. Prior to joining Avici Systems, she was a network scientist at GTE-BBN Technologies in their Department of Internetwork Research. At BBN she focused on router architectures for QoS and traffic engineering. She holds patents in the areas of real-time scheduling and protection mechanisms for IP networks. She holds a B.S.E.E. from MIT, and Master's and Ph.D. degrees in computer science from Boston University, where she focused on real-time systems.

FABIO M. CHIUSSI is founder and chief technical officer of Invento Networks. From 1993 to 2003 he was with Bell Laboratories, Lucent Technologies, where he was director, Data Networking and Wireless Systems, and a Bell Labs Fellow. During those years, he led the architectural design of three generations of the Lucent ATLANTA chipset (now a product of Agere Systems), an industry-leading silicon solution for ATM and IP switching and port processing; within the ATLANTA project, he also held various development responsibilities, including leading the development of the switch fabric devices for the third generation of the chipset. From 2000 to 2003 he and his department worked on the development of several MPLS-centric switching systems for applications ranging from Metro Gigabit Ethernet services to wireless access, and offering various flavors of VPNs and other advanced IP services. He has been conducting fundamental research in the area of scalable switch architectures, traffic management and scheduling, protocols and architectures for wireless land networks, congestion control, and VLSI design. He has written more than 80 technical papers, is active in the IETF where he has co-authored several drafts, and holds 15 patents, with 15 more pending. He received a Ph.D. in electrical engineering from Stanford University in 1993.

KIREETI KOMPELLA is a distinguished engineer at Juniper Networks. His current interests are all aspects of MPLS, including traffic engineering, GMPLS, and MPLS applications such as VPNs. He is active in the IETF where he is a co-chair of the CCAMP Working Group and the author of several Internet drafts and RFCs in the areas of CCAMP, IS-IS, L2VPN, MPLS, OSPF, and TE. He specializes in layer 2 VPNs, metro Ethernet, and virtual private LAN service. Previously, he worked in the area of file systems at Network Appliance and SGI, and earlier in the area of security and cryptography. He received his B.S. in electrical engineering and M.S. in computer science from the Indian Institute of Technology, Kanpur; and his Ph.D. in computer science from the University of Southern California.

GEORGE SWALLOW is a distinguished engineer at Cisco Systems where he is a member of the architecture team for label switching. He defined Cisco's architecture for applying MPLS to the problem of traffic engineering and fast reroute. Recently, he has been involved in pseudowire support of ATM and frame relay, and developing means of monitoring and diagnosing MPLS networks and MPLS-based network applications. He has also been involved in generalizing MPLS for optical and other technologies. Prior to Cisco, he was employed by BBN. There, he was involved in the design, deployment, and analysis of over 50 operational networks, including the Arpanet. This work involved extensive statistical measurement and analysis to investigate both network and protocol behavior. He was also involved in the design of packet and ATM switches. While at BBN, he held a number of positions ranging from senior network analyst to director of network engineering. He has been participating in the design and standardization of Internet and ATM standards since 1991. He is Co-Chair of the IETF Working Group on MPLS.

Because of the popularity of LDP as a label distribution protocol, LDP failures have great practical relevance and constitute a major reliability issue in MPLS networks. Means to combat them are receiving great attention in the industry and in the standard bodies.