spring   R. Geib, Ed.
Internet-Draft   Deutsche Telekom
Intended status: Informational   C. Filsfils
Expires: January 3, 2015   C. Pignataro
                                                              N. Kumar
                                                     Cisco Systems, Inc.
                                                          July 2, 2014

**Use case for a scalable and topology aware MPLS data plane monitoring
system**
draft-geib-spring-oam-usecase-02

Abstract

   This document describes features and a use case of a path monitoring
   system. Segment based routing enables a scalable and simple method
   to monitor data plane liveliness of the complete set of paths
   belonging to a single domain. Compared with legacy MPLS ping and
   path trace, MPLS topology awareness reduces management and control
   plane involvement of OAM measurements while enabling new OAM
   features.

Status of this Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF). Note that other groups may also distribute
   working documents as Internet-Drafts. The list of current Internet-
   Drafts is at http://datatracker.ietf.org/drafts/current/.

   Internet-Drafts are draft documents valid for a maximum of six months
   and may be updated, replaced, or obsoleted by other documents at any
   time. It is inappropriate to use Internet-Drafts as reference
   material or to cite them other than as "work in progress."

   This Internet-Draft will expire on January 3, 2015.

Copyright Notice

Table of Contents

1. Introduction

It is essential for a network operator to monitor all the forwarding
paths observed by the transported user packets. The monitoring flow
must be forwarded in dataplane in a similar way as user packets.
Problem localization is required.

This document describes a solution to this problem statement and
illustrates it with use-cases.

The solution is described for a single IGP MPLS domain.

The solution applies to monitoring of LDP LSP's as well as to
monitoring of Segment Routed LSP's. Segment Routing simplifies the
solution by the use of IGP-based signalled segments as specified by
[ID.sr-isis]. Thus a centralised monitoring unit is MPLS topology
aware in a Segment Routed domain and this topology awareness is used
for OAM purposes. The MPLS path monitoring system described by this
document can be realised with pre-Segment based Routing (SR)
technology. Making such a monitoring system aware of a domains
complete MPLS topology requires e.g. management plane access. To
avoid the use of stale MPLS label information, IGP must be monitored
and MPLS topology must be timely aligned with IGP topology.
Obviously, enhancing IGPs to exchange of MPLS topology information
significantly simplifies and stabilises such an MPLS path monitoring
system.

This document adopts the terminology and framework described in
[ID.sr-archi]. It further adopts the editorial simplification
explained in section 1.2 of the segment routing use-cases
[ID.sr-use].

The proposed solution offers several benefits for network monitoring.
A single centralized monitoring device is able to monitor the
complete set of a domains forwarding paths. OAM packets never leave
data plane. Legacy path trace is still required. In addition to
Segment Routing related IGP extensions, also RFC 4379 features should
be extended to support detection of SR routed paths. They further
should be enhanced to support all deployed IP/MPLS entropy options.
In an IPv6 domain, a MPLS like tree trace functionality is desirable.

Faults can be localized:

o by IGP LSA analysis.

o by correlation between different probes.

o by MPLS traceroute and adapted ping messages.

The proposed solution requires topology awareness as well as a
suitable security architecture. Topology awareness is an essential
part of link state IGPs. Adding MPLS topology awareness to an IGP
speaking device hence enables a simple and scaleable data plane
monitoring mechanism.

MPLS OAM offers flexible features to recognise an execute data paths
of an MPLS domain. By utilsing the ECMP related tool set of RFC 4379
[RFC4379], a segment based routing LSP monitoring system may:

o easily detect ECMP functionality and properties of paths at data
   level.

o construct monitoring packets executing desired paths also if ECMP
   is present.

o limit the MPLS label stack of an OAM packet to a minmum of 3
   labels.

MPLS OAM supports detection and execution of ECMP paths quite smart.
This document is foscused on MPLS path monitoring.

Alternatively, any path may be executed by building suitable label
stacks. This allows path execution without ECMP awareness.

The MPLS path monitoring system may be a specialised system residing
at a single interface of the domain to be monitored. As long as
measurement packets return to this or another interface to a
specialised OAM system, the MPLS monitoring system is the single
entity pushing monitoring packet label stacks. Concerns about router
label stack pushing capabilities don't apply in this case.

First drafts discussing requirements, extensions of RFC4379 and
possible solutions to allow SR usage as described by this document
are at hand, see [ID.sr-4379ext] and [ID.sr-oam_detect].

2. **An MPLS topology aware path monitoring system**

An MPLS path monitoring system (PMS) which is able to learn the IGP
LSDB (including the SID's) is able to build a measurement packet
which executes every arbitrary chain of paths. A node connected to
an SR domain is MPLS topology aware (the node knows all related IP
adresses, MPLS SIDs and labels).

Let us describe how the PMS can check the liveliness of the MPLS

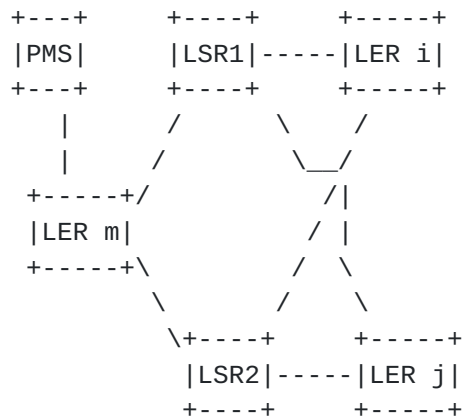transport path between LER i and LER j and then monitor it.

The PMS may do so by sending packets carrying the following MPLS
label stack infomation:

o Top Label: a path from PMS to LER i This is expressed as Node SID
   of LER i.

o Next Label: the path that needs to be monitored from LER i to LER
   j. If this path is a single physical interface (or a bundle of
   connected interfaces), it can be expressed by the related AdjSID.
   If the shortest path from LER i to LER j is supposed to be
   monitored, the Node-SID (LER j) can be used. Another option is to
   insert a list of segments expressing the desired path (hop by hop
   as an extreme case). If LER i pushes a stack of Labels based on a
   SR policy decision and this stack of LSPs is to be monitored, the
   PMS needs an interface to collect the information enabling it to
   address this SR created path.

o Next Label or address: the path back to the PMS. Likely, no
   further segment/label is required here. Indeed, once the packet
   reaches LER j, the 'steering' part of the solution is done and the
   probe just needs to return to the PMS. This is best achieved by
   popping the MPLS stack and revealing a probe packet with PMS as
   destination address (note that in this case, the source and
   destination addresses could be the same). If an IP address is
   applied, no SID/label has to be assigned to the PMS (if it is a
   host/server residing in an IP subnet outside the MPLS domain).

Note: if the PMS is an IP host not connected to the MPLS domain, the
PMS can send its probe with the list of SIDs/Labels onto a suitable
tunnel provding an MPLS access to a router which is part of the
monitored MPLS domain.

3. SR based OAM use case illustration

3.1. Use-case 1 - LSP dataplane liveliness detection and monitoring

```
          +---+      +----+      +-----+
          |PMS|      |LSR1|-----|LER i|
          +---+      +----+      +-----+
            |      /        \     /
            |     /          \__/
         +-----+/            /|
         |LER m|            / |
         +-----+\          /  \
               \         /     \
                \+----+      +-----+
                 |LSR2|-----|LER j|
                 +----+      +-----+
```

   Example of a PMS based LSP dataplane liveness detection and
   monitoring

                              Figure 1

   For the sake of simplicity, let's assume that all the nodes are
   configured with the same SRGB [ID.sr-archi]. as described by section
   1.2 of [ID.sr-use].

   Let's assign the following Node SIDs to the nodes of the figure: PMS
   = 10, LER i = 20, LER j = 30.

   The aim is to check liveliness of the path LER i to LER j and to
   monitor availability of that path afterwards. The PMS does this by
   creating a measurement packet with the following label stack (top to
   bottom): 20 - 30 - 10.

   LER m forwards the packet received from the PMS to LSR1. Assuming
   Pen-ultimate Hop Popping to be deployed, LSR1 pops the top label and
   forwards the packet to LER i. There the top label has a value 30 and
   LER i forwards it to LER j. This will be done transmitting the
   packet via LSR1 or LSR2. The LSR will again pop the top label. LER
   j will forward the packet now carrying the top label 10 to the PMS
   (and it will pass a LSR and LER m).

   A few observations on the example given in figure 1:

   o   The path PMS to LER i must be available. This path must be
       detectable, but it is usually sufficient to apply an SPF based
       path.

o If ECMP is deployed, it may be desired to measure along both
   possible paths, a packet may use between LER i and LER j. To do
   so, in a first step the PMS sends MPLS OAM packets to execute a so
   called tree trace between LER i and LER j and stores the IP
   destination addresses required to execute each detected path.
   This method of dealing with load balancing paths requires the
   smallest label stacks if long term monitoring of paths is applied
   after the tree trace completion.

o The path LER j to PMS to must be be available. This path must be
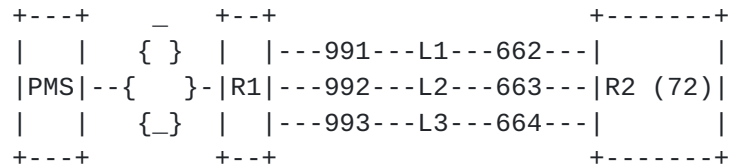   detectable, but it is usually sufficient to apply an SPF based
   path.

Once the MPLS paths (Node SIDs) and the required IP address
information has been detected, the LER i to LER j can be monitored by
the PMS. Monitoring doesn't require MPLS OAM functionality, it is
purely based on forwarding. To ensure reliable results, the PMS
should be aware of any changes in IGP or MPLS topology. Further
changes in ECMP functionality at LER i will impact results. Either
the PMS should be notified of such changes or they should be limited
to planned maintenance. After a topology change, MPLS OAM will be
useful to detect the impact of the change.

Determining a path to be executed prior to a measurement may also be
done by setting up a label including all node SIDs along that path
(if LER1 has Node SID 40 in the example and it should be passed
between LER i and LER j, the label stack is 20 - 40 - 30 - 10). The
advantage of this method is, that it does not involve MPLS OAM
functionality and it is independent of ECMP functionalities. The
method still is able to monitor all link combinations of all paths of
an MPLS domain. If correct forwarding along the desired paths has to
be checked, RFC4739 functionality should be applied also in this
case.

Obviously, the PMS is able to check and monitor data plane liveliness
of all LSPs in the domain. The PMS may be a router, but could also
be dedicated monitoring system. If measurement system reliability is
an issue, more than a single PMS may be connected to the MPLS domain.

Monitoring an MPLS domain by a PMS based on SR offers the option of
monitoring complete MPLS domains with little effort and very
excellent scalability. Data plane failure detection by circulating
monitoring packets can be executed at any time. The PMS further
executes MPLS OAM functions everywhere in the MPLS domain. It does
not require access to LSR/LER management interfaces to do so. MPLS
traceroutes as specified above should be executed only during off
peak times (and then with limited parallel MPLS ping/trace load).

3.2. Use-case 2 - Monitoring a remote bundle

```
        +---+    _    +--+                      +-------+
        |   |   { }   |  |---991---L1---662---|       |
        |PMS|--{   }-|R1|---992---L2---663---|R2 (72)|
        |   |   {_}   |  |---993---L3---664---|       |
        +---+         +--+                      +-------+
```

SR based probing of all the links of a remote bundle

                            Figure 2

R1 adresses Lx by the Adjacency SID 99x, while R2 adresses Lx by the
Adjacency SID 66(x+1).

In the above figure, the PMS needs to assess the dataplane
availability of all the links within a remote bundle connected to
routers R1 and R2.

The monitoring system retrieves the SID/Label information from the
IGP LSDB and appends the following segment list/label stack: {72,
662, 992, 664} on its IP probe (whose source and destination
addresses are the address of the PMS).

MS sends the probe to its connected router. If the connected router
is not SR compliant, a tunneling technique can be used to tunnel the
probe and its MPLS stack to the first SR router. The MPLS/SR domain
then forwards the probe to R2 (72 is the Node SID of R2). R2
forwards the probe to R1 over link L1 (Adjacency SID 662). R1
forwards the probe to R2 over link L2 (Adjacency SID 992). R2
forwards the probe to R1 over link L3 (Adjacency SID 664). R1 then
forwards the IP probe to PMS as per classic IP forwarding.

3.3. Use-Case 3 - Fault localization

In the previous example, a uni-directional fault on the middle link
from R1 to R2 would be localized by sending the following two probes
with respective segment lists:

o 72, 662, 992, 664

o 72, 663, 992, 664

The first probe would fail while the second would succeed.
Correlation of the measurements reveals that the only difference is

using the Adjacency SID 662 of the middle link from R1 to R2 in the
non successful measurement. Assuming the second probe has been
routed correctly, the fault must have been occurring in R2 which
didn't forward the packet to the interface identified by its
Adjacency SID 662.

[4](#). Failure Notification from PMS to LERi

PMS on detecting any failure in the path liveliness MAY use any out-
of-band mechanism to signal te\he failure to LERi. This document
does not not propose any specific mechanism and Operators can choose
any existing or new approach.

Alternately, the Operator may log the failure in local monitoring
system and take necessary action by manual intervention.

[5](#). Applying SR to monitor LDP paths

A SR based PMS connected to a MPLS domain consisting of LER and LSR
supporting SR and LDP in parrallel in all nodes may use SR paths to
transmit packets to and from start and end points of LDP paths to be
monitored. In the above example, the label stack top to bottom may
be as follows, when sent by the PMS:

o Top: SR based Node-SID of LER i at LER m.

o Next: LDP label identifying the path to LER j at LER i.

o Bottom: SR based Node-SID identifying the path to the PMS at LER j

While the mixed operation shown here still requires the PMS to be
aware of the LER LDP-MPLS topology, the PMS may learn the SR MPLS
topology by IGP and use this information.

[6](#). PMS monitoring of different Segment ID types

MPLS SR topology awareness should allow the SID to monitor liveliness
of most types of SIDs (this may not be recommendable if a SID
identifies an inter domain interface).

To match control plane information with data plane information,
[RFC4379](#) should be enhanced to allow collection of data relevant to
check all relevant types of Segment IDs.

7. Connectivity Verification using PMS

   While the PMS based use cases explained in Section 3 is sufficient to
   provide Continuity check between LER i and LER j, it may not help
   perform connectivity verification. So in some cases like data plane
   programming corruption, it is possible that a transit node between
   LER i and LER j erroneously remove the top segment ID and forward to
   PMS based on bottom segment ID leading to falsified path liveliness
   to PMS.

   There are various method to perform basic connectivity verification
   like intermittely setting the TTL to 1 in bottom label so LER j
   selectively perform connectivity verification. A detailed
   explanation will be added in later version.

8. Extensions of related standards

   RFC4379 functions should be extended to support Flow- and Entropy
   Label based ECMP. Further, an RFC4379 like functionality may be
   desirable for IPv6 networks.

9. IANA Considerations

   This memo includes no request to IANA.

10. Security Considerations

   As mentioned in the introduction, a PMS monitoring packet should
   never leave the domain where it originated. It therefore should
   never use stale MPLS or IGP routing information. Further, asigning
   different label ranges for different purposes may be useful. A well
   known global service level range may be excluded for utilisation
   within PMS measurement packets. These ideas shoulddn't start a
   discussion. They rather should point out, that such a discussion is
   required when SR based OAM mechanisms like a SR are standardised.

11. Acknowledgement

   The authors would like to thank Nobo Akiya for his cotribution.

12. References

12.1. Normative References

[RFC4379]   Kompella, K. and G. Swallow, "Detecting Multi-Protocol
            Label Switched (MPLS) Data Plane Failures", RFC 4379,
            February 2006.

12.2. Informative References

[ID.sr-4379ext]
            IETF, "Label Switched Path (LSP) Ping/Trace for Segment
            Routing Networks Using MPLS Dataplane", IETF, http://
            datatracker.ietf.org/doc/draft-kumar-mpls-spring-lsp-
            ping/, 2013.

[ID.sr-archi]
            IETF, "Segment Routing Architecture", IETF, https://
            datatracker.ietf.org/doc/
            draft-filsfils-rtgwg-segment-routing/, 2013.

[ID.sr-isis]
            IETF, "IS-IS Extensions for Segment Routing", IETF, http:
            //datatracker.ietf.org/doc/
            draft-previdi-isis-segment-routing-extensions/, 2013.

[ID.sr-oam_detect]
            IETF, "Detecting Multi-Protocol Label Switching (MPLS)
            Data Plane Failures in Source Routed LSPs", IETF, http:/
            /datatracker.ietf.org/doc/draft-kini-spring-mpls-lsp-
            ping/, 2013.

[ID.sr-use]
            IETF, "Segment Routing Use Cases", IETF, http://
            datatracker.ietf.org/doc/
            draft-filsfils-rtgwg-segment-routing-use-cases/, 2013.

Authors' Addresses

   Ruediger Geib (editor)
   Deutsche Telekom
   Heinrich Hertz Str. 3-7
   Darmstadt,   64295
   Germany

   Phone: +49 6151 5812747
   Email: Ruediger.Geib@telekom.de

Clarence Filsfils
Cisco Systems, Inc.
Brussels,
Belgium

Phone:
Email: cfilsfil@cisco.com

Carlos Pignataro
Cisco Systems, Inc.
7200 Kit Creek Road
Research Triangle Park, NC 27709-4987
US

Email: cpignata@cisco.com

Nagendra Kumar
Cisco Systems, Inc.
7200 Kit Creek Road
Research Triangle Park, NC 27709
US

Email: naikumar@cisco.com