A Connectivity Monitoring Metric for IPPM
draft-geib-ippm-connectivity-monitoring-01

Abstract

   Segment Routed measurement packets can be sent along pre-determined
   paths. This allows new kinds of measurements. Connectivity
   monitoring allows to supervise the state of a connection or a
   (sub)path from one or a few central monitoring systems. This
   document specifies a suitable type-P connectivity monitoring metric.

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 https://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 5, 2020.

Table of Contents

1. Introduction

   Segment Routing enables sending measurement packets along pre-
   determined segment routed paths [RFC8402]. A segment routed path may
   consist of pre-determined sub paths down to specific router-
   interfaces. It may also consist of sub paths spanning multiple
   routers, given that all segments to address a desired path are
   available and known at the SR domain edge interface.

   A Path Monitoring System or PMS (see [RFC8403]) is a dedicated rather
   central Segment Routing domain monitoring device (as compared to a
   distributed monitoring approach based on router data and functions
   only). Monitoring individual sub-paths or point-to-point connections
   is executed for different purposes. IGP routing exchanges hello
   messages between neighbors to keep alive routing and swiftly adapt to
   changes. Network Operators may be interested in monitoring
   connectivity and lasting congestion of interfaces or sub-paths at a
   higher timescale,e.g., on the order of seconds. This is still
   significantly faster than interface monitoring based on router
   information, which may be collected on a minute timescale to reduce
   the CPU load caused by monitoring.

   The IPPM architecture was a first step to that direction [RFC2330].
   Commodity IPPM solutions require dedicated measurement systems, a

large number of measurement agents and synchronised clocks.
Monitoring a domain from edge to edge by commodity IPPM solutions
helps to increase scalability of the monitoring system, but
localising a source cause of a detected change in network behaviour
then may require network tomography methods.

The IPPM Metrics for Measuring Connectivity offer generic
connectivity metrics [RFC2678]. These metrics allow to measure
connectivity between end nodes without making any assumption on the
paths between them. The metric and the type-p packet specified by
this document follow a different approach: they are designed to
monitor connectivity of a specific single link or a path segment.
The underlying definition of connectivity is partially the same, a
packet not reaching a destination indicates a loss of connectivity.
An IGP re-route may indicate a loss of a link, while it might not
cause loss of connectivity beween end systems. The metric specified
here is able to detect the loss of a link, if the change in end-to-
end delay along a new route are differing from that of the original
path.

A Segment Routing PMS which is part of an SR domain is IGP topology
aware, covering the IP and (if present) the MPLS layer topology
[RFC8402]. This allows to design a PMS which can steer packets along
arbitrary pre-determined concatenated sub-paths, identified by
suitable segments. Combining the SR measurement path configuration
with a priori network tomography assumptions and methods allows for
localisation of detected changes. The latter requires setting up
multiple measurement paths which share sub-paths following the
constraints derived from network tomography, and a suitable
evaluation of measurement results.

This document specifies a type-p metric determining properties of an
SR path which allows to monitor connectivity and congestion of
interfaces and further allows to locate the path or interface which
caused a change in the reported type-p metric. This document is
focussed on the MPLS layer, but the methodolgy may be applied within
SR domains or MPLS domains in general.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in RFC 2119 [RFC2119].

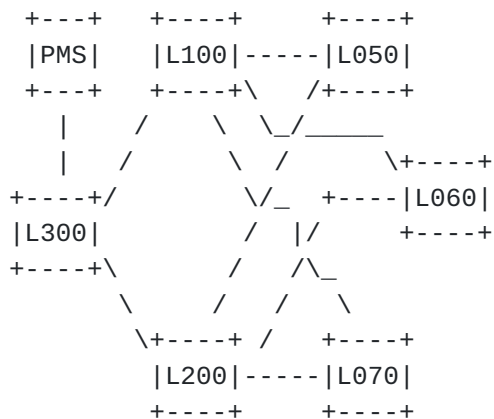## 2. A brief segment routing connectivity monitoring framework

The Segment Routing IGP topology information consists of the IP and
(if present) the MPLS layer topology. The minimum SR topology
information consists of Node-Segment-Identifiers (Node-SID),
identifying an SR router. The IGP exchange of Adjacency-SIDs [I-
D.draft-ietf-isis-segment-routing-extensions], which identify local
interfaces to adjacent nodes, is optional. It is RECOMMENDED to
distribute Adj-SIDs in a domain operating a PMS to monitor
connectivity as specified below. If Adj-SIDs aren't availbale,
[RFC8029] provides methods how to steer packets along desired paths
by the proper choice of an MPLS Echo-request IP-destination address.
A detailed description of [RFC8029] methods as a replacement of Adj-
SIDs is out of scope of this document.

A round trip measurement between two adjacent nodes is a simple
method to monitor connectivity of a connecting link. If multiple
links are operational between two adjacent nodes and only a single
one fails, a single plain round trip measurement may fail to identify
which link has failed. A round trip measurement also fails to
identify which inteface is congested, even if only a single link
connects two adjacent nodes.

Segment Routing enables the set-up of extended measurement loops.
Several different measurement loops can be set up. If these form a
partial overlay, any change in the network properties impacts more
than a single loops round trip time (or causes drops of packets of
more than one loop). Randomly chosen loop paths including the
interfaces or paths to be monitored may fail to produce unique result
patterns. The approach picked here uses specified measurement loop
and path overlay design. A centralised monitoring approach benefits
from keeping the number of required measurement loops low. This
improves scalability by minimising the number of measurement loops.
This also keeps the number of required packets and results to be
evaluated and correlated low.

An additional property of the measurement path set-up specified below
is that it allows to estimate the packet round trip and the one way
delay of a monitored link (or path). The delay along a single link
is not perfectly symmetric. Packet processing causes small delay
differences per interface and direction. These cause an error, which
can't be quantified or removed by the specified method. Quantifying
this error requires a different measurement set-up. As this will
introduce additional measurements loops, packets and evaluations, the
cost in terms of reduced scalability is not felt to be worth the
benefit in measurement accuracy. IPPM however honors precision more
than accuracy and the mentioned processing differences are relatively
stable, resulting in relatively precise delay estimates.

An example SR domain is shown below. The PMS shown should monitor
the connectivity of all 6 links between nodes L100 and L200 one one
side and the connected nodes L050, L060 and L070 on the other side.
The round trip times per measurement loop are assumed to exhibit
unique delays.

```
   +---+   +----+      +----+
   |PMS|   |L100|-----|L050|
   +---+   +----+\   /+----+
     |    /     \  \_/_____
     |   /       \  /      \+----+
  +----+/         \/_  +----|L060|
  |L300|          /  |/      +----+
  +----+\        /   /\_
         \      /   /   \
          \+----+ /    +----+
          |L200|-----|L070|
          +----+      +----+
```

Connectivity verification with a PMS

                          Figure 1

The SID values are picked for convenient reading only. Node-SID: 100
identifies L100, Node-SID: 300 identifies L300 and so on. Adj-SID
10050: Adjacency L100 to L050, Adj-SID 10060: Adjacency L100 to L060,
Adj-SID 60200: Adjacency L60 to L200

Monitoring the 6 links between Ln00 and L0m0 nodes requires 6
measurement loops, each of which has the following properties:

o Each loop follows a single round trip from one Ln00 to one L0m0
   (e.g., between L100 and L050).

o Each loop passes two more links: one between that Ln00 and another
   L0m0 and from there to the other Ln00 (e.g., between L100 and L060
   and then L060 to L200)

o Every link is passed by a single round trip per measurement loop
   only once and only once unidirectional by two other loops, and the
   latter two pass along opposing directions (that's three loops
   passing each single link, e.g., one having a round trip L100 to
   L050 and back, a second passing L100 to L050 only and a third loop
   passing L050 to L100 only).

Note that any 6 links between two to six nodes can be monitored that
way too (if multiple parallel links between two nodes are monitored,

the differences in delay may require a sufficiently high clock
resulotion, if applicable).

This results in 6 measurement loops for the given example (the start
and end of each measurement loop is PMS to L300 to L100 or L200 and a
similar sub-path on the return leg. It is ommitted here for
brevity):

1. M1 is the delay along L100 -> L050 -> L100 -> L060 -> L200

2. M2 is the delay along L100 -> L060 -> L100 -> L070 -> L200

3. M3 is the delay along L100 -> L070 -> L100 -> L050 -> L200

4. M4 is the delay along L200 -> L050 -> L200 -> L060 -> L100

5. M5 is the delay along L200 -> L060 -> L200 -> L070 -> L100

6. M6 is the delay along L200 -> L070 -> L200 -> L050 -> L100

An example for a stack of a loop consisting of Node-SID segments
allowing to caprture M1 is (top to bottom): 100 | 050 | 100 | 060 |
200 | PMS.

An example for a stack of Adj-SID segments the loop resulting in M1
is (top to bottom): 100 | 10050 | 50100 | 10060 | 60200 | PMS. As
can be seen, the Node-SIDs 100 and PMS are present at top and bottom
of the segment stack. Their purpose is to transport the packet from
the PMS to the start of the measurement loop at L100 and return it to
the PMS from its end.

The measurement loops set up as shown have the following properties:

o If the loops are set up using Node-SIDs only, any single complete
   loss of connectivity caused by a failing single link between any
   Ln00 and any L0m0 node briefly disturbs (and changes the measured
   delay) of three loops. Traffic to Node-SIDs is rerouted.

o If the loops are set up using Adj-SIDs only (and Node-SIDs only to
   send the packet from PMS to the loop starting point and from the
   loop end back to the PMS), any single complete loss of
   connectivity caused by a failing single link between any Ln00 and
   any L0m0 node terminates the traffic along three loops. The
   packets of these loops will be dropped, until the link gets back
   into service. Traffic to Adj-SIDs is not rerouted.

o Any congested single interface between any Ln00 and any L0m0 node
   only impacts the measured delay of two measurement loops.

o As an example, the formula for a single Round Trip Delay (RTD) is
   shown here 4 * RTD_L100-L050-L100 = 3 * M1 + M3 + M6 - M2 - M4 -
   M5

A closer look reveals that each single event of interest for the
proposed metric, which are a loss of connectivity or a case of
congestion, uniquely only impacts a single a-priori determinable set
of measurement loops. If, e.g., connectivity is lost between L200
and L050, measurement loops (3), (4) and (6) indicate a change in the
measured delay.

As a second example, if the interface L070 to L100 is congested,
measurement loops (3) and (5) indicate a change in the measured
delay. Without listing all events, all cases of single losses of
connectivity or single events of congestion influence only delay
measurements of a unique set of measurement loops.

A congestion event adding latency to two specific measurement loops
allows calculation of the delay added by the queue at the congested
interface. Thus, the resulting RTD increase can be assigned to a
single interface.

3. Singleton Definition for Type-P-SR-Path-Connectivity-and-Congestion

3.1. Metric Name

Type-P-SR-Path-Connectivity-and-Congestion

3.2. Metric Parameters

o Src, the IP address of a source host

o Dst, the IP address of a destination host if IP routing is
   applicable; in the case of MPLS routing, a diagnostic address as
   specified by [RFC8029]

o T, a time

o lambda, a rate in reciprocal seconds

o L, a packet length in bits. The packets of a Type P packet stream
   from which the sample Path-Connectivity-and-Congestion metric is
   taken MUST all be of the same length.

o MLA, a Monitoring Loop Address information ensuring that a
   singleton passes a single sub-path_a to be monitored
   bidirectional, a sub-path_b to be monitored unidirectional and a

sub-path_c to be monitored unidirectional, where sub-path_a, -_b
and -_c MUST NOT be identical.

o P, the specification of the packet type, over and above the source
   and destination addresses

o DS, a constant time interval between two type-P packets

3.3. Metric Units

A sequence of consecutive time values.

3.4. Definition

A moving average of AV time values per measurement path is compared
by a change point detection algorithm. The temporal packet spacing
value DS represents the smallest period within which a change in
connectivity or congestion may be detected.

A single loss of connectivity of a sub-path between two nodes affects
three different measurement paths. Depending on the value chosen for
DS, packet loss might occur (note that the moving average evaluation
needs to span a longer period than convergence time; alternatively,
packet-loss visible along the three measurement paths may serve as an
evaluation criterium). After routing convergence the type-p packets
along the three measurement paths show a change in delay.

A congestion of a single interface of a sub-path connecting two nodes
affects two different measurement paths. The the type-p packets
along the two congested measurement paths show an additional change
in delay.

3.5. Discussion

Detection of a multiple losses of monitored sub-path connectivity or
congestion of a multiple monitored sub-paths may be possible. These
cases have not been investigated, but may occur in the case of Shared
Risk Link Groups. Monitoring Shared Risk LinkGroups and sub-paths
with multiple failures abd congestion is not within scope of this
document.

3.6. Methodologies

For the given type-p, the methodology is as follows:

o The set of measurement paths MUST be routed in a way that each
   single loss of connectivity and each case of single interface
   congestion of one of the sub-paths passed by a type-p packet

creates a unique pattern of type-p packets belonging to a subset
of all configured measurement paths indicate a change in the
measured delay. As a minimum, each sub-path to be monitored MUST
be passed

o

   * by one measurement_path_1 and its type-p packet in
      bidirectional direction

   * by one measurement_path_2 and its type-p packet in "downlink"
      direction

   * by one measurement_path_3 and its type-p packet in "uplink"
      direction

o "Uplink" and "Downlink" have no architectural relevance. The
   terms are chosen to express, that the packets of
   measurement_path_2 and measuremnt_path_3 pass the monitored sub-
   path unidirectional in opposing direction. Measuremnt_path_1,
   measurement_path_2 and measurement_path_3 MUST NOT be identical.

o All measurement paths SHOULD terminate between identical sender
   and receiver interfaces. It is recommended to connect the sender
   and receiver as closely to the paths to be monitored as possible.
   Each intermediate sub-path between sender and receiver one one
   hand and sub-paths to be monitored is an additional source of
   errors requiring separate monitoring.

o Segment Routed domains supporting Node- and Adj-SIDs should enable
   the monitoring path set-up as specified. Other routing protocols
   may be used as well, but the monitoring path set up might be
   complex or impossible.

o Pre-compute how the two and three measurement path delay changes
   correlate to sub-path connectivity and congestion patterns.
   Absolute change valaues aren't required, a simultaneous change of
   two or three particular measurement paths is.

o Ensure that the temporal resolution of the measurement clock
   allows to reliably capture a unique delay value for each
   configured measurement path while sub-path connectivity is
   complete and no congestion is present.

o Synchronised clocks are not strictly required, as the metric is
   evaluating differences in delay. Changes in clock synchronisation
   SHOULD NOT be close to the time interval within which changes in
   connectivity or congestion should be monitored.

o At the Src host, select Src and Dst IP addresses, and address
   information to route the type-p packet along one of the configured
   measurement path. Form a test packet of Type-P with these
   addresses.

o Configure the Dst host access to receive the packet.

o At the Src host, place a timestamp, a sequence number and a unique
   identifier of the measurement path in the prepared Type-P packet,
   and send it towards Dst.

o Capture the one-way delay and determine packet-loss by the metrics
   specified by [RFC7679] and [RFC7680] respectively and store the
   result for the path.

o If two or three subpaths indicate a change in delay, report a
   change in connectivity or congestion status as pre-computed above.

o If two or three sub paths indicate a change in delay, report a
   change in connectivity or congestion status as pre-computed above.

Note that monitoring 6 sub paths requires setting up 6 monitoring
paths as shown in the figure above.

3.7. Errors and Uncertainties

Sources of error are:

o Measurement paths whose delays don't indicate a change after sub-
   path connectivity changed.

o A timestamps whose resolution is missing or inacurrate at the
   delays measured for the different monitoring paths.

o Multiple occurrences of sub path connectivity and congestion.

o Loss of connectivity and congestion along sub-paths connecting the
   measurement device(s) with the sub-paths to be monitored.

3.8. Reporting the Metric

The metric reports loss of connectivity of monitored sub-path or
congestion of an interface and identifies the sub-path and the
direction of traffic in the case of congestion.

4. **Singleton Definition for Type-P-SR-Path-Round-Trip-Delay-Estimate**

    This section will be added in a later version, if there's interest in
    picking up this work.

5. **IANA Considerations**

    If standardised, the metric will require an entry in the IPPM metric
    registry.

6. **Security Considerations**

    This draft specifies how to use methods specified or described within
    [RFC8402] and [RFC8403]. It does not introduce new or additional SR
    features. The security considerations of both references apply here
    too.

7. **References**

7.1. **Normative References**

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119,
              DOI 10.17487/RFC2119, March 1997,
              <https://www.rfc-editor.org/info/rfc2119>.

   [RFC2678] Mahdavi, J. and V. Paxson, "IPPM Metrics for Measuring
              Connectivity", RFC 2678, DOI 10.17487/RFC2678, September
              1999, <https://www.rfc-editor.org/info/rfc2678>.

   [RFC7679] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton,
              Ed., "A One-Way Delay Metric for IP Performance Metrics
              (IPPM)", STD 81, RFC 7679, DOI 10.17487/RFC7679, January
              2016, <https://www.rfc-editor.org/info/rfc7679>.

   [RFC7680] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton,
              Ed., "A One-Way Loss Metric for IP Performance Metrics
              (IPPM)", STD 82, RFC 7680, DOI 10.17487/RFC7680, January
              2016, <https://www.rfc-editor.org/info/rfc7680>.

   [RFC8029] Kompella, K., Swallow, G., Pignataro, C., Ed., Kumar, N.,
              Aldrin, S., and M. Chen, "Detecting Multiprotocol Label
              Switched (MPLS) Data-Plane Failures", RFC 8029,
              DOI 10.17487/RFC8029, March 2017,
              <https://www.rfc-editor.org/info/rfc8029>.

   [RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L.,
              Decraene, B., Litkowski, S., and R. Shakir, "Segment
              Routing Architecture", RFC 8402, DOI 10.17487/RFC8402,
              July 2018, <https://www.rfc-editor.org/info/rfc8402>.

7.2. Informative References

   [RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis,
              "Framework for IP Performance Metrics", RFC 2330,
              DOI 10.17487/RFC2330, May 1998,
              <https://www.rfc-editor.org/info/rfc2330>.

   [RFC8403] Geib, R., Ed., Filsfils, C., Pignataro, C., Ed., and N.
              Kumar, "A Scalable and Topology-Aware MPLS Data-Plane
              Monitoring System", RFC 8403, DOI 10.17487/RFC8403, July
              2018, <https://www.rfc-editor.org/info/rfc8403>.

Author's Address

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

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