Abstract

RFC 6374 describes methods of making loss and delay measurements on Label Switched Paths (LSPs) primarily as they are used in MPLS Transport Profile (MPLS-TP) networks. This document describes a method of extending the performance measurements (specified in RFC 6374) from flows carried over MPLS-TP to flows carried over generic MPLS LSPs. In particular, it extends the technique to allow loss and delay measurements to be made on multipoint-to-point LSPs and introduces some additional techniques to allow more sophisticated measurements to be made in both MPLS-TP and generic MPLS networks.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9571.

Copyright Notice

Copyright (c) 2024 IETF Trust and the persons identified as the document authors. All rights reserved.
This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction 3
2. Requirements Language 4
3. Packet Loss Measurement Using SFL 4
4. Single Packet Delay Measurement Using SFL 4
5. Data Service Packet Delay Measurement 4
6. Some Simplifying Rules 6
7. Multiple Packet Delay Characteristics 6
   7.1. Method 1: Time Buckets 7
   7.2. Method 2: Classic Standard Deviation 8
      7.2.1. Multi-packet Delay Measurement Message Format 9
   7.3. Per-Packet Delay Measurement 10
   7.4. Average Delay 10
8. Sampled Measurement 11
9. Carrying Packets over an LSP Using an SFL 12
   9.1. Extending RFC 6374 with SFL TLV 13
10. Combined Loss/Delay Measurement Using SFL 14
11. Privacy Considerations 14
12. Security Considerations 15
13. IANA Considerations 15
   13.1. Allocation of MPLS Generalized Associated Channel (G-ACh) Types 15
   13.2. Allocation of MPLS Loss/Delay TLV Object 15
14. References 16
   14.1. Normative References 16
   14.2. Informative References 16
1. Introduction

[RFC6374] was originally designed for use as an Operations, Administration, and Maintenance (OAM) protocol for use with MPLS Transport Profile (MPLS-TP) [RFC5921] LSPs. MPLS-TP only supports point-to-point and point-to-multipoint LSPs. This document describes how to use [RFC6374] in the generic MPLS case and also introduces a number of more sophisticated measurements of applicability to both cases.

[RFC8372] describes the requirement for introducing flow identities when using packet loss measurements described in [RFC6374]. In summary, [RFC6374] describes use of the loss measurement (LM) message as the packet accounting demarcation point. Unfortunately, this gives rise to a number of problems that may lead to significant packet accounting errors in certain situations. For example:

1. Where a flow is subjected to Equal-Cost Multipath (ECMP) treatment, packets can arrive out of order with respect to the LM packet.
2. Where a flow is subjected to ECMP treatment, packets can arrive at different hardware interfaces, thus requiring reception of an LM packet on one interface to trigger a packet accounting action on a different interface that may not be co-located with it. This is a difficult technical problem to address with the required degree of accuracy.
3. Even where there is no ECMP (for example, on RSVP-TE, MPLS-TP LSPs, and pseudowires (PWs)), local processing may be distributed over a number of processor cores, leading to synchronization problems.
4. Link aggregation techniques [RFC7190] may also lead to synchronization issues.
5. Some forwarder implementations have a long pipeline between processing a packet and incrementing the associated counter, again leading to synchronization difficulties.

An approach to mitigating these synchronization issues is described in [RFC9341] -- the packets are batched by the sender, and each batch is marked in some way such that adjacent batches can be easily recognized by the receiver.

An additional problem arises where the LSP is a multipoint-to-point LSP since MPLS does not include a source address in the packet. Network management operations require the measurement of packet loss between a source and destination. It is thus necessary to introduce some source-specific information into the packet to identify packet batches from a specific source.
[RFC8957] describes a method of encoding per-flow instructions in an MPLS label stack using a technique called Synonymous Flow Labels (SFLs), in which labels that mimic the behavior of other labels provide the packet batch identifiers and enable the per-batch packet accounting. This memo specifies how SFLs are used to perform packet loss and delay measurements as described in [RFC6374].

When the terms "performance measurement method," "Query," "packet," or "message" are used in this document, they refer to a performance measurement method, Query, packet, or message as specified in [RFC6374].

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Packet Loss Measurement Using SFL

The data service packets of the flow being instrumented are grouped into batches, and all the packets within a batch are marked with the SFL [RFC8372] corresponding to that batch. The sender counts the number of packets in the batch. When the batch has completed and the sender is confident that all of the packets in that batch will have been received, the sender issues a Query message to determine the number actually received and hence the number of packets lost. The Query message is sent using the same SFL as the corresponding batch of data service packets. The format of the Query and Response packets is described in Section 9.

4. Single Packet Delay Measurement Using SFL

[RFC6374] describes how to measure the packet delay by measuring the transit time of a packet over an LSP. Such a packet may not need to be carried over an SFL since the delay over a particular LSP should be a function of the Traffic Class (TC) bits.

However, where SFLs are being used to monitor packet loss or where label-inferred scheduling is used [RFC3270], then the SFL would be REQUIRED to ensure that the packet that was being used as a proxy for a data service packet experienced a representative delay. The format of a packet carried over the LSP using an SFL is shown in Section 9.

5. Data Service Packet Delay Measurement

Where it is desired to more thoroughly instrument a packet flow and to determine the delay of a number of packets, it is undesirable to send a large number of packets acting as proxy data service packets (see Section 4). A method of directly measuring the delay characteristics of a batch of packets is therefore needed.
Given the long intervals over which it is necessary to measure packet loss, it is not necessarily the case that the batch times for the two measurement types would be identical. Thus, we use a technique that permits the two measurements to be made concurrently and yet relatively independently from each other. The notion that they are relatively independent arises from the potential for the two batches to overlap in time, in which case either the delay batch time will need to be cut short or the loss time will need to be extended to allow correct reconciliation of the various counters.

The problem is illustrated in Figure 1.

---

**Figure 1: Query Packet with SFL**

In Case 1, we show where loss measurement alone is being carried out on the flow under analysis. For illustrative purposes, consider that 10 packets are used in each flow in the time interval being analyzed.

Now consider Case 2, where a small batch of packets need to be analyzed for delay. These are marked with a different SFL type, indicating that they are to be monitored for both loss and delay. The SFL=A indicates loss batch A, and SFL=D indicates a batch of packets that are to be instrumented for delay, but SFL D is synonymous with SFL A, which in turn is synonymous with the underlying Forwarding Equivalence Class (FEC). Thus, a packet marked "D" will be accumulated into the A loss batch, into the delay statistics, and will be forwarded as normal. Whether the packet is actually counted twice (for loss and delay) or whether the two counters are reconciled during reporting is a local matter.
Now consider Case 3, where a small batch of packets is marked for delay across a loss batch boundary. These packets need to be considered as a part of batch A or a part of batch B, and any Query needs to take place after all packets A or D (whichever option is chosen) have arrived at the receiving Label Switching Router (LSR).

Now consider Case 4. Here, we have a case where it is required to take a number of delay measurements within a batch of packets that we are measuring for loss. To do this, we need two SFLs for delay (C and D) and alternate between them (on a delay-batch-by-delay-batch basis) for the purposes of measuring the delay characteristics of the different batches of packets.

6. Some Simplifying Rules

It is possible to construct a large set of overlapping measurement types in terms of loss, delay, loss and delay, and batch overlap. If we allow all combinations of cases, this leads to configuration, testing, and implementation complexity and, hence, increased costs. The following simplifying rules represent the default case:

1. Any system that needs to measure delay **MUST** be able to measure loss.
2. Any system that is to measure delay **MUST** be configured to measure loss. Whether the loss statistics are collected or not is a local matter.
3. A delay measurement **MAY** start at any point during a loss measurement batch, subject to rule 4.
4. A delay measurement interval **MUST** be short enough that it will complete before the enclosing loss batch completes.
5. The duration of a second delay batch (D in Figure 1) must be such that all packets from the packets belonging to a first delay batch (C in Figure 1) will have been received before the second delay batch completes. This condition is satisfied when the time to send a batch is long compared to the network propagation time and is a parameter that can be established by the network operator.

Given that the sender controls both the start and duration of a loss and a delay packet batch, these rules are readily implemented in the control plane.

7. Multiple Packet Delay Characteristics

A number of methods are described that add to the set of measurements originally specified in [RFC6374]. Each of these methods has different characteristics and different processing demands on the packet forwarder. The choice of method will depend on the type of diagnostic that the operator seeks.

Three methods are discussed:

1. Time Buckets
2. Classic Standard Deviation
3. Average Delay
7.1. Method 1: Time Buckets

In this method, the receiving LSR measures the inter-packet gap, classifies the delay into a number of delay buckets, and records the number of packets in each bucket. As an example, if the operator were concerned about packets with a delay of up to 1 μs, 2 μs, 4 μs, 8 μs, and over 8 μs, then there would be five buckets, and packets that arrived up to 1 μs would cause the “up to 1 μs” bucket counter to increase. Likewise, for those that arrived between 1 μs and 2 μs, the “2 μs” bucket counter would increase, etc. In practice, it might be better in terms of processing and potential parallelism if both the “up to 1 μs” and “2 μs” counters were incremented when a packet had a delay relative to its predecessor of 2 μs, and any more detailed information was calculated in the analytics system.

This method allows the operator to see more structure in the jitter characteristics than simply measuring the average jitter and avoids the complication of needing to perform a per-packet multiply but will probably need the time intervals between buckets to be programmable by the operator.

The packet format of a Time Bucket Jitter Measurement message is shown below:

![Figure 2: Time Bucket Jitter Measurement Message Format](image)
The Version, Flags, Control Code, Message Length, Querier Timestamp Format (QTF), Responder Timestamp Format (RTF), Responder's Preferred Timestamp Format (RPTF), Session Identifier, Reserved, and Differentiated Services (DS) fields are as defined in Section 3.2 of [RFC6374]. The remaining fields, which are unsigned integers, are as follows:

- Number of Buckets in the measurement.
- Reserved 1 must be sent as zero and ignored on receipt.
- Interval (in 10 ns units) is the inter-packet interval for this bucket.
- Number of Pkts in Bucket 1 is the number of packets found in the first bucket.
- Number of Pkts in Bucket N is the number of packets found in the Nth bucket, where N is the value in the Number of Buckets field.

There will be a number of Interval/Number pairs depending on the number of buckets being specified by the Querier. If a message is being used to configure the buckets (i.e., the responder is creating or modifying the buckets according to the intervals in the Query message), then the responder **MUST** respond with 0 packets in each bucket until it has been configured for a full measurement period. This indicates that it was configured at the time of the last response message, and thus, the response is valid for the whole interval. As per the convention in [RFC6374], the Number of Pkts in Bucket fields are included in the Query message and set to zero.

Out-of-band configuration is permitted by this mode of operation.

Note this is a departure from the normal fixed format used in [RFC6374].

The Time Bucket Jitter Measurement message is carried over an LSP in the way described in [RFC6374] and over an LSP with an SFL as described in Section 9.

### 7.2. Method 2: Classic Standard Deviation

In this method, provision is made for reporting the following delay characteristics:

1. Number of packets in the batch (n)
2. Sum of delays in a batch (S)
3. Maximum delay
4. Minimum delay
5. Sum of squares of inter-packet delay (SumS)

Characteristics 1 and 2 give the mean delay. Measuring the delay of each pair in the batch is discussed in Section 7.3.

Characteristics 3 and 4 give the outliers.

Characteristics 1, 2, and 5 can be used to calculate the variance of the inter-packet gap, hence the standard deviation giving a view of the distribution of packet delays and hence the jitter. The equation for the variance (var) is given by:
There is some concern over the use of this algorithm for measuring variance because SumS and $S^2/n$ can be similar numbers, particularly where variance is low. However, the method is acceptable because it does not require a division in the hardware.

### 7.2.1. Multi-packet Delay Measurement Message Format

The packet format of a Multi-packet Delay Measurement message is shown below:

**Figure 3: Multi-packet Delay Measurement Message Format**

The Version, Flags, Control Code, Message Length, QTF, RTF, RPTF, Session Identifier, Reserved, and DS fields are as defined in Section 3.2 of [RFC6374]. The remaining fields are as follows:

- **Number of Packets** is the number of packets in this batch.
- **Sum of Delays for Batch** is the duration of the batch in the time measurement format specified in the RTF field.
- **Minimum Delay** is the minimum inter-packet gap observed during the batch in the time format specified in the RTF field.
- Maximum Delay is the maximum inter-packet gap observed during the batch in the time format specified in the RTF field.

The Multi-packet Delay Measurement message is carried over an LSP in the way described in [RFC6374] and over an LSP with an SFL as described in Section 9.

### 7.3. Per-Packet Delay Measurement

If detailed packet delay measurement is required, then it might be possible to record the inter-packet gap for each packet pair. In cases other than the exceptions of slow flows or small batch sizes, this would create a large (per-packet) demand on storage in the instrumentation system, a large bandwidth for such a storage system and large bandwidth for the analytics system. Such a measurement technique is outside the scope of this document.

### 7.4. Average Delay

Introduced in [RFC9341] is the concept of a one-way delay measurement in which the average time of arrival of a set of packets is measured. In this approach, the packet is timestamped at arrival, and the responder returns the sum of the timestamps and the number of timestamps. From this, the analytics engine can determine the mean delay. An alternative model is that the responder returns the timestamp of the first and last packet and the number of packets. This latter method has the advantage of allowing the average delay to be determined at a number of points along the packet path and allowing the components of the delay to be characterized. Unless specifically configured otherwise, the responder may return either or both types of response, and the analytics engine should process the response appropriately.

The packet format of an Average Delay Measurement message is shown below:
The Version, Flags, Control Code, Message Length, QTF, RTF, RPTF, Session Identifier, and DS fields are as defined in [RFC6374]. The remaining fields are as follows:

- Number of Packets is the number of packets in this batch.
- Time of First Packet is the time of arrival of the first packet in the batch.
- Time of Last Packet is the time of arrival of the last packet in the batch.
- Sum of Timestamps of Batch.

The Average Delay Measurement message is carried over an LSP in the way described in [RFC6374] and over an LSP with an SFL as described in Section 9. As is the convention with [RFC6374], the Query message contains placeholders for the Response message. The placeholders are sent as zero.

8. Sampled Measurement

In the discussion so far, it has been assumed that we would measure the delay characteristics of every packet in a delay measurement interval defined by an SFL of constant color. In [RFC9341], the concept of a sampled measurement is considered. That is, the responder only measures a packet at the start of a group of packets being marked for delay measurement by a particular color, rather than every packet in the marked batch. A measurement interval is not defined by...
the duration of a marked batch of packets but the interval between a pair of packets taking a
readout of the delay characteristic. This approach has the advantage that the measurement is not
impacted by ECMP effects.

This sampled approach may be used if supported by the responder and configured by the
operator.

9. Carrying Packets over an LSP Using an SFL

We illustrate the packet format of a Query message using SFLs for the case of an MPLS Direct
Loss Measurement in Figure 5.

Figure 5: Query Packet with SFL
The MPLS label stack is exactly the same as that used for the user data service packets being instrumented except for the inclusion of the Generic Associated Channel Label (GAL) [RFC5586] to allow the receiver to distinguish between normal data packets and OAM packets. Since the packet loss measurements are being made on the data service packets, an MPLS Direct Loss Measurement is being made, which is indicated by the type field in the Associated Channel Header (ACH) (Type = 0x000A).

The measurement message consists of up to three components as follows.

- The fixed-format portion of the message is carried over the ACH channel. The ACH channel type specifies the type of measurement being made (currently: loss, delay, or loss and delay).
- (Optional) The SFL TLV specified in Section 9.1 MAY be carried if needed. It is used to provide the implementation with a reminder of the SFL that was used to carry the message. This is needed because a number of MPLS implementations do not provide the MPLS label stack to the MPLS OAM handler. This TLV is required if messages are sent over UDP [RFC7876]. This TLV MUST be included unless, by some method outside the scope of this document, it is known that this information is not needed by the responder.
- (Optional) The Return Information MAY be carried if needed. It allows the responder send the response to the Querier. This is not needed if the response is requested in band and the MPLS construct being measured is a point-to-point LSP, but it otherwise MUST be carried. The Return Address TLV is defined in [RFC6374], and the optional UDP Return Object is defined in [RFC7876].

Where a measurement other than an MPLS Direct Loss Measurement is to be made, the appropriate measurement message is used (for example, one of the new types defined in this document), and this is indicated to the receiver by the use of the corresponding ACH type.

### 9.1. Extending RFC 6374 with SFL TLV

The [RFC6374] SFL TLV is shown in Figure 6. This contains the SFL that was carried in the label stack, the FEC that was used to allocate the SFL, and the index (into the batch of SFLs that were allocated for the FEC) that corresponds to this SFL.

![Figure 6: SFL TLV](image)

Where:
Type
Length
MBZ
SFL Batch
SFL Index
SFL
Reserved
FEC

Set to Synonymous Flow Label (SFL-TLV).
The length of the TLV is as specified in [RFC6374].
MUST be sent as zero and ignored on receive.
An identifier for a collection of SFLs grouped together for management and control purposes.
The index of this SFL within the list of SFLs that were assigned for the FEC.
Multiple SFLs can be assigned to a FEC, each with different actions. This index is an optional convenience for use in mapping between the TLV and the associated data structures in the LSRs. The use of this feature is agreed upon between the two parties during configuration. It is not required but is a convenience for the receiver if both parties support the facility.
The SFL used to deliver this packet. This is an MPLS label that is a component of a label stack entry as defined in Section 2.1 of [RFC3032].
MUST be sent as zero and ignored on receive.
The Forwarding Equivalence Class that was used to request this SFL. This is encoded as per Section 3.4.1 of [RFC5036].

This information is needed to allow for operation with hardware that discards the MPLS label stack before passing the remainder of the stack to the OAM handler. By providing both the SFL and the FEC plus index into the array of allocated SFLs, a number of implementation types are supported.

10. Combined Loss/Delay Measurement Using SFL

This mode of operation is not currently supported by this specification.

11. Privacy Considerations

The inclusion of originating and/or flow information in a packet provides more identity information and hence potentially degrades the privacy of the communication. While the inclusion of the additional granularity does allow greater insight into the flow characteristics, it does not specifically identify which node originated the packet other than by inspection of the network at the point of ingress or inspection of the control protocol packets. This privacy threat may be mitigated by encrypting the control protocol packets, regularly changing the synonymous labels, and by concurrently using a number of such labels.
12. Security Considerations

The security considerations documented in [RFC6374] and [RFC8372] (which in turn calls up [RFC5920] and [RFC7258]) are applicable to this protocol.

The issue noted in Section 11 is a security consideration. There are no other new security issues associated with the MPLS data plane. Any control protocol used to request SFLs will need to ensure the legitimacy of the request.

An attacker that manages to corrupt the SFL TLV in Section 9.1 could disrupt the measurements in a way that the responder is unable to detect. However, the network operator is likely to notice the anomalous network performance measurements, and in any case, normal MPLS network security procedures make this type of attack extremely unlikely.

13. IANA Considerations

13.1. Allocation of MPLS Generalized Associated Channel (G-ACh) Types

As per the IANA considerations in [RFC5586] updated by [RFC7026] and [RFC7214], IANA has allocated the following values in the "MPLS Generalized Associated Channel (G-ACh) Types" registry, in the "Generic Associated Channel (G-ACh) Parameters" registry group:

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0010</td>
<td>Time Bucket Jitter Measurement</td>
<td>RFC 9571</td>
</tr>
<tr>
<td>0x0011</td>
<td>Multi-packet Delay Measurement</td>
<td>RFC 9571</td>
</tr>
<tr>
<td>0x0012</td>
<td>Average Delay Measurement</td>
<td>RFC 9571</td>
</tr>
</tbody>
</table>

Table 1

13.2. Allocation of MPLS Loss/Delay TLV Object

IANA has allocated the following TLV from the 0-127 range of the "MPLS Loss/Delay Measurement TLV Object" registry in the "Generic Associated Channel (G-ACh) Parameters" registry group:

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Synonymous Flow Label</td>
<td>RFC 9571</td>
</tr>
</tbody>
</table>

Table 2
14. References

14.1. Normative References


14.2. Informative References

[RFC3270]
Acknowledgments

The authors thank Benjamin Kaduk and Elwyn Davies for their thorough and thoughtful review of this document.

Contributors

Zhenbin Li
Huawei
Email: lizhenbin@huawei.com

Siva Sivabalan
Ciena Corporation
Email: ssivabal@ciena.com
Authors' Addresses

Stewart Bryant (EDITOR)
University of Surrey
Email: sb@stewartbryant.com

George Swallow
Independent
Email: swallow.ietf@gmail.com

Mach(Guoyi) Chen
Huawei
Email: mach.chen@huawei.com

Giuseppe Fioccola
Huawei Technologies
Email: giuseppe.fioccola@huawei.com

Gregory Mirsky
ZTE Corp.
Email: gregimirsky@gmail.com