Packet Delay Measurement for SFC
draft-agv-sfc-packet-delay-measurement-00

Abstract

Service provider service level agreements (SLAs) depend on the capability to measure and monitor performance metrics for packet delay.

The common reasons for packet delay are Nodal processing (Algorithmic, Packetization etc...), Queuing, Transmission delay, Propagation delay and Service Function processing delay.

Packet Delay Measurement capability also provides operators with greater visibility into the performance characteristics of their networks, thereby facilitating planning, troubleshooting, and network performance evaluation.

This document specifies best possible efficient and accurate mechanism for passive packet delay measurement for Service Function Chains (SFCs) for a SFC network domain.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html
The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html
Copyright and License Notice

Copyright (c) 2015 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 (http://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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1 Introduction .................................................. 5
  1.1 Terminology ............................................. 6
  1.2 Terms & Definition ..................................... 6
  1.3 Applicability and Scope ................................. 9
2 Message Format ............................................... 10
  2.1 Performance Measurement Context Header ................. 10
  2.2 Packet Delay PM Context Header ......................... 10
  2.3 Performance Measurement Flow ID ....................... 12
3 Operations .................................................. 13
  3.1 Delay Measurement Options .............................. 13
  3.2 Packet Delay Parameters for Option 1 .................. 13
  3.3 Packet Delay Parameters for Option 2 .................. 14
  3.4 Initiating a Packet Delay Measurement .................. 15
  3.5 Encapsulation of Classified Packets .................... 15
  3.6 Incoming Packets Processing at MA ..................... 16
  3.7 Outgoing Packets Processing at MA ..................... 17
  3.8 MA Reporting the statistics to Collector .............. 17
  3.9 Packet Delay Calculation at Collector ................. 18
    3.9.1 Packet Delay Calculation at Collector (Option 1) .... 18
    3.9.1 Packet Delay Calculation at Collector (Option 2) .... 18
4 Security Considerations .................................... 19
5 IANA Considerations ....................................... 19
  5.1 TLV Class .............................................. 19
6 References .................................................. 20
  6.1 Normative References ................................. 20
  6.2 Informative References ............................... 20
Authors’ Addresses ........................................... 21
1 Introduction

The delivery of end-to-end services often requires various service functions. These include traditional network service functions such as firewalls and traditional IP Network Address Translators (NATs), as well as application-specific functions. The definition and instantiation of an ordered set of service functions and subsequent "steering" of traffic through them is termed Service Function Chaining (SFC).

The purpose of this memo is to define a mechanism for packet delay measurement for Service Function Chains (SFCs).

As mentioned in [SFC-PM-arch] packet delay measurement could be performed using Active Measurement or Passive Measurement method. This document describes the Passive Measurement method & active measurement method is out of scope of this document. The passive packet delay measurement described in this document is realized by adding a new Context Header in the Network Service Header.
1.1 Terminology

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].

1.2 Terms & Definition

Network Service: An offering provided by an operator that is delivered using one or more service functions. This may also be referred to as a "composite service". The term "service" is used to denote a "network service" in the context of this document.

Note: Beyond this document, the term "service" is overloaded with varying definitions. For example, to some a service is an offering composed of several elements within the operator’s network, whereas for others a service, or more specifically a network service, is a discrete element such as a "firewall". Traditionally, such services (in the latter sense) host a set of service functions and have a network locator where the service is hosted.

Classification: Locally instantiated matching of traffic flows against policy for subsequent application of the required set of network service functions. The policy may be customer/network/service specific.

Classifier: An element that performs Classification.

Service Function Chain (SFC): A service function chain defines an ordered set of abstract service functions and ordering constraints that must be applied to packets and/or frames and/or flows selected as a result of classification. An example of an abstract service function is "a firewall". The implied order may not be a linear progression as the architecture allows for SFCs that copy to more than one branch, and also allows for cases where there is flexibility in the order in which service functions need to be applied. The term "service chain" is often used as shorthand for service function chain.

Service Function (SF): A function that is responsible for specific treatment of received packets. A Service Function can act at various layers of a protocol stack (e.g., at the network layer or other OSI layers). As a logical component, a service function can be realized as a virtual element or be
embedded in a physical network element. One or more Service Functions can be embedded in the same network element. Multiple occurrences of the service function can exist in the same administrative domain.

One or more service functions can be involved in the delivery of added-value services. A non-exhaustive list of abstract service functions includes: firewalls, WAN and application acceleration, Deep Packet Inspection (DPI), Lawful Intercept (LI), server load balancing, NAT44 [RFC3022], NAT64 [RFC6146], NPTv6 [RFC6296], HOST_ID injection, HTTP Header Enrichment functions, and TCP optimizer.

An SF may be SFC encapsulation aware (that is, it receives and acts on information in the SFC encapsulation) or unaware (in which case, data forwarded to the SF does not contain the SFC encapsulation). This is often referred to as "SFC aware" and "SFC unaware", respectively.

Service Function Forwarder (SFF): A service function forwarder is responsible for forwarding traffic to one or more connected service functions according to information carried in the SFC encapsulation, as well as handling traffic coming back from the SF. Additionally, an SFF is responsible for delivering traffic to a classifier when needed and supported, transporting traffic to another SFF (in the same or different type of overlay), and terminating the Service Function Path (SFP).

Metadata: Provides the ability to exchange context information between classifiers and SFs, and among SFs.

Service Function Path (SFP): The service function path is a constrained specification of where packets assigned to a certain service function path must go. While it may be so constrained as to identify the exact locations, it can also be less specific. The SFP provides a level of indirection between the fully abstract notion of service chain as a sequence of abstract service functions to be delivered, and the fully specified notion of exactly which SFF/SFs the packet will visit when it actually traverses the network. By allowing the control components to specify this level of indirection, the operator may control the degree of SFF/SF selection authority that is delegated to the network.

SFC Encapsulation: The SFC encapsulation provides, at a minimum, SFP identification, and is used by the SFC-aware functions,
such as the SFF and SFC-aware SFs. The SFC encapsulation is not used for network packet forwarding. In addition to SFP identification, the SFC encapsulation carries metadata including data-plane context information.

Rendered Service Path (RSP): Within an SFP, packets themselves are of course transmitted from and to specific places in the network, visiting a specific sequence of SFFs and SFs. This sequence of actual visits by a packet to specific SFFs and SFs in the network is known as the Rendered Service Path (RSP). This definition is included here for use by later documents, such as when solutions may need to discuss the actual sequence of locations the packets visit.

SFC-Enabled Domain: A network or region of a network that implements SFC. An SFC-enabled domain is limited to a single network administrative domain.

SFC Proxy: Removes and inserts SFC encapsulation on behalf of an SFC-unaware service function. SFC proxies are logical elements.

Network Node/Element: Device that forwards packets or frames based on outer header information. In most cases is not aware of the presence of NSH.

Network Overlay: Logical network built on top of existing network (the underlay). Packets are encapsulated or tunneled to create the overlay network topology.

Network Service Header: Data plane header added to frames/packets. The header contains information required for service chaining, as well as metadata added and consumed by network nodes and service elements.

NSH Proxy: Acts as a gateway: removes and inserts SH on behalf of a service function that is not NSH aware.

Service Classifier: Function that performs classification and imposes an NSH. Creates a service path. Non-initial (i.e. subsequent) classification can occur as needed and can alter, or create a new service path.

Measurement Collector: An operational function that collects measurement data from a Measurement Agent. Measurement collector is responsible for collecting the Performance measurement Data from Measurement Agent. Measurement collector functionality could be integrated with one of
the MA or with the Controller itself.

**Measurement Agent:** An operational function that contains one or more Measurement functions. Measurement Agents is responsible for understand and analyze Performance measurement control information encoded in NSH metadata and perform the performance data collecting and report the same to Measurement collector with key information to identify performance measurement instance along with data collected.

**Measurement Controller:** An operational function that controls running, scheduling, and general coordination of Measurement functions by instructing a Measurement Agent using NSH metadata. Measurement Controller is responsible for Configuring the Performance measurement Instance. Optionally Performance measurement instance can be configured manually at the Ingress in which case Controller is not required.

**Base Header:** Provides information about service header and payload protocol

**Service Path Header:** Provides Path identification and location within a path

**Context Header:** Carries opaque metadata and variable length encoded information

**MD Type:** Indicates format of NSH beyond Mandatory Base Header and Service Path Header.

### 1.3 Applicability and Scope

This document defines the implementation mechanism for the Packet Delay Measurement as per Performance Measurement Architecture [SFC-PM-arch]. This document defines a new NSH Message Format for carrying Packet Delay Measurement related control information. It also defines Operations to be carried out for Packet Loss Measurement. Communication mechanism between Measurement Controller, Measurement Collector and MA is out of scope of this document.
2 Massage Format

2.1 Performance Measurement Context Header

The format of the Performance measurement context headers, is as described below.

```
+-----------------------------------------------+-----------------------------------------------+-----------------------------------------------+
| TLV Class = (IANA PM Class)                  | PM Type                                      | Length                                       |
+-----------------------------------------------+-----------------------------------------------+-----------------------------------------------+
| PM Metadata                                  |                                               |                                              |
+-----------------------------------------------+-----------------------------------------------+-----------------------------------------------+
```

Figure 1: Performance Measurement Context Header

TLV Class: As per [I-D.ietf-sfc-nsh] the TLV class value for Performance measurement needs to be applied from IANA.

PM Type: indicates the type of performance measurement that needs to be carried out by the MA.

Length: Length of the variable PM metadata, in 4-byte words.

2.2 Packet Delay PM Context Header

The format of the Packet Delay PM context headers, is as described below.

```
+-----------------------------------------------+-----------------------------------------------+-----------------------------------------------+
| TLV Class = (IANA PM Class)                  | PM Type                                      | Length                                       |
+-----------------------------------------------+-----------------------------------------------+-----------------------------------------------+
| Measurement Window Index                     | Reserved                                     |                                              |
+-----------------------------------------------+-----------------------------------------------+-----------------------------------------------+
| SI(1)                                        |                                              | SI(n)                                        |
+-----------------------------------------------+-----------------------------------------------+-----------------------------------------------+
```

Figure 2: Packet Delay PM Context Header

TLV Class: As per [I-D.ietf-sfc-nsh] the TLV class value for Performance measurement needs to be applied from IANA.

PM Type: Indicates the type of performance measurement that needs to be carried out by the MA. (IANA assigned PM Type PM Delay value)
This memo defines the following values.
0x6 - Packet Delay: indicates PM meta data contains info for packet delay. Specified SF(s) SHOULD participate in Packet Delay PM.

0x7 - SF Hop by Hop Packet Delay: indicates PM meta data contains info for packet delay. All the SF(s) in the SFP SHOULD participate in the Packet Delay PM.

0x8 - SFF Hop by Hop Packet Delay: indicates PM meta data contains info for packet delay. All the SFF(s) in the SFP SHOULD participate in the Packet Delay PM.

0x9 - All Hop by Hop Packet Delay: indicates PM meta data contains info for packet delays. All the SF(s) and SFF(s) in the SFP SHOULD participate in the Packet Delay PM.

0xA - End to End: indicates PM meta data contains info for packet delay. Classifier MA and SF domain boundary SFF SHOULD participate in the Packet Delay PM.

Additional values can be added for future PM types and scenarios.

Length: Length of the variable PM metadata, in 4-byte words. It will have a minimum length of 1 (mention the measurement window index).

Measurement Window Index: 16 bit number to denote the current measurement window to which the packet belongs.

Reserved: Reserved 16 bits for future purpose.

SI: List of participating Service functions in the SFP. It SHOULD be in decreasing order of the SI for optimized traversal of the SI participation.

If the number of participating SF is not multiples of 4, then the last word of the PM context header needs to pad 0x0 to make it 4 byte aligned.

The Length of the Packet Delay PM context headers is fixed to 1 as described below, when the PM Type is 2 to 5. Since all SF’s in the SFP has to perform the Delay measurement specified in PM Type field.
2.3 Performance Measurement Flow ID

The method of encoding the PMF id is done using the flow id defined in [I-D.ietf-sfc-nsh].

Sample encoding of PMF using flow ID

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Ver|O|C|R|R|R|R|   Length  |  MD-type=0x2  | Next Protocol |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Service Path ID |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| TLV Class=IANA   |C|  Type=0x2   |R|R|   Len   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Measurement Window Index |   Reserved   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| TLV Class = 0x0   |C|  Type=0x7   |R|R|   L=0x1 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Performance Measurement Flow ID |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
3 Operations

For packet delay measurement all the MA’s time should be synchronized from a centralized stable clock. Among many available time synchronization protocols all MAs’s must use same protocol and synchronize time to the same centralized stable clock. Selection of type of time synchronization protocol is for MAs is out of scope of this document.

Every MA needs to maintain the packet delay statistics which they immediately report it to collector or may accumulate and report to collector at a reporting interval which depends on local policy.

3.1 Delay Measurement Options

There are 2 options for packet delay measurement

Option 1: Delay Measurement is performed using single packet in the Measurement Window.

Option 2: Delay Measurement is performed using multiple packet in a measurement window.

In case of Option 2 is used for delay measurement, MA sums the measured time for the received window and reports the summed time together with the count of samples to the Collector. Collector will co-relate the time info received across multiple report and compute the average time for a window.

3.2 Packet Delay Parameters for Option 1

Parameters for Option 1:

1) Rx-Time[P][W]

2) Tx-Time[P][W]

These parameters will be created and maintained at every MA

Rx-Time[P][W]: This 2 dimensional Parameter is maintained at every MA. In this P stands for PMF ID and W stands for Window ID. It will contain the system time (NTP UTC time format) when packets is received for a given PMF ID + Window.

PMF ID is unique within a SFC Domain, for a given PMF ID there could be multiple Windows. This Parameter is created when a Packet Delay PM packet is received at a first time within a Reporting interval. The Parameter is deleted upon expiry of Reporting interval at which it
will be sent to Collector and deleted from MA.

**Tx-Time[P][W]:** This 2 dimensional Parameter is maintained at every MA. In this P stands for PMF ID and W stands for Window ID. It will contain the system time (NTP UTC time format) when packets is sent out for a given PMF ID + Window.

PMF ID is unique within a SFC Domain, for a given PMF ID there could be multiple Windows. This Parameter is created when a Packet Delay PM packet is sent at a first time within a Reporting interval. The Parameter is deleted upon expiry of Reporting interval at which it will be sent to Collector and deleted from MA.

### 3.3 Packet Delay Parameters for Option 2

Parameter for Option 2:

1) **Rx-Sum-Time[P][W]**
2) **Rx-Sample-Count[P][W]**
3) **Tx-Sum-Time[P][W]**
4) **Tx-Sample-Count[P][W]**

These parameters will be created and maintained at every MA

**Rx-Time[P][W]:** This 2 dimensional Parameter is maintained at every MA. In this P stands for PMF ID and W stands for Window ID. It will contain the sum of system time when packets are received for a given PMF ID + Window.

PMF ID is unique within a SFC Domain, for a given PMF ID there could be multiple Windows. This Parameter is created when a Packet Delay PM packet is received at a first time within a Reporting interval. The Parameter is deleted upon expiry of Reporting interval at which it will be sent to Collector and deleted from MA.

**Rx-Sample-Count[P][W]:** This 2 dimensional Parameter is maintained at every MA. In this P stands for PMF ID and W stands for Window ID. It will contain the count of the PM Delay Sample packets received for a given PMF ID + Window.

**Tx-Time[P][W]:** This 2 dimensional Parameter is maintained at very MA. In this P stands for PMF ID and W stands for Window ID. It will contain the sum of system time when packets are sent for a given PMF ID + Window.
PMF ID is unique within a SFC Domain, for a given PMF ID there could be multiple Windows. This Parameter is created when a Packet Delay PM packet is sent at a first time within a Reporting interval. The Parameter is deleted upon expiry of Reporting interval at which it will be sent to Collector and deleted from MA.

Tx-Sample-Count[P][W]: This 2 dimensional Parameter is maintained at every MA. In this P stands for PMF ID and W stands for Window ID. It will contain the count of the PM Delay Sample packets sent for a given PMF ID + Window.

3.4 Initiating a Packet Delay Measurement

Measurement Controller Programs the PM Instance at the Classifier. Within the PM Schedule as programmed in PM Instance, Classifier starts the packet classification (based on the classification rule programmed in PM Instance). The packet classification policy may be customer/network/service specific.

3.5 Encapsulation of Classified Packets

For the classified packet, PM Context Header is prepared with following information

- Set the Type Class value to the IANA Allocated PM Class.
- Set the Type Value to the required Packet Delay Measurement type.
- If Type Value = 0x6
  - Encode the list of Service Index that need to participate in Packet Delay Measurement.
- If Value = 0x7, 0x8, 0x9, 0x10
  - No need to encode SI, since the type dictates the involved participating MA
- Set the Window Identifier as the current running Window number
- Encode the 32 bit globally unique PMF ID using the Flow Id Context Header as defined in [I-D quinn-sfc-nsh-tlv].
3.6 Incoming Packets Processing at MA

On receiving the packet with NSH header following operations are carried out:

Step 1: Detection of PM Context Header in a packet, by verifying the PM TLV Class as allocated by IANA. (If not detected, move to step 6)

Step 2: Check if PM Type field value is 6 to 10. (If not move to step 6).

Step 3: If PM Type Value = 7 to 10 move directly to Step 5

Step 4: Check Presence of self Service index in Service Index List (If not present, move to step 6)

Step 5: If Option 1 is used then go to Step 5a, if Option 2 is used then go to Step 5b

Step 5a: Record the value for PMF ID and Window ID; Record the time when the packet is received Rx-Time[P][W].

Step 5b: Record the value for PMF ID and Window ID; Record the time when the packet is received and sum it up in Rx-Sum-Time[P][W]. If Rx-Sum-Time[P][W] doesn’t exist then create it and initialize it with recorded time value.

Also Increment the value of statistics counter Rx-Sample-Count[P][W] by 1. If statistics counter doesn’t exist, create it and initialize it with 1.

Step 6: Packet is sent for Normal Processing
3.7 Outgoing Packets Processing at MA

At outgoing port of MA following operations are carried out.

Step 1: If Option 1 is used then go to Step 2a, if Option 2 is used then go to Step 2b

Step 2: If Option 1 is used then go to Step 2a, if Option 2 is used then go to Step 2b

Step 2a: Record the value for PMF ID and Window ID; Record the time when the packet is sent Tx-Time[P][W].

Step 2b: Record the value for PMF ID and Window ID; Record the time when the packet is sent and sum it up in Tx-Sum-Time[P][W]. If Tx-Sum-Time[P][W] doesn’t exist then create it and initialize it with recorded time value.

Also Increment the value of statistics counter Tx-Sample-Count[P][W] by 1. If statistics counter doesn’t exist, create it and initialize it with 1.

Step 3: Packet is sent for Normal Processing

3.8 MA Reporting the statistics to Collector

Reporting timer will run on Each MA. Consistency of this timer should be ensured across the entire MA in the SFP, ensuring the same is out of scope of this document.

On expiry of this timer following information needs to be sent to the Collector.

- Service Index
- PM Type Value
- Value of all the collected Rx-Time[P][W], Rx-Sum-Time[P][W], & Rx-Sample-Count[P][W] parameters along with the corresponding PMF ID and Window Id
- Value of all the collected Tx-Time[P][W], Tx-Sum-Time[P][W], & Tx-Sample-Count[P][W] parameters along with the corresponding PMF ID and Window Id

MA may delete the parameters after sending the same to collector.
3.9 Packet Delay Calculation at Collector

Collector accumulates the parameters per PMF per MA per Window.
Collector computes the Packet Delay using the accumulated parameters per PMF per MA per Window.

3.9.1 Packet Delay Calculation at Collector (Option 1)

Packet Delay between MA (for example from MA1 to MA2)
= (MA2)Rx-Time[P][W] - (MA1)Tx-Time[P][W]


3.9.1 Packet Delay Calculation at Collector (Option 2)

On receipt of report from a MA Collector performs following operation

1) For a PMF if it has Rx-Sum-Time[P][W]/Tx-Sum-Time[P][W] related for an already collected window, received value will be summed up with an existing one, otherwise a new entry is created.

2) For a PMF if it has Rx-Sample-Count[P][W]/Tx-Sample-Count[P][W] related for an already collected window, received value will be summed up with an existing one, otherwise a new entry is created.

3) Average Time computation by Collector for every MA
Average-Rx-Time[P][W] = Rx-Sum-Time[P][W]/Rx-Sample-Count[P][W]
Average-Tx-Time[P][W] = Tx-Sum-Time[P][W]/Tx-Sample-Count[P][W]

4) Packet Delay between MA (for example from MA1 to MA2)
= (MA2)Average-Rx-Time[P][W] - (MA1)Average-Tx-Time[P][W]

Packet Delay at MA
= (MA)Average-Tx-Time[P][W] - (MA)Average-Rx-Time[P][W]
4 Security Considerations

No specific security considerations for this document

5 IANA Considerations

5.1 TLV Class

IANA is requested to allocate a new class value for performance measurement from "Network Service Header (NSH) Parameters" registry.

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>Performance Measurement</td>
<td>[SFC-PM.arch]</td>
</tr>
</tbody>
</table>
6 References

6.1 Normative References

[I-D.ietf-sfc-nsh]


6.2 Informative References


Authors’ Addresses

Anil Kumar S N
Huawei Technologies India Pvt. Ltd,
Near EPIP Industrial Area,
Kundalahalli Village,
Whitefield,
Bangalore - 560037
EMail: anil.sn@huawei.com

Gaurav Agrawal
Huawei Technologies India Pvt. Ltd,
Near EPIP Industrial Area,
Kundalahalli Village,
Whitefield,
Bangalore - 560037
EMail: gaurav.agrawal@huawei.com

Vinod Kumar S
Huawei Technologies India Pvt. Ltd,
Near EPIP Industrial Area,
Kundalahalli Village,
Whitefield,
Bangalore - 560037
EMail: vinods.kumar@huawei.com