Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 24, 2014

M. Chen
H. Liu
Y. Yin
R. Papneja
Huawei Technologies
S. Abhyankar
Vodafone
G. Deng
CNNIC
Y. Huang
China Unicom

October 21, 2013

Coloring based IP Flow Performance Measurement Framework
draft-chen-ippm-coloring-based-ipfpm-framework-01

Abstract

By changing one or more bits of packets to "color" the packets into different color blocks, it naturally gives a way to measure the real packet loss and delay without inserting any extra OAM packets. This is called "coloring" based IP Flow Performance Measurement (IPFPM). This document specifies a framework for this "coloring" based IPFPM and defines a new application to the IPFIX for exporting the performance measurement statistic data.

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].

This Internet-Draft will expire on April 24, 2014.

Copyright Notice

Table of Contents

1. Introduction

Performance Measurement (PM) is an important tool that can not only provide Service Level Agreement (SLA) verification but facilitate in troubleshooting (e.g., fault localization or fault delimitation) and network visualization.

There are two typical types of performance measurement: one is active performance measurement, and the other is passive performance measurement.

In active performance measurement the receiver measures the injected packets to evaluate the performance of a path. The active measurement measures the performance of the extra injected packets, the rate, numbers and interval of the injected packets will largely affect the accuracy of the results. In addition, it also requires that the injected packets have to follow the same path as the real traffic; this normally cannot be guaranteed in the pure IP network. The One-Way Active Measurement Protocol (OWAMP) [RFC4656] and Two-Way Active Measurement Protocol (TWAMP) [RFC5357] are tools to enable active performance measurement.

In passive performance measurement, no artificial traffic is injected into the flow and measurements are taken to record the performance metrics of the real traffic. The Multiprotocol Label Switching (MPLS) PM protocol [RFC6374] for packet loss is an example of passive performance measurement. By periodically inserting auxiliary Operations, Administration and Maintenance (OAM) packets, the traffic is delimited by the OAM packets into consecutive blocks, and the receivers count the packets and calculate the packets loss each block.

But, when the OAM channel is in-band, solutions like [RFC6374] are not pure passive measurement as the OAM packets are inserted into the data stream. Furthermore because solutions like [RFC6374] depend on the fixed positions of the delimiting OAM packets for packets counting, they are vulnerable to out-of-order arrival of packets. This could happen particularly with out-of-band OAM channels, but might also happen with in-band OAM because of the presence of multipath forwarding within the network. Out of order delivery of data and the delimiting OAM can give rise to inaccuracies in the performance measurement figures. The scale of these inaccuracies will depend on data speeds and the variation in delivery, but with out-of-band OAM, this could result in significant differences between real and reported performance.

This document describes a mechanism where data packets are marked or "colored" so that they form blocks of data. No additional delimiting

OAM is needed and the performance can be measured in-service without the insertion of additional traffic. Furthermore, because coloring based IP performance measurement does not require extra OAM packets for traffic delimitation, it can be used in situations where there is packets re-ordering. This document specifies a framework for the "coloring" based IP Flow Performance Measurement (IPFPM) . This document also defines a new application of and some extensions to the IP Flow Information eXport (IPFIX) for exporting the performance measurement statistic data.

2. Overview and Concept

The concept of "coloring" IP packets for performance measurement is described in [I-D.tempia-opsawg-p3m]. Coloring of packets in a specific IP flow to different colors divides the flows into different consecutive blocks. Packets in a block have same color and consecutive blocks will have different colors. This enables the measuring node to count and calculate packet loss and/or delay based for each color block without any additional auxiliary OAM packets. The following figure (Figure 1) is an example that illustrates the different colors in a single IP flow in interleaved red and blue blocks.

```
| Red Block  | Blue Block | Red Block  | Blue Block |
 RRRRRRRRRRRR BBBBBBBBBBBB RRRRRRRRRRRR BBBBBBBBBBBB
```

Figure 1: Packet Coloring

For packet loss measurement, there are two ways to color packets: fixed packet numbers or fixed time period for each color block. This document considers only fixed time period. The sender and receiver nodes count the transmitted and received packets/octets based on each color block. By counting and comparing the transmitted and received packets/octets, the packet loss can be detected.

For packet delay measurement, there are two solutions. One is similar to the packet loss, that it still colors the IP flows to different color blocks and uses the time of the color change as the reference time for delay calculations. This solution requires that there must not be any out-of-order packets; otherwise, the result will not be accurate. Because it uses the first packet of each color block for delay measurement, if there is packet reordering, the first packet of each block at the sender will be probably different from the first packet of the block at the receiver. The alternate way is to periodically coloring a single packet in the IP flow. Within a given time period, there is only one packet that can be colored. The

sender records the timestamp when the colored packet is transmitted, the receiver records the timestamp when detecting the colored packet. With the two timestamps, the packet delay can be computed.

To make the above solutions work, two preconditions are required. The first is that there has to be a way to collect the packet counts and timestamps from the senders and receivers to a centralized calculation element. The IP Flow Information eXport (IPFIX) [RFC5101] protocol is used for collecting the performance measurement statistic information (Section 5 of this document). The second is that the centralized calculation element has to know exactly what packet pair counts (one from the sender and the other is from the receiver) are based on the same color block and a pair of timestamps (one from the sender and the other is from the receiver) are based on the same colored packet. The "Period Number" based solution (Section 4.2 of this document) is introduced to achieve this.

3. Reference Model and Functional Components

3.1. Reference Model

A Multipoint-to-Multipoint (MP2MP) reference model (as shown in Figure 2) is introduced in this document. For a specific IP flow, there may be one or more upstream and downstream Measurement Points (MPs). An IP flow can be identified by the Source IP (SIP) and Destination IP (DIP) addresses, and it may combine the SIP and DIP with any or all of the Protocol number, the Source port, the Destination port, and the Type of Service (TOS) to identify an IP flow. For each flow, there will be a flow identifier that is unique within a certain administrative domain.

An MP is a network node. From the measurement point of view, it consists of two parts (as shown in Figure 3): Data Collecting Point (DCP), and Target Logical Port (TLP). For an MP, there is only one DCP and may be one or more TLPs. The Measurement Control Point (MCP) is a centralized calculation element, MPs periodically report their measurement data to the MCP for final performance calculation. The report protocol is defined in Section 5 of this document.

The reason for choosing MP2MP model is that it can satisfy all the scenarios that include Point-to-Point (P2P), Point-to-Multipoint (P2MP), Multipoint-to-Point (MP2P), and MP2MP. P2P scenario is obvious and can be used anywhere. P2MP and MP2P are very common in mobile backhaul networks. For example, a Cell Site Gateway (CSG) multi-homing to two Radio Network Controller (RNC) Site Gateways (RSGs) is a typical network design. When there is a failure, there is a requirement to monitor the flows between the CSG and the two RSGs hence to determine whether the fault is in the transport network

or in the wireless network (typically called "fault delimitation"). This is especially useful in the situation where the transport network belongs to one service provider and wireless network belongs to other service providers.

```
                          +-----+
                   +------| MCP |------+
                   |      +-----+      |
         +-----+   |  +---/     \---+  |   +-----+
         | MP1 |---+  |             |  +---| MP3 |
         +-----+      |             |      +-----+
         +-----+      |             |      +-----+
         | MP2 |------+             +------| MP4 |
         +-----+                           +-----+
                 Figure 2: MP2MP based Model
```

```
                 +-----------------------+
                 |      +---------+      |
                 |      |   DCP   |      | Control Plane
                 |      +---------+      |
       ----------|-----/----------\-----|--------------
                 | +--+--+      +--+--+ |
                 | | TLP1|      | TLP2| | Data plane
                 | +-----+      +-----+ |
                 +-----------------------+
                 Figure 3: Measurement Point
```

3.2. Functional Components

3.2.1. Measurement Control Point

The MCP is responsible for calculating the final performance metrics according to the received measurement data from the MPs (actually from the DCPs). For packet loss, based on each color block, the difference between the total counts received from all upstream MPs and the total counts received from all downstream MPs are the lost packet numbers. The MCP must make sure that the counts from the upstream MPs and downstream MPs are related to the same color/packets block. For packet delay (e.g., one way delay), the difference between the timestamps from the downstream MP and upstream MP is the packet delay. Similarly to packet loss, the MCP must make sure the two timestamps are based on the same colored packet.

This document introduces a Period Number (PN) based synchronization mechanism to help the MCP to determine whether any two or more packet counts (from distributed MPs) are related to the same color block or

any two timestamps are related to the same colored packet. The PN is generated each time a DCP reads the packet counts or timestamps from the TLP, and is equal to the modulo of the local time (when the counts or timestamps are read) and the interval of the color time period. Each packet count and timestamp has a PN when reported to the MCP, and the same PN means that they are related to the same color block or colored packet. This requires that the upstream and downstream MPs having a certain time synchronization capability (e.g., supporting the Network Time Protocol (NTP) [RFC5905], or the IEEE 1588 Precision Time Protocol (PTP) [IEEE1588].) and assumes that the upstream and downstream MPs have already time synchronized. Since there is the intention to measure packet delay, this requirement for time synchronization is already present.

3.2.2. Data Collecting Point

The DCP is responsible for periodically collecting the measurement data from the TLPs and reporting the data to the MCP. In addition, when to change the color, when to color a packet (for packet delay measurement), and when to read the packet counts and timestamps are also controlled by DCP. Each DCP will maintain two timers, one (C-timer, used at upstream DCP) is for color changing, the other (R-timer, used at downstream DCP) is for reading the packet counts and timestamps. The two timers have the same time interval but are started at different times. A DCP can be either an upstream or a downstream DCP: the role is specific to an IP flow. For a specific IP flow, the upstream DCP will change the color and read the packet counts and timestamps when the C-timer expires, the downstream DCP just reads the packets counts and timestamps when the R-timer expires.

In order to allow for a certain degree of packets re-ordering, the R-timer should be started later than a defined period of time (Δt) after the C-timer is started, so as long as the delta-t satisfies the following conditions:

(Time-L - Time-MRO) < Δt < (Time-L + Time-MRO)

where:

Time-L: the link delay time between the sender and receiver;

Time-MRO: the maximum re-ordering time difference; if a packet is expected to arrive at t1 but actually arrives at t2, then the Time-MRO = | t2 - t1|.

So, the R-timer should be started at "t + Δt" (where the t is the time when C-timer started).

For simplicity, the C-timer should be started at the beginning of each time period. This document recommends the implementation to support at least these time periods (1s, 10s, 1min, 10min and 1h). So, if the time period is 10s, the C-timer should be started at the time of any multiples of 10 in seconds (e.g., 0s, 10s, 20s, etc.), then the R-timer should be started (e.g., 0s+Δt, 10s+Δt, 20s+Δt, etc.). With this method, each DCP can independently start its C-timer and R-timer given that the clocks have been synchronized.

3.2.3. Target Logical Port

The TLP is a logical entity that actually executes the final measurement actions (e.g., colors the packets, counts the packets, records the timestamps, etc.). Normally, a physical interface corresponds to a TLP, and the TLP resides in the data plane. For a measurement instance (corresponding to an IP flow), a TLP will maintain a pairs of packet counters and a timestamp counter for each color block. As for the pair of packet counters, one is for counting packets and the other is for counting octets.

4. Principles of Coloring based IPFPM

To simplify the process description, the flows discussed in this document are all unidirectional. A bidirectional flow can be seen as two unidirectional flows. For a specific flow, there will be upstream and downstream TLPs and upstream and downstream packet counts/timestamp accordingly.

4.1. Packet Loss Measurement

For packet loss measurement, this document defines the following counters and quantities:

U-CountP[n][m]: U-CountP is a two-dimensional array that stores the number of packets transmitted by each upstream TLP in each color time period. Specifically, parameter "n" is the "period number" of measured color blocks while parameter "m" refers to the m-th TLP of the upstream TLPs.

D-CountP[n][m]: D-CountP is a two-dimensional array that stores the number of packets received by each downstream TLP in each color time period. Specifically, parameter "n" is the "period number" of measured color blocks while parameter "m" refers to the m-th TLP of the downstream TLPs.

U-CountO[n][m]: U-CountO is a two-dimensional array that stores the number of octets transmitted by each upstream TLP in each color time

period. Specifically, parameter "n" is the "period number" of measured color blocks while parameter "m" refers to the m-th TLP of the upstream TLPs.

D-CountO[n][m]: D-CountO is a two-dimensional array that stores the number of octets received by each downstream TLP in each color time period. Specifically, parameter "n" is the "period number" of measured color blocks while parameter "m" refers to the m-th TLP of the downstream TLPs.

LossP: the number of packets transmitted by the upstream TLPs but not received at the downstream TLPs.

LossO: the total octets transmitted by the upstream TLPs but not received at the downstream TLPs.

The total packet loss of a flow can be computed as follows:

LossP = U-CountP[1][1] + U-CountP[1][2] + ... + U-CountP[n][m] - D-CountP[1][1] - D-CountP[1][2] - .... - D-CountP[n][m'].

LossO = U-CountO[1][1] + U-CountO[1][2] + ... + U-CountO[n][m] - D-CountO[1][1] - D-CountO[1][2] - .... - D-CountO[n][m'].

Where the m and m' are the number of upstream TLPs and downstream TLPs, respectively.

4.2. Packet Delay Measurement

For packet delay measurement, there will be only one upstream TLP and may be one or more (P2MP) downstream TLPs. Although the coloring based IPFPM supports P2MP model, this document only discusses P2P model, the P2MP model is left for future study. This document defines the following timestamps and quantities:

U-Time[n]: U-Time is a one-dimension array that stores the time when colored packets are sent; parameter "n" is the "period number" of colored packets.

D-Time[n]: D-Time is a one-dimension array that stores the time when colored packets are received; parameter "n" is the "period number" of colored packets. This is only for P2P model.

D-Time[n][m]: D-Time a two-dimension array that stores the time when the colored packet is received by downstream TLPs at each color time period. Here, parameter "n" is the "period number" of colored packets while parameter "m" refers to the m-th TLP of the downstream TLPs. This is for P2MP model which is left for future study.

One-way Delay[n]: The one-way delay metric for packet networks is described in [RFC2679]. The "n" identifies the "period number" of the colored packet.

One-way Delay[1] = D-Time[1] - U-Time[1].

One-way Delay[2] = D-Time[2] - U-Time[2].

...

One-way Delay[n] = D-Time[n] - U-Time[n].

In the case of two-way delay, the delay is the sum of the two one-way delays of the two flows that have the same TLPs but have opposite directions.

Two-way Delay[1] = (D-Time[1] - U-Time[1]) + (D-Time'[1] - U-Time'[1]).

Two-way Delay[2] = (D-Time[2] - U-Time[2]) + (D-Time'[2] - U-Time'[2]).

...

Two-way Delay[n] = (D-Time[n] - U-Time[n]) + (D-Time'[n] - U-Time'[n]).

Where the D-Time and U-Time are for one forward flow, the D-Time' and U-Time' are for reverse flow.

5. Consideration on Color Bits Selection

This document does not specify which bits in IP header should be used for coloring; it primarily introduces options that the operator can choose for packet coloring. This document introduces the following options:

1. For IPv4, there is only one bit (the last reserved bit of the Flag field of the IPv4 header) that can be used for coloring. With one bit, at any time it can only be used for loss or delay measurement and cannot be used for packet loss and delay measurement simultaneously;

2. For IPv6, it can leverage the IPv6 extension header for coloring, for example, adding a new option to the Hop-by-Hop Options header[RFC2460] for coloring. More detail will be added in a future version or in a separate document.

For the above options, the operators should carefully think of the color bits selection to make sure that the setting or changing of the color bits SHOULD NOT affect the normal packet forwarding and process.

The implementations SHOULD provide knobs for operators to configure and change the color bits according to their network design and policies.

6. Statistic Information Report

As described in Section 4 of this document, during the performance measurement period, each DCP periodically reports performance measurement statistic information to the MCP, and the MCP will compute the final performance measurement results according to the received statistic information.

For a specific IP flow, for either packet delay or loss measurement, there will be at least one upstream and one downstream DCP. For accurate measurements, time synchronization is required and the Period Number is used by MCP to uniquely identify and correlate the packet counts/ timestamps between the upstream and downstream DCPs for a specific color block or colored packet.

For packet loss measurement, the following information is required to report to the MCP:

DCP Identifier

TLP Identifier

Flow Identifier

Period Number

Packet Number Count

Packet Octets Count

For packet delay measurement, the following information is required to report to the MCP:

DCP Identifier

TLP Identifier

Flow Identifier

Period Number

Timestamp

In addition, a DCP may report some status (e.g., whether a DCP is time synchronized) to the MCP, hence to help the MCP to determine whether measurement data from a DCP is valid and can be used for computation. The following information may be included:

DCP Identifier

DCP Status

TLP Status

6.1. IPFIX for Coloring based IPFPM

The IPFIX protocol [RFC5101] defines how IP Flow information can be exported from routers, measurement probes, or other devices. It defines many Information Elements [RFC5102] that can be used to carry and export the above information from DCPs to MCP. Except the Period Number, DCP Status and TLP Status, all the above information can be identified with the existing Information Elements.

DCP Identifier: exporterIPv4Address/exporterIPv6Address (130/131)

TLP Identifier: portId (142)

Flow Identifier: flowId (148)

Packet Number Count: packetTotalCount (86)

Packet Octets Count: octetTotalCount (85)

Timestamp: flowStartMicroseconds (154)

6.1.1. Information Element for IPFPM

6.1.1.1. periodNumber

Description: The periodNumber is used to identify a packet count or timestamp that belongs to a specific color block or colored packet. The MCP uses it to determine whether any two or more packet counts (from distributed DCPs) are related to the same color block or any two timestamps are related to the same colored packet. The PN is generated each time a DCP reads the packet counts and timestamp from the TLP, and is equal to the modulo of the local time (when the counts and timestamps are read) and the interval of the color time period.

Abstract Data Type: unsigned32

ElementId: TBD1

Status: current

6.1.1.2. dcpStatus

Description: The dcpStatus is used to carry some status of a DCP (For example, whether a DCP has already time synchronized).

Abstract Data Type: unsigned16

ElementId: TBD2

Status: current

The dcpStaus is defined as follows:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Reserved              |T|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

One bit (the Time synchronized bit) is defined in this document, it is used to indicate whether a DCP is time synchronized. When the T bit set, the DCP is time synchronized, otherwise, the DCP is not time synchronized. The MCP MUST calculate the results when all related DCPs of a flow are time synchronized, otherwise, the results will not correct.

6.1.1.3. tlpStatus

Description: The tlpStatus is used to carry some status of a TLP.

Abstract Data Type: unsigned8

ElementId: TBD3

Status: current

The dcpStaus is defined as follows:

```
+-+-+-+-+-+-+-+-+
|   Reserved  |U|
+-+-+-+-+-+-+-+-+
```

One bit (the Upstream TLP bit) is defined in this document; it is used to indicate whether a TLP is the upstream or downstream TLP for an IP flow. When the U bit set, the TLP is the upstream TLP of the flow; otherwise, the TLP is the downstream TLP of the flow.

6.1.2. Templates for IPFPM

6.1.2.1. DCP Status Options Template

The DCP Status Options Template is used to report the status of a DCP; it SHOULD contain the following Information Elements:

meteringProcessId (scope) [IPFIX-IANA]

DCP Identifier

This is the identifier of a DCP, any of the following Information Elements can be used:

exporterIPv4Address [IPFIX-IANA]

exporterIPv6Address [IPFIX-IANA]

dcpStatus

The dcpStatus is as defined in Section 6.1.1.2of this document.

The Data Records specified by the DCP Status Options Template SHOULD be exported once the IPFIX session established or when status changed.

An example of the DCP Status Options Template Set is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
|         Set ID = 3            |          Length = 24          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Template ID XXX         |        Field Count = 3        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Scope Field Count = 1     |0|   meteringProcessId = 143   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Scope 1 Field Length = 4    |0|  exporterIPv4Address = 130  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Scope 1 Field Length = 4    |0|      dcpStatus = TBD3       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 1        |            Padding            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

6.1.2.2. Flow Options Template

The Flow Options Template is used to report all the configured flows on a DCP. It SHOULD include the following the Information Elements:

meteringProcessId (scope) [IPFIX-IANA]

TLP Identifier

The identifier of the TLP that is responsible for measuring the flow; portId is used to carry the TLP Identifier:

portId [IPFIX-IANA]

TLP Name

The name of the TLP that is responsible for measuring the flow; interfaceName is used to carry the TLP Name:

interfaceName [IPFIX-IANA]

tlpStatus [Section 6.1.1.3]

flowId [IPFIX-IANA]

protocolIdentifier [IPFIX-IANA]

Source IP address

The source IP address or prefix of an IP flow, for this address, any of the following Information Elements can be used:

sourceIPv4Address [IPFIX-IANA]

sourceIPv6Address [IPFIX-IANA]

sourceIPv4Prefix [IPFIX-IANA]

sourceIPv6Prefix [IPFIX-IANA]

Source IP prefix length

The source IP prefix length of a prefix, any of the following Information Elements can be used:

sourceIPv4PrefixLength [IPFIX-IANA]

sourceIPv6PrefixLength [IPFIX-IANA]

Source port

The source port of an IP flow, any of the following Information Elements can be used:

udpSourcePort [IPFIX-IANA]

tcpSourcePort [IPFIX-IANA]

Destination IP address

The destination IP address or prefix of an IP flow, for this address, any of the following Information Elements can be used:

destinationIPv4Address [IPFIX-IANA]

destinationIPv6Address [IPFIX-IANA]

destinationIPv4Prefix [IPFIX-IANA]

destinationIPv6Prefix [IPFIX-IANA]

Destination IP prefix length

The destination IP prefix length of a prefix, any of the following Information Elements can be used:

destinationIPv4PrefixLength [IPFIX-IANA]

destinationIPv6PrefixLength [IPFIX-IANA]

Destination port

The destination port of an IP flow, any of the following Information Elements can be used:

udpDestinationPort [IPFIX-IANA]

tcpDestinationPort [IPFIX-IANA]

The Data Records specified by the Flow Options Template SHOULD be exported once the IPFIX session established or when the configured flows changed (e.g., a new flow is added for measurement or a flow deleted to stop the measurement).

An example of the Flow Options Template Set is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Set ID = 3            |          Length = 52          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Template ID XXX         |        Field Count = 10       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Scope Field Count = 1     |0|  meteringProcessId=143      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Scope 1 Field Length = 4    |0|  portId = 142               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 4        |0|  interfaceName = 82         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 4        |0|  tlpStatus = TBD3           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 1        |0|  flowID = 148               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 4        |0|  protocolIdentifier = 4     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 1        |0|  sourceIPv4Address = 8      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 4        |0|  udpSourcePort = 4          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 2        |0|  destinationIPv4Address = 4 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 4        |0|  udpDestinationPort = 4     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 4        |0|  udpDestinationPort = 4     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Field Length = 2        |           Padding             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

6.1.2.3. Packet Loss Template

The Packet Loss Template is used by a DCP to report the packet loss measurement statistic of a flow to the MCP; it SHOULD contain the following Information Elements:

TLP Identifier

This is the identifier of a TLP, portId is used to carry the TLP Identifier:

portId[IPFIX-IANA]

flowId[IPFIX-IANA]

periodNumber [Section 6.1.1.1]

packetTotalCount[IPFIX-IANA]

octetTotalCount[IPFIX-IANA]

An example of the Packet Loss Data Set is as follows:

```
     0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Set ID = 2           |      Length = 33 octets       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        Template ID XXX        |       Field Count = 7         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0|        portId = 142         |       Field Length = 4        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0|        flowId = 148         |       Field Length = 4        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0|      periodNumber = TBD1    |       Field Length = 4        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0|    packetTotalCount = 86    |       Field Length = 8        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0|    octetTotalCount = 85     |       Field Length = 8        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The portId is used to identify and carry the TLP ID.

The flowId is a identifier that is unique within a specific administrative domain (e.g., an Autonomous System). The TLP, DCP and MCP have to agree a flow identifier related to a specific flow. For

example, the flow identifier can be generated and maintained by a centralized element. How to generate and maintain the flowId is out the scope of this document.

The flowId has the following structure, the Reserved field that is left for future extensions, the Identifier field is 24-bit in length.

```
     0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Reserved    |                  Identifier                   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The periodNumber is as defined in Section 6.1.1.1 of this document.

The packetTotalCount is used to carry the total transmitted/received packets of a flow since the measurement start.

The octetTotalCount is used to carry the total transmitted/received octets of a flow since the measurement start.

6.1.2.4. Packet Delay Template

The Packet Delay Template is used by a DCP to report the packet delay measurement statistic of a flow to the MCP; it SHOULD contain the following Information Elements:

TLP Identifier

  This is the identifier of a TLP, portId is used to carry the TLP Identifier:

  portId [IPFIX-IANA]

flowId [IPFIX-IANA]

periodNumber [Section 6.1.1.1]

timestamp

  The time when color a packet, flowStartMicroseconds is used to carry the timestamp:

  flowStartMicroseconds[IPFIX-IANA]

An example of the Packet Delay Data Set is as follows:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         Set ID = 2            |      Length = 25 octets       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |       Template ID 258         |       Field Count = 6         |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0|       portId = 142          |       Field Length = 4        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0|       flowId = 148          |       Field Length = 4        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0|     periodNumber = TBD1     |       Field Length = 4        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |0| flowStartMicroseconds = 154 |       Field Length = 8        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The flowId is used to carry the flow identifier of a flow; the structure is defined in Section 6.1.2.3 of this document.

The periodNumber is as defined in Section 6.1.1.1 of this document.

The flowStartMicroseconds is used to carry the timestamp of a colored packet of a specific flow.

7. IANA Considerations

The IANA is required to allocate 3 new Information Elements codes for the Information Elements defined in Section 6.1.1 from the IPFIX Information Elements registry.

8. Security Considerations

This document specifies a passive mechanism for measuring packet loss and delay within a Service Provider's network where the IP packets are marked or "colored" with the unused bits in IP head field, and then inserting additional OAM packets during the measurement is avoided. Obviously, such mechanism does not directly affect other applications running on the Internet but may lead to potential affects to the measurement itself.

First, the measurement itself may be affected by routers (or other network devices) along the path of IP packets intentionally altering the value of color bits of packets. Just as mentioned before, the mechanism specified in this document is just in the context of one Service Provider's network, so the routers (or other network devices) are controllable and thus this kind of attack can be omitted.

Second, the measurement can be harmed by attackers injecting artificial traffic. Then authentication techniques, like digital signatures, may be used to guard against such kind of attack.

9. Acknowledgements

The authors would like to thank Adrian Farrel for his review, suggestion and comments to this document.

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC5101] Claise, B., "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information", RFC 5101, January 2008.

[RFC5102] Quittek, J., Bryant, S., Claise, B., Aitken, P., and J. Meyer, "Information Model for IP Flow Information Export", RFC 5102, January 2008.

10.2. Informative References

[I-D.tempia-opsawg-p3m]
Capello, A., Cociglio, M., Castaldelli, L., and A. Bonda, "A packet based method for passive performance monitoring", draft-tempia-opsawg-p3m-03 (work in progress), February 2013.

[IEEE1588]
IEEE, "1588-2008 IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", March 2008.

[IPFIX-IANA]
, "http://www.iana.org/assignments/ipfix/ipfix.xml", .

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[RFC2679] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Delay Metric for IPPM", RFC 2679, September 1999.

[RFC3260] Grossman, D., "New Terminology and Clarifications for Diffserv", RFC 3260, April 2002.

[RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)", RFC 4656, September 2006.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, October 2008.

[RFC5905] Mills, D., Martin, J., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, June 2010.

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, September 2011.

Authors' Addresses

Mach(Guoyi) Chen
Huawei Technologies

Email: mach.chen@huawei.com

Hongming Liu
Huawei Technologies

Email: liuhongming@huawei.com

Yuanbin Yin
Huawei Technologies

Email: yinyuanbin@huawei.com

Rajiv Papneja
Huawei Technologies

Email: Rajiv.Papneja@huawei.com

Shailesh Abhyankar
Vodafone
Vodafone House, Ganpat Rao kadam Marg Lower Parel
Mumbai 40003
India

Email: shailesh.abhyankar@vodafone.com

Guangqing Deng
CNNIC
4 South 4th Street, Zhongguancun, Haidian District
Beijing
China

Email: dengguangqing@cnnic.cn

Yongliang Huang
China Unicom

Email: huangyl@dipmt.com