Abstract

By setting one unused bit of the IP header 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 and protocol for this "coloring" based IPFPM.

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].

Status of this Memo

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

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

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

This Internet-Draft will expire on August 29, 2013.

Copyright Notice

Copyright (c) 2013 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
Table of Contents

1. Introduction .................................................. 3
2. Overview and Concept ......................................... 4
3. Reference Model and Functional Components ................... 5
   3.1. Reference Model ........................................... 5
   3.2. Functional Components .................................... 6
      3.2.1. Measurement Control Point ............................ 6
      3.2.2. Data Collecting Point ................................. 7
      3.2.3. Target Logical Port ................................... 7
4. Principles of Coloring based IPFPM ............................ 8
   4.1. Packet Loss Measurement ................................... 8
   4.2. Packet Delay Measurement .................................. 9
5. Statistic Information Report ................................... 10
6. IPFIX for coloring based IPFPM ............................... 11
   6.1. Information Element for IPFPM ............................ 12
   6.1.1. DCP Status Template Set ................................. 13
   6.1.2. Packet Loss Template Set ............................... 13
   6.1.3. Packet Delay Template Set .............................. 15
7. IANA Considerations .......................................... 15
8. Security Considerations ....................................... 16
9. Acknowledgements ............................................. 16
10. References .................................................. 16
    10.1. Normative References ................................... 16
    10.2. Informative References ................................. 16
Authors’ Addresses .............................................. 17
1. Introduction

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

There are two 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 and protocol for the "coloring" based IP Flow Performance Measurement (IPFPM).

## 2. Overview and Concept

The concept of "coloring" IP packets for performance measurement is described in [I-D.tempia-opsawg-p3m]. By "coloring" the packets of a specific IP flow to different colors, it naturally splits the IP flow into different consecutive blocks.

For packet loss measurement, there are two ways to color packets: fixed packet numbers or fixed time period for each color block. This document only talks about the way of fixed time period. The sender and receiver nodes count the transmitted and received packets/octets based on each color block. By collecting and comparing the transmitted and received packets/octets, it can easily detect whether there is packet loss and how many packets/octets get lost.

For packet delay measurement, there are two solutions. One is similar to packet loss, it still colors the IP flow to different color blocks and uses the time when color changing as the reference time for delay calculation. 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 other way is to periodically color a single packet of the IP flow. Within a time period, there is only one packet 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 conditions are required. The first one is that there have 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 what exactly a pair of packet counts (one from the sender and the other 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

An Multipoint-to-Multipoint (MP2MP) reference model (as shown in Figure 1) 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 2): 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 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 (this is normally 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.
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 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 and timestamps 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. 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 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 for 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 reordering, the R-timer should be started later than a defined period of time after the C-timer is started (e.g., 1/3 or 2/3 T, where T is the interval of the C-timer). It recommends that: for packet loss measurement, the R-timer should be started at 1/3 T after the C-Timer is started, and for packet delay measurement, the R-timer should be started at 2/3 T after the C-Timer is started.

To make the implementation simple, 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, for example, at the time of T+1/3 or 2/3 T. 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. One packet counter is for counting packets and the other is for counting octets.

4. Principles of Coloring based IPFPM

The flows discussed in this document are all unidirectional. 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 identifies the packets transmitted by a upstream TLP, the "n" identifies the "period number" of the measured color block, the "m" identifies the No. m TLP of the upstream TLPs.

D-CountP[n][m]: U-CountP identifies the packets received by a downstream TLP, the "n" identifies the "period number" of the measured color block, the "m" identifies the No. m TLP of the downstream TLPs.

U-CountO[n][m]: U-CountO identifies the octets transmitted by a upstream TLP, the "n" identifies the "period number" of the measured color block, the "m" identifies the No. m TLP of the upstream TLPs.

D-CountO[n][m]: U-CountO identifies the packets received by a downstream TLP, the "n" identifies the "period number" of the measured color block, the "m" identifies the No. m 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 the total packet loss of a flow can be computed as follows:

\[
\text{LossP} = \text{U-CountP}[1][1] + \text{U-CountP}[1][2] + ... + \text{U-CountP}[n][m] - \text{D-CountP}[1][1] - \text{D-CountP}[1][2] - ... - \text{D-CountP}[n][m'] .
\]

\[
\text{LossO} = \text{U-CountO}[1][1] + \text{U-CountO}[1][2] + ... + \text{U-CountO}[n][m] - \text{D-CountO}[1][1] - \text{D-CountO}[1][2] - ... - \text{D-CountO}[n][m'] .
\]
Where the m is the number of the upstream TLPs, and the m' is the number of the downstream TLPs.

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 identifies the time when sent a colored packet, the "n" identifies the "period number" of the colored packet.

D-Time[n]: D-Time identifies the time when received a colored packet, the "n" identifies the "period number" of the colored packet. This is only for P2P model.

D-Time[n][m]: D-Time identifies the time when received a colored packet, the "n" identifies the "period number" of the colored packet, the "m" identifies the No. m TLP of the downstream TLPs. This is for P2MP model and 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.


...


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[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. Color Bits Selection

This document does not specify which bits should be used for coloring, it just introduces some options that the operators can select for packet coloring usage. There are not too much bits in the IP header that can be used for IP packet coloring, this document introduces two options:

One is to use the reserved bit of the Flag field. There is only one bit that is left, so it cannot be used for packet loss and delay measurement simultaneously.

The other option is to use some "unused" bits of the Type Of Service (TOS) field.

For both 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 some 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, when the performance measurement started, each DCP will periodically report 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, whatever for packet delay or loss measurement, there will be at least one upstream and one downstream DCP. As described above section, time synchronization is required and the Period Number is used for MCP to identify and correlate the packet counts and timestamps from the upstream and downstream DCPs to a specific color block or colored packet.

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

In addition, a DCP may report some status and statistic information to the MCP, the following information may be included:

- DCP Identifier
- DCP Status
- Others

### 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 DCP to MCP. Except the Period Number and DCP Status, all the above information can be identified with the existing Information Elements.

- **DCP Identifier**: exporterIPv4Address/exporterIPv6Address (130/131)
- **TLP Identifier**: meteringProcessId (143)
Flow Identifier: flowId (148)
Packet Number Count: packetTotalCount (86)
Packet Octets Count: octetTotalCount (85)
Timestamp: flowStartMilliseconds (152)

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: TBD
Status: current

6.1.1.2. dcpStatus

Description: The dcpStatus is used to carry some status of the DCP, for example, whether the DCP has already time synchronized.

Abstract Data Type: unsigned8
ElementId: TBD
Status: current

The dcpStaus is defined as follows:

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

One bit (Time synchronized bit) is defined in this document, it is used to indicate whether a DCP is time synchronized. When 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.

6.1.2. DCP Status Template Set

The DCP Status Template is an Option Template Set, it is used to report the status and statistic information of the DCP to MCP.

```
|         Set ID = 3          |          Length = 24          |
|-----------------------------------------------|
|       Template ID 256          |        Field Count = 3        |
|-----------------------------------------------|
|     Scope Field Count = 1      |0|      dcpStatus = TBD        |
|-----------------------------------------------|
|   Scope 1 Field Length = 1    |0|exportedMessageTotalCount=41 |
|-----------------------------------------------|
|   Scope 2 Field Length = 2    |0|exportedFlowRecordTotalCo.=42 |
|-----------------------------------------------|
|       Field Length = 2        |           Padding             |
```

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

The exportedMessageTotalCount field is used to report how many IPFIX messages that the DCP has sent to the MCP.

The exportedFlowRecordTotalCount is used to report how many IPFIX flow records that the DCP has sent to the MCP.

6.1.3. Packet Loss Template Set

The Packet Loss Template Set is a Data Set, it is used to report the packet loss measurement statistic of a flow to the MCP.

The Data Set will be as follows:
The exporterIPv4Address is used to carry the DCP ID.

The meteringProcessId is used to 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 of the scope of this document.

The flowId has the following structure, it consists of two parts: one the Reserved field that is left for future extensions, the other is the Identifier field is 24-bit in length.

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.

6.1.4. Packet Delay Template Set

The Packet Delay Template Set is a Data Set, it is used to report the packet delay measurement statistic of a flow to the MCP.

The Data Set will be as follows:

```
0                   1                   2                   3
+---------------------------+---------------------------+
| Set ID = 2                | Length = 24 octets       |
+---------------------------+---------------------------+
| Template ID 258           | Field Count = 5          |
+---------------------------+---------------------------+
| exporterIPv4Address = 130 | Field Length = 4         |
| meteringProcessId = 143   | Field Length = 4         |
| flowId = 148              | Field Length = 4         |
+---------------------------+---------------------------+
| PeriodNumber = TBD        | Field Length = 4         |
+---------------------------+---------------------------+
| flowStartMilliseconds = 152| Field Length = 8         |
```

The exporterIPv4Address is used to carry the DCP ID.

The meteringProcessId is used to carry the TLP ID.

The flowId is used to carry the flow identifier of a flow, the structure is defined in Section 6.1.3 of this document.

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

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

7. IANA Considerations

TBD.
8. Security Considerations

TBD.

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


10.2. Informative References


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