Network working group
Internet Draft
Intended status: Proposed Standard
Expires: January 12, 2011
July 12, 2010

H. Liu
M. Chen
L. Zheng
Huawei Technologies

IP Multicast Inline Real Stream Monitoring

draft-liu-mboned-multicast-realstream-monitor-02.txt

Status of this Memo

Abstract

This document defines an efficient IP multicast performance monitoring method through packet loss and packet delay measurement. It has the characteristics of monitoring real IP multicast stream with the measurement packets following the actual multicast forwarding path and it enables the fault detection and isolation in IP multicast network.

Table of Contents

1. Introduction

With the deployment of IP video multicast service, there is an increasing demand for the performance monitoring for providers' multicast network. The benefits of performance monitoring are to guarantee the service level agreement (SLA) provided to the customers, to discover the network performance defects proactively, to react in response to the failure quickly, and further to optimize the network resources utilizations.

This document describes an IP multicast network performance monitoring solution referred to as Inline Real Stream Monitoring (IRSM) based on the requirements given in [3]. IRSM is proposed to meet the service provider's manageability requirements on increasingly deployed multicast network. It enables efficient measurement of performance metrics of a multicast channel and provides diagnostic information in case of performance degradation or failure.

2. Conventions used in this document

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.

2.1. Terminologies

IRSM: Inline Real Stream Monitoring

(S,G): a source address and group address pair to identify a multicast forwarding channel or multicast forwarding state.

(*,G): a notation for multicast forwarding state to receive from all the sources sending to this group.

MEG: Maintenance Entity Group

MEP: MEG Maintenance Point

MEP_I: MEP Ingress

MEP_E: MEP Egress

MIP: MEG Intermediate Point

On-demand: OAM operation manner initiated manually and for a limited amount of time, usually for fault diagnostics.

Proactively: preconfigured OAM operation manner either running periodically and continuously, or acting on certain events such as alarm signals.

3. Characteristics of IRSM

IRSM currently utilizes packet Loss Measurement (LM) and packet Delay Measurement (DM) to accomplish performance monitoring. It has following features desirable as a carrier-grade monitoring scheme, as required in [3]:

o Independency - It is operated independently from multicast forwarding plane and control plane and it does not have bad influences on the running of the two planes.

o Real stream - The data to be monitored is from real multicast stream, usually specified by (*,G) or (S,G) pairs.

o Inline - It enables the on-the-spot measurement or monitoring when carrier network is loaded with customer multicast traffic.

o Inband - The OAM measurement packet is routed following strictly the same multicast forwarding path as the monitored multicast stream, which help gathering the true network forwarding metrics.

o End-to-End and per segment measurement - It is capable of monitoring the whole end-to-end forwarding path from one multicast root to one or more leaf nodes, which provides the path performance for a particular multicast stream as a whole. It also supports measurement for a segment (i.e. a forwarding branch, a forwarding node or the combination of them) of a multicast forwarding path. The features enable the monitoring of a whole multicast tree, of one or more forwarding paths, and of parts of the tree or path(s).

o Proactive and on-demand modes - It is capable of carrying out a measurement session proactively or on demand according to the configured management policy.

4. IRSM Message

Two IRSM message types are defined: Loss Measurement (LM) message and Delay Measurement (DM) message. An example format of both LM and DM messages are given in section 5.2 and 5.3 respectively.

4.1. Encapsulation

IRSM measurement messages are encapsulated in IP packets. Same SA, DA and DSCP value as the multicast stream monitored are used for IRSM packets. By this means, it is ensured that IRSM packets for a specific (S,G) or (*,G) follow the exact same data path as user traffic, i.e. fate sharing. IRSM packets can be distinguished from the data traffic using a dedicated IP protocol type in IP header. In another way, an UDP port number could be assigned to make this identification. Using UDP port requires an intermediate MIP node to look deeper into an OAM packet, which introduces additional processing burden on MIP nodes.

The data part of an IRSM message adopts the popular Type-Length-Value structure, as shown in figure 1.

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |     Length    |          Value                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                          Value  Continued                     ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1. The Type-Length-Value Structure of IRSM Messages

Type - the type of the OAM message.

0, LM - Loss Measurement message

1, DM - Delay Measurement message

2-255, reserved for future use

Length - the length of the OAM message, not including the common IP header, the Type field and the Length field.

Value - the content of a specific OAM messages except for the Type and the Length fields.

4.2. Loss Measurement Message

An example format of LM message is shown in figure 2.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type = LM   |    Length     |    Version    |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Session ID                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Transmission Period                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Sequence Number                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Transmitted Packet Count                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Received Packet Count                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                            Optional                           ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2. The Format of an LM Message

Version - The version of the message. Its current value is 0.

Session ID - The identification of this measurement session.

Transmission Period - The period of LM message within this measurement session.

Sequence Number - A unique identification of an IRSM message. It is increased by 1 when a new LM message is generated within a measurement session.

Transmitted Packet Count - the accumulated number of data packets transmitted since the last LM message was generated. This field is filled by MEP-I when generating a LM message.

Received Packet Count - the accumulated number of received data packets received by this MEP_E or MIP entity since the previous LM packet was received. It is an optional field.

Optional - This is an optional field, which is reserved for future use to carry other information (e.g., Authentication info) if needed.

4.3. Delay Measurement Message

The DM message can be defined as shown in figure 3.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  DM Type = 1  |    Length     |    Version    |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           Session ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Transmission Period                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Sequence Number                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Transmission Timestamp                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Receiving Timestamp                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                            Optional                           ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3. The Format of DM message

The Version, Session ID, Transmission Period, and Sequence Number fields have the same meaning as those defined in LM message.

Transmission Timestamp - The local timestamp when generating this DM message.

Receiving Timestamp - The local timestamp when receiving this DM message. This field is optional and reserved for future use.

Optional - This is an optional field, which is reserved for future use to carry other information (e.g., Authentication info) if needed.

5. Principle of IRSM

5.1. Measurement Architecture

Three functional entities are defined in IRSM measurement architecture: MEP-I, MIP and MEP-E[2]. They are logical entities that can be configured on the incoming or outgoing interfaces of the monitoring equipments. The relationship of these entities is illustrated in figure 4.

```
+------+     +------+     +------+     +-------+    +------+
| root <> - <>router<> - <>router<> - <>router<> - <> leaf |
+------+     +------+     +------+     +-------+    +------+
        .    .       .    .       .    .       .    .
        .    .       .    .       .    .       .    .
        .    .       .    .       .    .       .    .
     MEP-I  MIP1    MIP2 MIP3   MIP4  MIP5   MIP6  MEP-E

   <>     Interface
   ------  Link
```

Figure 4. The relationship of MEP and MIP entities

The basic function of MEP-I is to initiate a measurement session. It generates OAM measurement packets for certain (S,G) or (*,G) and injects it into the data traffic. The measurement session could be initiated either proactively or on-demand. During a packet loss measurement, MEP-I takes count of the transmitted packets from a specific multicast stream and sends out the Loss Measurement (LM) message along the multicast path carrying this packet count. For delay measurement, MEP-I generates the Delay Measurement (DM) message recording the locally generated timestamp.

MEP-E is the end point of the measurement path. It terminates the OAM measurement packets and measures the packet loss or delay from MEP-I to MEP-E. MEP-E calculates the packet loss from MEP-I according to local packet count of the stream and the packet count in the received LM packet sent by MEP-I, and calculates the delay from MEP-I by comparing the timestamp carried in the DM message and local time for receiving this message.

MIP entity locates between MEP-I and MEP-E and forward OAM packets from upstream MEP-I to downstream MEP-E(s). It is optionally configured on the intermediate node of a multicast forwarding path. If MIP function is enabled on an intermediate node, it can perform the measurement for a certain network segment. MIP entity snoops the

OAM measurement packets and calculates the packet loss and delay from MEP-I to itself in the same way as an MEP-E does.

If MEP-I function is configured on the root node and MEP-E configured on the leaf node, then the monitored path is a complete multicast forward path, as depicted in figure 1. If MEP-I or MIP-E is configured on an intermediate node, then part of the multicast path could be monitored.

5.2. Packet Loss Measurement

Packet loss measurement could be performed proactively or on demand according to the configuration of a measurement session. In proactive mode, LM packet will be transmitted by MEP-I continuously with a specific time interval. For on-demand mode, LM will be sent periodically, or based on a pre-arranged schedule. If the schedule is for a single measurement, then two LM messages are required to be generated, the reason is given as below.

When a measurement session starts, MEP-I counts the transmitted packets from a multicast data stream for a specified time interval. If the timer expires, MEP-I generates an LM packet carrying this transmitted packet count (say Tx_Count) of the multicast data packets having been sent. The LM packet is injected into the data stream, and forwarded in the same way as a real multicast data packet.

MEP_E counts the number of the received packet from a multicast stream for a specified time (denoted as Rx_Count). The packet loss is the difference between the two counters (i.e. Tx_Count - Rx_Count) for this time interval. This calculation is incorrect when the packet counter(s) of MEP_I and/or MEP_E wrap after reaching their maximum values. Two successive measurements are used to eliminate this effect, with the calculation taken as:

Packet Loss = (Tx_Count2 - Tx_Count1)-(Rx_Count2 - Rx_Count1)

Where Tx_Count1 and Tx_Count2 are respectively packet count values carried in two successive packets, and Rx_Count1 and Rx_Count2 are packet counts locally accumulated by MEP_E during the same time interval.

If MIP function is enabled on an intermediate node, it will snoop the measurement packets and count the received data packets locally. On receiving an LM packet, it records the current local packet count (say Rx_Count1') and the transmitted packet count Tx_Count1' in the received LM packet. And after receiving a subsequent LM packet, it takes the same action as above to acquire the local packet count (say

Rx_Count2') and the transmitted packet count carried in this LM packet (say Tx_Count2'). The packet loss is calculated as:

Packet Loss = (Tx_Count2' - Tx_Count1')-(Rx_Count2' -Rx_Count1')

The calculation could be performed in a process component within (e.g., a dedicated process component) or outside (e.g., NMS) the MIP node.

Each MIP and MEP-E node on the path could obtain the packet loss statistics of the path from MEP-I to itself by this means and both per-segment and end-to-end performance monitoring are available within a measurement session. The integration of the overall measurement results could help to detect and locate the failure point(s) and the performance bottle point(s) along the forwarding path, which is shown in section 6.1.

5.3. Packet Delay Measurement

Time synchronization among the measuring entities is required for packet delay measurement. The measurement process of DM is almost the same as packet loss measurement. It can be operated proactively or on-demand. The only difference is that the process is based on a timestamp value other than a packet counter.

When a measurement session starts, MEP-I generates DM packets carrying the local timestamps (say Tx_Time) and sends them onto the multicast path. The DM packets are forwarded in the same way as the normal multicast data.

When MEP_E receives a DM packet, it records the local timestamp that identifies the arrival time of the DM packet (Rx_Time). The delay measurement result could be calculated by:

Packet Delay = Rx_Time - Tx_Time

Similarly, if MIP function is enabled in an intermediate node, it will snoop the DM packet to get the timestamps of the time when a DM packets are sent at MEP-I. On receipt of a DM packet, the MIP node records the local timestamp (say Rx_Time1) and the timestamp (say Tx_Time1) carried in the DM packet. The packet delay between MEP-I and MIP could be calculated as (Rx_Time1- Tx_Time1). Similar to LM measurement, the calculation could be performed in a process component within (e.g., a dedicated process component) or outside (e.g., NMS) the MIP node.

Each MIP and MEP-E entity could acquire the packet delay information from MEP-I to itself. By this means it is able to perform per-segment and end-end delay measurement within a single measurement session, which helps detection and isolation of performance defect points, as shown in section 6.2

6. Application in multicast network monitoring

This section describes how packet loss measurement and packet delay measurement are used in IRSM to accomplish multicast network performance monitoring. To simplify the illustration, a small multicast tree is depicted in figure 5. In this example, the downstream interface of the root node acts as MEP-I, upstream and downstream interfaces of intermediate nodes are configured as MIPs, and upstream interfaces of leaf nodes are assigned as MEP-Es.

```
                                                          I +--------+
                                                       +---<> Leaf 2 |
                                      +----------+ G   |    +--------+
                   +----------+ C   E |          <>----+
   +---#--+ A    B |          <>-----<> router 2 |
   | root <>------<> router 1 |       |          <>----+
   +------+        |          <>---+  +----------+ H   |  J +--------+
           .       +----------+ D  |                   +---<> Leaf 3 |
           .       .           .   | F +--------+           +--------+
           .       .           .   +--<> Leaf 1 |           .
           .       .           .       +--------+           .
           .       .           .       .          .         .
         MEP-I    MIP1        MIP2    MIP5       MIP6     MEP-E2
                              MIP3    MEP-E1     MIP7     MEP-E3

      <>      Interface
      ------  Link
```

Figure 5. An example of multicast forwarding tree to be monitored

6.1. Fault Detection and Localization Based on LM

The sending and processing of LM packets by MEP and MIP enable the fault detection and location for performance monitoring, both for network link and network node. Suppose in figure 5, the link C-E between router1 and router2 has performance bottleneck which causes packet losses. Based on the principle given in section 4.2, all the monitoring entities on the downstream of the link will perceive this loss by packet loss calculation, while the upstream will not. Because the downstream entities (i.e. MIP5, MIP6, MIP7, MEP-E2, MEP-

E3) will detect the packet losses, whereas the upstream entities (i.e. MIP1 and MIP2) will not, it can be inferred that the link between C and E is suffering from performance problem.

The degradation of an intermediate node can also be detected and isolated in this way. For example if router2 has forwarding defect which introduces packet loss, by snooping LM messages and counting locally received packets, MIP6, MIP7, MEP-E2, and MEP-E3 will detect the packet loss while the upstream MIP1, MIP2 and MIP5 will not. The fault point of router2 will be easily located.

It is even possible to detect multiple point failures along the multicast forwarding path, if such errors occur. In figure 5, if link C-E and H-J both undergo packet losses, all the entities down from the C-E will detect the defects. But as MIP7 and MEP-E3 have different packet loss values from MIP5, in which case packet loss detected in MIP7 and MEP-E3 are equal but are greater than those measured in MIP5, then additional fault point of link H-J could be easily picked out.

6.2. Fault Detection and Localization Based on DM

The fault detection and location principle of the packet delay measurement is the same as that of packet loss measurement given in section 6.1. If the link or node has defects that cause packet delay increasing, their downstream MIPs and MEP-Es will perceive them. If the delay value rises over a reasonable threshold level, then it can be judged that the link or node is undergoing performance abnormities. The DM could be operated to support single-point link and node detection, multipoint link and node detection, as long as the forwarding path is equipped with enough monitoring entities.

7. Deployment Considerations

When IRSM is deployed in practical network, many issues should be considered to enable the scheme to work efficiently. Here are emulated some of the key aspects that should be paid special attention to when implementing IRSM.

7.1. Acquisition of monitored stream

Because LM needs to take count of the data packet from a real stream, an IRSM-enabled node must provide the means to acquire the multicast stream to be monitored. In practice this could be implemented by an ACL method. If the stream to be monitored is specified by an (*,G) or an (S,G) pair, the ACL policy could be set to permit the packets belong to this stream to be processed by the IRSM module.

7.2. Alarm and Reporting Processing

The alarming and reporting method in case of abnormal and normal status should be in the scope of network planning and should be designed according to the local policy of the management and the scale of the multicast tree. To prevent alarm and report from overburdening the network and the NMS systems, the amount of these messages generated should be minimized.

In IRSM performance monitoring, a feasible scheme is to let only the MEP-E entities alarm the exception state when packet loss or unacceptable packet delay is detected. MIP nodes only log the exception and will only send report to the management system regularly or passively in response to the queries. It needs further study on how to design in detail the alarming and reporting functions of an IRSM system.

7.3. Configuration of Monitoring Nodes

IRSM performance monitoring system should be flexible enough for the provider to operate. For example, the provider may at this time prefer proactively monitoring and in other occasions need to take some discrete tests on demand. He may choose to monitor the whole multicast tree, only some important forwarding paths or branches, or even merely several nodes prone to performance degradation. An IRSM-capable node should be able to be enabled or disabled for its monitoring function as required by a configuration operation to support this flexibility.

The configuration should also be used to provide the parameters of a monitoring session, such as the OAM execution frequency, the starting or ending of a monitoring session, the multicast stream to be monitored, and etc. The configuration could be implemented as the manual manner by an administrator, or by a control plane protocol. The latter has the advantages of flexibility and scalability. It is for further study on how to realize a practical configuration control.

7.4. Topology Discovery

Topology discovery is the precondition of the monitoring of a multicast tree. IRSM administrator could make use of current available multicast tree topology discovery tool in a multicast management system. If this is unavailable, it is possible to define a lightweight tool specific for IRSM uses.

7.5. Interoperation among Different Vendors

If equipments from different vendors are deployed in the same multicast tree or on the same multicast forwarding path, cares should be taken to interoperate them well to fulfill the performance monitoring task. Because the LM and DM packets are treated normally as the multicast data packets, the only interoperability requirement is that those intermediate IRSM-incapable equipments do not discard the LM or DM packets.

8. Security Considerations

It should be recognized that conducting performance monitoring measurements can raise security concerns. IRSM system, in which traffic is injected into the network, can be abused for denial-of-service attacks disguised as legitimate measurement activity. Authentication, authorization and encryption techniques may be used where appropriate to guard against injected traffic attacks. These aspects will be discussed in the future version of the memo.

9. IANA Considerations

If IP Protocol in IP header is used as the Identification of IRSM OAM packets, then a new IP protocol value is required to be assigned by IANA.

If UDP port number in UDP header is used as the identification of IRSM OAM packets, then a new UDP value is required to be assigned by IANA.

10. References

10.1. Normative References

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

[2] ITU-T Recommendation Y.1731 (02/2008), " OAM functions and mechanisms for Ethernet based networks ", Feb,2008.

10.2. Informative References

[3] M. Bianchetti, G. Picciano, M. Chen, and L. Zheng, " Requirements for IP multicast performance monitoring ", draft-bipi-mboned-ip-multicast-pm-requirement-01.txt, March 2010.

11. Acknowledgments

Special thanks should be given to Guo Xinchun and Wang Yan for their valuable comments of the draft.

Authors' Addresses

Liu Hui
Huawei Technologies Co., Ltd.
Huawei Building, No.3 Xinxi Road,
Hai-Dian District,
Beijing 100085
China

Email: liuhui47967@huawei.com

Mach(Guoyi) Chen
Huawei Technologies Co., Ltd.
Huawei Building, No.3 Xinxi Road,
Hai-Dian District,
Beijing 100085
China

Email: mach@huawei.com

Lianshu Zheng
Huawei Technologies Co., Ltd.
Huawei Building, No.3 Xinxi Road,
Hai-Dian District,
Beijing 100085
China

Email: verozheng@huawei.com