

Network working group  
Internet Draft  
Intended status: Informational

A. Tempia Bonda  
G. Picciano  
Telecom Italia  
M. Chen  
L. Zheng  
Huawei Technologies Co., Ltd  
July 1, 2011

Expires: January 1, 2012

**Requirements for IP multicast performance monitoring  
draft-ietf-mboned-ip-multicast-pm-requirement-01.txt**

Abstract

This document describes the requirement for an IP multicast performance monitoring system for service provider IP multicast networks. This system enables efficient performance monitoring in Service Providers' production networks and provides diagnostic information in case of performance degradation or failure.

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/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on January 1, 2011.

Copyright Notice

Copyright (c) 2009 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 ..... [3](#)
- [2](#). Conventions used in this document ..... [4](#)
- [3](#). Terminologies ..... [4](#)
- [4](#). Functional Requirements ..... [6](#)
  - [4.1](#). Topology discovery and monitoring ..... [6](#)
  - [4.2](#). Performance measurement ..... [6](#)
    - [4.2.1](#). Loss rate ..... [6](#)
    - [4.2.2](#). One-way delay ..... [7](#)
    - [4.2.3](#). Jitter ..... [7](#)
    - [4.2.4](#). Throughput ..... [8](#)
  - [4.3](#). Measurement session management ..... [8](#)
    - [4.3.1](#). Segment v.s. Path ..... [8](#)
    - [4.3.2](#). Static v.s. Dynamic configuration ..... [8](#)
    - [4.3.3](#). Proactive v.s. on-demand ..... [9](#)
  - [4.4](#). Measurement result report ..... [9](#)
    - [4.4.1](#). Performance reports ..... [9](#)
    - [4.4.2](#). Exceptional alarms ..... [9](#)
- [5](#). Design considerations ..... [10](#)
  - [5.1](#). Inline data-plane measurement ..... [10](#)
  - [5.2](#). Scalability ..... [11](#)
  - [5.3](#). Robustness ..... [11](#)
  - [5.4](#). Security ..... [11](#)
  - [5.5](#). Device flexibility..... [11](#)
  - [5.6](#). Extensibility ..... [12](#)
- [6](#). Security Considerations ..... [12](#)
- [7](#). IANA Considerations ..... [12](#)
- [8](#). References ..... [12](#)
  - [8.1](#). Normative References ..... [12](#)
  - [8.2](#). Informative References ..... [12](#)
- [9](#). Acknowledgments ..... [14](#)



## **1. Introduction**

Service providers (SPs) have been leveraging IP multicast to provide revenue-generating services, such as IP television (IPTV), video conferencing, as well as the distribution of stock quotes or news. These services are usually loss-sensitive or delay-sensitive, and their data packets need to be delivered over a large scale IP network in real-time. Meanwhile, these services demand relatively strict service-level agreements (SLAs). For example, loss rate over 5% is generally considered unacceptable for IPTV delivery. Video conferencing normally demands delays no more than 150 milliseconds. However, the real-time nature of the traffic and the deployment scale of service make it very challenging for IP multicast performance monitoring in a SP's production network. With increasing deployment of multicast service in SP networks, it becomes mandatory to develop an efficient system that is designed for SPs to accommodate the following functions.

- o SLA monitoring and verification: verify whether the performance of a production multicast network meets SLA requirements.
- o Network optimization: identify bottlenecks when the performance metrics do not meet the SLA requirements.
- o Fault localization: pin-point impaired components in case of performance degradation and service disruption.

These functions alleviate the OAM cost of IP multicast network for SPs, and ensure the quality of services.

However, the existing IP multicast monitoring tools and systems, which were mostly designed either for primitive connectivity diagnosis or for experimental evaluations, do not suit an SP production network, given the following facts:

- o Most of them provide end-to-end reachability check only [2][4][6]. They cannot provide sophisticated measurement metrics such as packet loss, one-way delay, and jitter, for the purpose of SLA verification.
- o Most of them can perform end-to-end measurements only. For example, RTCP-based monitoring system [5] can report end-to-end packet loss rate and jitter. End-to-end measurements are usually inadequate for fault localization, which needs finer grain measurement data to pin-point exact root causes.



- o Most of them use probing packets to probe network performance [2] [4]. The approach might yield biased or even irrelevant results because the probing results are sampled and the out-of-band probing packets might be forwarded differently from the monitored user traffic.
- o Most of them are not scalable in a large deployment like an SPs' production network. For example, in IPTV deployment, the number of group members might be in the order of thousands. In this scale, an RTCP-based multicast monitoring system [5] becomes almost unusable because RTCP report intervals of each receiver might be delayed up to minutes or even hours because of over-crowded reporting multicast channel [12].
- o Some of them rely on the information from external protocols, which make their capabilities and deployment scenarios limited by the external protocols. The examples are passive measurement tools that collect and analyze messages from protocols such as multicast routing protocols [7], IGMP [9], or RTCP [5], etc. Another example is a SNMP-based system [8] that collects and analyzes relevant multicast MIB information.

This document describes the requirement for an IP multicast performance monitoring system for service provider (SP) IP multicast networks. This system should enable efficient monitoring of performance metrics of any given multicast channel (\*,G) or (S,G) and provides diagnostic information in case of performance degradation or failure, which help SPs to do SLA verification, network optimization, and fault localizations in a large production network.

## **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](#) [1].

## **3. Terminologies**

- o SSM (source specific multicast): When a multicast group is operating in SSM mode, only one designated node is eligible to send traffic through the multicast channel. An SSM multicast group with the designated source address *s* and group address *G* is denoted by (*s*, *G*).



- o ASM (any source multicast): When a multicast group is operating in ASM mode, any node can multicast packets through the multicast channel to other group members. An ASM multicast group with group address  $G$  is denoted by  $(*, G)$ .
- o Root (of a multicast group): In an SSM multicast group  $(s, G)$ , the root of this group is the first-hop router next to the source node  $s$ . In an ASM multicast group  $(*, G)$ , the root of this group is the selected rendezvous point router.
- o Receiver: The term receiver refers to any node in the multicast group that should receive multicast traffic.
- o Internal forwarding path: Given a multicast group and a forwarding node in the group, the internal forwarding path inside the node refers to the data path between the upstream interface towards the root and one of the downstream interfaces toward a receiver.
- o Multicast forwarding path: Given a multicast group, a multicast forwarding path refers to the sequence of the interfaces, links and internal forwarding paths from the downstream interface at the root until the upstream interface at a receiver.
- o Multicast forwarding tree: Given a multicast group  $G$ , the union of all multicast forwarding paths composes the multicast forwarding tree.
- o Segment (of multicast forwarding path): The segment of a multicast forwarding path refers to part of the path between any two given interfaces.
- o Measurement session: A measurement session refers to the period of time in which certain performance metrics over a segment of multicast forwarding path is monitored and measured.
- o Monitoring node: A monitoring node is a node on a multicast forwarding path that is capable of performing traffic performance measurements on its interfaces.
- o Active interface: An interface of a monitoring node that is turned on to start a measurement session is said to be active.
- o Measurement session control packets: The packets are used for dynamic configuration for active interface to coordinate measurement sessions.



Figure 1 shows a multicast forwarding tree rooted at a root's interface A. Within router 1, B-C and B-D are two internal forwarding paths. Path A-B-C-E-G-I is a multicast forwarding path, which starts at root's downstream interface A and ends at receiver 2's upstream interface I. A-B, B-C-E are two segments of this forwarding path. When a measurement session for a metric such as loss rate is turned on over segment A-B, interfaces A and B are active interfaces.

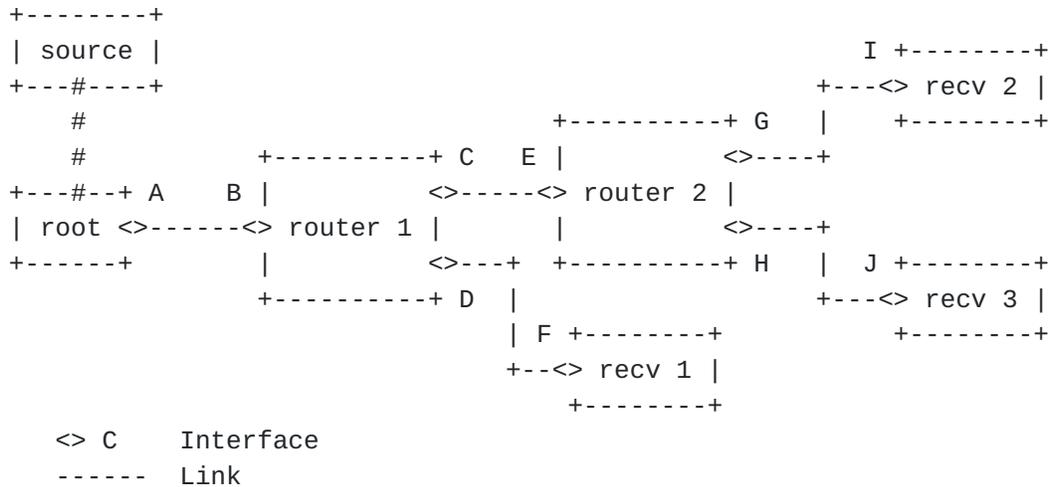


Figure 1. Example of multicast forwarding tree

#### 4. Functional Requirements

##### 4.1. Topology discovery and monitoring

The monitor system SHOULD have mechanisms to collect topology information of the multicast forwarding trees for any given multicast group. The function can be an integrated part of this monitoring system. Alternatively, the function might rely on other tools and protocols, such as mtrace [3], MANTRA[7], etc. The topology information will be referenced by network operators to decide where to enable measurement sessions.

##### 4.2. Performance measurement

The performance metrics that a monitoring node needs to collect include, but are not limit to, the following.

###### 4.2.1. Loss rate

Loss rate over a segment is the ratio of user packets not delivered to the total number of user packets delivered over this segment during a given interval. The number of user packets not delivered



over a segment is the difference between the number of packets transmitted at the starting interface of the segment and received at the ending interface of this segment. Loss rate is crucial for multimedia streaming, such as IPTV, video/audio conferencing.

Loss rate over any segment of a multicast forwarding path **MUST** be provided. The measurement interval **MUST** be configurable.

#### **[4.2.2.](#) One-way delay**

One-way delay over a segment is the average time that user packets take to traverse this segment of forwarding path during a given interval. The time that a user packet traversing a segment is the difference between the time when the user packet leaves the starting interface of this segment and the time when the same user packet arrives at the ending interface of this segment. The one-way delay metric is essential for real-time interactive applications, such as video/audio conferencing, multiplayer gaming.

One-way delay over any segment of a multicast forwarding path **SHOULD** be able to be measured. The measurement interval **MUST** be configurable.

To get accurate one-way delay measurement results, the two end monitoring nodes of the investigated segments might need to have clock synchronized. The one-way delay metric for packet networks is described in [RFC2679](#) [13].

#### **[4.2.3.](#) Jitter**

Jitter over a segment is the variance of one-way delay over this segment during a given interval. The metric is of great importance for real-time streaming and interactive applications, such as IPTV, audio/video conferencing.

One-way delay jitter over any segment of a multicast forwarding path **SHOULD** be able to be measured. The measurement interval **MUST** be configurable.

Same as One-way delay measurement, to get accurate jitter, the clock frequencies at the two end monitoring nodes might need to be synchronized so that the clocks at two systems will proceed at the same pace. The jitter metric for packet networks is described in [RFC2681](#) [14].



#### **4.2.4. Throughput**

Throughput of multicast traffic for a group over a segment is the average number of bytes of user packets of this multicast group transmitted over this segment in unit time during a given interval. The information might be useful for resource management.

Throughput of multicast traffic over any segment of a multicast forwarding path MAY be measured. The measurement interval MUST be configurable.

#### **4.3. Measurement session management**

A measurement session refers to the period of time in which measurement for certain performance metrics is enabled over a segment of multicast forwarding path or over a complete multicast forwarding path. During a measurement session, the two end interfaces are said active. When an interface is activated, the interfaces start collecting statistics, such as number or timestamps of user packets which belongs to the given multicast group and pass through the interface. When both interfaces are activated, the measurement session starts. During a measurement session, data from two active interfaces are periodically correlated and the performance metrics, such as loss rate or delay, are derived. The correlation can be done either on the downstream interface if the upstream interface passes its data to it or on a third-party if the raw data on two active interfaces are reported to it. When one of the two interfaces is deactivated, the measurement session stops.

##### **4.3.1. Segment v.s. Path**

Network operators SHOULD be able to turn on or off measurements sessions for specific performance metrics over either a segment of multicast forwarding path or over a complete multicast forwarding path at any time. For example in Figure 1, network operator can turn on the measurement session of loss rate over path A-B-D-F and segment A-B-C as well as jitter over segment C-E-G-I simultaneously. This feature allows network operators to zoom into the suspicious components when degradation or failure occurs.

##### **4.3.2. Static v.s. Dynamic configuration**

A measurement session can be configured statically. In this case, network operators activate the two interfaces or configure their parameter settings on the relevant nodes either manually or automatically through agents of network management system (NMS).



Optionally, a measurement session can be configured dynamically. In this case, an interface may coordinate another interface on its forwarding path to start or stop a session. Accordingly, the format and process routines of the measurement session control packets need to be specified. The delivery of such packets SHOULD be reliable and it MUST be possible to secure the delivery of such packets.

#### **4.3.3. Proactive v.s. on-demand**

A measurement session can be started either proactively or on demand. Proactive monitoring is either configured to be carried out periodically and continuously or preconfigured to act on certain events such as alarm signals. To save resources, operators may turn on measurement sessions proactively for critical performance metrics over the backbone segments of multicast forwarding tree only. This keeps the overall monitoring overhead minimal during normal network operations.

In contrast to proactive monitoring, on-demand monitoring is initiated manually and for a limited amount of time to carry out diagnostics. When network performance degradation or service disruption occurs, operators might turn on measurement sessions on-demand over the interested segments to facilitate fault localization.

#### **4.4. Measurement result report**

The measurement results might be present in two forms: reports or alarms.

##### **4.4.1. Performance reports**

Performance reports contain streams of measurement data over a period of time. A data collection agent MAY actively poll the monitoring nodes and collect the measurement reports from all active interfaces. Alternatively, the monitoring nodes might be configured to upload the reports to the specific data collection agents once the data become available. To save bandwidth, the content of the reports might be aggregated and compressed. The period of reporting SHOULD be able to be configured or controlled by rate limitation mechanisms (e.g., exponentially increasing).

##### **4.4.2. Exceptional alarms**

On the other hand, the active interfaces of a monitoring node or a third-party MAY be configured to raise alarms when exceptional events such as performance degradation or service disruption occur. Alarm thresholds and the management should be specified for each of the



performance metric when the measurement session is configured on this interface. During measurement session, once the value of certain performance metric exceeds the threshold, alarm will be raised and reported to the configured nodes. To prevent huge volume of alarms from overloading the management nodes and network congestion, alarm suppression and aggregation mechanisms SHOULD be employed on the interfaces to limit the rate of alarm report and the volume of data.

## **5. Design considerations**

To make the monitoring system feasible and optimal for a SP production network, the following considerations should take into account when design the system.

### **5.1. Inline data-plane measurement**

Measurement results collected by probing packets might be biased or even totally irrelevant given the facts that (1) probing packets collect sampled results only and might not capture the real statistic characteristics of the monitored user traffic. Experiments have demonstrated that the measurement sampled by the probing packets, such as ping probes, might be incorrect if sampling interval is too long [1]; (2) probing packets introduce extra load onto the network. In order to improve accuracy, sampling frequency has to be high enough, which in turn increase network overhead and further bias the measurement results; (3) probing packets are usually not in the same multicast group as user packets and might take different forwarding path given that equal cost multi-path routing (ECMP) and link aggregation (LAG) have been widely adopted in SP network. An out-of-band probing packet might take a path totally different from the user packets of the multicast group that it is monitoring. Even if the forwarding path is the same, the intermediate node might apply different queuing and scheduling strategy for the probing packets. As a result, the measured results might be irrelevant.

The performance measurement should be "inline" in the sense that the measurement statistics are derived directly from user packets, instead of probing packets. At the same time, unlike offline packet analysis, the measurement is counting user packets at line-speed in real-time without any packet duplication or buffering.

To accomplish the inline measurement, some extra packets might need to be injected into user traffic to coordinate measurement across nodes. The volume of these packets SHOULD be keep minimal such that the injection of such packets will not impact measurement accuracy.



## **5.2. Scalability**

The measurement methodology and system architecture **MUST** be scalable. A multicast network for an SP production network usually comprises of thousands of nodes. Given the scale, the collecting, processing and reporting overhead of performance measurement data **SHOULD NOT** overwhelm either monitoring nodes or management nodes. The volume of reporting traffic should be reasonable and not cause any network congestion.

## **5.3. Robustness**

The measurements **MUST** be independent of the failure of the underlying multicast network. For example, the monitor **SHOULD** generate correct measurement result even if some measurement coordinating packets are lost; invalid performance reports should be able to be identified in case that the underlying multicast network is undergoing drastic changes.

If dynamic configuration is supported, the delivery of measurement session control packets **SHOULD** be reliable so that the measurement sessions can be started, ended and performed in a predictable manner. Meanwhile, the control packets **SHOULD** not be delivered based on the multicast routing decision. This multicast independent characteristic guarantees that the active interfaces are still under control even if the multicast service is malfunctioning.

Similarly, if an NMS is used to control the monitoring nodes remotely, the communication between monitoring nodes and the NMS **SHOULD** be reliable.

## **5.4. Security**

The monitoring system **MUST** not impose security risks on the network. For example, the monitoring nodes should be prevented from being exploited by third parties to control measurement sessions arbitrarily, which might make the nodes vulnerable for DDoS attacks.

If dynamic configuration is supported, the measurement session control packets need to be encrypted and authenticated.

## **5.5. Device flexibility**

Both the software and hardware deployment requirement for the monitoring system **SHOULD** be reasonable. For example, one-way delay measurement needs clock synchronization across nodes. To require the installation of expensive hardware clock synchronization devices on



all monitoring nodes might be too costly to make the monitoring system infeasible for large deployment.

The monitor system SHOULD be incrementally deployable, which means that the system can enable monitoring functionality even if some of the nodes in the network are not equipped with the required software and hardware or does not meet the software and hardware deployment requirements.

The non-monitoring nodes without the monitoring capabilities SHOULD be able to coexist with monitoring nodes and function. The packets exchanged between monitoring nodes SHOULD be transparent to other nodes and MUST not cause any malfunction of the non-monitoring nodes.

### **5.6. Extensibility**

The system should be easy to be extended for new functionalities. For example, the system should be easily extended to collect newly defined performance metrics.

## **6. Security Considerations**

The security issues have been taken into account in design considerations (see [Section 5.4](#)).

## **7. IANA Considerations**

There is no IANA action required by this draft.

## **8. References**

### **8.1. Normative References**

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

### **8.2. Informative References**

- [2] Venaas, S., "Multicast Ping Protocol", [draft-ietf-mboned-ssmping-07](#), December 2008.
- [3] Asaeda, H., Jinmei, T., Fenner, W., and S. Casner, "Mtrace Version 2: Traceroute Facility for IP Multicast", [draft-ietf-mboned-mtrace-v2-03](#), March 2009.

- [4] Almeroth, K., Wei, L., and D. Farinacci, "Multicast Reachability Monitor (MRM)", [draft-ietf-mboned-mrm-01](#), July 2000.
- [5] Bacher, D., Swan, A., and L. Rowe, "rtpmon: a third-party RTCP monitor", Conference 4th ACM International conference on multimediuM, 1997.
- [6] Sarac, K. and K. Almeroth, "Application Layer Reachability Monitoring for IP Multicast", Journal Computer Networks Journal, Vol.48, No.2, pp.195-213, June 2005.
- [7] Rajvaidya, P., Almeroth, K., and k. claffy, "A Scalable Architecture for Monitoring and Visualizing Multicast Statistics", Conference IFIP/IEEE Workshop on Distributed Systems: Operations & Management (DSOM), Austin, Texas, USA, December 2000.
- [8] Sharma, P., Perry, E., and R. Malpani, "IP Multicast Operational Network Management: Design, Challenges and Experiences", Journal IEEE Network, Volume 17, Issue 2, Mar/Apr 2003 Page(s): 49 - 55, Mar/Apr 2003.
- [9] Al-Shaer, E. and Y. Tang, "MRMON: Remote Multicast Monitoring", Conference NOMS, 2004.
- [10] Sarac, K. and K. Almeroth, "Supporting Multicast Deployment Efforts: A Survey of Tools for Multicast Monitoring", Journal Journal of High Speed Networks, Vol.9, No.3-4, pp.191-211, 2000.
- [11] Sarac, K. and K. Almeroth, "Monitoring IP Multicast in the Internet: Recent Advances and Ongoing Challenges", Journal IEEE Communication Magazine, 2005.
- [12] Vit Novotny, Dan Komosny, "Optimization of Large-Scale RTCP Feedback Reporting in Fixed and Mobile Networks," icwmc, pp.85, Third International Conference on Wireless and Mobile Communications (ICWMC'07), 2007
- [13] Almes, G., Kalidindi,S. and M. Zekauskas, "A One-way Delay Metric for IPPM", [RFC 2679](#), September 1999
- [14] Almes, G., Kalidindi,S. and M. Zekauskas, "A Round-trip Delay Metric for IPPM", [RFC 2681](#), September 1999.



## **9. Acknowledgments**

The authors would like to thank Wei Cao, Xinchun Guo, and Hui Liu for their helpful comments and discussions.

Authors' Addresses

Alberto Tempia Bonda  
Telecom Italia  
Via Reiss Romoli, 274  
Torino 10148  
Italy

Email: [alberto.tempiabonda@telecomitalia.it](mailto:alberto.tempiabonda@telecomitalia.it)

Giovanni Picciano  
Telecom Italia  
Via Di Val Cannuta 250  
Roma 00166  
Italy

Email: [giovanni.picciano@telecomitalia.it](mailto:giovanni.picciano@telecomitalia.it)

Mach(Guoyi) Chen  
Huawei Technologies Co., Ltd  
No. 3 Xixi Road, Shang-di, Hai-dian District  
Beijing 100085  
China

Email: [mach@huawei.com](mailto:mach@huawei.com)

Lianshu Zheng  
Huawei Technologies Co., Ltd  
No. 3 Xixi Road, Shang-di, Hai-dian District  
Beijing 100085  
China

Email: [verozheng@huawei.com](mailto:verozheng@huawei.com)