Network working group
Internet Draft
Intended status: Informational
Expires: January 1, 2010

M. Bianchetti
G. Picciano
Telecom Italia
M. Chen
J. Qiu
Huawei Technologies Co., Ltd.
July 6, 2009

Requirements for IP multicast performance monitoring
draft-bipi-mboned-ip-multicast-pm-requirement-00.txt

Status of this Memo

This Internet-Draft will expire on August 15, 2009.

Copyright Notice

Abstract

With increasing deployment of IP multicast in service provider (SP) network, SPs need a carrier-grade IP multicast performance monitoring solution. This document describes the requirements for such a system for a SP network. This system enables efficient performance monitoring in SPs' production network and provides diagnostic information in case of performance degradation or failure.

Table of Contents

1. Introduction

This document describes the requirement for an IP multicast performance monitoring system for service provide (SP) IP multicast

network. This system enables 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.

Increasing deployment of IP multicast in SP network calls for a carrier-grade IP multicast performance monitoring solution. 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 IP network in real-time. Accordingly, these services demands very 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 solution that is designed for SPs to solve the following problems.

- SLA monitoring and verification: verify whether the performance of production multicast network meets SLA requirements.

- Network optimization: identify bottlenecks when the performance metrics do not meet the SLA requirements.

- Fault localization: pin-point impaired components in case of performance degradation and service disruption.

These capabilities ease the management and maintenance costs of IP multicast network for SPs, and ensure 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, does not suit for a SP production network, given the following facts:

- Most of them provide end-to-end reachability check only [4][6][8]. They cannot provide sophisticated measurement metrics such as packet loss, one-way delay, and jitter, for the purpose of SLAs verification.

- Most of them can perform end-to-end measurements only. For example, RTCP-based monitoring system [7] can report end-to-end packet loss rate and jitter. End-to-end measurements are usually inadequate for fault localization, which needs fine-grain measurement data to pin-point exact root causes.

- Most of them use probing packets to probe network performance [4] [6]. 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.

- Most of them are not scalable in a large deployment like SPs' production network. For example, in IPTV deployment, the number of group member might be in the order of thousands. In this scale, a RTCP-based multicast monitoring system [7] 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 [14].

- 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 [9], IGMP [11], or RTCP [7], etc. Another example is a SNMP-based system [10] that collects and analyzes relevant multicast MIB information.

This document specifies the requirements for an IP multicast traffic monitor for a carrier-grade IP multicast network, 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

- 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. A SSM multicast group with the designated source address s and group address G is denoted by (s, G).

- 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).

- Root (of a multicast group): In a 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.

- Receiver: The term receiver refers to any node in the multicast group that receives multicast traffic.

- 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 interface towards the receivers.

- 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 root until the upstream interface at a receiver.

- Multicast forwarding tree: Given a multicast group G, the union of all multicast forwarding paths composes the multicast forwarding tree.

- Segment (of multicast forwarding path): The segment of a multicast forwarding path refers to part of the path between any two given interfaces.

- 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.

- Monitoring node: A monitoring node is a node on a multicast forwarding path that is capable of performing traffic performance measurements on its interfaces.

- Active interface: An interface of a monitoring node that is turned on to start a measurement session is said active.

- 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.

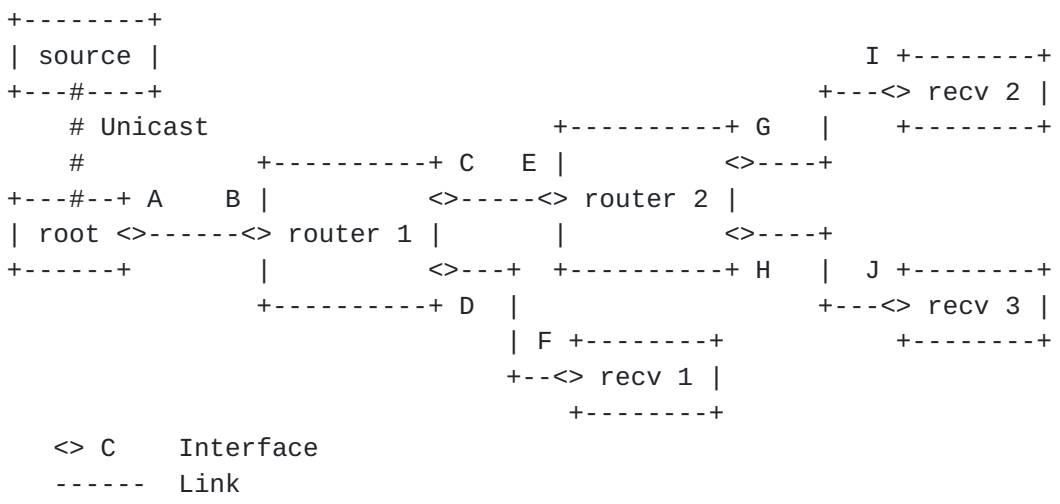

Figure 1. Example of multicast forwarding tree

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 relies on other tools and protocols, such as mtrace [5], MANTRA[9], etc. The topology information will be referred 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 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 IP, 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 correct one-way delay measurements, the two end monitoring nodes of the investigated segments might need to have clock synchronized.

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.

To get correct jitter statistics, 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.

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, statistics 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 statistics to it or on a third-party if the raw statistics 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 MUST be secured.

4.3.3. Proactive v.s. on-demand

A measurement session can be started either proactively or on demand. 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. However, 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 continuous 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 MAY be configured to raise alarms if exceptional events such as performance degradation or service disruption occur. During the meantime, the monitoring node remains quiescent. In this configuration, alarm thresholds and the nodes to which alarms are reported should be specified for each of the performance metric when the measurement session is configured on this interface. During measurement session, once the 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 congest the network, 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 for a SP production network, the designers should take into account the following requirements.

5.1. Inline data-plane measurement

The system is monitoring the performance of multicast traffic, and thus is operating on data plane. 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.

Probing packet SHOULD be avoided in performance 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.

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 SP is usually composed of thousands of nodes. Given the scale, the collecting, processing and reporting overhead of performance statistic 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 fault-free, namely, 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 packets SHOULD be multicast independent as the packets should not use multicast to delivery. This guarantees that the active interfaces are still under control even if the multicast service is malfunctioning.

Similarly, if NMS are used to control the monitoring nodes remotely, the communication between them SHOULD be reliable and multicast independent.

5.4. Security

The monitoring system MUST not impose security risks on the network. For example, the monitoring nodes should not be exploited by third parties to control measurement sessions arbitrarily. This might leave holes for DDoS attacks.

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

5.5. Device flexibility

The deployment requirement in terms of both software and hardware SHOULD be reasonable. For example, one-way delay measurement needs clock synchronization across nodes. To require all nodes to install expensive hardware clock synchronization devices 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 extend 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.

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.

[2] Crocker, D. and Overell, P.(Editors), "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, Internet Mail Consortium and Demon Internet Ltd., November 1997.

8.2. Informative References

[3] "Trading Floor Architecture", White Paper , Online http://www.cisco.com/en/US/docs/solutions/Verticals/Trading_Floor_Architecture-E.html.

[4] Venaas, S., "Multicast Ping Protocol", draft-ietf-mboned-ssmping-07, December 2008.

[5] 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.

[6] Almeroth, K., Wei, L., and D. Farinacci, "Multicast Reachability Monitor (MRM)", draft-ietf-mboned-mrm-01, July 2000.

[7] Bacher, D., Swan, A., and L. Rowe, "rtpmon: a third-party RTCP monitor", Conference 4th ACM International conference on multimedium, 1997.

[8] 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.

[9] Rajvaidya, P., Almeroth, K., and k. claffy, "A Scalable Architecture for Monitoring and Visualizating Multicast Statistics", Conference IFIP/IEEE Workshop on Distributed Systems: Operations & Management (DSOM), Austin, Texas, USA, December 2000.

[10] 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.

[11] Al-Shaer, E. and Y. Tang, "MRMON: Remote Multicast Monitoring", Conference NOMS, 2004.

[12] 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.

[13] Sarac, K. and K. Almeroth, "Monitoring IP Multicast in the Internet: Recent Advances and Ongoing Challenges", Journal IEEE Communication Magazine, 2005.

[14] 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

9. Acknowledgments

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

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

Mario Bianchetti
Broadband Network Services Innovation, Telecom Italia

Email: mario.bianchetti@telecomitalia.it

Giovanni Picciano
Access Network Engineering, Telecom Italia

Email: giovanni.picciano@telecomitalia.it

Mach(Guoyi) Chen
Huawei Technologies Co.,Ltd
KuiKe Building, No.9 Xinxi Rd.,
Hai-Dian District
Beijing, 100085
P.R. China

EMail: mach@huawei.com

Jian Qiu
Huawei Technology CO. LTD.
No. 9 Xinxi Road
Shangdi Information Industry Base
Hai-Dian District, Beijing 100085
China

Email: qj@huawei.com