Passive One-way Delay Measurement

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

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.

Abstract

This document describes a non-intrusive method for measuring one-way delay of IP packets.

Table of Contents

1. Introduction..................................................2
2. Terminology..................................................2
3. General Architecture........................................3
4. Observation Point Selection..................................5
5. Packet Capturing............................................5
6. Timestamping................................................5
7. Filtering/Classification.....................................5
8. Packet ID Generation........................................6
9. OWD Calculation Process.....................................7
10. Measurement Result Transfer.................................7
11. Security Considerations....................................8
11.1 Measurement Infrastructure................................8
11.2 Exchange of Measurement Data..............................8
11.3 Sensitivity..............................................9
12. References...............................................9
1. Introduction

There exist a variety of motivations for performing passive measurements in IP networks. Many applications require the measurement of the Quality of Service (QoS) for the transport of specific IP flows or traffic aggregates. Network providers and customers are interested whether negotiated QoS values in SLAs are met (SLA validation). Measurements also provide the basis for usage-based accounting. Furthermore, measurement results are an important input for traffic engineering decisions.

Special motivations for One-way delay measurements are listed in RFC2679. The advantage of passive measurements is that these measurements are based on existing traffic within the network. They provide a statement about the delay of the current traffic in the observed network section. Since no test traffic is generated, passive measurements can only be applied in cases where the kind of traffic we are interested in is already present in the network. This is the case for most applications where a statements about the actual situation in the network is required (like SLA validation, traffic engineering).

The goodness of the results of passive QoS measurements depend on the choice of the right observation points because the measurement results are only valid between the observation points.

This draft adopts the terminology of IPFIX, PSAMP and IPPM.

2. Terminology

Collecting Process
The collecting process receives records of flow or packet information. The data is stored for later processing (by the calculation process).

Exporting Process
The exporting processes send flow and packet records to the collecting processes. The records are generated by the measurement process.

IP Packet Selection Process
An IP packet selection process takes IP packets or parts of IP packets (e.g. header) as input and extracts a subset of these packets by applying a selection function.

Filtering
Filtering selects a subset of packets by applying deterministic functions on parts of the packet content like header fields or parts of the payload. A filtering process needs to process the packet (look at packet header and/or payload) in order to make the selection decision.

Measurement Device
A measurement device has access to at least one observation
point. It is hosting at least one measurement process and one export process.

Metering/Measurement Process
The measurement process generates records of packet and flow information. Packets passing the observation point are captured, timestamped, filtered and classified. The measurement process calculates the packet IDs.
3. General Architecture

Passive one-way delay measurements require two measurement processes at two observation points — at a reference and at a monitor observation point. We consider a unidirectional flow of packets as determined by source and destination address of its packets. The mentioned reference observation point is closer to the source of the packets while the monitor observation point is closer to the destination of the flow. The two measurement processes at reference and monitor observation point generate a timestamp and a unique packet ID for each packet as it passes an observation point. The information is sent to a common collecting process. A calculation process uses the data to calculate the one-way delay. (This structure is shown in [Figure 3-1: Passive OWD Measurement Setup].)

For each packet that is measured a timestamp and a packet ID has to be generated, stored and transmitted to a collecting process. Then the delay calculation takes place, based on correlating results from the different observation points. Therefore the amount of measurement result data that arises per second depends
on the number of measured packets $n$ per second, the number of bits $l_t$ used for the representation of the timestamp and the number of bits $l_{id}$ for the packet ID.

The location of the collecting process does not need to be necessarily physically different from the measurement process. Instead, the collecting process could be co-located with one of
the measurement processes in order to reduce the effort for transmitting measurement data.

[Figure 3-2] presents a refined view of the OWD Measurement Processes.

The processes involved are packet capturing, timestamping, filtering and classification, generation of a packet ID and transfer of measurement data. The requirements for these processes are examined in detail in the following subsections.

Further issues that must be dealt with are the optimal selection of the observation points, privacy issues when capturing public traffic, difficulties in packet event correlation when packets are lost or duplicated, and the overall amount of measurement data captured and transmitted for evaluation.
A problem that has to be solved when using multiple measurement points is clock synchronization between these points. Current solutions are based on the Network Time Protocol (NTP) [RFC1305], the Global Positioning System (GPS), and radio signals (e.g. DCF77). Each solution has its own drawbacks and advantages. [RFC2679] explains time keeping related terms like \textit{synchronization}, \textit{accuracy}, \textit{resolution} and \textit{skew} -
furthermore, it examines errors and uncertainties caused by imperfect synchronization and clocks.

4. Observation Point Selection

The placement of measurement processes is crucial in the sense that the observation point determines the endpoints for the measurement of the delay. When we aim to measure the delay between a source and destination host the locations for the observation points have to be as close as possible to source and destination, respectively. Although strictly speaking we never are able to measure a true end-to-end delay, practically we will have a very good estimation if we obeyed an appropriate observation point selection.

There is a second point that has a direct impact on observation point determination. We need a priori knowledge of the physical path that packets or flows, that we intend to examine, will follow. Then one has to select measurement processes, which are located at observation points along this path.

5. Packet Capturing

A certain amount of bytes needs to be captured per packet as basis for the generation of a packet ID. The packet ID collision probability depends on the generation function and the number of bytes that are used as input. In [DuGr00] it is stated that for IPv4 the first 40 Bytes starting at the IP header are sufficient.

6. Timestamping

A timestamp can be represented as absolute time. With this, the number of bits needed for the representation of the timestamp depends only on the desired accuracy for the measured metric. A possibility to reduce the number of bits $l_t$ used for the timestamp is to use relative timestamps. One approach is to make an assumption on the maximum time $t_{max}$ a packet needs to traverse the network from the ingress to the egress measurement point. With this upper limit, the timestamp needs to be non-ambiguous only within this limit. In this case the value $l_t$ depends not only on the desired accuracy of the time representation but also on the predetermined limit for the maximum time the packet needs to traverse the network. Another possibility is to use an absolute timestamp only for the first packet in a given time interval $[0,t_{int}]$ and use timestamps relative to this for successive packets that arrive in the same interval. Further issues are that the time that is needed for the timestamping process can differ for subsequent packets, which can also lead to inaccuracy [C1DG00].

7. Filtering/Classification

A filtering or classification of packets is required if only
selected packets are used for the measurement. A filter is useful to reduce the amount of resulting measurement data and the required processing time for the subsequent processes (like packet ID generation). The classification can filter out packets with specific characteristics as to choose packets from the population of interest. These can be for example all packets that belong to a specific flow or traffic aggregate (characterized by a common DiffServ Codepoint) in order to
determine the quality for specific applications or traffic classes.
In certain cases it is important to maintain the information to which flow or class the measured packet belongs to. For instance if the quality for different DiffServ traffic aggregates is measured simultaneously it is desired to keep the information about the class together with packet ID and timestamp.

In some cases, the packet ID already contains additional information because it has been calculated with a bijective function on the fields of information (e.g. a lossless compression function that compresses the IP header can be decompressed to determine the information on source and destination). In all other cases, the wanted information has to be transferred to the analysis application in addition to the packet ID.

8. Packet ID Generation

Passive measurements are based on methods where packets are neither marked nor modified. Consequently the recognition has to be based on fields that already exist in the packet. In order to get the same packet ID for one packet at both measurement points the packet ID generation should be based only on fields that are invariant or predictable during the transport. Fields that are highly variable between the packets (e.g. the datagram ID for IPv4) are more suitable than fields that are nearly constant or vary only between a few values (e.g. version field). The generation of a packet ID should be based on fields that

- already exist in the packet (no modification of the packet),
- are invariant or predictable during the transport (at least on the path from the ingress to the egress second measurement point) and
- are highly variable between the different packets.

This topic is covered in [RFC2402], in [GrDM98], [DuGr00] and [ZsZC01].

The request for low collisions (uniqueness of the ID) contradicts the request for a small packet ID; because the more bits are used for representing the packet ID the lower is the probability of collisions. The collision probability within a traffic trace depends on

- the distribution of the bit sequences taken as input to the packet ID generation (that means it is highly dependent on the considered traffic mix),
- the packet ID generation function,
- the size of the packet ID l_id,
- the used Operating System (OS) (if the datagram ID is considered)

The goal is to achieve an acceptable low probability of collisions with a packet ID that does not exceed the available capacity for the measurement result data transfer. As for the
timestamp, the packet ID only needs to be unique in the given
time interval \([0, t_{\text{max}}]\). This limits the number of possible
combinations to the number of packets \(n_{\text{max}}\) that can be observed
within this interval. For example for a 155 Mbits/s link with an
average packet size of 512 Bytes and a maximum time to traverse
the network of 10s, \(n_{\text{max}}\) would be 378,420 packets. This amount
of combinations can be represented by 19 bits.
IPv4 and IPv6 packets have different packet headers. That implies that the fields of the IP packets, which can be used for packet ID generation differ between IPv4 and IPv6 packets.

IPv4: ip-len, ip-ID, ip-prot, srcaddr, dstaddr, X bytes payload
IPv6: payload-len, ip-prot, srcaddr, dstaddr, Y bytes payload

There are different possibilities to generate an ID from the considered fields:
- unprocessed plain
- one-way hash function (e.g. MD5)
- checksum CRC
- compression function

Functions that map a large bit sequence (the selected fields of the packet) to a smaller bit sequence (the packet ID) can always increase the collision probability. Using the selected fields of the packet without further processing means to use the calculation basis itself instead of a derived ID. This would lead to the minimum collision probability that ever can be achieved with the given traffic mix. Furthermore it would reduce the required processing power because no function has to be performed on the selected fields. Nevertheless this method would result in a packet ID size m that is equal to the sum of bits in the selected fields. Especially for small packets this would increase the rate required for the measurement report messages up to the rate for the data flow itself.

9. OWD Calculation Process

To estimate the one-way delay of an IP packet between two observation points Ref (reference) and Mon (monitor) the difference of the arrival times of the packet at the two observation points is calculated:

\[ T_{\text{owd}} = T_{\text{ref}} - T_{\text{mon}} \]

The correlation between the two timestamps to the same IP packet is done via the packet ID in conjunction with a time window \( T_{\text{delta}} \). If a packet represented by its specific ID is captured at the reference point but its ID is not detected at the monitor point within \( T_{\text{delta}} \) the packet is considered as lost.

Note, that the difference in time, i.e. One-way delay, for a specific packet between two monitor points is semantically equivalent to the singleton metric defined in [RFC2679].

The IPPM Framework [RFC2330] refers to packet properties (packet type) as \( \text{"type-P}\)\(^\text{A}\). A \( \text{"type-P}\)\(^\text{A}\) might subsume properties such as protocol (UDP/TCP), size, TOS/DSCP, and so on. For passive One-way delay measurements, types of packets that are considered within the metric depend on the applied filter. Therefore, a description of set filters should be provided as part of any result.

10. Measurement Result Transfer
In order to calculate QoS parameters like delay, two timestamps have to be compared. Measurement results (timestamps and packet ID) from different observation points have to be collected at a common location in order to calculate the delay value. This collection point can be located on a separate host. It also can be co-located with one of the measurement processes in order to reduce the amount of data that has to be transferred over the
network. The following possibilities to transfer the measurement results have to be distinguished:

a) In-band: The measurement results are sent directly on the same path as the data.

b) Out-of-band: The measurement results are sent on a separate path.

Solution a) leads to additional load on the network under test. That means measurement result data packets can influence and distort the original data flow and we might end up with disadvantages that are inherent in active measurements. The in-band sending of packets with measurement results therefore somehow contradicts the passive approach where no influence on the original traffic is desired. Nevertheless, there is a difference between sending test traffic for active measurements and the transmission of measurement results. The amount, type and timeframe for the sending of test traffic for active measurements are dictated by the measurement task. In contrast to this, the sending of measurement results can be controlled by other means. For instance, it could be sent with a lower priority (e.g. lower-than-best-effort class) or only at times when the network is lightly loaded, or routed over paths that are currently not loaded (e.g. via MPLS). Which alternative is to be preferred depends on the policy for the evaluation of the metric (e.g. real-time or non-real-time).

Solution b) requires a separate path to the analysis application or collection point. This can be achieved for instance via a second interface card and a separate measurement network that connects all measurement points. This approach does not influence the data traffic but requires capacity in a different network.

In all cases, additional transfer capacity (either within the examined or a separate network) is required. For economic reasons even a separate reporting network would probably have a lower capacity than the ‘production network’. As a result, to save resources (storage capacity and bandwidth) the measurement result traffic should be kept as low as possible.

11. Security Considerations

11.1 Measurement Infrastructure

We have to distinguish between two different interfaces of a measurement device: there is one interface to control the devices and to exchange measurement data; a second interface is used to capture the traffic.

The access to the measurement infrastructure and the measurement devices in particular must be secured in a manner indicated by best known praxis in order to prevent unintended malicious use of the measurement infrastructure.

Malicious injection of packets into the network under test through the capture interface can be prevented by properly
utilizing network taps or optical splitters or by proper design of the physical interface.

11.2 Exchange of Measurement Data

The exchange of measurement data is vulnerable and appropriate actions have to be taken to hinder injection of faked measurement data. (However, the design of a data exchange
11.3 Sensitivity

When only packet identifier and timestamps are stored then the measurement data itself do not contain sensitive content as long as the packet identifier generation process is irreversible, so that for instance IP addresses cannot be retrieved.

12. References


13. Author’s Addresses

Lutz Mark
Fraunhofer Institute for Open Communication Systems
Kaiserin-Augusta-Allee 31
10589 Berlin
Germany
Phone: +49-30-34 63 7306
Fax: +49-30-34 53 8306
Email: mark@fokus.fraunhofer.de

Guido Pohl
Fraunhofer Institute for Open Communication Systems
Kaiserin-Augusta-Allee 31
10589 Berlin
Germany
Phone: +49-30-34 63 7164
Fax: +49-30-34 53 8164
Email: pohl@fokus.fraunhofer.de

Tanja Zseby
Fraunhofer Institute for Open Communication Systems
Kaiserin-Augusta-Allee 31
10589 Berlin
Germany
Phone: +49-30-34 63 7153
Fax: +49-30-34 53 8153
Email: zseby@fokus.fraunhofer.de

Kiminori Sugauchi
Hitachi, ltd., System Development Laboratory
1099, Ohzenji, Asao-ku, Kawasaki-shi, Kanagawa-ken,
215-0013 Japan
Phone: +81-44-959-0266
Fax: +81-44-959-0860
E-mail: sugauchi@sdl.hitachi.co.jp

14. Full Copyright Statement

Copyright (C) The Internet Society (2002). All Rights Reserved.
This document and translations of it may be copied and furnished
to others, and derivative works that comment on or otherwise
explain it or assist in its implementation may be prepared,
copied, published and distributed, in whole or in part, without
restriction of any kind, provided that the above copyright
notice and this paragraph are included on all such copies and
derivative works. However, this document itself may not be
modified in any way, such as by removing the copyright notice or
references to the Internet Society or other Internet
organizations, except as needed for the purpose of developing
Internet standards in which case the procedures for copyrights
defined in the Internet Standards process must be followed, or
as required to translate it into languages other than
English.
The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE
OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY
IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A
PARTICULAR PURPOSE.