Internet Engineering Task Force (IETF)
Request for Comments: 9341
Obsoletes: 8321
Category: Standards Track
ISSN: 2070-1721

G. Fioccola, Ed.
Huawei Technologies
M. Cociglio
Telecom Italia
G. Mirsky
Ericsson
T. Mizrahi
T. Zhou
Huawei Technologies

December 2022

Alternate-Marking Method

Abstract

This document describes the Alternate-Marking technique to perform packet loss, delay, and jitter measurements on live traffic. This technology can be applied in various situations and for different protocols. According to the classification defined in RFC 7799, it could be considered Passive or Hybrid depending on the application. This document obsoletes RFC 8321.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9341.

Table of Contents

1. Introduction

Most Service Providers' networks carry traffic with contents that are highly sensitive to packet loss [RFC7680], delay [RFC7679], and jitter [RFC3393].

Methodologies and tools are therefore needed to monitor and accurately measure network performance, in order to constantly control the quality of experience perceived by the end customers. Performance monitoring also provides useful information for improving network management (e.g., isolation of network problems, troubleshooting, etc.).

[RFC7799] defines Active, Passive, and Hybrid Methods of Measurement. In particular, Active Methods of Measurement depend on a dedicated measurement packet stream; Passive Methods of Measurement are based solely on observations of an undisturbed and unmodified packet stream of interest; Hybrid Methods are Methods of Measurement that use a combination of Active Methods and Passive Methods.

This document proposes a performance monitoring technique, called the "Alternate-Marking Method", which is potentially applicable to any kind of packet-based traffic, both point-to-point unicast and multicast, including Ethernet, IP, and MPLS. The method primarily addresses packet-loss measurement, but it can be easily extended to one-way or two-way delay and delay variation measurements as well.

The Alternate-Marking methodology, described in this document, allows the synchronization of the measurements at different points by dividing the packet flow into batches. So it is possible to get

coherent counters and timestamps in every marking period and therefore measure the Performance Metrics for the monitored flow.

The method has been explicitly designed for Passive or Hybrid measurements as stated in [RFC8321]. But, according to the definitions of [RFC7799], the Alternate-Marking Method can be classified as Hybrid Type I. Indeed, Alternate Marking can be implemented by using reserved bits in the protocol header, and the change in value of these marking bits at the domain edges (and not along the path) is formally considered a modification of the stream of interest.

It is worth mentioning that this is a methodology document, so the mechanism that can be used to transmit the counters and the timestamps is out of scope here. Additional details about the applicability of the Alternate-Marking methodology are described in [IEEE-NETWORK-PNPM].

1.1. Summary of Changes from RFC 8321

This document defines the Alternate-Marking Method, addressing ambiguities and building on its experimental phase that was based on the original specification [RFC8321].

The relevant changes are:

* Added the recommendations about the methods to employ in case one or two flag bits are available for marking (Section 7).

* Changed the structure to improve the readability.

* Removed the wording about the initial experiments of the method and considerations that no longer apply.

* Extended the description of detailed aspects of the methodology, e.g., synchronization, timing, packet fragmentation, and marked and unmarked traffic handling.

It is important to note that all the changes are totally backward compatible with [RFC8321] and no new additional technique has been introduced in this document compared to [RFC8321].

1.2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Overview of the Method

In order to perform packet-loss measurements on a production traffic flow, different approaches exist. The most intuitive one consists in numbering the packets so that each router that receives the flow can immediately detect a packet that is missing. This approach, though

very simple in theory, is not simple to achieve: it requires the insertion of a sequence number into each packet, and the devices must be able to extract the number and check it in real time. Such a task can be difficult to implement on live traffic: if UDP is used as the transport protocol, the sequence number is not available; on the other hand, if a higher-layer sequence number (e.g., in the RTP header) is used, extracting that information from each packet and processing it in real time could overload the device.

An alternate approach is to count the number of packets sent on one end, count the number of packets received on the other end, and compare the two values. This operation is much simpler to implement, but it requires the devices performing the measurement to be in sync: in order to compare two counters, it is required that they refer exactly to the same set of packets. Since a flow is continuous and cannot be stopped when a counter has to be read, it can be difficult to determine exactly when to read the counter. A possible solution to overcome this problem is to virtually split the flow in consecutive blocks by periodically inserting a delimiter so that each counter refers exactly to the same block of packets. The delimiter could be, for example, a special packet inserted artificially into the flow. However, delimiting the flow using specific packets has some limitations. First, it requires generating additional packets within the flow and requires the equipment to be able to process those packets. In addition, the method is vulnerable to out-of-order reception of delimiting packets and, to a lesser extent, to their loss.

The method proposed in this document follows the second approach, but it doesn't use additional packets to virtually split the flow in blocks. Instead, it "marks" the packets so that the packets belonging to the same block will have the same notional "color", whilst consecutive blocks will have different colors. Each change of color represents a sort of auto-synchronization signal that enhances the consistency of measurements taken by different devices along the path.

Figure 1 represents a very simple network and shows how the method can be used to measure packet loss on different network segments: by enabling the measurement on several interfaces along the path, it is possible to perform link monitoring, node monitoring, or end-to-end monitoring. The method is flexible enough to measure packet loss on any segment of the network and can be used to isolate the faulty element.

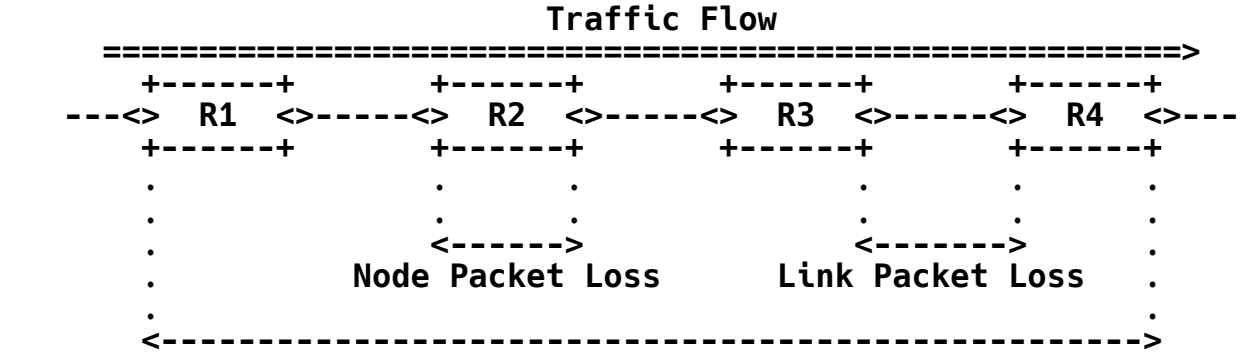

Figure 1: Available Measurements

3. Detailed Description of the Method

This section describes, in detail, how the method operates. A special emphasis is given to the measurement of packet loss, which represents the core application of the method, but applicability to delay and jitter measurements is also considered.

3.1. Packet-Loss Measurement

The basic idea is to virtually split traffic flows into consecutive blocks: each block represents a measurable entity unambiguously recognizable by all network devices along the path. By counting the number of packets in each block and comparing the values measured by different network devices along the path, it is possible to measure if packet loss occurred in any single block between any two points.

As discussed in the previous section, a simple way to create the blocks is to "color" the traffic (two colors are sufficient) so that packets belonging to alternate consecutive blocks will have different colors. Whenever the color changes, the previous block terminates and the new one begins. Hence, all the packets belonging to the same block will have the same color, and packets of different consecutive blocks will have different colors. The number of packets in each block depends on the criterion used to create the blocks:

* if the color is switched after a fixed number of packets, then each block will contain the same number of packets (except for any losses); and

* if the color is switched according to a fixed timer, then the number of packets may be different in each block depending on the packet rate.

The use of a fixed timer for the creation of blocks is REQUIRED when implementing this specification. The switching after a fixed number of packets is an additional possibility, but its detailed specification is out of scope. An example of application is in [EXPLICIT-FLOW-MEASUREMENTS].

The following figure shows how a flow appears when it is split into traffic blocks with colored packets.

A: packet with A coloring
B: packet with B coloring

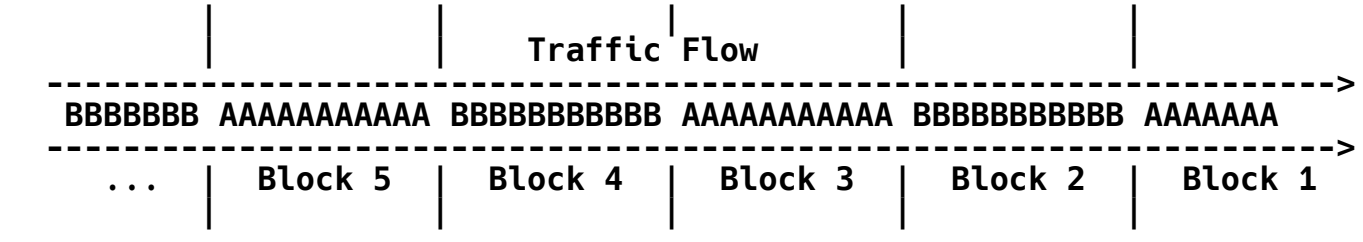

Figure 2: Traffic Coloring

Figure 3 shows how the method can be used to measure link packet loss between two adjacent nodes.

Referring to the figure, let's assume we want to monitor the packet loss on the link between two routers: router R1 and router R2. According to the method, the traffic is colored alternatively with two different colors: A and B. Whenever the color changes, the transition generates a sort of square-wave signal, as depicted in the following figure.

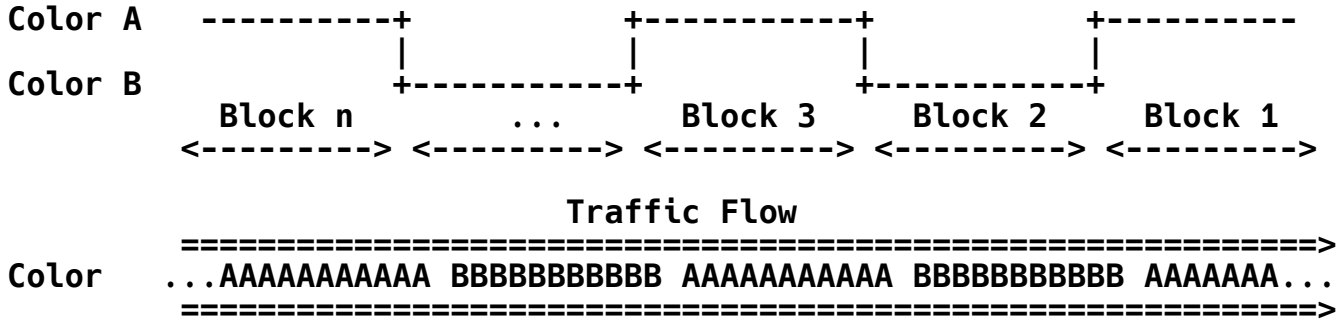

Figure 3: Computation of Link Packet Loss

Traffic coloring can be done by R1 itself if the traffic is not already colored. R1 needs two counters, C(A)R1 and C(B)R1, on its egress interface: C(A)R1 counts the packets with color A and C(B)R1 counts those with color B. As long as traffic is colored as A, only counter C(A)R1 will be incremented, while C(B)R1 is not incremented; conversely, when the traffic is colored as B, only C(B)R1 is incremented. C(A)R1 and C(B)R1 can be used as reference values to determine the packet loss from R1 to any other measurement point down the path. Router R2, similarly, will need two counters on its ingress interface, C(A)R2 and C(B)R2, to count the packets received on that interface and colored with A and B, respectively. When an A block ends, it is possible to compare C(A)R1 and C(A)R2 and calculate the packet loss within the block; similarly, when the successive B block terminates, it is possible to compare C(B)R1 with C(B)R2, and so on, for every successive block.

Likewise, by using two counters on the R2 egress interface, it is possible to count the packets sent out of the R2 interface and use them as reference values to calculate the packet loss from R2 to any measurement point downstream from R2.

The length of the blocks can be chosen large enough to simplify the collection and the comparison of measures taken by different network devices. It's preferable to read the value of the counters not immediately after the color switch: some packets could arrive out of order and increment the counter associated with the previous block (color), so it is worth waiting for some time. A safe choice is to wait L/2 time units (where L is the duration for each block) after the color switch, to read the counter of the previous color

(Section 5). The drawback is that the longer the duration of the block, the less frequently the measurement can be taken.

Two different strategies that can be used when implementing the method are:

flow-based: the flow-based strategy is used when well-defined traffic flows need to be monitored. According to this strategy, only the specified flow is colored. Counters for packet-loss measurements can be instantiated for each single flow, or for the set as a whole, depending on the desired granularity. With this approach, it is necessary to know in advance the path followed by flows that are subject to measurement. Path rerouting and traffic load balancing need to be taken into account.

link-based: measurements are performed on all the traffic on a link-by-link basis. The link could be a physical link or a logical link. Counters could be instantiated for the traffic as a whole or for each traffic class (in case it is desired to monitor each class separately), but in the second case, two counters are needed for each class.

The flow-based strategy is REQUIRED when implementing this specification. It requires the identification of the flow to be monitored and the discovery of the path followed by the selected flow. It is possible to monitor a single flow or multiple flows grouped together, but in this case, measurement is consistent only if all the flows in the group follow the same path. Moreover, if a measurement is performed by grouping many flows, it is not possible to determine exactly which flow was affected by packet loss. In order to have measures per single flow, it is necessary to configure counters for each specific flow. Once the flow(s) to be monitored has been identified, it is necessary to configure the monitoring on the proper nodes. Configuring the monitoring means configuring the rule to intercept the traffic and configuring the counters to count the packets. To have just an end-to-end monitoring, it is sufficient to enable the monitoring on the first- and last-hop routers of the path: the mechanism is completely transparent to intermediate nodes and independent of the path followed by traffic flows. On the contrary, to monitor the flow on a hop-by-hop basis along its whole path, it is necessary to enable the monitoring on every node from the source to the destination. In case the exact path followed by the flow is not known a priori (i.e., the flow has multiple paths to reach the destination), it is necessary to enable the monitoring on every path: counters on interfaces traversed by the flow will report packet count, whereas counters on other interfaces will be null.

3.2. One-Way Delay Measurement

The same principle used to measure packet loss can be applied also to one-way delay measurement. There are two methodologies, as described hereinafter.

Note that, for all the one-way delay alternatives described in the next sections, by summing the one-way delays of the two directions of a path, it is always possible to measure the two-way delay (round-

trip "virtual" delay). The Network Time Protocol (NTP) [RFC5905] or the IEEE 1588 Precision Time Protocol (PTP) [IEEE-1588] (as discussed in the previous section) can be used for the timestamp formats depending on the needed precision.

3.2.1. Single-Marking Methodology

The alternation of colors can be used as a time reference to calculate the delay. Whenever the color changes (which means that a new block has started), a network device can store the timestamp of the first packet of the new block; that timestamp can be compared with the timestamp of the same packet on a second router to compute packet delay. When looking at Figure 2, R1 stores the timestamp TS(A1)R1 when it sends the first packet of block 1 (A-colored), the timestamp TS(B2)R1 when it sends the first packet of block 2 (B-colored), and so on for every other block. R2 performs the same operation on the receiving side, recording TS(A1)R2, TS(B2)R2, and so on. Since the timestamps refer to specific packets (the first packet of each block), in the case where no packet loss or misordering exists, we would be sure that timestamps compared to compute delay refer to the same packets. By comparing TS(A1)R1 with TS(A1)R2 (and similarly TS(B2)R1 with TS(B2)R2, and so on), it is possible to measure the delay between R1 and R2. In order to have more measurements, it is possible to take and store more timestamps, referring to other packets within each block. The number of measurements could be increased by considering multiple packets in the block; for instance, a timestamp could be taken every N packets, thus generating multiple delay measurements. Taking this to the limit, in principle, the delay could be measured for each packet by taking and comparing the corresponding timestamps (possible but impractical from an implementation point of view).

In order to coherently compare timestamps collected on different routers, the clocks on the network nodes MUST be in sync (Section 5). Furthermore, a measurement is valid only if no packet loss occurs and if packet misordering can be avoided; otherwise, the first packet of a block on R1 could be different from the first packet of the same block on R2 (for instance, if that packet is lost between R1 and R2 or it arrives after the next one). Since packet misordering is generally undetectable, it is not possible to check whether the first packet on R1 is the same on R2, and this is part of the intrinsic error in this measurement.

3.2.1.1. Mean Delay

The method previously exposed for measuring the delay is sensitive to out-of-order reception of packets. In order to overcome this problem, an approach based on the concept of mean delay can be considered. The mean delay is calculated by considering the average arrival time of the packets within a single block. The network device locally stores a timestamp for each packet received within a single block: summing all the timestamps and dividing by the total number of packets received, the average arrival time for that block of packets can be calculated. By subtracting the average arrival times of two adjacent devices, it is possible to calculate the mean delay between those nodes. This method greatly reduces the number of

timestamps that have to be collected (only one per block for each network device), and it is robust to out-of-order packets with only a small error introduced in case of packet loss. But, when computing the mean delay, the measurement error could be augmented by accumulating the measurement error of a lot of packets. Additionally, it only gives one measure for the duration of the block, and it doesn't give the minimum, maximum, and median delay values [RFC6703]. This limitation could be overcome by reducing the duration of the block (for instance, from minutes to seconds), which implies a highly optimized implementation of the method. For this reason, the mean delay calculation may not be so viable in some cases.

3.2.2. Double-Marking Methodology

As mentioned above, the Single-Marking methodology for one-way delay measurement has some limitations, since it is sensitive to out-of-order reception of packets, and even the mean delay calculation is limited because it doesn't give information about the delay value's distribution for the duration of the block. Actually, it may be useful to have not only the mean delay but also the minimum, maximum, and median delay values and, in wider terms, to know more about the statistical distribution of delay values. So, in order to have more information about the delay and to overcome out-of-order issues, a different approach can be introduced, and it is based on a Double-Marking methodology.

Basically, the idea is to use the first marking to create the alternate flow and, within this colored flow, a second marking to select the packets for measuring delay/jitter. The first marking is needed for packet loss and may be used for mean delay measurement. The second marking creates a new set of marked packets that are fully identified over the network so that a network device can store the timestamps of these packets. These timestamps can be compared with the timestamps of the same packets on the next node to compute packet delay values for each packet. The number of measurements can be easily increased by changing the frequency of the second marking. But the frequency of the second marking must not be too high in order to avoid out-of-order issues. Between packets with the second marking, there should be an adequate time gap to avoid out-of-order issues and also to have a number of measurement packets that are rate independent. This gap may be, at the minimum, the mean network delay calculated with the previous methodology. Therefore, it is possible to choose a proper time gap to guarantee a fixed number of double-marked packets uniformly spaced in each block. If packets with the second marking are lost, it is easy to recognize the loss since the number of double-marked packets is known for each block. Based on the spacing between these packets, it can also be possible to understand which packet of the second marking sequence has been lost and perform the measurements only for the remaining packets. But this may be complicated if more packets are lost. In this case, an implementation may simply discard the delay measurements for the corrupted block and proceed with the next block.

An efficient and robust mode is to select a single packet with the second marking for each block; in this way, there is no time gap to

consider between the double-marked packets to avoid their reorder. In addition, it is also easier to identify the only double-marked packet in each block and skip the delay measurement for the block if it is lost.

The Double-Marking methodology can also be used to get more statistics of delay extent data, e.g., percentiles, variance, and median delay values. Indeed, a subset of batch packets is selected for extensive delay calculation by using the second marking, and it is possible to perform a detailed analysis on these double-marked packets. It is worth noting that there are classic algorithms for median and variance calculation, but they are out of the scope of this document. The conventional range (maximum-minimum) should be avoided for several reasons, including stability of the maximum delay due to the influence by outliers. In this regard, Section 6.5 of [RFC5481] highlights how the 99.9th percentile of delay and delay variation is more helpful to performance planners.

3.3. Delay Variation Measurement

Similar to one-way delay measurement (both for Single Marking and Double Marking), the method can also be used to measure the inter-arrival jitter. We refer to the definition in [RFC3393]. The alternation of colors, for a Single-Marking Method, can be used as a time reference to measure delay variations. In case of Double Marking, the time reference is given by the second-marked packets. Considering the example depicted in Figure 2, R1 stores the timestamp TS(A)R1 whenever it sends the first packet of a block, and R2 stores the timestamp TS(B)R2 whenever it receives the first packet of a block. The inter-arrival jitter can be easily derived from one-way delay measurement, by evaluating the delay variation of consecutive samples.

The concept of mean delay can also be applied to delay variation, by evaluating the average variation of the interval between consecutive packets of the flow from R1 to R2.

4. Alternate-Marking Functions

4.1. Marking the Packets

The coloring operation is fundamental in order to create packet blocks and marked packets. This implies choosing where to activate the coloring and how to color the packets.

In case of flow-based measurements, the flow to monitor can be defined by a set of selection rules (e.g., header fields) used to match a subset of the packets; in this way, it is possible to control the number of nodes involved, the path followed by the packets, and the size of the flows. It is possible, in general, to have multiple coloring nodes or a single coloring node that is easier to manage and doesn't raise any risk of conflict. Coloring in multiple nodes can be done, and the requirement is that the coloring must change periodically between the nodes according to the timing considerations in Section 5; so every node that is designated as a measurement point along the path should be able to identify unambiguously the colored

packets. Furthermore, [RFC9342] generalizes the coloring for multipoint-to-multipoint flow. In addition, it can be advantageous to color the flow as close as possible to the source because it allows an end-to-end measure if a measurement point is enabled on the last-hop router as well.

For link-based measurements, all traffic needs to be colored when transmitted on the link. If the traffic had already been colored, then it has to be re-colored because the color must be consistent on the link. This means that each hop along the path must (re-)color the traffic; the color is not required to be consistent along different links.

Traffic coloring can be implemented by setting specific flags in the packet header and changing the value of that bit periodically. How to choose the marking field depends on the application and is out of scope here.

4.2. Counting and Timestamping Packets

For flow-based measurements, assuming that the coloring of the packets is performed only by the source nodes, the nodes between source and destination (inclusive) have to count and timestamp the colored packets that they receive and forward: this operation can be enabled on every router along the path or only on a subset, depending on which network segment is being monitored (a single link, a particular metro area, the backbone, or the whole path). Since the color switches periodically between two values, two counters (one for each value) are needed for each flow and for every interface being monitored. The number of timestamps to be stored depends on the method for delay measurement that is applied. Furthermore, [RFC9342] generalizes the counting for multipoint-to-multipoint flow.

In case of link-based measurements, the behavior is similar except that coloring, counting, and timestamping operations are performed on a link-by-link basis at each endpoint of the link.

Another important consideration is when to read the counters or when to select the packets to be double-marked for delay measurement. It involves timing aspects to consider that are further described in Section 5.

4.3. Data Collection and Correlation

The nodes enabled to perform performance monitoring collect the value of the counters and timestamps, but they are not able to directly use this information to measure packet loss and delay, because they only have their own samples.

Data collection enables the transmission of the counters and timestamps as soon as it has been read. Data correlation is the mechanism to compare counters and timestamps for packet loss, delay, and delay variation calculation.

There are two main possibilities to perform both data collection and correlation depending on the Alternate-Marking application and use

case:

* Use of a centralized solution using the Network Management System (NMS) to correlate data. This can be done in Push Mode or Polling Mode. In the first case, each router periodically sends the information to the NMS; in the latter case, it is the NMS that periodically polls routers to collect information.

* Definition of a protocol-based distributed solution to exchange values of counters and timestamps between the endpoints. This can be done by introducing a new protocol or by extending the existing protocols (e.g., the Two-Way Active Measurement Protocol (TWAMP) as defined in [RFC5357] or the One-Way Active Measurement Protocol (OWAMP) as defined in [RFC4656]) in order to communicate the counters and timestamps between nodes.

In the following paragraphs, an example data correlation mechanism is explained and could be used independently of the adopted solutions.

When data is collected on the upstream and downstream nodes, e.g., packet counts for packet-loss measurement or timestamps for packet delay measurement, and is periodically reported to or pulled by other nodes or an NMS, a certain data correlation mechanism SHOULD be in use to help the nodes or NMS tell whether any two or more packet counts are related to the same block of markers or if any two timestamps are related to the same marked packet.

The Alternate-Marking Method described in this document literally splits the packets of the measured flow into different measurement blocks. An implementation MAY use a Block Number (BN) for data correlation. The BN MUST be assigned to each measurement block and associated with each packet count and timestamp reported to or pulled by other nodes or NMSs. When the nodes or NMS see, for example, the same BNs associated with two packet counts from an upstream and a downstream node, respectively, it considers that these two packet counts correspond to the same block. The assumption of this BN mechanism is that the measurement nodes are time synchronized. This requires the measurement nodes to have a certain time synchronization capability (e.g., the NTP [RFC5905] or the IEEE 1588 PTP [IEEE-1588]).

5. Synchronization and Timing

Color switching is the reference for all the network devices acting as measurement points, and the only requirement to be achieved is that they have to recognize the right batch along the path in order to get the related information of counters and timestamps.

In general, clocks in network devices are not accurate and for this reason, there is a clock error between the measurement points R1 and R2. And, to implement the methodology, they must be synchronized to the same clock reference with an adequate accuracy in order to guarantee that all network devices consistently match the marking bit to the correct block. Additionally, in practice, besides clock errors, packet reordering is also common in a packet network due to equal-cost multipath (ECMP). In particular, the delay between

measurement points is the main cause of out-of-order packets because each packet can be delayed differently. If the block is sufficiently large, packet reordering occurs only at the edge of adjacent blocks, and it can be easy to assign reordered packets to the right interval blocks.

In summary, we need to take into account two contributions: clock error between network devices and the interval we need to wait to avoid packets being out of order because of network delay.

The following figure explains both issues:

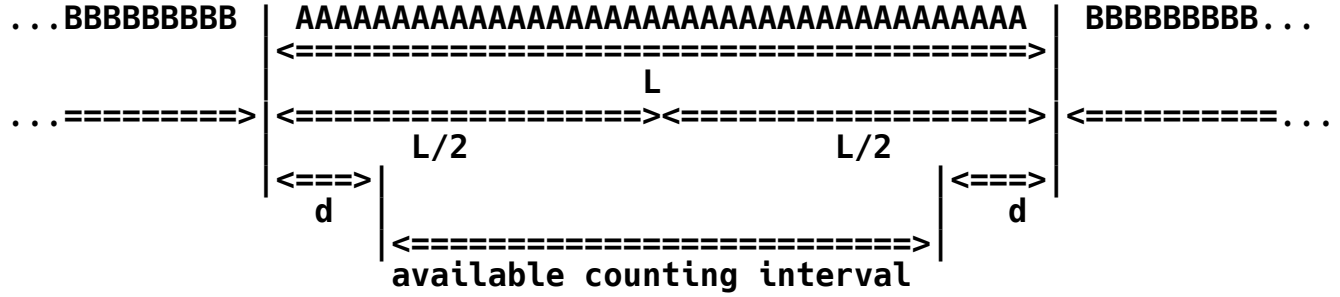

Figure 4: Timing Aspects

where L is the time duration of each block.

It is assumed that all network devices are synchronized to a common reference time with an accuracy of +/- A/2. Thus, the difference between the clock values of any two network devices is bounded by A.

The network delay between the network devices can be represented as a normal distribution and 99.7% of the samples are within 3 standard deviations of the average.

The guard band d is given by:

d = A + D_avg + 3*D_stddev,

where A is the clock accuracy, D_avg is the average value of the network delay between the network devices, and D_stddev is the standard deviation of the delay.

The available counting interval is L - 2d, which must be > 0.

The condition that MUST be satisfied and is a requirement on the synchronization accuracy is:

d < L/2.

This is the fundamental rule for deciding when to read the counters and when to select the packets to be double-marked; indeed, packet counters and double-marked packets MUST respectively be taken and chosen within the available counting interval that is not affected by error factors.

If the time duration L of each block is not so small, the

synchronization requirement could be satisfied even with a relatively inaccurate synchronization method.

6. Packet Fragmentation

Fragmentation can be managed with the Alternate-Marking Method using the following guidance:

> Marking nodes MUST mark all fragments if there are flag bits to use (i.e., it is in the specific encapsulation), as if they were separate packets.
>
> Nodes that fragment packets within the measurement domain SHOULD, if they have the capability to do so, ensure that only one resulting fragment carries the marking bit(s) of the original packet. Failure to do so can introduce errors into the measurement.
>
> Measurement points SHOULD simply ignore unmarked fragments and count marked fragments as full packets. However, if resources allow, measurement points MAY make note of both marked and unmarked initial fragments and only increment the corresponding counter if (a) other fragments are also marked or (b) it observes all other fragments and they are unmarked.

The proposed approach allows the marking node to mark all the fragments except in the case of fragmentation within the network domain; in that event, it is suggested to mark only the first fragment.

7. Recommendations for Deployment

The methodology described in the previous sections can be applied to various performance measurement problems. The only requirement is to select and mark the flow to be monitored; in this way, packets are batched by the sender, and each batch is alternately marked such that it can be easily recognized by the receiver. [RFC8321] reports experimental examples, and [IEEE-NETWORK-PNPM] also includes some information about the deployment experience.

Either one or two flag bits might be available for marking in different deployments:

One flag: packet-loss measurement MUST be done as described in Section 3.1, while delay measurement MUST be done according to the Single-Marking Method described in Section 3.2.1. Mean delay (Section 3.2.1.1) MAY also be used but it could imply more computational load.

Two flags: packet-loss measurement MUST be done as described in Section 3.1, while delay measurement MUST be done according to the Double-Marking Method as described in Section 3.2.2. In this case, Single Marking MAY also be used in combination with Double Marking, and the two approaches provide slightly different pieces of information that can be combined to have a more robust data set.

There are some operational guidelines to consider for the purpose of deciding to follow the recommendations above and to use one or two flags.

* The Alternate-Marking Method utilizes specific flags in the packet header, so an important factor is the number of flags available for the implementation. Indeed, if there is only one flag available, then there is no other way; if two flags are available, then the option with two flags is certainly more complete.

* The duration of the Alternate-Marking period affects the frequency of the measurement, and this is a parameter that can be decided on the basis of the required temporal sampling. But it cannot be freely chosen, as explained in Section 5.

* The Alternate-Marking methodologies enable packet loss, delay, and delay variation calculation, but in accordance with the method used (e.g., Single Marking or Double Marking), there is a different kind of information that can be derived. For example, to get more statistics of extent data, the option with two flags is desirable. For this reason, the type of data needed in the specific scenario is an additional element to take into account.

* The Alternate-Marking Methods imply different computational load depending on the method employed. Therefore, the available computational resources on the measurement points can also influence the choice. As an example, mean delay calculation may require more processing, and it may not be the best option to minimize the computational load.

The experiment with Alternate-Marking methodologies confirmed the benefits already described in [RFC8321].

A deployment of the Alternate-Marking Method should also take into account how to handle and recognize marked and unmarked traffic. Since Alternate Marking normally employs a marking field that is dedicated, reserved, and included in a protocol extension, the measurement points can learn whether the measurement is activated or not by checking if the specific extension is included or not within the packets.

It is worth mentioning some related work; in particular, [IEEE-NETWORK-PNPM] explains the Alternate-Marking Method together with new mechanisms based on hashing techniques.

7.1. Controlled Domain Requirement

The Alternate-Marking Method is an example of a solution limited to a controlled domain [RFC8799].

A controlled domain is a managed network that selects, monitors, and controls access by enforcing policies at the domain boundaries in order to discard undesired external packets entering the domain and to check internal packets leaving the domain. It does not necessarily mean that a controlled domain is a single administrative

domain or a single organization. A controlled domain can correspond to a single administrative domain or multiple administrative domains under a defined network management. It must be possible to control the domain boundaries and use specific precautions to ensure authentication, encryption, and integrity protection if traffic traverses the Internet.

For security reasons, the Alternate-Marking Method MUST only be applied to controlled domains.

8. Compliance with Guidelines from RFC 6390

[RFC6390] defines a framework and a process for developing Performance Metrics for protocols above and below the IP layer (such as IP-based applications that operate over reliable or datagram transport protocols).

This document doesn't aim to propose a new Performance Metric but rather a new Method of Measurement for a few Performance Metrics that have already been standardized. Nevertheless, it's worth applying guidelines from [RFC6390] to the present document, in order to provide a more complete and coherent description of the proposed method. The mechanisms described in this document use a combination of the Performance Metric Definition template defined in Section 5.4 of [RFC6390] and the Dependencies laid out in Section 5.5 of that document.

* Metric Name / Metric Description: as already stated, this document doesn't propose any new Performance Metrics. On the contrary, it describes a novel method for measuring packet loss [RFC7680]. The same concept, with small differences, can also be used to measure delay [RFC7679] and jitter [RFC3393]. The document mainly describes the applicability to packet-loss measurement.

* Method of Measurement or Calculation: according to the method described in the previous sections, the number of packets lost is calculated by subtracting the value of the counter on the source node from the value of the counter on the destination node. Both counters must refer to the same color. The calculation is performed when the value of the counters is in a steady state. The steady state is an intrinsic characteristic of the marking method counters because the alternation of color makes the counter associated with a color inactive for the duration of a marking period.

* Units of Measurement: the method calculates and reports the exact number of packets sent by the source node and not received by the destination node.

* Measurement Point(s) with Potential Measurement Domain: the measurement can be performed between adjacent nodes, on a per-link basis, or along a multi-hop path, provided that the traffic under measurement follows that path. In case of a multi-hop path, the measurements can be performed both end to end and hop by hop.

* Measurement Timing: the method has a constraint on the frequency

of measurements. This is detailed in Section 5, where it is specified that the marking period and the guard band interval are strictly related to each other to avoid out-of-order issues. That is because, in order to perform a measurement, the counter must be in a steady state, and this happens when the traffic is being colored with the alternate color.

* Implementation: the method uses one or two marking bits to color the packets; this enables the use of policy configurations on the router to color the packets and accordingly configure the counter for each color. The path followed by traffic being measured should be known in advance in order to configure the counters along the path and be able to compare the correct values.

* Verification: the methodology has been tested and deployed experimentally in both lab and operational network scenarios performing packet loss and delay measurements on traffic patterns created by traffic generators together with precision test instruments and network emulators.

* Use and Applications: the method can be used to measure packet loss with high precision on live traffic; moreover, by combining end-to-end and per-link measurements, the method is useful to pinpoint the single link that is experiencing loss events.

* Reporting Model: the value of the counters has to be sent to a centralized management system that performs the calculations; such samples must contain a reference to the time interval they refer to so that the management system can perform the correct correlation. The samples have to be sent while the corresponding counter is in a steady state (within a time interval); otherwise, the value of the sample should be stored locally.

* Dependencies: the values of the counters have to be correlated to the time interval they refer to.

* Organization of Results: the Method of Measurement produces singletons, according to the definition of [RFC2330].

* Parameters: the main parameters of the method are the information about the flow or the link to be measured, the time interval chosen to alternate the colors and to read the counters, and the type of method selected for packet-loss and delay measurements.

9. IANA Considerations

This document has no IANA actions.

10. Security Considerations

This document specifies a method to perform measurements that does not directly affect Internet security nor applications that run on the Internet. However, implementation of this method must be mindful of security and privacy concerns.

There are two types of security concerns: potential harm caused by

the measurements and potential harm to the measurements.

* Harm caused by the measurement: the measurements described in this document are Passive, so there are no new packets injected into the network causing potential harm to the network itself and to data traffic. Nevertheless, the method implies modifications on the fly to a header or encapsulation of the data packets: this must be performed in a way that doesn't alter the quality of service experienced by packets subject to measurements and that preserves stability and performance of routers doing the measurements. One of the main security threats in Operations, Administration, and Maintenance (OAM) protocols is network reconnaissance; an attacker can gather information about the network performance by passively eavesdropping on OAM messages. The advantage of the methods described in this document is that the marking bits are the only information that is exchanged between the network devices. Therefore, Passive eavesdropping on data plane traffic does not allow attackers to gain information about the network performance.

* Harm to the Measurement: the measurements could be harmed by routers altering the marking of the packets or by an attacker injecting artificial traffic. Authentication techniques, such as digital signatures, may be used where appropriate to guard against injected traffic attacks. Since the measurement itself may be affected by routers (or other network devices) along the path of IP packets intentionally altering the value of marking bits of packets, as mentioned above, the mechanism specified in this document can be applied just in the context of a controlled domain; thus, the routers (or other network devices) are locally administered, and this type of attack can be avoided.

An attacker that does not belong to the controlled domain can maliciously send marked packets. However, no problems occur if Alternate Marking is not supported in the controlled domain. If Alternate Marking is supported in the controlled domain, it is necessary to keep the measurements from being affected; therefore, externally marked packets must be checked to see if they are marked and eventually filtered or cleared.

The precondition for the application of the Alternate-Marking Method is that it MUST be applied in specific controlled domains, thus confining the potential attack vectors within the network domain. A limited administrative domain provides the network administrator with the means to select, monitor, and control the access to the network, making it a trusted domain. In this regard, it is expected to enforce policies at the domain boundaries to filter both external marked packets entering the domain and internal marked packets leaving the domain. Therefore, the trusted domain is unlikely subject to the hijacking of packets since marked packets are processed and used only within the controlled domain. But despite that, leakages may happen for different reasons, such as a failure or a fault. In this case, nodes outside the domain are expected to ignore marked packets since they are not configured to handle it and should not process it.

It might be theoretically possible to modulate the marking to serve as a covert channel to be used by an on-path observer. This may affect both the data and management plane, but, here too, the application to a controlled domain helps to reduce the effects.

It is worth highlighting that an attacker can't gain information about network performance from a single monitoring point; they must use synchronized monitoring points at multiple points on the path because they have to do the same kind of measurement and aggregation that Service Providers using Alternate Marking must do.

Attacks on the data collection and reporting of the statistics between the monitoring points and the NMS can interfere with the proper functioning of the system. Hence, the channels used to report back flow statistics MUST be secured.

The privacy concerns of network measurement are limited because the method only relies on information contained in the header or encapsulation without any release of user data. Although information in the header or encapsulation is metadata that can be used to compromise the privacy of users, the limited marking technique in this document seems unlikely to substantially increase the existing privacy risks from header or encapsulation metadata. It might be theoretically possible to modulate the marking to serve as a covert channel, but it would have a very low data rate if it is to avoid adversely affecting the measurement systems that monitor the marking.

Delay attacks are another potential threat in the context of this document. Delay measurement is performed using a specific packet in each block, marked by a dedicated color bit. Therefore, an on-path attacker can selectively induce synthetic delay only to delay-colored packets, causing systematic error in the delay measurements. As discussed in previous sections, the methods described in this document rely on an underlying time synchronization protocol. Thus, by attacking the time protocol, an attacker can potentially compromise the integrity of the measurement. A detailed discussion about the threats against time protocols and how to mitigate them is presented in [RFC7384].

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC3393] Demichelis, C. and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)", RFC 3393, DOI 10.17487/RFC3393, November 2002, <https://www.rfc-editor.org/info/rfc3393>.

[RFC7679] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Delay Metric for IP Performance Metrics (IPPM)", STD 81, RFC 7679, DOI 10.17487/RFC7679, January

2016, <https://www.rfc-editor.org/info/rfc7679>.

[RFC7680] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Loss Metric for IP Performance Metrics (IPPM)", STD 82, RFC 7680, DOI 10.17487/RFC7680, January 2016, <https://www.rfc-editor.org/info/rfc7680>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

11.2. Informative References

[EXPLICIT-FLOW-MEASUREMENTS]
Cociglio, M., Ferrieux, A., Fioccola, G., Lubashev, I., Bulgarella, F., Nilo, M., Hamchaoui, I., and R. Sisto, "Explicit Flow Measurements Techniques", Work in Progress, Internet-Draft, draft-ietf-ippm-explicit-flow-measurements-02, 13 October 2022, <https://datatracker.ietf.org/doc/html/draft-ietf-ippm-explicit-flow-measurements-02>.

[IEEE-1588]
IEEE, "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", IEEE Std 1588-2008, DOI 10.1109/IEEESTD.2008.4579760, July 2008, <https://doi.org/10.1109/IEEESTD.2008.4579760>.

[IEEE-NETWORK-PNPM]
Mizrahi, T., Navon, G., Fioccola, G., Cociglio, M., Chen, M., and G. Mirsky, "AM-PM: Efficient Network Telemetry using Alternate Marking", IEEE Network Vol. 33, Issue 4, DOI 10.1109/MNET.2019.1800152, July 2019, <https://doi.org/10.1109/MNET.2019.1800152>.

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics", RFC 2330, DOI 10.17487/RFC2330, May 1998, <https://www.rfc-editor.org/info/rfc2330>.

[RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)", RFC 4656, DOI 10.17487/RFC4656, September 2006, <https://www.rfc-editor.org/info/rfc4656>.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, DOI 10.17487/RFC5357, October 2008, <https://www.rfc-editor.org/info/rfc5357>.

[RFC5481] Morton, A. and B. Claise, "Packet Delay Variation Applicability Statement", RFC 5481, DOI 10.17487/RFC5481, March 2009, <https://www.rfc-editor.org/info/rfc5481>.

[RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms

Specification", RFC 5905, DOI 10.17487/RFC5905, June 2010, <https://www.rfc-editor.org/info/rfc5905>.

[RFC6390] Clark, A. and B. Claise, "Guidelines for Considering New Performance Metric Development", BCP 170, RFC 6390, DOI 10.17487/RFC6390, October 2011, <https://www.rfc-editor.org/info/rfc6390>.

[RFC6703] Morton, A., Ramachandran, G., and G. Maguluri, "Reporting IP Network Performance Metrics: Different Points of View", RFC 6703, DOI 10.17487/RFC6703, August 2012, <https://www.rfc-editor.org/info/rfc6703>.

[RFC7384] Mizrahi, T., "Security Requirements of Time Protocols in Packet Switched Networks", RFC 7384, DOI 10.17487/RFC7384, October 2014, <https://www.rfc-editor.org/info/rfc7384>.

[RFC7799] Morton, A., "Active and Passive Metrics and Methods (with Hybrid Types In-Between)", RFC 7799, DOI 10.17487/RFC7799, May 2016, <https://www.rfc-editor.org/info/rfc7799>.

[RFC8321] Fioccola, G., Ed., Capello, A., Cociglio, M., Castaldelli, L., Chen, M., Zheng, L., Mirsky, G., and T. Mizrahi, "Alternate-Marking Method for Passive and Hybrid Performance Monitoring", RFC 8321, DOI 10.17487/RFC8321, January 2018, <https://www.rfc-editor.org/info/rfc8321>.

[RFC8799] Carpenter, B. and B. Liu, "Limited Domains and Internet Protocols", RFC 8799, DOI 10.17487/RFC8799, July 2020, <https://www.rfc-editor.org/info/rfc8799>.

[RFC9342] Fioccola, G., Ed., Cociglio, M., Sapio, A., Sisto, R., and T. Zhou, "Clustered Alternate-Marking Method", RFC 9342, DOI 10.17487/RFC9342, December 2022, <https://www.rfc-editor.org/info/rfc9342>.

Acknowledgements

The authors would like to thank Alberto Tempia Bonda, Luca Castaldelli, and Lianshu Zheng for their contribution to the experimentation of the method.

The authors would also like to thank Martin Duke and Tommy Pauly for their assistance and their detailed and precious reviews.

Contributors

Xiao Min
ZTE Corp.
Email: xiao.min2@zte.com.cn

Mach(Guoyi) Chen
Huawei Technologies
Email: mach.chen@huawei.com

Alessandro Capello
Telecom Italia
Email: alessandro.capello@telecomitalia.it

Authors' Addresses

Giuseppe Fioccola (editor)
Huawei Technologies
Riesstrasse, 25
80992 Munich
Germany
Email: giuseppe.fioccola@huawei.com

Mauro Cociglio
Telecom Italia
Email: mauro.cociglio@outlook.com

Greg Mirsky
Ericsson
Email: gregimirsky@gmail.com

Tal Mizrahi
Huawei Technologies
Email: tal.mizrahi.phd@gmail.com

Tianran Zhou
Huawei Technologies
156 Beiqing Rd.
Beijing
100095
China
Email: zhoutianran@huawei.com