Network Working Group
Internet-Draft
Obsoletes: 8321 (if approved)
Intended status: Standards Track
Expires: August 27, 2022

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

February 23, 2022

Alternate-Marking Method
draft-fioccola-rfc8321bis-03

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. It could be considered Passive or Hybrid depending on the application. This document obsoletes [RFC8321].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on August 27, 2022.

Copyright Notice

Table of Contents

1. Introduction

Nowadays, most Service Providers' networks carry traffic with

contents that are highly sensitive to packet loss [RFC7680], delay [RFC7679], and jitter [RFC3393].

In view of this scenario, Service Providers need methodologies and tools to monitor and measure network performance with an adequate

accuracy, in order to constantly control the quality of experience perceived by their customers. Performance monitoring also provides useful information for improving network management (e.g., isolation of network problems, troubleshooting, etc.).

A lot of work related to Operations, Administration, and Maintenance (OAM), which also includes performance monitoring techniques, has been done by Standards Developing Organizations (SDOs): [RFC7276] provides a good overview of existing OAM mechanisms defined in the IETF, ITU-T, and IEEE. In the IETF, a lot of work has been done on fault detection and connectivity verification, while a minor effort has been thus far dedicated to performance monitoring. The IPPM WG has defined standard metrics to measure network performance; however, the methods developed in this WG mainly refer to focus on Active measurement techniques. More recently, the MPLS WG has defined mechanisms for measuring packet loss, one-way and two-way delay, and delay variation in MPLS networks [RFC6374], but their applicability to Passive measurements has some limitations, especially for pure connection-less networks.

The lack of adequate tools to measure packet loss with the desired accuracy drove an effort to design a new method for the performance monitoring of live traffic, which is easy to implement and deploy. The effort led to the method described in this document: basically, it is a Passive performance monitoring technique, potentially applicable to any kind of packet-based traffic, including Ethernet, IP, and MPLS, both unicast and multicast. The method addresses primarily packet loss measurement, but it can be easily extended to one-way or two-way delay and delay variation measurements as well.

The method has been explicitly designed for Passive measurements, but it can also be used with Active probes. Passive measurements are usually more easily understood by customers and provide much better accuracy, especially for packet loss measurements.

RFC 7799 [RFC7799] defines Passive and Hybrid Methods of Measurement.

In particular, 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.

Taking into consideration these definitions, the Alternate-Marking Method could be considered Hybrid or Passive, depending on the case. In the case where the marking method is obtained by changing existing field values of the packets the technique is Hybrid. In the case where the marking field is dedicated, reserved, and included in the protocol specification, the Alternate-Marking technique can be considered as Passive.

1.1. 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.

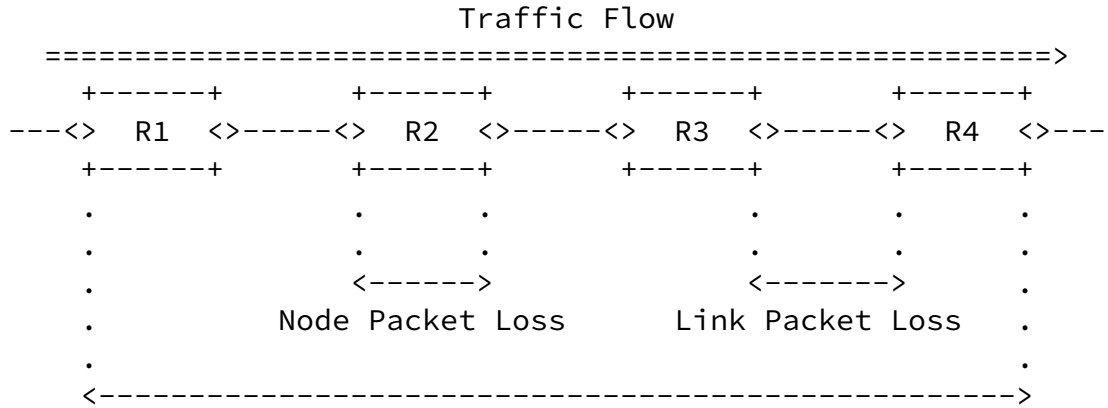
An alternate approach is to count the number of packets sent on one end, count the number of packets received on the other end, and compare the two values. This operation is much simpler to implement, but it requires the devices performing the measurement to be in sync: in order to compare two counters, 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 color, whilst consecutive blocks will have different colors. Each change of color represents a sort of auto-synchronization signal that guarantees 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.



End-to-End Packet Loss

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 different 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 rest of the document assumes that the blocks are created according to a fixed timer. The switching after a fixed number of packets is an additional possibility but its detailed specification is out of scope.

The following figure shows how a flow looks like when it is split in traffic blocks with colored packets.

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

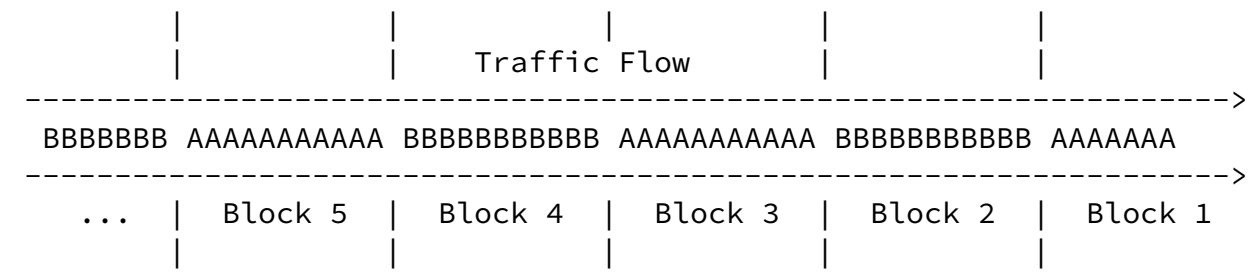
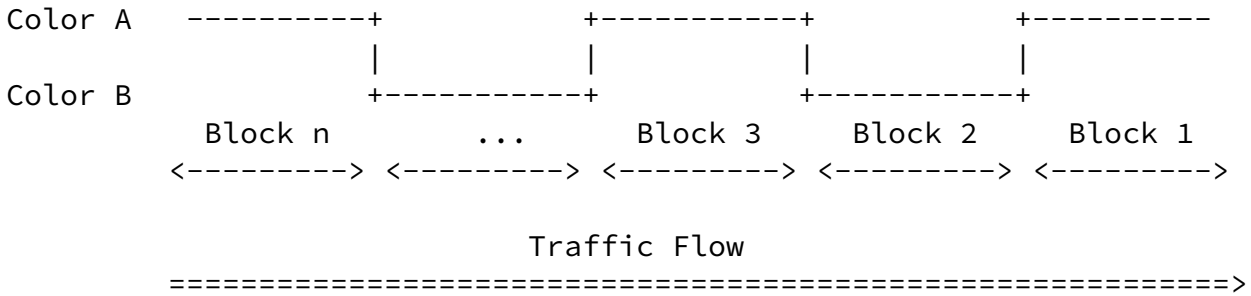


Figure 2: Traffic Coloring

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.



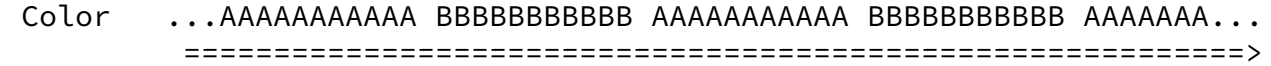


Figure 3: Computation of Link Packet Loss

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 down R2.

Using a fixed timer for color switching offers better control over the method: the (time) 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 still counter of the previous color, so the possibility of reading a running counter instead of a still one is minimized. The drawback is that the longer the duration of the block, the less frequent the measurement can be taken.

The timer-based batches is preferable because it is more

deterministic that the counter-based batches and it will be considered hereafter.

It's worth mentioning two different strategies that can be used when implementing the method:

- flow-based: the flow-based strategy is used when only a limited number of traffic flows need to be monitored. According to this strategy, only a subset of the flows 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. A relevant problem with this approach is the necessity to know in advance the path followed by flows that are subject to measurement. Path rerouting and traffic load-balancing increase the issue complexity, especially for unicast traffic. The problem is easier to solve for multicast traffic, where load-balancing is seldom used and static joins are frequently used to force traffic forwarding and replication.

- 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.

As mentioned, the flow-based measurement 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 from 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 system 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 three alternatives, 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).

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), we are 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. 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

As mentioned before, the method previously exposed for measuring the delay is sensitive to out-of-order reception of packets. In order to overcome this problem, a different approach has been considered: it is based on the concept of mean delay. 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. When computing the mean delay, the measurement error could be augmented by accumulating the measurement error of a lot of packets. This method is robust to out-of-order packets and also to packet loss (only a small error is introduced). Moreover, it greatly reduces the number of timestamps (only one per block for each network device) that have to be collected by the management system. On the other hand, 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 5 minutes to a few seconds), which implies a highly optimized implementation of the method.

3.2.2. Double-Marking Methodology

The Single-Marking methodology for one-way delay measurement is sensitive to out-of-order reception of packets. The first approach to overcome this problem has been described before and is based on the concept of mean delay. But the limitation of mean delay is that it doesn't give information about the delay value's distribution for the duration of the block. Additionally, 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 statistic

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; 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 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 a second router 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 a security time gap (e.g., this gap could be, at the minimum, the mean network delay calculated with the previous methodology) to avoid out-of-order issues and also to have a number of measurement packets that are rate independent. If a second-marking packet is lost, the delay measurement for the considered block is corrupted and should be discarded.

Mean delay is calculated on all the packets of a sample and is a simple computation to be performed for a Single-Marking Method. In some cases, the mean delay measure is not sufficient to characterize the sample, and more statistics of delay extent data are needed, e.g., percentiles, variance, and median delay values. The conventional range (maximum-minimum) should be avoided for several reasons, including stability of the maximum delay due to the influence by outliers. RFC 5481 [RFC5481], Section 6.5 highlights how the 99.9th percentile of delay and delay variation is more helpful to performance planners. To overcome this drawback, the idea is to couple the mean delay measure for the entire batch with a Double-Marking Method, where a subset of batch packets is selected for extensive delay calculation by using a second marking. In this way, it is possible to perform a detailed analysis on these double-marked packets. Please note that there are classic algorithms for median and variance calculation, but they are out of the scope of this document. The comparison between the mean delay for the entire

batch and the mean delay on these double-marked packets gives useful information since it is possible to understand if the Double-Marking measurements are actually representative of the delay trends.

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 RFC 3393 [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 involved nodes, 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, [I-D.fioccola-rfc8889bis] 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 (included) 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: one counter for packets with color A and one counter for packets with color B. For each flow (or group of flows) being monitored and for every interface where the monitoring is Active, two counters are needed. For example, in order to separately monitor three flows on a router with four interfaces involved, 24 counters are needed (two counters for each of the three flows on each of the four interfaces). The number of timestamps to be stored depends on the method for delay measurement that is applied.

Furthermore, [I-D.fioccola-rfc8889bis] 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 aspect to take into consideration is when to read the counters: in order to count the exact number of packets of a block, the routers must perform this operation when that block has ended; in other words, the counter for color A must be read when the current block has color B, in order to be sure that the value of the counter is stable. This task can be accomplished in two ways. The general approach suggests reading the counters periodically, many times during a block duration, and comparing these successive readings: when the counter stops incrementing, it means that the current block has ended, and its value can be elaborated safely. Alternatively, if the coloring operation is performed on the basis of a fixed timer, it is possible to configure the reading of the counters according to that timer: for example, reading the counter for color A every period in the middle of the subsequent block with color B is a safe choice. A sufficient margin should be considered between the end of a block and the reading of the counter, in order to take into account any out-of-order packets. Regarding the selection of the packet to be double-marked for delay measurement, the same considerations for packet loss measurement apply also here and it is reasonable to choose the double-marked packet in the middle of the block. The timing aspects 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. While, 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 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. In any case, the NMS has to collect all the relevant values from all the routers within one cycle of the timer.

- 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 RFC 5357 [RFC5357] or the One-Way Active Measurement Protocol (OWAMP) as defined in RFC 4656 [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; in addition, a Block Number (BN) could be assigned to each such measurement block. The BN is generated each time a node reads the data (packet counts or timestamps) and is associated with each

packet count and timestamp reported to or pulled by other nodes or NMSs. The value of a BN could be calculated as the modulo of the local time (when the data are read) and the interval of the marking time period.

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, i.e., these two packet counts belong to the same block of markers from the upstream and downstream nodes. 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 Network Time Protocol (NTP) [RFC5905] or the IEEE 1588 Precision Time Protocol (PTP) [IEEE-1588]).

5. Synchronization and Timing

This document introduces two color-switching methods: one is based on a fixed number of packets, and the other is based on a fixed timer. But the method based on a fixed timer is preferable because it is more deterministic, and it is considered in the document.

Color switching is the reference for all the network devices, and the only requirement to be achieved is that all network devices have to recognize the right batch along the path.

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. But, to implement the methodology, they must be synchronized to the same clock reference with an accuracy of +/- L/2 time units, where L is the fixed time duration of the block. So each colored packet can be assigned to the right batch by each router. This is because the minimum time distance between two packets of the same color but that belong to different batches is L time units. This level of accuracy guarantees that all network devices consistently match the marking bit to the correct block.

If the value of L is not too small, this synchronization requirement could be satisfied even with a relatively inaccurate synchronization method. This is true for packet loss and two-way delay measurement, but not for one-way delay measurement, where clock synchronization must be accurate. Therefore, a system that uses only packet loss and two-way delay measurement does not require a very precise synchronization. This is because the value of the clocks of network devices does not affect the computation of the two-way delay measurement.

But, in practice, besides clock errors, packet reordering is also very 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 because each packet can be delayed differently. For this reason, the accuracy of the Alternate-Marking Method, especially for packet loss measurement, is affected by packet reordering.

If the time duration L of each block is too small, it may be difficult to determine to which block the reordered packets belong. However, if the value of L is sufficiently large, packet reordering occurs only at the edge of adjacent blocks and it becomes easy to assign reordered packets to the right interval blocks. This means that, without considering clock error, we can wait L/2 after color switching to be sure to take a still counter and mitigate the reordering issues.

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.

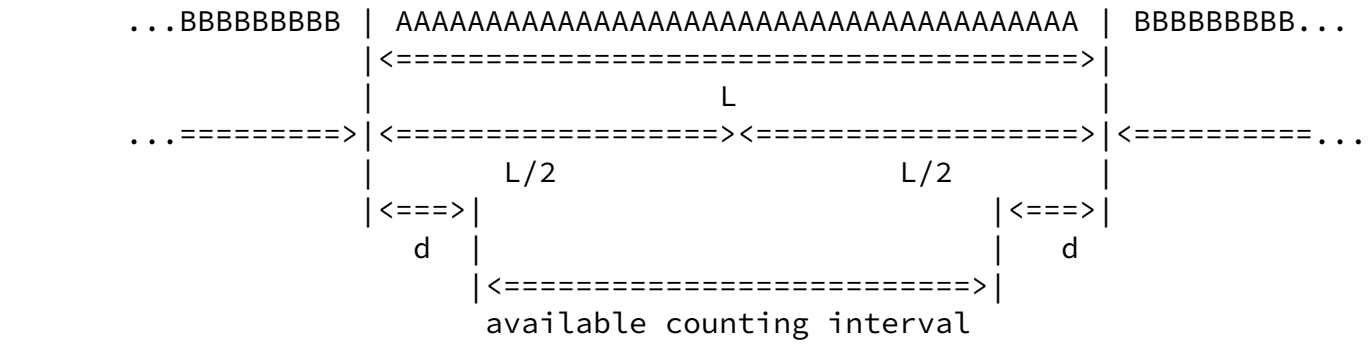


Figure 4: Timing Aspects

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 data set and 99.7% of the samples are within 3 standard deviation 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 that must be > 0.

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

d < L/2.

6. Packet Fragmentation

Fragmentation can be managed with the Alternate-Marking Method and in particular it is possible to give 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 MAY 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. In addition it could be possible to take the counters properly in order to keep track of both marked and unmarked fragments.

7. Results of the Alternate Marking Experiment

The methodology described in the previous sections can be applied to various performance measurement problems, as explained in [RFC8321]. 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.

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

One flag: packet loss measurement SHOULD be done as described in Section 3.1, while delay measurement MAY be done according to the single-marking method described in Section 3.2.1. Mean delay (Section 3.2.1.1) is NOT RECOMMENDED since it implies more computational load.

Two flags: packet loss measurement SHOULD be done as described in Section 3.1, while delay measurement SHOULD be done according to double-marking method 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.

The experiment with Alternate Marking methodologies confirmed the following benefits:

- easy implementation: it can be implemented by using features already available on major routing platforms, or by applying an optimized implementation of the method for both legacy and newest technologies;
- low computational effort: the additional load on processing is negligible;
- accurate loss and delay measurements: single packet loss granularity is achieved with a Passive measurement;
- potential applicability to any kind of packet-based or frame-based traffic: Ethernet, IP, MPLS, etc., and both unicast and multicast;

- robustness: the method can easily tolerate out-of-order packets, and it's not based on "special" packets whose loss could have a negative impact;

- flexibility: all the timestamp formats are allowed, because they are managed out of band. The format (the Network Time Protocol (NTP) [RFC5905] or the IEEE 1588 Precision Time Protocol (PTP) [IEEE-1588]) depends on the precision you want; and

- no interoperability issues: the features required are available on all current routing platforms. Both a centralized or distributed solution can be used to harvest data from the routers.

A deployment of the Alternate-Marking Method SHOULD also take into account how to handle and recognize marked and unmarked traffic

depending on whether the technique is applied as Hybrid or Passive. In the case where the marking method is applied by changing existing fields of the packets, it is RECOMMENDED to use an additional flag or some out-of-band signaling to indicate if the measurement is activated or not in order to inform the measurement points. While, in the case where the marking field 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 as also further described in [I-D.mizrahi-ippm-marking]; while [I-D.zhou-ippm-enhanced-alternate-marking] extends the Alternate-Marking Data Fields, to provide enhanced capabilities and allow advanced functionalities.

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

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 if traffic traverses the Internet.

For security reasons, the Alternate Marking Method is RECOMMENDED only for controlled domains.

8. Compliance with Guidelines from RFC 6390

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. We used 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 counters associated with each color still one at a time 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 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: both in the lab and in the operational network, the methodology has been tested and experimented for packet loss and delay measurements by using 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.

- Parameters: currently, the main parameter of the method is the time interval used to alternate the colors and read the counters.

9. IANA Considerations

This document has no IANA actions.

10. Security Considerations

This document specifies a method to perform measurements in the context of a Service Provider's network and has not been developed to conduct Internet measurements, so it 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 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. In addition, an attacker can't gain information about network performance from a single monitoring point; it 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 network management system 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, a

man-in-the-middle 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 RFC 7384 [RFC7384].

11. Contributors

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

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

12. 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 thank Martin Duke and Tommy Pauly for their assistance and their detailed and precious reviews.

13. References

13.1. Normative References

[IEEE-1588] IEEE, "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", IEEE Std 1588-2008.

[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>.

[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>.

[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>.

13.2. Informative References

[I-D.fioccola-rfc8889bis]
Fioccola, G., Cociglio, M., Sapio, A., Sisto, R., and T. Zhou, "Multipoint Alternate-Marking Method", draft-fioccola-rfc8889bis-02 (work in progress), February 2022.

[I-D.mizrahi-ippm-marking]
Mizrahi, T., Fioccola, G., Cociglio, M., Chen, M., and G. Mirsky, "Marking Methods for Performance Measurement", draft-mizrahi-ippm-marking-00 (work in progress), October 2021.

[I-D.zhou-ippm-enhanced-alternate-marking]
Zhou, T., Fioccola, G., Liu, Y., Lee, S., Cociglio, M., and W. Li, "Enhanced Alternate Marking Method", draft-zhou-ippm-enhanced-alternate-marking-08 (work in progress), January 2022.

[IEEE-Network-PNPM]
IEEE Network, "AM-PM: Efficient Network Telemetry using Alternate Marking", DOI 10.1109/MNET.2019.1800152, 2019.

[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>.

[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>.

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, DOI 10.17487/RFC6374, September 2011, <https://www.rfc-editor.org/info/rfc6374>.

[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>.

[RFC7276] Mizrahi, T., Sprecher, N., Bellagamba, E., and Y. Weingarten, "An Overview of Operations, Administration, and Maintenance (OAM) Tools", RFC 7276, DOI 10.17487/RFC7276, June 2014, <https://www.rfc-editor.org/info/rfc7276>.

[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>.

[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>.

[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>.

Appendix A. Changes Log

Changes from RFC 8321 include:

- Minor editorial changes
- Replacement of the section on "Applications, Implementation, and Deployment" with "Finding of the Alternate Marking Implementations and Deployments"
- Moved advantages and benefits of the method from "Introduction" to the new section on "Finding of the Alternate Marking Implementations and Deployments"
- Removed section on "Hybrid Measurement"

Changes in v-(01) include:

- Considerations on the reference: [IEEE-Network-PNPM]
- Clarified that the method based on a fixed timer is specified in this document while the method based on a fixed number of packets is only mentioned but not detailed.
- Explanation of the the intrinsic error in section 3.3.1 on "Single-Marking Methodology"
- Deleted some parts in section 4 "Considerations" that no longer

apply

- New section on "Packet Fragmentation"

Changes in v-(02) include:

- Considerations on how to handle unmarked traffic in section 5 on "Results of the Alternate Marking Experiment"

- Minor rewording in section 4.4 on "Packet Fragmentation"

Changes in v-(03) include:

- Deleted numeric examples in sections on "Packet Loss Measurement" and on "Single-Marking Methodology"

- New section on "Alternate Marking Functions"

- Moved sections 3.1.1 on "Coloring the Packets", 3.1.2 on "Counting the Packets" and 3.1.3 on "Collecting Data and Calculating Packet Loss" into the new section on "Alternate Marking Functions"

- Renamed sections 4.1 as "Marking the Packets", 4.2 as "Counting and Timestamping Packets" and 4.3 as "Data Collection and Correlation"

- Merged old section on "Data Correlation" with section 4.3 on "Data Collection and Correlation"

- Moved and renamed section on "Timing Aspects" as "Synchronization and Timing"

- Merged old section on "Synchronization" with section on "Synchronization and Timing"

- Merged old section on "Packet Reordering" with section on "Synchronization and Timing"

Authors' Addresses

Giuseppe Fioccola (editor)
Huawei Technologies
Riesstrasse, 25
Munich 80992
Germany

Email: giuseppe.fioccola@huawei.com

Mauro Cociglio
Telecom Italia
Via Reiss Romoli, 274
Torino 10148
Italy

Email: mauro.cociglio@telecomitalia.it

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

Xiao Min
ZTE Corp.

Email: xiao.min2@zte.com.cn