Internet Engineering Task Force (IETF)
Request for Comments: 8889
Category: Experimental
ISSN: 2070-1721

G. Fioccola, Ed.
Huawei Technologies
M. Cociglio
Telecom Italia
A. Sapio
Intel Corporation
R. Sisto
Politecnico di Torino

August 2020

Multipoint Alternate-Marking Method for Passive and Hybrid Performance Monitoring

Abstract

The Alternate-Marking method, as presented in RFC 8321, can only be applied to point-to-point flows, because it assumes that all the packets of the flow measured on one node are measured again by a single second node. This document generalizes and expands this methodology to measure any kind of unicast flow whose packets can follow several different paths in the network -- in wider terms, a multipoint-to-multipoint network. For this reason, the technique here described is called "Multipoint Alternate Marking".

Status of This Memo

This document is not an Internet Standards Track specification; it is published for examination, experimental implementation, and evaluation.

This document defines an Experimental Protocol for the Internet community. 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). Not all documents approved by the IESG are candidates for any level of Internet Standard; see 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/rfc8889.

Table of Contents

1. Introduction

The Alternate-Marking method, as described in RFC 8321 [RFC8321], is applicable to a point-to-point path. The extension proposed in this document applies to the most general case of multipoint-to-multipoint path and enables flexible and adaptive performance measurements in a managed network.

The Alternate-Marking methodology described in RFC 8321 [RFC8321] allows the synchronization of the measurements in different points by dividing the packet flow into batches. So it is possible to get coherent counters and show what is happening in every marking period for each monitored flow. The monitoring parameters are the packet counter and timestamps of a flow for each marking period. Note that additional details about the applicability of the Alternate-Marking methodology are described in RFC 8321 [RFC8321] while implementation details can be found in the paper "AM-PM: Efficient Network Telemetry using Alternate Marking" [IEEE-Network-PNPM].

There are some applications of the Alternate-Marking method where there are a lot of monitored flows and nodes. Multipoint Alternate Marking aims to reduce these values and makes the performance monitoring more flexible in case a detailed analysis is not needed. For instance, by considering n measurement points and m monitored flows, the order of magnitude of the packet counters for each time interval is n*m*2 (1 per color). The number of measurement points

and monitored flows may vary and depends on the portion of the network we are monitoring (core network, metro network, access network) and the granularity (for each service, each customer). So if both n and m are high values, the packet counters increase a lot, and Multipoint Alternate Marking offers a tool to control these parameters.

The approach presented in this document is applied only to unicast flows and not to multicast. Broadcast, Unknown Unicast, and Multicast (BUM) traffic is not considered here, because traffic replication is not covered by the Multipoint Alternate-Marking method. Furthermore, it can be applicable to anycast flows, and Equal-Cost Multipath (ECMP) paths can also be easily monitored with this technique.

In short, RFC 8321 [RFC8321] applies to point-to-point unicast flows and BUM traffic, while this document and its Clustered Alternate-Marking method is valid for multipoint-to-multipoint unicast flows, anycast, and ECMP flows.

Therefore,the Alternate-Marking method can be extended to any kind of multipoint-to-multipoint paths, and the network-clustering approach presented in this document is the formalization of how to implement this property and allow a flexible and optimized performance measurement support for network management in every situation.

Without network clustering, it is possible to apply Alternate Marking only for all the network or per single flow. Instead, with network clustering, it is possible to use the partition of the network into clusters at different levels in order to perform the needed degree of detail. In some circumstances, it is possible to monitor a multipoint network by analyzing the network clustering, without examining in depth. In case of problems (packet loss is measured or the delay is too high), the filtering criteria could be specified more in order to perform a detailed analysis by using a different combination of clusters up to a per-flow measurement as described in RFC 8321 [RFC8321].

This approach fits very well with the Closed-Loop Network and Software-Defined Network (SDN) paradigm, where the SDN orchestrator and the SDN controllers are the brains of the network and can manage flow control to the switches and routers and, in the same way, can calibrate the performance measurements depending on the desired accuracy. An SDN controller application can orchestrate how accurately the network performance monitoring is set up by applying the Multipoint Alternate Marking as described in this document.

It is important to underline that, as an extension of RFC 8321 [RFC8321], this is a methodology document, so the mechanism that can be used to transmit the counters and the timestamps is out of scope here, and the implementation is open. Several options are possible -- e.g., see "Enhanced Alternate Marking Method" [ENHANCED-ALTERNATE-MARKING].

Note that the fragmented packets case can be managed with the Alternate-Marking methodology only if fragmentation happens outside

the portion of the network that is monitored. This is always true for both RFC 8321 [RFC8321] and Multipoint Alternate Marking, as explained here.

2. Terminology

The definitions of the basic terms are identical to those found in Alternate Marking [RFC8321]. It is to be remembered that RFC 8321 [RFC8321] is valid for point-to-point unicast flows and BUM traffic.

The important new terms that need to be explained are listed below:

Multipoint Alternate Marking: Extension to RFC 8321 [RFC8321], valid for multipoint-to-multipoint unicast flows, anycast, and ECMP flows. It can also be referred to as Clustered Alternate Marking.

Flow definition: The concept of flow is generalized in this document. The identification fields are selected without any constraints and, in general, the flow can be a multipoint-to-multipoint flow, as a result of aggregate point-to-point flows.

Monitoring network: Identified with the nodes of the network that are the measurement points (MPs) and the links that are the connections between MPs. The monitoring network graph depends on the flow definition, so it can represent a specific flow or the entire network topology as aggregate of all the flows.

Cluster: Smallest identifiable subnetwork of the entire monitoring network graph that still satisfies the condition that the number of packets that go in is the same as the number that go out.

Multipoint metrics: Packet loss, delay, and delay variation are extended to the case of multipoint flows. It is possible to compute these metrics on the basis of multipoint paths in order to associate the measurements to a cluster, a combination of clusters, or the entire monitored network. For delay and delay variation, it is also possible to define the metrics on a single-packet basis, and it means that the multipoint path is used to easily couple packets between input and output nodes of a multipoint path.

The next section highlights the correlation with the terms used in RFC 5644 [RFC5644].

2.1. Correlation with RFC 5644

RFC 5644 [RFC5644] is limited to active measurements using a single source packet or stream. Its scope is also limited to observations of corresponding packets along the path (spatial metric) and at one or more destinations (one-to-group) along the path.

Instead, the scope of this memo is to define multiparty metrics for passive and hybrid measurements in a group-to-group topology with multiple sources and destinations.

RFC 5644 [RFC5644] introduces metric names that can be reused here

but have to be extended and rephrased to be applied to the Alternate-Marking schema:

a. the multiparty metrics are not only one-to-group metrics but can be also group-to-group metrics;

b. the spatial metrics, used for measuring the performance of segments of a source to destination path, are applied here to group-to-group segments (called clusters).

3. Flow Classification

A unicast flow is identified by all the packets having a set of common characteristics. This definition is inspired by RFC 7011 [RFC7011].

As an example, by considering a flow as all the packets sharing the same source IP address or the same destination IP address, it is easy to understand that the resulting pattern will not be a point-to-point connection, but a point-to-multipoint or multipoint-to-point connection.

In general, a flow can be defined by a set of selection rules used to match a subset of the packets processed by the network device. These rules specify a set of Layer 3 and Layer 4 header fields (identification fields) and the relative values that must be found in matching packets.

The choice of the identification fields directly affects the type of paths that the flow would follow in the network. In fact, it is possible to relate a set of identification fields with the pattern of the resulting graphs, as listed in Figure 1.

A TCP 5-tuple usually identifies flows following either a single path or a point-to-point multipath (in the case of load balancing). On the contrary, a single source address selects aggregate flows following a point-to-multipoint, while a multipoint-to-point can be the result of a matching on a single destination address. In the case where a selection rule and its reverse are used for bidirectional measurements, they can correspond to a point-to-multipoint in one direction and a multipoint-to-point in the opposite direction.

So the flows to be monitored are selected into the monitoring points using packet selection rules, which can also change the pattern of the monitored network.

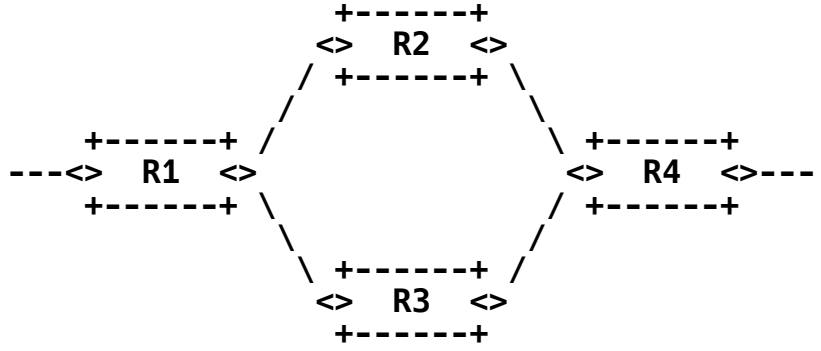
Note that, more generally, the flow can be defined at different levels based on the potential encapsulation, and additional conditions that are not in the packet header can also be included as part of matching criteria.

The Alternate-Marking method is applicable only to a single path (and partially to a one-to-one multipath), so the extension proposed in this document is suitable also for the most general case of multipoint-to-multipoint, which embraces all the other patterns of

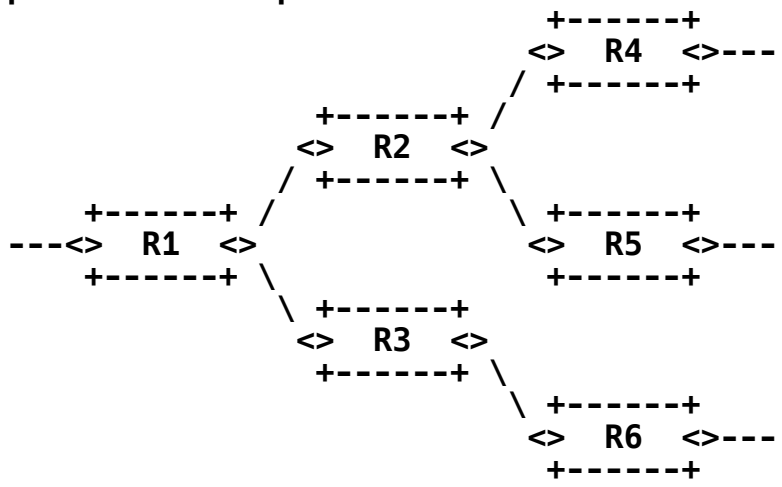
Figure 1.

point-to-point single path

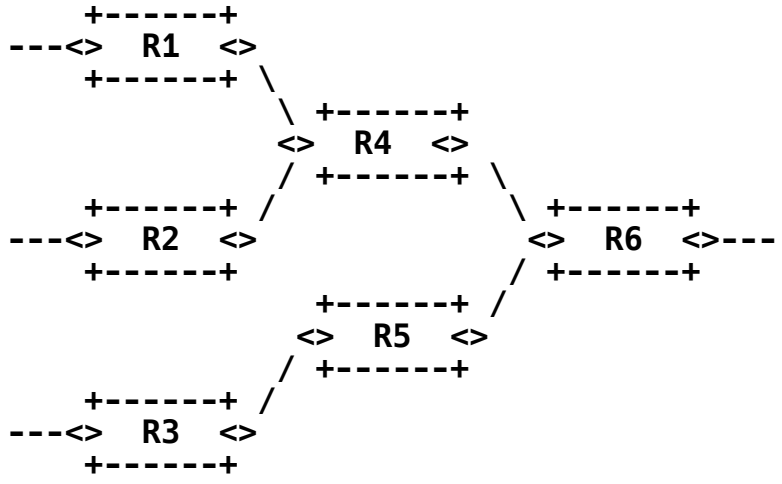
point-to-point multipath



point-to-multipoint



multipoint-to-point



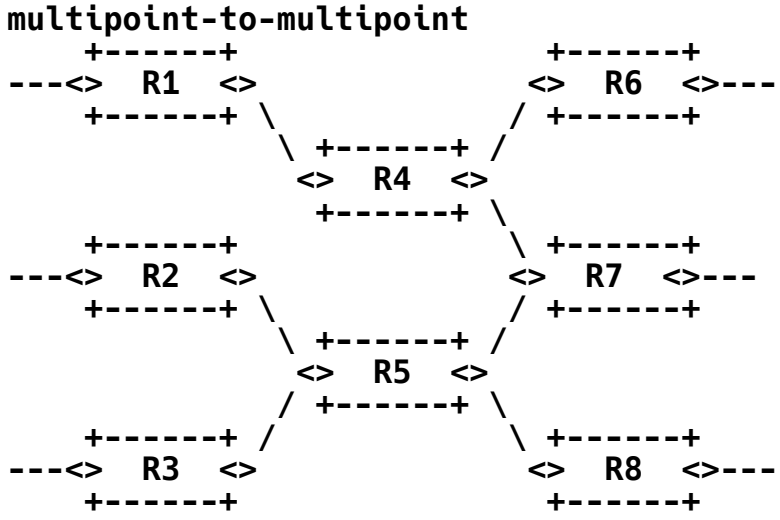


Figure 1: Flow Classification

The case of unicast flow is considered in Figure 1. The anycast flow is also in scope, because there is no replication and only a single node from the anycast group receives the traffic, so it can be viewed as a special case of unicast flow. Furthermore, an ECMP flow is in scope by definition, since it is a point-to-multipoint unicast flow.

4. Multipoint Performance Measurement

By using the Alternate-Marking method, only point-to-point paths can be monitored. To have an IP (TCP/UDP) flow that follows a point-to-point path, we have to define, with a specific value, 5 identification fields (IP Source, IP Destination, Transport Protocol, Source Port, Destination Port).

Multipoint Alternate Marking enables the performance measurement for multipoint flows selected by identification fields without any constraints (even the entire network production traffic). It is also possible to use multiple marking points for the same monitored flow.

4.1. Monitoring Network

The monitoring network is deduced from the production network by identifying the nodes of the graph that are the measurement points, and the links that are the connections between measurement points.

There are some techniques that can help with the building of the monitoring network (as an example, see [ROUTE-ASSESSMENT]). In general, there are different options: the monitoring network can be obtained by considering all the possible paths for the traffic or periodically checking the traffic (e.g. daily, weekly, monthly) and updating the graph as appropriate, but this is up to the Network Management System (NMS) configuration.

So a graph model of the monitoring network can be built according to the Alternate-Marking method: the monitored interfaces and links are

identified. Only the measurement points and links where the traffic has flowed have to be represented in the graph.

Figure 2 shows a simple example of a monitoring network graph:

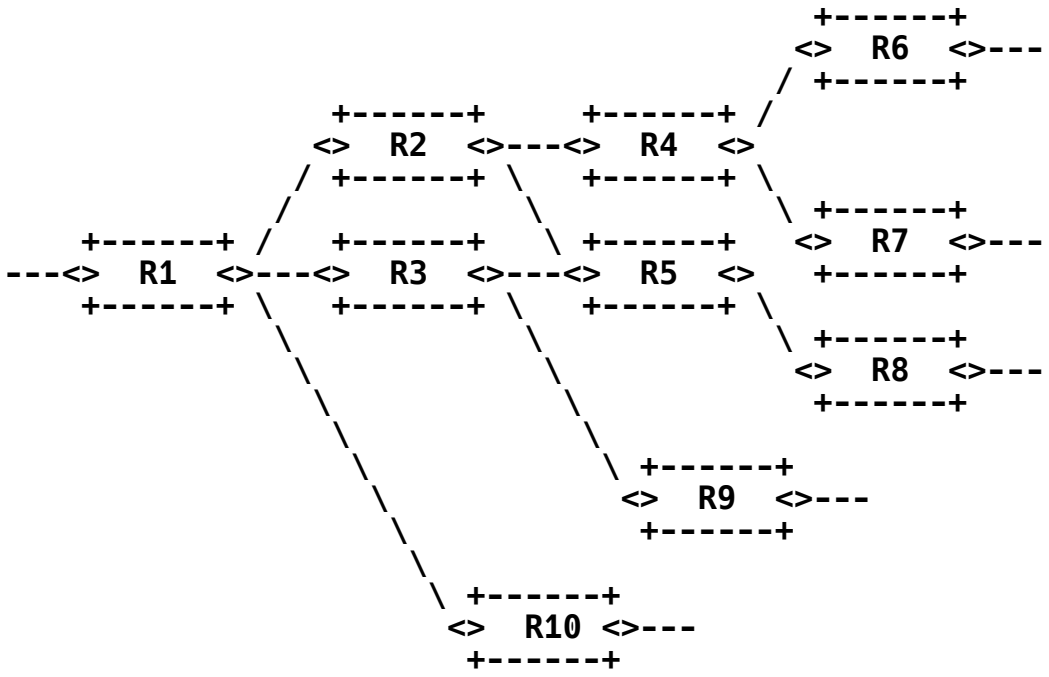


Figure 2: Monitoring Network Graph

Each monitoring point is characterized by the packet counter that refers only to a marking period of the monitored flow.

The same is also applicable for the delay, but it will be described in the following sections.

5. Multipoint Packet Loss

Since all the packets of the considered flow leaving the network have previously entered the network, the number of packets counted by all the input nodes is always greater than, or equal to, the number of packets counted by all the output nodes. Noninitial fragments are not considered here.

The assumption is the use of the Alternate-Marking method. In the case of no packet loss occurring in the marking period, if all the input and output points of the network domain to be monitored are measurement points, the sum of the number of packets on all the ingress interfaces equals the number on egress interfaces for the monitored flow. In this circumstance, if no packet loss occurs, the intermediate measurement points only have the task of splitting the measurement.

It is possible to define the Network Packet Loss of one monitored flow for a single period. In a packet network, the number of lost packets is the number of packets counted by the input nodes minus the number of packets counted by the output nodes. This is true for

every packet flow in each marking period.

The monitored network packet loss with n input nodes and m output nodes is given by:

PL = (PI1 + PI2 +...+ PIn) - (PO1 + PO2 +...+ POm)

where:

PL is the network packet loss (number of lost packets)

PIi is the number of packets flowed through the i-th input node in this period

POj is the number of packets flowed through the j-th output node in this period

The equation is applied on a per-time-interval basis and a per-flow basis:

> The reference interval is the Alternate-Marking period, as defined in RFC 8321 [RFC8321].
>
> The flow definition is generalized here. Indeed, as described before, a multipoint packet flow is considered, and the identification fields can be selected without any constraints.

6. Network Clustering

The previous equation can determine the number of packets lost globally in the monitored network, exploiting only the data provided by the counters in the input and output nodes.

In addition, it is also possible to leverage the data provided by the other counters in the network to converge on the smallest identifiable subnetworks where the losses occur. These subnetworks are named "clusters".

A cluster graph is a subnetwork of the entire monitoring network graph that still satisfies the packet loss equation (introduced in the previous section), where PL in this case is the number of packets lost in the cluster. As for the entire monitoring network graph, the cluster is defined on a per-flow basis.

For this reason, a cluster should contain all the arcs emanating from its input nodes and all the arcs terminating at its output nodes. This ensures that we can count all the packets (and only those) exiting an input node again at the output node, whatever path they follow.

In a completely monitored unidirectional network (a network where every network interface is monitored), each network device corresponds to a cluster, and each physical link corresponds to two clusters (one for each device).

Clusters can have different sizes depending on the flow-filtering

criteria adopted.

Moreover, sometimes clusters can be optionally simplified. For example, when two monitored interfaces are divided by a single router (one is the input interface, the other is the output interface, and the router has only these two interfaces), instead of counting exactly twice, upon entering and leaving, it is possible to consider a single measurement point. In this case, we do not care about the internal packet loss of the router.

It is worth highlighting that it might also be convenient to define clusters based on the topological information so that they are applicable to all the possible flows in the monitored network.

6.1. Algorithm for Clusters Partition

A simple algorithm can be applied in order to split our monitoring network into clusters. This can be done for each direction separately. The clusters partition is based on the monitoring network graph, which can be valid for a specific flow or can also be general and valid for the entire network topology.

It is a two-step algorithm:

1. Group the links where there is the same starting node;

2. Join the grouped links with at least one ending node in common.

Considering that the links are unidirectional, the first step implies listing all the links as connections between two nodes and grouping the different links if they have the same starting node. Note that it is possible to start from any link, and the procedure will work. Following this classification, the second step implies eventually joining the groups classified in the first step by looking at the ending nodes. If different groups have at least one common ending node, they are put together and belong to the same set. After the application of the two steps of the algorithm, each one of the composed sets of links, together with the endpoint nodes, constitutes a cluster.

In our monitoring network graph example, it is possible to identify the clusters partition by applying this two-step algorithm.

The first step identifies the following groups:

1. Group 1: (R1-R2), (R1-R3), (R1-R10)

2. Group 2: (R2-R4), (R2-R5)

3. Group 3: (R3-R5), (R3-R9)

4. Group 4: (R4-R6), (R4-R7)

5. Group 5: (R5-R8)

And then, the second step builds the clusters partition (in

particular, we can underline that Groups 2 and 3 connect together, since R5 is in common):

1. Cluster 1: (R1-R2), (R1-R3), (R1-R10)

2. Cluster 2: (R2-R4), (R2-R5), (R3-R5), (R3-R9)

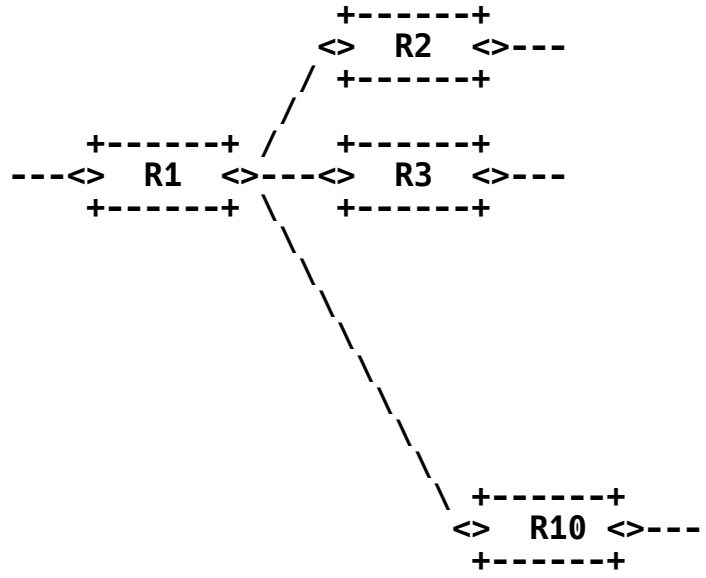
3. Cluster 3: (R4-R6), (R4-R7)

4. Cluster 4: (R5-R8)

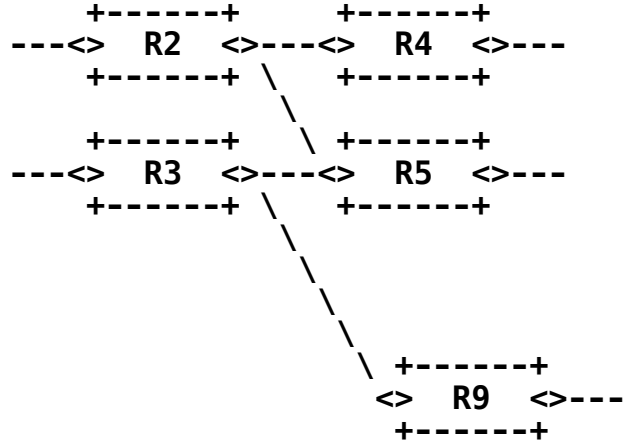
The flow direction here considered is from left to right. For the opposite direction, the same reasoning can be applied, and in this example, you get the same clusters partition.

In the end, the following 4 clusters are obtained:

Cluster 1



Cluster 2



Cluster 3

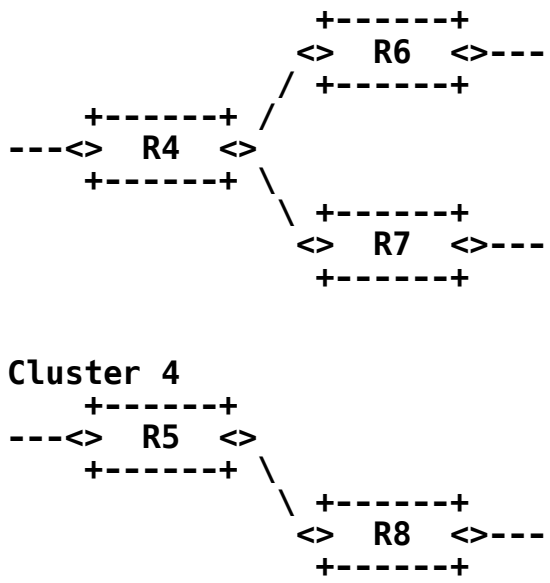


Figure 3: Clusters Example

There are clusters with more than two nodes as well as two-node clusters. In the two-node clusters, the loss is on the link (Cluster 4). In more-than-two-node clusters, the loss is on the cluster, but we cannot know in which link (Cluster 1, 2, or 3).

In this way, the calculation of packet loss can be made on a cluster basis. Note that the packet counters for each marking period permit calculating the packet rate on a cluster basis, so Committed Information Rate (CIR) and Excess Information Rate (EIR) could also be deduced on a cluster basis.

Obviously, by combining some clusters in a new connected subnetwork (called a "super cluster"), the packet-loss rule is still true.

In this way, in a very large network, there is no need to configure detailed filter criteria to inspect the traffic. You can check a multipoint network and, in case of problems, go deep with a step-by-step cluster analysis, but only for the cluster or combination of clusters where the problem happens.

In summary, once a flow is defined, the algorithm to build the clusters partition is based on topological information; therefore, it considers all the possible links and nodes crossed by the given flow, even if there is no traffic. So, if the flow does not enter or traverse all the nodes, the counters have a nonzero value for the involved nodes and a zero value for the other nodes without traffic; but in the end, all the formulas are still valid.

The algorithm described above is an iterative clustering algorithm, but it is also possible to apply a recursive clustering algorithm by using the node-node adjacency matrix representation [IEEE-ACM-ToN-MPNPM].

The complete and mathematical analysis of the possible algorithms for clusters partition, including the considerations in terms of

efficiency and a comparison between the different methods, is in the paper [IEEE-ACM-ToN-MPNPM].

7. Timing Aspects

It is important to consider the timing aspects, since out-of-order packets happen and have to be handled as well, as described in RFC 8321 [RFC8321]. However, in a multisource situation, an additional issue has to be considered. With multipoint path, the egress nodes will receive alternate marked packets in random order from different ingress nodes, and this must not affect the measurement.

So, if we analyze a multipoint-to-multipoint path with more than one marking node, it is important to recognize the reference measurement interval. In general, the measurement interval for describing the results is the interval of the marking node that is more aligned with the start of the measurement, as reported in Figure 4.

Note that the mark switching approach based on a fixed timer is considered in this document.

Figure 4: Measurement Interval

In Figure 4, it is assumed that the node with the earliest clock (R1) identifies the right starting and ending times of the measurement, but it is just an assumption, and other possibilities could occur. So, in this case, T(R1) is the measurement interval, and its recognition is essential in order to make comparisons with other active/passive/hybrid Packet Loss metrics.

When we expand to multipoint-to-multipoint flows, we have to consider that all source nodes mark the traffic, and this adds more complexity.

Regarding the timing aspects of the methodology, RFC 8321 [RFC8321] already describes two contributions that are taken into account: the clock error between network devices and the network delay between measurement points.

But we should now consider an additional contribution. Since all source nodes mark the traffic, the source measurement intervals can be of different lengths and with different offsets, and this mismatch m can be added to d, as shown in Figure 5.

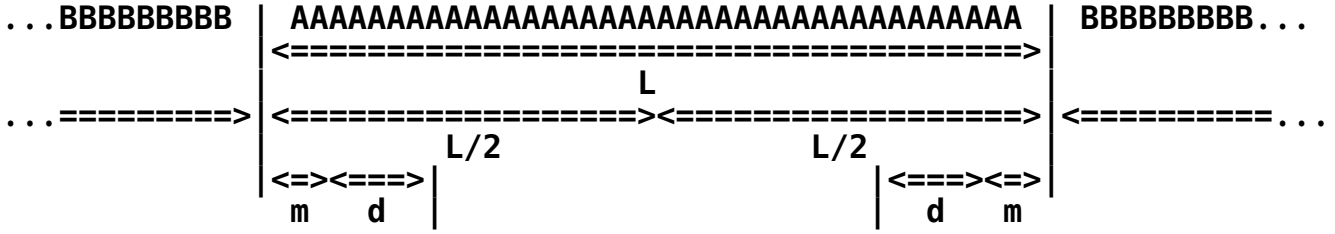


Figure 5: Timing Aspects for Multipoint Paths

So the misalignment between the marking source routers gives an additional constraint, and the value of m is added to d (which already includes clock error and network delay).

Thus, three different possible contributions are considered: clock error between network devices, network delay between measurement points, and the misalignment between the marking source routers.

In the end, the condition that must be satisfied to enable the method to function properly is that the available counting interval must be > 0, and that means:

L - 2m - 2d > 0.

This formula needs to be verified for each measurement point on the multipoint path, where m is misalignment between the marking source routers, while d, already introduced in RFC 8321 [RFC8321], takes into account clock error and network delay between network nodes. Therefore, the mismatch between measurement intervals must satisfy this condition.

Note that the timing considerations are valid for both packet loss and delay measurements.

8. Multipoint Delay and Delay Variation

The same line of reasoning can be applied to delay and delay variation. Similarly to the delay measurements defined in RFC 8321 [RFC8321], the marking batches anchor the samples to a particular period, and this is the time reference that can be used. It is important to highlight that both delay and delay-variation measurements make sense in a multipoint path. The delay variation is calculated by considering the same packets selected for measuring the delay.

In general, it is possible to perform delay and delay-variation measurements on the basis of multipoint paths or single packets:

* Delay measurements on the basis of multipoint paths mean that the delay value is representative of an entire multipoint path (e.g., the whole multipoint network, a cluster, or a combination of clusters).

* Delay measurements on a single-packet basis mean that you can use a multipoint path just to easily couple packets between input and output nodes of a multipoint path, as described in the following sections.

8.1. Delay Measurements on a Multipoint-Paths Basis

8.1.1. Single-Marking Measurement

Mean delay and mean delay-variation measurements can also be generalized to the case of multipoint flows. It is possible to compute the average one-way delay of packets in one block, a cluster, or the entire monitored network.

The average latency can be measured as the difference between the weighted averages of the mean timestamps of the sets of output and input nodes. This means that, in the calculation, it is possible to weigh the timestamps by considering the number of packets for each endpoints.

8.2. Delay Measurements on a Single-Packet Basis

8.2.1. Single- and Double-Marking Measurement

Delay and delay-variation measurements relative to only one picked packet per period (both single and double marked) can be performed in the multipoint scenario, with some limitations:

- Single marking based on the first/last packet of the interval would not work, because it would not be possible to agree on the first packet of the interval.

- Double marking or multiplexed marking would work, but each measurement would only give information about the delay of a single path. However, by repeating the measurement multiple times, it is possible to get information about all the paths in the multipoint flow. This can be done in the case of a point-to-multipoint path, but it is more difficult to achieve in the case of a multipoint-to-multipoint path because of the multiple source routers.

If we would perform a delay measurement for more than one picked packet in the same marking period, and especially if we want to get delay measurements on a multipoint-to-multipoint basis, neither the single- nor the double-marking method is useful in the multipoint scenario, since they would not be representative of the entire flow. The packets can follow different paths with various delays, and in general it can be very difficult to recognize marked packets in a multipoint-to-multipoint path, especially in the case when there is more than one per period.

A desirable option is to monitor simultaneously all the paths of a multipoint path in the same marking period; for this purpose, hashing can be used, as reported in the next section.

8.2.2. Hashing Selection Method

RFCs 5474 [RFC5474] and 5475 [RFC5475] introduce sampling and filtering techniques for IP packet selection.

The hash-based selection methodologies for delay measurement can work in a multipoint-to-multipoint path and can be used either coupled to mean delay or stand-alone.

[ALTERNATE-MARKING] introduces how to use the hash method (RFCs 5474 [RFC5474] and 5475 [RFC5475]) combined with the Alternate-Marking method for point-to-point flows. It is also called Mixed Hashed Marking: the coupling of a marking method and hashing technique is very useful, because the marking batches anchor the samples selected with hashing, and this simplifies the correlation of the hashing packets along the path.

It is possible to use a basic-hash or a dynamic-hash method. One of the challenges of the basic approach is that the frequency of the sampled packets may vary considerably. For this reason, the dynamic approach has been introduced for point-to-point flows in order to have the desired and almost fixed number of samples for each measurement period. Using the hash-based sampling, the number of samples may vary a lot because it depends on the packet rate that is variable. The dynamic approach helps to have an almost fixed number of samples for each marking period, and this is a better option for making regular measurements over time. In the hash-based sampling, Alternate Marking is used to create periods, so that hash-based samples are divided into batches, which allows anchoring the selected samples to their period. Moreover, in the dynamic hash-based sampling, by dynamically adapting the length of the hash value, the number of samples is bounded in each marking period. This can be realized by choosing the maximum number of samples (NMAX) to be caught in a marking period. The algorithm starts with only a few hash bits, which permits selecting a greater percentage of packets (e.g., with 0 bits of hash all the packets are sampled, with 1 bit of hash half of the packets are sampled, and so on). When the number of selected packets reaches NMAX, a hashing bit is added. As a consequence, the sampling proceeds at half of the original rate, and also the packets already selected that do not match the new hash are discarded. This step can be repeated iteratively. It is assumed that each sample includes the timestamp (used for delay measurement) and the hash value, allowing the management system to match the samples received from the two measurement points. The dynamic process statistically converges at the end of a marking period, and the final number of selected samples is between NMAX/2 and NMAX. Therefore, the dynamic approach paces the sampling rate, allowing to bound the number of sampled packets per sampling period.

In a multipoint environment, the behavior is similar to a point-to-point flow. In particular, in the context of a multipoint-to-multipoint flow, the dynamic hash could be the solution for performing delay measurements on specific packets and overcoming the single- and double-marking limitations.

The management system receives the samples, including the timestamps and the hash value, from all the MPs, and this happens for both point-to-point and multipoint-to-multipoint flows. Then, the longest hash used by the MPs is deduced and applied to couple timestamps from either the same packets of 2 MPs of a point-to-point path, or the input and output MPs of a cluster (or a super cluster or the entire network). But some considerations are needed: if there isn't packet loss, the set of input samples is always equal to the set of output samples. In the case of packet loss, the set of output samples can be a subset of input samples, but the method still works because, at

the end, it is easy to couple the input and output timestamps of each caught packet using the hash (in particular, the "unused part of the hash" that should be different for each packet).

Therefore, the basic hash is logically similar to the double-marking method, and in the case of a point-to-point path, double-marking and basic-hash selection are equivalent. The dynamic approach scales the number of measurements per interval. It would seem that double marking would also work well if we reduced the interval length, but this can be done only for a point-to-point path and not for a multipoint path, where we cannot couple the picked packets in a multipoint path. So, in general, if we want to get delay measurements on the basis of a multipoint-to-multipoint path, and want to select more than one packet per period, double marking cannot be used because we could not be able to couple the picked packets between input and output nodes. On the other hand, we can do that by using hashing selection.

9. A Closed-Loop Performance-Management Approach

The Multipoint Alternate-Marking framework that is introduced in this document adds flexibility to Performance Management (PM), because it can reduce the order of magnitude of the packet counters. This allows an SDN orchestrator to supervise, control, and manage PM in large networks.

The monitoring network can be considered as a whole or split into clusters that are the smallest subnetworks (group-to-group segments), maintaining the packet-loss property for each subnetwork. The clusters can also be combined in new, connected subnetworks at different levels, depending on the detail we want to achieve.

An SDN controller or a Network Management System (NMS) can calibrate performance measurements, since they are aware of the network topology. They can start without examining in depth. In case of necessity (packet loss is measured or the delay is too high), the filtering criteria could be immediately reconfigured in order to perform a partition of the network by using clusters and/or different combinations of clusters. In this way, the problem can be localized in a specific cluster or a single combination of clusters, and a more detailed analysis can be performed step by step by successive approximation up to a point-to-point flow detailed analysis. This is the so-called "closed loop".

This approach can be called "network zooming" and can be performed in two different ways:

1) change the traffic filter and select more detailed flows;

2) activate new measurement points by defining more specified clusters.

The network-zooming approach implies that some filters or rules are changed and that therefore there is a transient time to wait once the new network configuration takes effect. This time can be determined by the Network Orchestrator/Controller, based on the network

conditions.

For example, if the network zooming identifies the performance problem for the traffic coming from a specific source, we need to recognize the marked signal from this specific source node and its relative path. For this purpose, we can activate all the available measurement points and better specify the flow filter criteria (i.e., 5-tuple). As an alternative, it can be enough to select packets from the specific source for delay measurements; in this case, it is possible to apply the hashing technique, as mentioned in the previous sections.

[IFIT-FRAMEWORK] defines an architecture where the centralized Data Collector and Network Management can apply the intelligent and flexible Alternate-Marking algorithm as previously described.

As for RFC 8321 [RFC8321], it is possible to classify the traffic and mark a portion of the total traffic. For each period, the packet rate and bandwidth are calculated from the number of packets. In this way, the network orchestrator becomes aware if the traffic rate surpasses limits. In addition, more precision can be obtained by reducing the marking period; indeed, some implementations use a marking period of 1 sec or less.

In addition, an SDN controller could also collect the measurement history.

It is important to mention that the Multipoint Alternate Marking framework also helps Traffic Visualization. Indeed, this methodology is very useful for identifying which path or cluster is crossed by the flow.

10. Examples of Application

There are application fields where it may be useful to take into consideration the Multipoint Alternate Marking:

VPN: The IP traffic is selected on an IP-source basis in both directions. At the endpoint WAN interface, all the output traffic is counted in a single flow. The input traffic is composed of all the other flows aggregated for source address. So, by considering n endpoints, the monitored flows are n (each flow with 1 ingress point and (n-1) egress points) instead of n*(n-1) flows (each flow, with 1 ingress point and 1 egress point).

Mobile Backhaul: LTE traffic is selected, in the Up direction, by the EnodeB source address and, in the Down direction, by the EnodeB destination address, because the packets are sent from the Mobile Packet Core to the EnodeB. So the monitored flow is only one per EnodeB in both directions.

Over The Top (OTT) services: The traffic is selected, in the Down direction, by the source addresses of the packets sent by OTT servers. In the opposite direction (Up), it is selected by the destination IP addresses of the same servers. So the monitoring is based on a single flow per OTT server in both directions.

Enterprise SD-WAN: SD-WAN allows connecting remote branch offices to data centers and building higher-performance WANs. A centralized controller is used to set policies and prioritize traffic. The SD-WAN takes into account these policies and the availability of network bandwidth to route traffic. This helps ensure that application performance meets Service Level Agreements (SLAs). This methodology can also help the path selection for the WAN connection based on per-cluster and per-flow performance.

Note that the preceding list is just an example and is not exhaustive. More applications are possible.

11. Security Considerations

This document specifies a method of performing measurements that does not directly affect Internet security or applications that run on the Internet. However, implementation of this method must be mindful of security and privacy concerns, as explained in RFC 8321 [RFC8321].

12. IANA Considerations

This document has no IANA actions.

13. References

13.1. Normative References

[RFC5474] Duffield, N., Ed., Chiou, D., Claise, B., Greenberg, A., Grossglauser, M., and J. Rexford, "A Framework for Packet Selection and Reporting", RFC 5474, DOI 10.17487/RFC5474, March 2009, <https://www.rfc-editor.org/info/rfc5474>.

[RFC5475] Zseby, T., Molina, M., Duffield, N., Niccolini, S., and F. Raspall, "Sampling and Filtering Techniques for IP Packet Selection", RFC 5475, DOI 10.17487/RFC5475, March 2009, <https://www.rfc-editor.org/info/rfc5475>.

[RFC5644] Stephan, E., Liang, L., and A. Morton, "IP Performance Metrics (IPPM): Spatial and Multicast", RFC 5644, DOI 10.17487/RFC5644, October 2009, <https://www.rfc-editor.org/info/rfc5644>.

[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>.

13.2. Informative References

[ALTERNATE-MARKING] Mizrahi, T., Arad, C., Fioccola, G., Cociglio, M., Chen, M., Zheng, L., and G. Mirsky, "Compact Alternate Marking Methods for Passive and Hybrid Performance Monitoring", Work in Progress, Internet-Draft, draft-mizrahi-ippm-

compact-alternate-marking-05, 6 July 2019, <https://tools.ietf.org/html/draft-mizrahi-ippm-compact-alternate-marking-05>.

[ENHANCED-ALTERNATE-MARKING]
Zhou, T., Fioccola, G., Lee, S., Cociglio, M., and W. Li, "Enhanced Alternate Marking Method", Work in Progress, Internet-Draft, draft-zhou-ippm-enhanced-alternate-marking-05, 13 July 2020, <https://tools.ietf.org/html/draft-zhou-ippm-enhanced-alternate-marking-05>.

[IEEE-ACM-ToN-MPNPM]
Cociglio, M., Fioccola, G., Marchetto, G., Sapio, A., and R. Sisto, "Multipoint Passive Monitoring in Packet Networks", IEEE/ACM Transactions on Networking vol. 27, no. 6, pp. 2377-2390, DOI 10.1109/TNET.2019.2950157, December 2019, <https://doi.org/10.1109/TNET.2019.2950157>.

[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, no. 4, pp. 155-161, DOI 10.1109/MNET.2019.1800152, July 2019, <https://doi.org/10.1109/MNET.2019.1800152>.

[IFIT-FRAMEWORK]
Song, H., Qin, F., Chen, H., Jin, J., and J. Shin, "In-situ Flow Information Telemetry", Work in Progress, Internet-Draft, draft-song-opsawg-ifit-framework-12, 14 April 2020, <https://tools.ietf.org/html/draft-song-opsawg-ifit-framework-12>.

[RFC7011] Claise, B., Ed., Trammell, B., Ed., and P. Aitken, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information", STD 77, RFC 7011, DOI 10.17487/RFC7011, September 2013, <https://www.rfc-editor.org/info/rfc7011>.

[ROUTE-ASSESSMENT]
Alvarez-Hamelin, J., Morton, A., Fabini, J., Pignataro, C., and R. Geib, "Advanced Unidirectional Route Assessment (AURA)", Work in Progress, Internet-Draft, draft-ietf-ippm-route-10, 13 August 2020, <https://tools.ietf.org/html/draft-ietf-ippm-route-10>.

Acknowledgements

The authors would like to thank Al Morton, Tal Mizrahi, and Rachel Huang for the precious contributions.

Authors' Addresses

Giuseppe Fioccola (editor)
Huawei Technologies
Riesstrasse, 25

80992 Munich
Germany

Email: giuseppe.fioccola@huawei.com

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

Email: mauro.cociglio@telecomitalia.it

Amedeo Sapio
Intel Corporation
4750 Patrick Henry Dr.
Santa Clara, CA 95054
United States of America

Email: amedeo.sapio@intel.com

Riccardo Sisto
Politecnico di Torino
Corso Duca degli Abruzzi, 24
10129 Torino
Italy

Email: riccardo.sisto@polito.it