IPPM Working Group
Internet-Draft
Intended status: Experimental
Expires: September 24, 2020

G. Fioccola, Ed.
Huawei Technologies
M. Cociglio
Telecom Italia
A. Sapio
R. Sisto
Politecnico di Torino
March 23, 2020

Multipoint Alternate Marking method for passive and hybrid performance monitoring
draft-ietf-ippm-multipoint-alt-mark-09

Abstract

The Alternate Marking method, as presented in RFC 8321, can be applied only 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 flows, 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 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 September 24, 2020.

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 both in RFC 8321 [RFC8321] and in the paper [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 on 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.

The Alternate Marking method can therefore 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 analysing 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 accurate the network performance monitoring is setup by applying the Multipoint Alternate Marking as described in this document.

It is important to underline that, as extension of RFC 8321 [RFC8321], this is a methodology draft, 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. [I-D.zhou-ippm-enhanced-alternate-marking].

Note that, as for RFC 8321 [RFC8321], the fragmented packets case can be managed with this methodology if fragmentation happens outside the portion of the monitored network.

2. Terminology

The definitions of the basic terms are identical to those found in Alternate Marking (RFC 8321 [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 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: it is 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 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 goes in is the same that goes 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 multipoint paths basis in order to associate the measurements to a cluster, to a combination of clusters or to 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 RFC5644

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

or both.

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 also 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

An 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 headers 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.

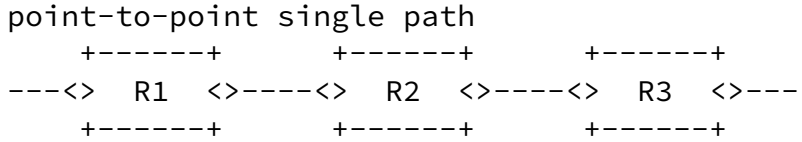
A TCP 5-tuple usually identifies flows following either a single path

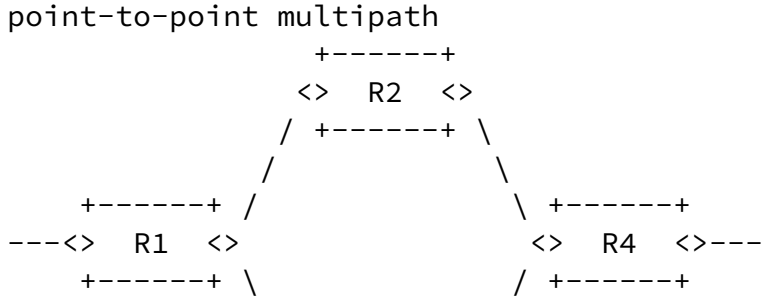
or a point-to-point multipath (in case of load balancing). On the contrary, a single source address selects aggregate flows following a point-to-multipoint, while a multipoint-to-point can be the result of a matching on a single destination address. In case a selection rule and its reverse are used for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

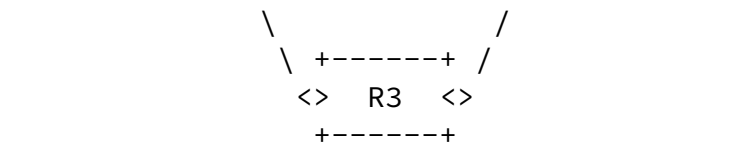
So the flows to be monitored are selected into the monitoring points using packet selection rules, that can also change the pattern of the monitored network.

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

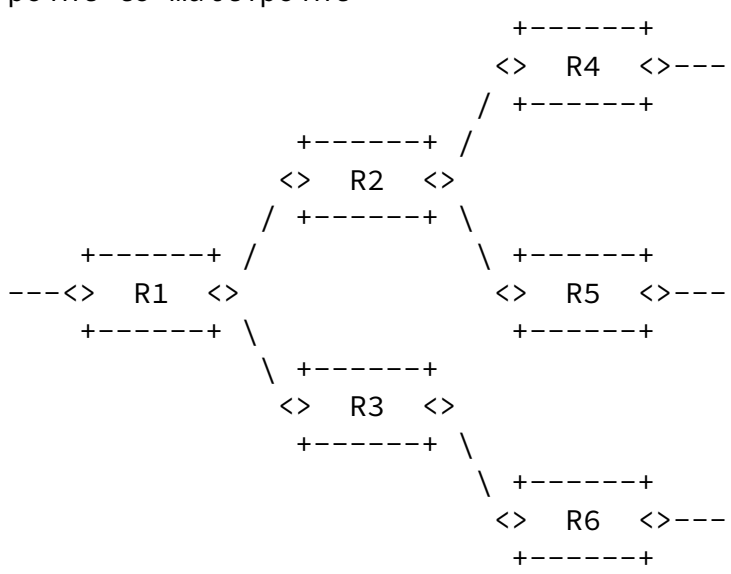
The Alternate Marking method is applicable only to a single path (and partially to a one-to-one multipath), so the extension proposed in this document is suitable also for the most general case of multipoint-to-multipoint, which embraces all the other patterns of Figure 1.



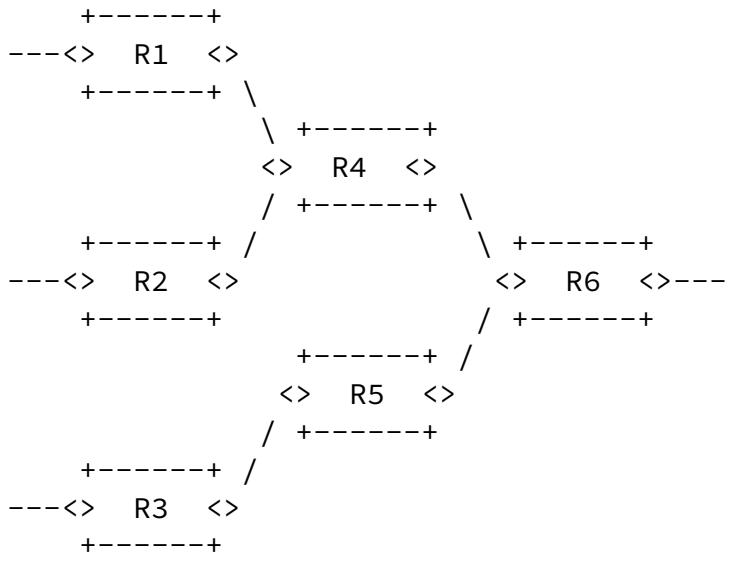




point-to-multipoint



multipoint-to-point



multipoint-to-multipoint

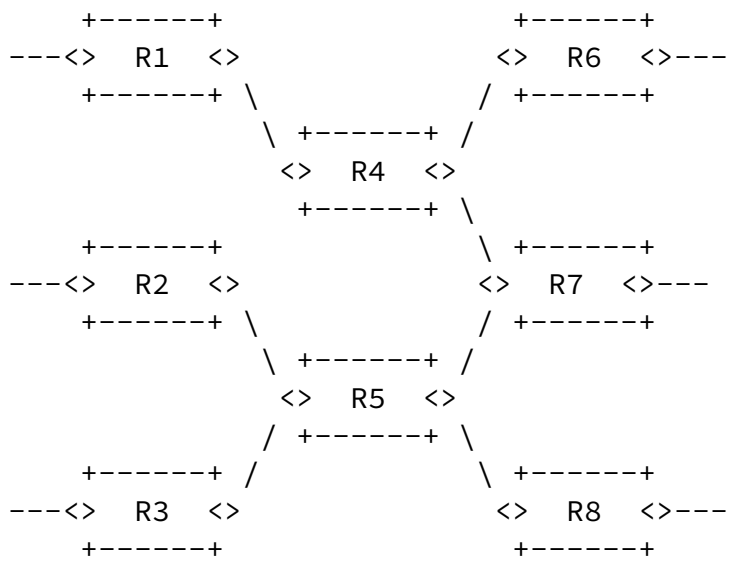


Figure 1: Flow classification

The case of unicast flow is considered in the previous figure. Anyway 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 it is possible to mention

[I-D.ietf-ippm-route]). In general there are different options: the monitoring network can be obtained by considering all the possible paths for the traffic or also by periodically checking the traffic (e.g. daily, weekly, monthly) and update 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.

The following figure shows a simple example of a Monitoring Network graph:

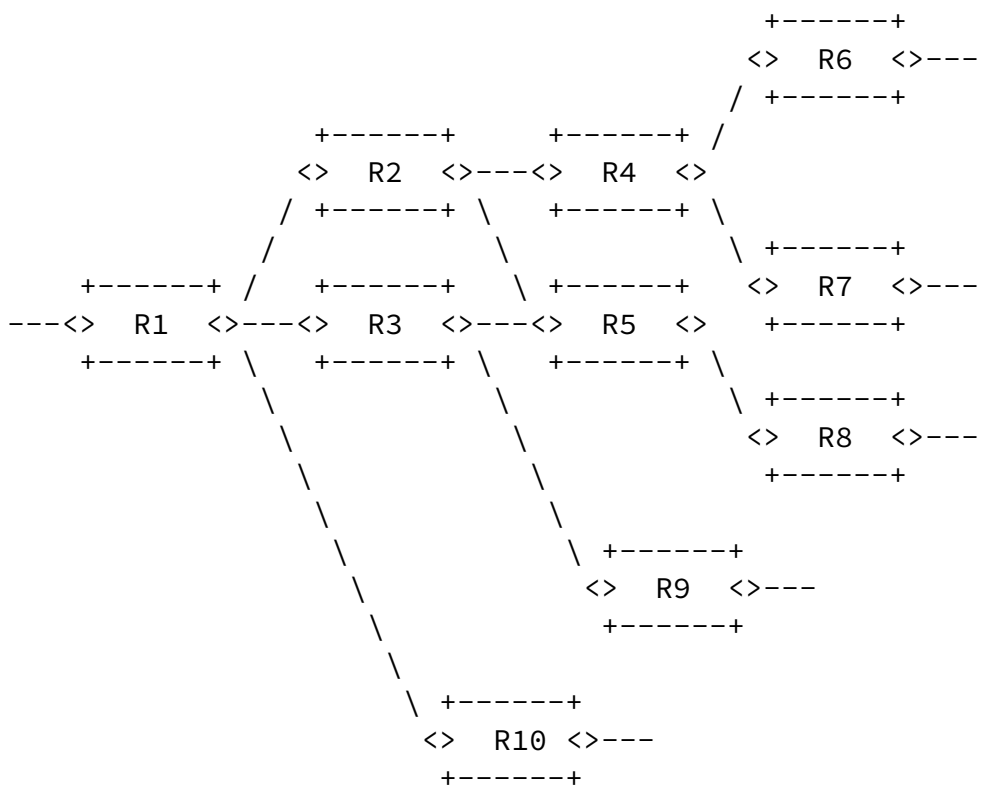


Figure 2: Monitoring Network Graph

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

The same is applicable also 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 or equal than the number of packets counted by all the output nodes. Non-initial fragments are not considered here.

The assumption is the use of the Alternate Marking method. And in 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 have only the task to split 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 on an 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 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 and 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 of 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 and applicable to all the possible flows in the monitored network.

6.1. Algorithm for Cluster 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 Cluster partition is based on the Monitoring Network

Graph that 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:

- Group the links where there is the same starting node;
- Join the grouped links with at least one ending node in common.

Considering that the links are unidirectional, the first step implies to list all the links as connection between two nodes and to group the different links if they have the same starting node. Note that it is possible to start from any link and the procedure works anyway. Following this classification, the second step implies to eventually join 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 Group 2 and Group 3 connect together, since R5 is in common):

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

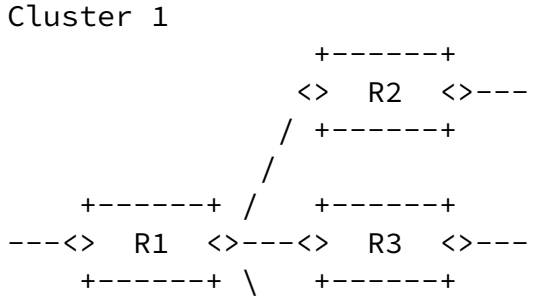
2. Cluster 2: (R2-R4), (R2-R5), (R3-R5), (R3-R9)

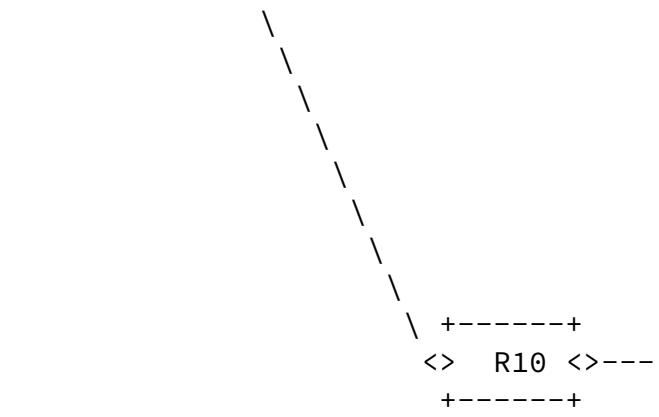
3. Cluster 3: (R4-R6), (R4-R7)

4. Cluster 4: (R5-R8)

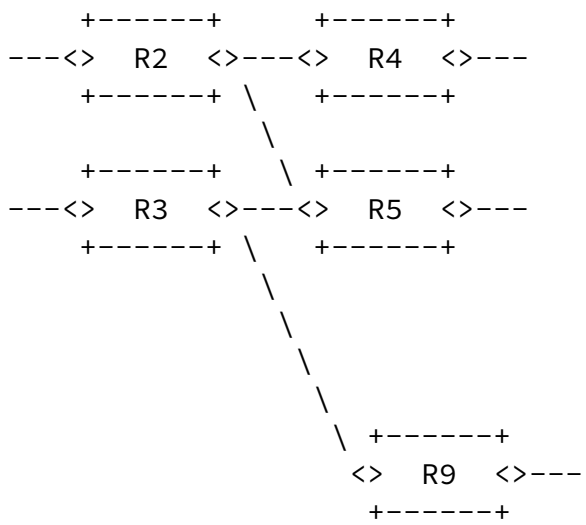
The flow direction here considered is from left to right. For the opposite direction the same way of reasoning can be applied and, in this example, you get the same Clusters partition.

In the end the following 4 Clusters are obtained:

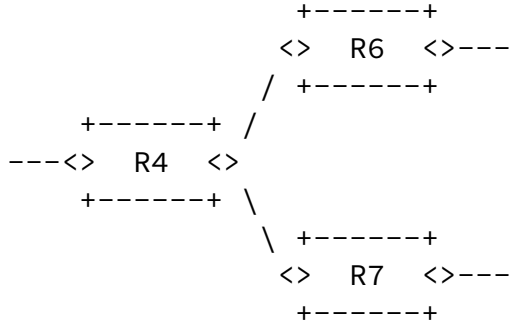




Cluster 2



Cluster 3



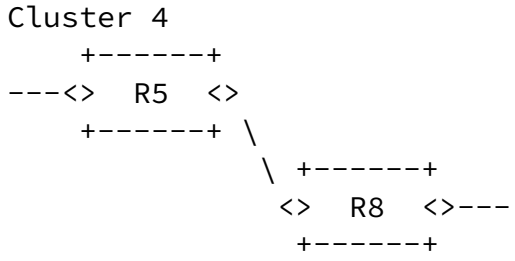


Figure 3: Clusters example

There are Clusters with more than 2 nodes and two-nodes Clusters. In the two-nodes Clusters the loss is on the link (Cluster 4). In more-than-2-nodes Clusters the loss is on the Cluster but we cannot know in which link (Cluster 1, 2, 3).

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

Obviously, by combining some Clusters in a new connected subnetwork (called 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, you can 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 defined a flow, the algorithm to build the Cluster Partition considers all the possible links and nodes crossed by the given flow, even if there is no traffic. It is based on topological information. So, if the flow does not enter or traverse all the nodes, the counters have a non-zero value for the involved nodes,

while 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 Cluster 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]. But, in a multi-source 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 analyse 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 the following figure.

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

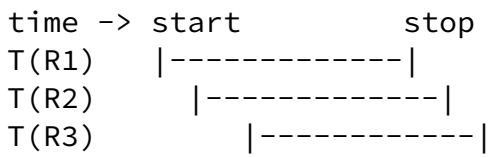


Figure 4: Measurement Interval

In the figure it is assumed that the node with the earliest clock (R1) identifies the right starting and ending time 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 be compatible and make comparison 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.

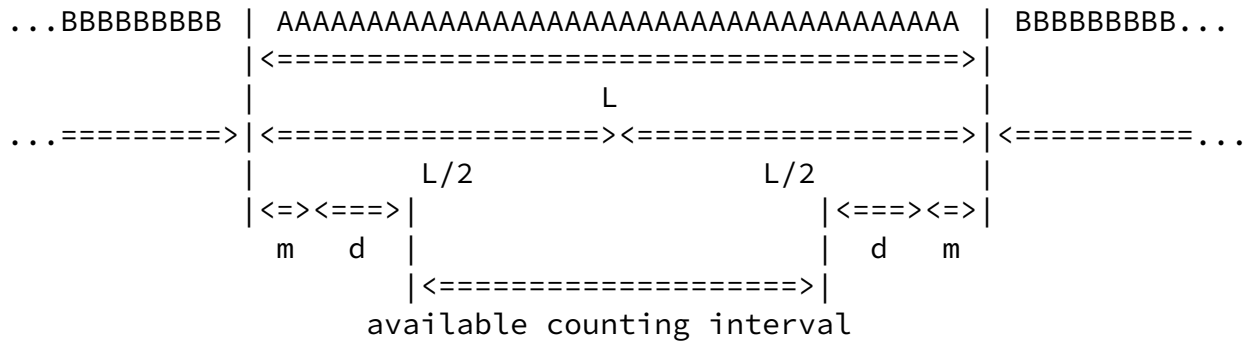


Figure 5: Timing Aspects for Multipoint paths

So the misalignment between the marking source routers gives an additional constraint and the value of m is added to d (that 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 multipoint paths basis or on single packets basis:

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

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

8.1. Delay measurements on 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, in a cluster or in 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 single packets 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 case of point-to-multipoint path but it is more difficult to achieve in case of 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 multipoint-to-multipoint basis, both single and double marking method are not 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 and, for this purpose, hashing can be used as reported in the next Section.

8.2.2. Hashing selection method

RFC 5474 [RFC5474] and RFC 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 both coupled to mean delay or stand alone.

[I-D.mizrahi-ippm-compact-alternate-marking] introduces how to use the Hash method (RFC 5474 [RFC5474] and RFC 5475 [RFC5475]) combined with Alternate Marking method for point-to-point flows. It is also called Mixed Hashed Marking: the coupling of 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 flow in order to have

the desired and almost fixed number of samples for each measurement period. In the hash-based sampling, Alternate Marking is used to create periods, so that hash-based samples are divided into batches, allowing to anchor 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 few hash bits, that permit to select a greater percentage of packets (e.g. with 0 bit 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 behaviour 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 to perform

delay measurements on specific packets and to overcome 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 both for point-to-point and for multipoint-to-multipoint flows. Then the longest hash used by MPs is deduced and it is applied to couple timestamps of the same packets of 2 MPs of a point-to-point path or of 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 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 case of point-to-point path double marking and basic hash selection are equivalent. The dynamic approach scales the number of measurements per interval, and it would seem that double marking would also work well if we reduced the interval length, but

this can be done only for point-to-point path and not for multipoint path, where we cannot couple the picked packets in a multipoint paths. So, in general, if we want to get delay measurements on multipoint-to-multipoint path basis 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 can be split in Clusters, that are the smallest subnetworks (group-to-group segments), maintaining the packet loss property for each subnetwork.

They 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 in 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 the some filters or rules are changed and there is a transient time to wait once the new network configuration takes effect and it 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 specify better 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, and in this case it is possible to apply the hashing technique as mentioned in the previous sections.

[I-D.song-opsawg-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 overcomes limits. In addition more precision can be obtained by reducing the marking period, indeed some implementations use a marking period of 1 sec and 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 to identify which path or which 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 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 by all the other flows aggregated for source address. So, by considering n end-points, 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 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) by the destination IP addresses of the same Servers. So the monitoring is based on a single flow per OTT Servers in both directions.

- Enterprise SD-WAN: SD-WAN allows to connect remote branch offices to Data Centers and build 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 list is just an example and it is not exhaustive. More applications are possible.

11. Security Considerations

This document specifies a method to perform measurements that does not directly affect Internet security nor applications that run on the Internet. However, implementation of this method must be mindful of security and privacy concerns, as explained in RFC 8321 [RFC8321].

12. Acknowledgements

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

13. IANA Considerations

This memo makes no requests of IANA.

14. References

14.1. 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>.

14.2. Informative References

[I-D.ietf-ippm-route]
Alvarez-Hamelin, J., Morton, A., Fabini, J., Pignataro, C., and R. Geib, "Advanced Unidirectional Route Assessment (AURA)", draft-ietf-ippm-route-07 (work in progress), December 2019.

[I-D.mizrahi-ippm-compact-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", draft-mizrahi-ippm-compact-alternate-marking-05 (work in progress), July 2019.

[I-D.song-opsawg-ifit-framework]
Song, H., Qin, F., Chen, H., Jin, J., and J. Shin, "In-situ Flow Information Telemetry", draft-song-opsawg-ifit-framework-11 (work in progress), March 2020.

[I-D.zhou-ippm-enhanced-alternate-marking]
Zhou, T., Fioccola, G., Li, Z., Lee, S., and M. Cociglio, "Enhanced Alternate Marking Method", draft-zhou-ippm-enhanced-alternate-marking-04 (work in progress), October 2019.

[IEEE-ACM-ToN-MPNPM]
IEEE/ACM TRANSACTION ON NETWORKING, "Multipoint Passive Monitoring in Packet Networks", DOI 10.1109/TNET.2019.2950157, 2019.

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

[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>.

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

Amedeo Sapio
Politecnico di Torino
Corso Duca degli Abruzzi, 24
Torino 10129
Italy

Email: amedeo.sapio@polito.it

Riccardo Sisto
Politecnico di Torino
Corso Duca degli Abruzzi, 24

Torino 10129
Italy

Email: riccardo.sisto@polito.it