Internet Engineering Task Force (IETF)
Request for Comments: 9198
Updates: 2330
Category: Standards Track
ISSN: 2070-1721

J. Alvarez-Hamelin
Universidad de Buenos Aires
A. Morton
AT&T Labs
J. Fabini
TU Wien
C. Pignataro
Cisco Systems, Inc.
R. Geib
Deutsche Telekom

May 2022

# Advanced Unidirectional Route Assessment (AURA)

## Abstract

This memo introduces an advanced unidirectional route assessment (AURA) metric and associated measurement methodology based on the IP Performance Metrics (IPPM) framework (RFC 2330). This memo updates RFC 2330 in the areas of path-related terminology and path description, primarily to include the possibility of parallel subpaths between a given Source and Destination pair, owing to the presence of multipath technologies.

## Status of This Memo

This is an Internet Standards Track document.

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

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9198.



## Table of Contents

## 1. Introduction

The IETF IP Performance Metrics (IPPM) Working Group first created a framework for metric development in [RFC2330]. This framework has stood the test of time and enabled development of many fundamental metrics. It has been updated in the area of metric composition [RFC5835] and in several areas related to active stream measurement of modern networks with reactive properties [RFC7312].

The framework in [RFC2330] motivated the development of "performance and reliability metrics for paths through the Internet"; Section 5 of [RFC2330] defines terms that support description of a path under test. However, metrics for assessment of paths and related performance aspects had not been attempted in IPPM when the framework in [RFC2330] was written.

This memo takes up the Route measurement challenge and specifies a new Route metric, two practical frameworks for methods of measurement (using either active or hybrid active-passive methods [RFC7799]), and Round-Trip Delay and link information discovery using the results of measurements. All Route measurements are limited by the willingness of Hosts along the path to be discovered, to cooperate with the methods used, or to recognize that the measurement operation is taking place (such as when tunnels are present).

### 1.1. Issues with Earlier Work to Define a Route Metric

Section 7 of [RFC2330] presents a simple example of a "Route" metric along with several other examples. The example is reproduced below (where the reference is to Section 5 of [RFC2330]):

> | route: The path, as defined in Section 5, from A to B at a given
> | time.

This example provides a starting point to develop a more complete definition of Route. Areas needing clarification include:

Time: In practice, the Route will be assessed over a time interval because active path detection methods like Paris-traceroute [PT] rely on Hop Limits for their operation and cannot accomplish discovery of all Hosts using a single packet.

Type-P: The legacy Route definition lacks the option to cater for packet-dependent routing. In this memo, we assess the Route for a specific packet of Type-P and reflect this in the metric definition. The methods of measurement determine the specific Type-P used.

Parallel Paths: Parallel paths are a reality of the Internet and a strength of advanced Route assessment methods, so the metric must acknowledge this possibility. Use of Equal-Cost Multipath (ECMP) and Unequal-Cost Multipath (UCMP) technologies are common sources of parallel subpaths.

Cloud Subpath: Cloud subpaths may contain Hosts that do not decrement the Hop Limit but may have two or more exchange links connecting "discoverable" Hosts or routers. Parallel subpaths contained within clouds cannot be discovered. The assessment methods only discover Hosts or routers on the path that decrement Hop Limit or cooperate with interrogation protocols. The presence of tunnels and nested tunnels further complicate assessment by hiding Hops.

Hop: The definition of Hop in [RFC2330] was a link-Host pair. However, only Hosts that were discoverable and cooperated with interrogation protocols (where link information may be exposed) provided both link and Host information.

Note that the actual definitions appear in Section 3.1.

### 1.2. Requirements Language

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

## 2. Scope

The purpose of this memo is to add new Route metrics and methods of measurement to the existing set of IPPM metrics.

The scope is to define Route metrics that can identify the path taken by a packet or a flow traversing the Internet between two Hosts. Although primarily intended for Hosts communicating on the Internet, the definitions and metrics are constructed to be applicable to other network domains, if desired. The methods of measurement to assess the path may not be able to discover all Hosts comprising the path, but such omissions are often deterministic and explainable sources of error.

This memo also specifies a framework for active methods of measurement that uses the techniques described in [PT] as well as a framework for hybrid active-passive methods of measurement, such as the Hybrid Type I method [RFC7799] described in [RFC9197]. Methods using [RFC9197] are intended only for single administrative domains that provide a protocol for explicit interrogation of Nodes on a path. Combinations of active methods and hybrid active-passive methods are also in scope.

Further, this memo provides additional analysis of the Round-Trip Delay measurements made possible by the methods in an effort to discover more details about the path, such as the link technology in use.

This memo updates Section 5 of [RFC2330] in the areas of path-related terminology and path description, primarily to include the

possibility of parallel subpaths between a given Source and Destination address pair (possibly resulting from ECMP and UCMP technologies).

There are several simple non-goals of this memo. There is no attempt to assess the reverse path from any Host on the path to the Host attempting the path measurement. The reverse path contribution to delay will be that experienced by ICMP packets (in active methods) and may be different from delays experienced by UDP or TCP packets. Also, the Round-Trip Delay will include an unknown contribution of processing time at the Host that generates the ICMP response. Therefore, the ICMP-based active methods are not supposed to yield accurate, reproducible estimations of the Round-Trip Delay that UDP or TCP packets will experience.

## 3. Route Metric Specifications

This section sets requirements for the components of the route metric.

### 3.1. Terms and Definitions

Host
: A Host (as defined in [RFC2330]) is a computer capable of IP communication, including routers (aka an RFC 2330 Host).

Node
: A Node is any network function on the path capable of IP-layer Communication, including RFC 2330 Hosts.

Node Identity
: The Node identity is the unique address for Nodes communicating within the network domain. For Nodes communicating on the Internet with IP, it is the globally routable IP address that the Node uses when communicating with other Nodes under normal or error conditions. The Node identity revealed (and its connection to a Node name through reverse DNS) determines whether interfaces to parallel links can be associated with a single Node or appear to identify unique Nodes.

Discoverable Node
: Discoverable Nodes are Nodes that convey their Node identity according to the requirements of their network domain, such as when error conditions are detected by that Node. For Nodes communicating with IP packets, compliance with Section 3.2.2.4 of [RFC1122], when discarding a packet due to TTL or Hop Limit Exceeded condition, MUST result in sending the corresponding Time Exceeded message (containing a form of Node identity) to the source. This requirement is also consistent with Section 5.3.1 of [RFC1812] for routers.

Cooperating Node
: Cooperating Nodes are Nodes that respond to direct queries for their Node identity as part of a previously established and agreed upon interrogation protocol. Nodes SHOULD also provide information such as arrival/departure interface identification, arrival timestamp, and any relevant information about the Node or specific link that delivered the query to the Node.

Hop specification
: A Hop specification MUST contain a Node identity and MAY contain arrival and/or departure interface identification, Round-Trip Delay, and an arrival timestamp.

Routing Class

Routing Class is a Route that treats a class of different types of packets, designated "C" (unrelated to address classes of the past) equally ([RFC2330] [RFC8468]). Knowledge of such a class allows any one of the types of packets within that class to be used for subsequent measurement of the Route. The designator "class C" is used for historical reasons; see [RFC2330].

### 3.2. Formal Name

The formal name of the metric is:

Type-P-Route-Ensemble-Method-Variant

abbreviated as Route Ensemble.

Note that Type-P depends heavily on the chosen method and variant.

### 3.3. Parameters

This section lists the REQUIRED input factors to define and measure a Route metric, as specified in this memo.

Src: the address of a Node (such as the globally routable IP address).

Dst: the address of a Node (such as the globally routable IP address).

i: the limit on the number of Hops a specific packet may visit as it traverses from the Node at Src to the Node at Dst (such as the TTL or Hop Limit).

MaxHops: the maximum value of i used (i=1,2,3,...MaxHops).

T0: a time (start of measurement interval).

Tf: a time (end of measurement interval).

MP(address): the Measurement Point at address, such as Src or Dst, usually at the same Node stack layer as "address".

T: the Node time of a packet as measured at MP(Src), meaning Measurement Point at the Source.

Ta: the Node time of a reply packet’s *arrival* as measured at MP(Src), assigned to packets that arrive within a "reasonable" time (see parameter below).

Tmax: a maximum waiting time for reply packets to return to the source, set sufficiently long to disambiguate packets with long delays from packets that are discarded (lost), such that the distribution of Round-Trip Delay is not truncated.

F: the number of different flows simulated by the method and variant.

flow: the stream of packets with the same n-tuple of designated header fields that (when held constant) result in identical treatment in a multipath decision (such as the decision taken in load balancing). Note: The IPv6 flow label MAY be included in the flow definition if the MP(Src) is a Tunnel Endpoint (TEP) complying with the guidelines in [RFC6438].

Type-P: the complete description of the packets for which this assessment applies (including the flow-defining fields).

## 3.4. Metric Definitions

This section defines the REQUIRED measurement components of the Route metrics (unless otherwise indicated):

M: the total number of packets sent between T0 and Tf.

N: the smallest value of i needed for a packet to be received at Dst (sent between T0 and Tf).

Nmax: the largest value of i needed for a packet to be received at Dst (sent between T0 and Tf). Nmax may be equal to N.

Next, define a *singleton* for a Node on the path with sufficient indexes to identify all Nodes identified in a measurement interval (where *singleton* is part of the IPPM Framework [RFC2330]).

singleton: A Hop specification, designated h(i,j), the IP address and/or identity of Discoverable Nodes (or Cooperating Nodes) that are i Hops away from the Node with address = Src and part of Route j during the measurement interval T0 to Tf. As defined here, a Hop singleton measurement MUST contain a Node identity, hid(i,j), and MAY contain one or more of the following attributes:

* a(i,j) Arrival Interface ID (e.g., when [RFC5837] is supported)

* d(i,j) Departure Interface ID (e.g., when [RFC5837] is supported)

* t(i,j) arrival timestamp, where t(i,j) is ideally supplied by the Hop (note that t(i,j) might be approximated from the sending time of the packet that revealed the Hop, e.g., when the round-trip response time is available and divided by 2)

* Measurements of Round-Trip Delay (for each packet that reveals the same Node identity and flow attributes, then this attribute is computed; see next section)

Node identities and related information can be ordered by their distance from the Node with address Src in Hops h(i,j). Based on this, two forms of Routes are distinguished:

A Route Ensemble is defined as the combination of all Routes traversed by different flows from the Node at Src address to the Node at Dst address. A single Route traversed by a single flow (determined by an unambiguous tuple of addresses Src and Dst and other identical flow criteria) is a member of the Route Ensemble and called a Member Route.

Using h(i,j) and components and parameters further define:

When considering the set of Hops in the context of a single flow, a Member Route j is an ordered list {h(1,j), ... h(Nj, j)} where h(i-1, j) and h(i, j) are one Hop away from each other and Nj satisfying h(Nj,j)=Dst is the minimum count of Hops needed by the packet on member Route j to reach Dst. Member Routes must be unique. The uniqueness property requires that any two Member Routes, j and k, that are part of the same Route Ensemble differ either in terms of minimum Hop count Nj and Nk to reach the destination Dst or, in the case of identical Hop count Nj=Nk, they have at least one distinct Hop: h(i,j) != h(i,k) for at least one i (i=1..Nj).

All the optional information collected to describe a Member Route, such as the arrival interface, departure interface, and Round-Trip Delay at each Hop, turns each list item into a rich structure. There may be information on the links between Hops, possible information on the routing (arrival interface and departure interface), an estimate

of distance between Hops based on Round-Trip Delay measurements and calculations, and a timestamp indicating when all these additional details were measured.

The Route Ensemble from Src to Dst, during the measurement interval T0 to Tf, is the aggregate of all m distinct Member Routes discovered between the two Nodes with Src and Dst addresses. More formally, with the Node having address Src omitted:

```
Route Ensemble = {
{h(1,1), h(2,1), h(3,1), ... h(N1,1)=Dst},
{h(1,2), h(2,2), h(3,2),..., h(N2,2)=Dst},
...
{h(1,m), h(2,m), h(3,m), ....h(Nm,m)=Dst}
}
```

where the following conditions apply: i <= Nj <= Nmax (j=1..m)

Note that some h(i,j) may be empty (null) in the case that systems do not reply (not discoverable or not cooperating).

h(i-1,j) and h(i,j) are the Hops on the same Member Route one Hop away from each other.

Hop h(i,j) may be identical with h(k,l) for i!=k and j!=l, which means there may be portions shared among different Member Routes (parts of Member Routes may overlap).

### 3.5. Related Round-Trip Delay and Loss Definitions

RTD(i,j,T) is defined as a singleton of the [RFC2681] Round-Trip Delay between the Node with address = Src and the Node at Hop h(i,j) at time T.

RTL(i,j,T) is defined as a singleton of the [RFC6673] Round-Trip Loss between the Node with address = Src and the Node at Hop h(i,j) at time T.

### 3.6. Discussion

Depending on the way that the Node identity is revealed, it may be difficult to determine parallel subpaths between the same pair of Nodes (i.e., multiple parallel links). It is easier to detect parallel subpaths involving different Nodes.

* If a pair of discovered Nodes identify two different addresses (IP or not), then they will appear to be different Nodes. See item below.

* If a pair of discovered Nodes identify two different IP addresses and the IP addresses resolve to the same Node name (in the DNS), then they will appear to be the same Node.

* If a discovered Node always replies using the same network address, regardless of the interface a packet arrives on, then multiple parallel links cannot be detected in that network domain. This condition may apply to traceroute-style methods but may not apply to other hybrid methods based on In situ Operations, Administration, and Maintenance (IOAM). For example, if the ICMP extension mechanism described in [RFC5837] is implemented, then parallel links can be detected with the discovery traceroute-style methods.

* If parallel links between routers are aggregated below the IP layer, then, from the Node’s point of view, all these links share the same pair of IP addresses. The existence of these parallel

links can't be detected at the IP layer. This applies to other network domains with layers below them as well. This condition may apply to traceroute-style methods but may not apply to other hybrid methods based on IOAM.

When a Route assessment employs IP packets (for example), the reality of flow assignment to parallel subpaths involves layers above IP. Thus, the measured Route Ensemble is applicable to IP and higher layers (as described in the methodology's packet of Type-P and flow parameters).

### 3.7. Reporting the Metric

An Information Model and an XML Data Model for Storing Traceroute Measurements is available in [RFC5388]. The measured information at each Hop includes four pieces of information: a one-dimensional Hop index, Node symbolic address, Node IP address, and RTD for each response.

The description of Hop information that may be collected according to this memo covers more dimensions, as defined in Section 3.4. For example, the Hop index is two-dimensional to capture the complexity of a Route Ensemble, and it contains corresponding Node identities at a minimum. The models need to be expanded to include these features as well as Arrival Interface ID, Departure Interface ID, and arrival timestamp, when available. The original sending Timestamp from the Src Node anchors a particular measurement in time.

## 4. Route Assessment Methodologies

There are two classes of methods described in this section, active methods relying on the reaction to TTL or Hop Limit Exceeded condition to discover Nodes on a path and hybrid active-passive methods that involve direct interrogation of Cooperating Nodes (usually within a single domain). Description of these methods follow.

### 4.1. Active Methodologies

This section describes the method employed by current open-source tools, thereby providing a practical framework for further advanced techniques to be included as method variants. This method is applicable for use across multiple administrative domains.

Internet routing is complex because it depends on the policies of thousands of Autonomous Systems (ASes). Most routers perform load balancing on flows using a form of ECMP. [RFC2991] describes a number of flow-based or hashed approaches (e.g., Modulo-N Hash, Hash-Threshold, and Highest Random Weight (HRW)) and makes some good suggestions. Flow-based ECMP avoids increased packet Delay Variation and possibly overwhelming levels of packet reordering in flows.

A few routers still divide the workload through packet-based techniques, such as a round-robin scheme to distribute every new outgoing packet to multiple links, as explained in [RFC2991]. The methods described in this section assume flow-based ECMP.

Taking into account that Internet protocol was designed under the "end-to-end" principle, the IP payload and its header do not provide any information about the Routes or path necessary to reach some destination. For this reason, the popular tool, traceroute, was developed to gather the IP addresses of each Hop along a path using ICMP [RFC0792]. Traceroute also measures RTD from each Hop. However, the growing complexity of the Internet makes it more challenging to develop an accurate traceroute implementation. For instance, the early traceroute tools would be inaccurate in the current network,

mainly because they were not designed to retain a flow state. However, evolved traceroute tools, such as Paris-traceroute ([PT] [MLB]) and Scamper ([SCAMPER]), expect to encounter ECMP and achieve more accurate results when they do, where Scamper ensures traceroute packets will follow the same path in 98% of cases ([SCAMPER]).

Today's traceroute tools send Type-P of packets, which are either ICMP, UDP, or TCP. UDP and TCP are used when a particular characteristic needs to be verified, such as filtering or traffic shaping on specific ports (i.e., services). UDP and TCP traceroute are also used when ICMP responses are not received. [SCAMPER] supports IPv6 traceroute measurements, keeping the Flow Label constant in all packets.

Paris-traceroute allows its users to measure the RTD to every Node of the path for a particular flow. Furthermore, either Paris-traceroute or Scamper is capable of unveiling the many available paths between a source and destination (which are visible to active methods). This task is accomplished by repeating complete traceroute measurements with different flow parameters for each measurement; Paris-traceroute provides an "exhaustive" mode, while Scamper provides "tracelb" (which stands for "traceroute load balance"). "Framework for IP Performance Metrics" [RFC2330], updated by [RFC7312], has the flexibility to require that the Round-Trip Delay measurement [RFC2681] uses packets with the constraints to assure that all packets in a single measurement appear as the same flow. This flexibility covers ICMP, UDP, and TCP. The accompanying methodology of [RFC2681] needs to be expanded to report the sequential Hop identifiers along with RTD measurements, but no new metric definition is needed.

The advanced Route assessment methods used in Paris-traceroute [PT] keep the critical fields constant for every packet to maintain the appearance of the same flow. When considering IPv6 headers, it is necessary to ensure that the IP Source and Destination addresses and Flow Label are constant (but note that many routers ignore the Flow Label field at this time); see [RFC6437]. Use of IPv6 Extension Headers may add critical fields and SHOULD be avoided. In IPv4, certain fields of the IP header and the first 4 bytes of the IP payload should remain constant in a flow. In the IPv4 header, the IP Source and Destination addresses, protocol number, and Diffserv fields identify flows. The first 4 payload bytes include the UDP and TCP ports and the ICMP type, code, and checksum fields.

Maintaining a constant ICMP checksum in IPv4 is most challenging, as the ICMP sequence number or identifier fields will usually change for different probes of the same path. Probes should use arbitrary bytes in the ICMP data field to offset changes to the sequence number and identifier, thus keeping the checksum constant.

Finally, it is also essential to Route the resulting ICMP Time Exceeded messages along a consistent path. In IPv6, the fields above are sufficient. In IPv4, the ICMP Time Exceeded message will contain the IP header and the first 8 bytes of the IP payload, both of which affect its ICMP checksum calculation. The TCP sequence number, UDP length, and UDP checksum will affect this value and should remain constant.

Formally, to maintain the same flow in the measurements to a particular Hop, the Type-P-Route-Ensemble-Method-Variant packets should have the following attributes (see [PT]):

TCP case: For IPv4, the fields Src, Dst, port-Src, port_Dst, sequence number, and Diffserv SHOULD be the same. For IPv6, the fields Flow Label, Src, and Dst SHOULD be the same.

UDP case: For IPv4, the fields Src, Dst, port-Src, port-Dst, and Diffserv should be the same, and the UDP checksum SHOULD change to keep the IP checksum of the ICMP Time Exceeded reply constant. Then, the data length should be fixed, and the data field is used to make it so (consider that ICMP checksum uses its data field, which contains the original IP header plus 8 bytes of UDP, where TTL, IP identification, IP checksum, and UDP checksum changes). For IPv6, the field Flow Label and Source and Destination addresses SHOULD be the same.

ICMP case: For IPv4, the data field SHOULD compensate variations on TTL or Hop Limit, IP identification, and IP checksum for every packet. There is no need to consider ICMPv6 because only Flow Label of IPv6 and Source and Destination addresses are used, and all of them SHOULD be constant.

Then, the way to identify different Hops and attempts of the same IPv4 flow is:

TCP case: The IP identification field.

UDP case: The IP identification field.

ICMP case: The IP identification field and ICMP sequence number.

#### 4.1.1. Temporal Composition for Route Metrics

The active Route assessment methods described above have the ability to discover portions of a path where ECMP load balancing is present, observed as two or more unique Member Routes having one or more distinct Hops that are part of the Route Ensemble. Likewise, attempts to deliberately vary the flow characteristics to discover all Member Routes will reveal portions of the path that are flow invariant.

Section 9.2 of [RFC2330] describes the Temporal Composition of metrics and introduces the possibility of a relationship between earlier measurement results and the results for measurement at the current time (for a given metric). There is value in establishing a Temporal Composition relationship for Route metrics; however, this relationship does not represent a forecast of future Route conditions in any way.

For Route-metric measurements, the value of Temporal Composition is to reduce the measurement iterations required with repeated measurements. Reduced iterations are possible by inferring that current measurements using fixed and previously measured flow characteristics:

* will have many common Hops with previous measurements.

* will have relatively time-stable results at the ingress and egress portions of the path when measured from user locations, as opposed to measurements of backbone networks and across inter-domain gateways.

* may have greater potential for time variation in path portions where ECMP load balancing is observed (because increasing or decreasing the pool of links changes the hash calculations).

Optionally, measurement systems may take advantage of the inferences above when seeking to reduce measurement iterations after exhaustive measurements indicate that the time-stable properties are present. Repetitive active Route measurement systems:

1. SHOULD occasionally check path portions that have exhibited

stable results over time, particularly ingress and egress portions of the path (e.g., daily checks if measuring many times during a day).

2. SHOULD continue testing portions of the path that have previously exhibited ECMP load balancing.

3. SHALL trigger reassessment of the complete path and Route Ensemble if any change in Hops is observed for a specific (and previously tested) flow.

#### 4.1.2. Routing Class Identification

There is an opportunity to apply the notion from [RFC2330] of equal treatment for a class of packets, "...very useful to know if a given Internet component treats equally a class C of different types of packets", as it applies to Route measurements. The notion of class C was examined further in [RFC8468] as it applied to load-balancing flows over parallel paths, which is the case we develop here. Knowledge of class C parameters (unrelated to address classes of the past) on a path potentially reduces the number of flows required for a given method to assess a Route Ensemble over time.

First, recognize that each Member Route of a Route Ensemble will have a corresponding class C. Class C can be discovered by testing with multiple flows, all of which traverse the unique set of Hops that comprise a specific Member Route.

Second, recognize that the different classes depend primarily on the hash functions used at each instance of ECMP load balancing on the path.

Third, recognize the synergy with Temporal Composition methods (described above), where evaluation intends to discover time-stable portions of each Member Route so that more emphasis can be placed on ECMP portions that also determine class C.

The methods to assess the various class C characteristics benefit from the following measurement capabilities:

* flows designed to determine which n-tuple header fields are considered by a given hash function and ECMP Hop on the path and which are not. This operation immediately narrows the search space, where possible, and partially defines a class C.

* a priori knowledge of the possible types of hash functions in use also helps to design the flows for testing (major router vendors publish information about these hash functions; examples are in [LOAD_BALANCE]).

* ability to direct the emphasis of current measurements on ECMP portions of the path, based on recent past measurement results (the Routing Class of some portions of the path is essentially "all packets").

#### 4.1.3. Intermediate Observation Point Route Measurement

There are many examples where passive monitoring of a flow at an Observation Point within the network can detect unexpected Round-Trip Delay or Delay Variation. But how can the cause of the anomalous delay be investigated further *from the Observation Point* possibly located at an intermediate point on the path?

In this case, knowledge that the flow of interest belongs to a

specific Routing Class C will enable measurement of the Route where anomalous delay has been observed. Specifically, Round-Trip Delay assessment to each Hop on the path between the Observation Point and the Destination for the flow of interest may discover high or variable delay on a specific link and Hop combination.

The determination of a Routing Class C that includes the flow of interest is as described in the section above, aided by computation of the relevant hash function output as the target.

### 4.2. Hybrid Methodologies

The Hybrid Type I methods provide an alternative for Route assessment. The "Scope, Applicability, and Assumptions" section of [RFC9197] provides one possible set of data fields that would support Route identification.

In general, Nodes in the measured domain would be equipped with specific abilities:

* Store the identity of Nodes that a packet has visited in header data fields in the order the packet visited the Nodes.

* Support of a "Loopback" capability where a copy of the packet is returned to the encapsulating Node and the packet is processed like any other IOAM packet on the return transfer.

In addition to Node identity, Nodes may also identify the ingress and egress interfaces utilized by the tracing packet, the absolute time when the packet was processed, and other generic data (as described in Section 3 of [RFC9197]). Interface identification isn’t necessarily limited to IP, i.e., different links in a bundle (Link Aggregation Control Protocol (LACP)) could be identified. Equally well, links without explicit IP addresses can be identified (like with unnumbered interfaces in an IGP deployment).

Note that the Type-P packet specification for this method will likely be a partial specification because most of the packet fields are determined by the user traffic. The packet encapsulation header or headers added by the hybrid method can certainly be specified in Type-P, in unpopulated form.

### 4.3. Combining Different Methods

In principle, there are advantages if the entity conducting Route measurements can utilize both forms of advanced methods (active and hybrid) and combine the results. For example, if there are Nodes involved in the path that qualify as Cooperating Nodes but not as Discoverable Nodes, then a more complete view of Hops on the path is possible when a hybrid method (or interrogation protocol) is applied and the results are combined with the active method results collected across all other domains.

In order to combine the results of active and hybrid/interrogation methods, the network Nodes that are part of a domain supporting an interrogation protocol have the following attributes:

1. Nodes at the ingress to the domain SHOULD be both Discoverable and Cooperating.

2. Any Nodes within the domain that are both Discoverable and Cooperating SHOULD reveal the same Node identity in response to both active and hybrid methods.

3. Nodes at the egress to the domain SHOULD be both Discoverable and

Cooperating and SHOULD reveal the same Node identity in response to both active and hybrid methods.

When Nodes follow these requirements, it becomes a simple matter to match single-domain measurements with the overlapping results from a multidomain measurement.

In practice, Internet users do not typically have the ability to utilize the Operations, Administrations, and Maintenance (OAM) capabilities of networks that their packets traverse, so the results from a remote domain supporting an interrogation protocol would not normally be accessible. However, a network operator could combine interrogation results from their access domain with other measurements revealing the path outside their domain.

## 5. Background on Round-Trip Delay Measurement Goals

The aim of this method is to use packet probes to unveil the paths between any two End-Nodes of the network. Moreover, information derived from RTD measurements might be meaningful to identify:

1. Intercontinental submarine links
2. Satellite communications
3. Congestion
4. Inter-domain paths

This categorization is widely accepted in the literature and among operators alike, and it can be trusted with empirical data and several sources as ground of truth (e.g., [RTTSub]), but it is an inference measurement nonetheless [bdrmap] [IDCong].

The first two categories correspond to the physical distance dependency on RTD, the next one binds RTD with queuing delay on routers, and the last one helps to identify different ASes using traceroutes. Due to the significant contribution of propagation delay in long-distance Hops, RTD will be on the order of 100 ms on transatlantic Hops, depending on the geolocation of the vantage points. Moreover, RTD is typically higher than 480 ms when two Hops are connected using geostationary satellite technology (i.e., their orbit is at 36000 km). Detecting congestion with latency implies deeper mathematical understanding, since network traffic load is not stationary. Nonetheless, as the first approach, a link seems to be congested if observing different/varying statistical results after sending several traceroute probes (e.g., see [IDCong]). Finally, to recognize distinctive ASes in the same traceroute path is challenging because more data is needed, like AS relationships and Regional Internet Registry (RIR) delegations among others (for more details, please consult [bdrmap]).

## 6. RTD Measurements Statistics

Several articles have shown that network traffic presents a self-similar nature [SSNT] [MLRM] that is accountable for filling the queues of the routers. Moreover, router queues are designed to handle traffic bursts, which is one of the most remarkable features of self-similarity. Naturally, while queue length increases, the delay to traverse the queue increases as well and leads to an increase on RTD. Due to traffic bursts generating short-term overflow on buffers (spiky patterns), every RTD only depicts the queueing status on the instant when that packet probe was in transit. For this reason, several RTD measurements during a time window could begin to describe the random behavior of latency. Loss must also be accounted for in the methodology.

To understand the ongoing process, examining the quartiles provides a nonparametric way of analysis. Quartiles are defined by five values: minimum RTD (m), RTD value of the 25% of the Empirical Cumulative Distribution Function (ECDF) (Q1), the median value (Q2), the RTD value of the 75% of the ECDF (Q3), and the maximum RTD (M). Congestion can be inferred when RTD measurements are spread apart; consequently, the Interquartile Range (IQR), i.e., the distance between Q3 and Q1, increases its value.

This procedure requires the algorithm presented in [P2] to compute quartile values "on the fly".

This procedure allows us to update the quartile values whenever a new measurement arrives, which is radically different from classic methods of computing quartiles, because they need to use the whole dataset to compute the values. This way of calculus provides savings in memory and computing time.

To sum up, the proposed measurement procedure consists of performing traceroutes several times to obtain samples of the RTD in every Hop from a path during a time window (W) and compute the quartiles for every Hop. This procedure could be done for a single Member Route flow, for a non-exhaustive search with parameter E (defined below) set to False, or for every detected Route Ensemble flow (E=True).

The identification of a specific Hop in a traceroute is based on the IP origin address of the returned ICMP Time Exceeded packet and on the distance identified by the value set in the TTL (or Hop Limit) field inserted by traceroute. As this specific Hop can be reached by different paths, the IP Source and Destination addresses of the traceroute packet also need to be recorded. Finally, different return paths are distinguished by evaluating the ICMP Time Exceeded TTL (or Hop Limit) of the reply message; if this TTL (or Hop Limit) is constant for different paths containing the same Hop, the return paths have the same distance. Moreover, this distance can be estimated considering that the TTL (or Hop Limit) value is normally initialized with values 64, 128, or 255. The 5-tuple (origin IP, destination IP, reply IP, distance, and response TTL or Hop Limit) unequivocally identifies every measurement.

This algorithm below runs in the origin of the traceroute. It returns the Qs quartiles for every Hop and Alt (alternative paths because of balancing). Notice that the "Alt" parameter condenses the parameters of the 5-tuple (origin IP, destination IP, reply IP, distance, and response TTL), i.e., one for each possible combination.

```
================================================================
0  input:   W (window time of the measurement)
1           i_t (time between two measurements, set the i_t time
2                long enough to avoid incomplete results)
3           E (True: exhaustive, False: a single path)
4           Dst (destination IP address)
5  output:  Qs (quartiles for every Hop and Alt)
----------------------------------------------------------------
6  T := start_timer(W)
7  while T is not finished do:
8  |       start_timer(i_t)
9  |       RTD(Hop,Alt) = advanced-traceroute(Dst,E)
10 |       for each Hop and Alt in RTD do:
11 |       |     Qs[Dst,Hop,Alt] := ComputeQs(RTD(Hop,Alt))
12 |       done
13 |       wait until i_t timer is expired
14 done
15  return (Qs)
================================================================
```

During the time W, lines 6 and 7 assure that the measurement loop is made. Lines 8 and 13 set a timer for each cycle of measurements. A cycle comprises the traceroutes packets, considering every possible Hop and the alternatives paths in the Alt variable (ensured in lines 9-12). In line 9, the advanced-traceroute could be either Paris-traceroute or Scamper, which will use the "exhaustive" mode or "tracelb" option if E is set to True, respectively. The procedure returns a list of tuples (m, Q1, Q2, Q3, and M) for each intermediate Hop, or "Alt" in as a function of the 5-tuple, in the path towards the Dst. Finally, lines 10 through 12 store each measurement into the real-time quartiles computation.

Notice there are cases where even having a unique Hop at distance h from the Src to Dst, the returning path could have several possibilities, yielding different total paths. In this situation, the algorithm will return another "Alt" for this particular Hop.

## 7. Security Considerations

The security considerations that apply to any active measurement of live paths are relevant here as well. See [RFC4656] and [RFC5357].

The active measurement process of changing several fields to keep the checksum of different packets identical does not require special security considerations because it is part of synthetic traffic generation and is designed to have minimal to zero impact on network processing (to process the packets for ECMP).

Some of the protocols used (e.g., ICMP) do not provide cryptographic protection for the requested/returned data, and there are risks of processing untrusted data in general, but these are limitations of the existing protocols where we are applying new methods.

For applicable hybrid methods, the security considerations in [RFC9197] apply.

When considering the privacy of those involved in measurement or those whose traffic is measured, the sensitive information available to potential observers is greatly reduced when using active techniques that are within this scope of work. Passive observations of user traffic for measurement purposes raise many privacy issues. We refer the reader to the privacy considerations described in the Large-scale Measurement of Broadband Performance (LMAP) Framework [RFC7594], which covers active and passive techniques.

## 8. IANA Considerations

This document has no IANA actions.

## 9. References

### 9.1. Normative References


[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, DOI 10.17487/RFC0792, September 1981, <https://www.rfc-editor.org/info/rfc792>.

[RFC1122] Braden, R., Ed., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, DOI 10.17487/RFC1122, October 1989, <https://www.rfc-editor.org/info/rfc1122>.

[RFC1812] Baker, F., Ed., "Requirements for IP Version 4 Routers", RFC 1812, DOI 10.17487/RFC1812, June 1995, <https://www.rfc-editor.org/info/rfc1812>.

[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>.

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

[RFC2681] Almes, G., Kalidindi, S., and M. Zekauskas, "A Round-trip Delay Metric for IPPM", RFC 2681, DOI 10.17487/RFC2681, September 1999, <https://www.rfc-editor.org/info/rfc2681>.

[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>.

[RFC5388] Niccolini, S., Tartarelli, S., Quittek, J., Dietz, T., and M. Swany, "Information Model and XML Data Model for Traceroute Measurements", RFC 5388, DOI 10.17487/RFC5388, December 2008, <https://www.rfc-editor.org/info/rfc5388>.

[RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", RFC 6438, DOI 10.17487/RFC6438, November 2011, <https://www.rfc-editor.org/info/rfc6438>.

[RFC6673] Morton, A., "Round-Trip Packet Loss Metrics", RFC 6673, DOI 10.17487/RFC6673, August 2012, <https://www.rfc-editor.org/info/rfc6673>.

[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>.

[RFC8029] Kompella, K., Swallow, G., Pignataro, C., Ed., Kumar, N., Aldrin, S., and M. Chen, "Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures", RFC 8029, DOI 10.17487/RFC8029, March 2017, <https://www.rfc-editor.org/info/rfc8029>.

[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>.

[RFC8468] Morton, A., Fabini, J., Elkins, N., Ackermann, M., and V. Hegde, "IPv4, IPv6, and IPv4-IPv6 Coexistence: Updates for the IP Performance Metrics (IPPM) Framework", RFC 8468, DOI 10.17487/RFC8468, November 2018, <https://www.rfc-editor.org/info/rfc8468>.

[RFC9197] Brockners, F., Ed., Bhandari, S., Ed., and T. Mizrahi, Ed., "Data Fields for In Situ Operations, Administration, and Maintenance (IOAM)", RFC 9197, DOI 10.17487/RFC9197, May 2022, <https://www.rfc-editor.org/info/rfc9197>.

## 9.2. Informative References

[bdrmap] Luckie, M., Dhamdhere, A., Huffaker, B., Clark, D., and KC. Claffy, "bdrmap: Inference of Borders Between IP Networks", Proceedings of the 2016 ACM on Internet Measurement Conference, pp. 381-396, DOI 10.1145/2987443.2987467, November 2016,

<https://doi.org/10.1145/2987443.2987467>.

[IDCong] Luckie, M., Dhamdhere, A., Clark, D., and B. Huffaker, "Challenges in Inferring Internet Interdomain Congestion", Proceedings of the 2014 Conference on Internet Measurement Conference, pp. 15-22, DOI 10.1145/2663716.2663741, November 2014, <https://doi.org/10.1145/2663716.2663741>.

[LOAD_BALANCE] Sanguanpong, S., Pittayapitak, W., and K. Kasom Koht-Arsa, "COMPARISON OF HASH STRATEGIES FOR FLOW-BASED LOAD BALANCING", International Journal of Electronic Commerce Studies, Vol.6, No.2, pp.259-268, DOI 10.7903/ijecs.1346, December 2015, <https://doi.org/10.7903/ijecs.1346>.

[MLB] Augustin, B., Friedman, T., and R. Teixeira, "Measuring load-balanced paths in the internet", Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pp. 149-160, DOI 10.1145/1298306.1298329, October 2007, <https://doi.org/10.1145/1298306.1298329>.

[MLRM] Fontugne, R., Mazel, J., and K. Fukuda, "An empirical mixture model for large-scale RTT measurements", 2015 IEEE Conference on Computer Communications (INFOCOM), pp. 2470-2478, DOI 10.1109/INFOCOM.2015.7218636, April 2015, <https://doi.org/10.1109/INFOCOM.2015.7218636>.

[P2] Jain, R. and I. Chlamtac, "The P 2 algorithm for dynamic calculation of quartiles and histograms without storing observations", Communications of the ACM 28.10 (1985): 1076-1085, DOI 10.1145/4372.4378, October 1985, <https://doi.org/10.1145/4372.4378>.

[PT] Augustin, B., Cuvellier, X., Orgogozo, B., Viger, F., Friedman, T., Latapy, M., Magnien, C., and R. Teixeira, "Avoiding traceroute anomalies with Paris traceroute", Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, pp. 153-158, DOI 10.1145/1177080.1177100, October 2006, <https://doi.org/10.1145/1177080.1177100>.

[RFC2991] Thaler, D. and C. Hopps, "Multipath Issues in Unicast and Multicast Next-Hop Selection", RFC 2991, DOI 10.17487/RFC2991, November 2000, <https://www.rfc-editor.org/info/rfc2991>.

[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>.

[RFC5835] Morton, A., Ed. and S. Van den Berghe, Ed., "Framework for Metric Composition", RFC 5835, DOI 10.17487/RFC5835, April 2010, <https://www.rfc-editor.org/info/rfc5835>.

[RFC5837] Atlas, A., Ed., Bonica, R., Ed., Pignataro, C., Ed., Shen, N., and JR. Rivers, "Extending ICMP for Interface and Next-Hop Identification", RFC 5837, DOI 10.17487/RFC5837, April 2010, <https://www.rfc-editor.org/info/rfc5837>.

[RFC6437] Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", RFC 6437, DOI 10.17487/RFC6437, November 2011, <https://www.rfc-editor.org/info/rfc6437>.

[RFC7312] Fabini, J. and A. Morton, "Advanced Stream and Sampling Framework for IP Performance Metrics (IPPM)", RFC 7312,

DOI 10.17487/RFC7312, August 2014,
<https://www.rfc-editor.org/info/rfc7312>.

[RFC7325] Villamizar, C., Ed., Kompella, K., Amante, S., Malis, A., and C. Pignataro, "MPLS Forwarding Compliance and Performance Requirements", RFC 7325, DOI 10.17487/RFC7325, August 2014, <https://www.rfc-editor.org/info/rfc7325>.

[RFC7594] Eardley, P., Morton, A., Bagnulo, M., Burbridge, T., Aitken, P., and A. Akhter, "A Framework for Large-Scale Measurement of Broadband Performance (LMAP)", RFC 7594, DOI 10.17487/RFC7594, September 2015, <https://www.rfc-editor.org/info/rfc7594>.

[RFC8403] Geib, R., Ed., Filsfils, C., Pignataro, C., Ed., and N. Kumar, "A Scalable and Topology-Aware MPLS Data-Plane Monitoring System", RFC 8403, DOI 10.17487/RFC8403, July 2018, <https://www.rfc-editor.org/info/rfc8403>.

[RTTSub] Bischof, Z., Rula, J., and F. Bustamante, "In and out of Cuba: Characterizing Cuba's Connectivity", Proceedings of the 2015 ACM Conference on Internet Measurement Conference, pp. 487-493, DOI 10.1145/2815675.2815718, October 2015, <https://doi.org/10.1145/2815675.2815718>.

[SCAMPER] Matthew Luckie, M., "Scamper: a scalable and extensible packet prober for active measurement of the internet", Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, pp. 239-245, DOI 10.1145/1879141.1879171, November 2010, <https://doi.org/10.1145/1879141.1879171>.

[SSNT] Park, K. and W. Willinger, "Self-Similar Network Traffic and Performance Evaluation (1st ed.)", DOI 10.1002/047120644X, John Wiley & Sons, Inc., New York, NY, USA, August 2000, <https://doi.org/10.1002/047120644X>.

## Appendix A. MPLS Methods for Route Assessment

A Node assessing an MPLS path must be part of the MPLS domain where the path is implemented. When this condition is met, [RFC8029] provides a powerful set of mechanisms to detect "correct operation of the data plane, as well as a mechanism to verify the data plane against the control plane".

MPLS routing is based on the presence of a Forwarding Equivalence Class (FEC) Stack in all visited Nodes. Selecting one of several Equal-Cost Multipaths (ECMPs) is, however, based on information hidden deeper in the stack. Late deployments may support a so-called "Entropy label" for this purpose. State-of-the-art deployments base their choice of an ECMP member interface on the complete MPLS label stack and on IP addresses up to the complete 5-tuple IP header information (see Section 2.4 of [RFC7325]). Load sharing based on IP information decouples this function from the actual MPLS routing information. Thus, an MPLS traceroute is able to check how packets with a contiguous number of ECMP-relevant IP addresses (and an identical MPLS label stack) are forwarded by a particular router. The minimum number of equivalent MPLS paths traceable at a router should be 32. Implementations supporting more paths are available.

The MPLS echo request and reply messages offering this feature must support the Downstream Detailed Mapping TLV (was Downstream Mapping initially, but the latter has been deprecated). The MPLS echo response includes the incoming interface where a router received the MPLS echo request. The MPLS echo reply further informs which of the n addresses relevant for the load-sharing decision results in a

particular next-hop interface and contains the next Hop's interface address (if available). This ensures that the next Hop will receive a properly coded MPLS echo request in the next step Route of assessment.

[RFC8403] explains how a central Path Monitoring System could be used to detect arbitrary MPLS paths between any routers within a single MPLS domain. The combination of MPLS forwarding, Segment Routing, and MPLS traceroute offers a simple architecture and a powerful mechanism to detect and validate (segment-routed) MPLS paths.

## Acknowledgements

The original three authors (Ignacio, Al, Joachim) acknowledge Ruediger Geib for his penetrating comments on the initial document and his initial text for the appendix on MPLS. Carlos Pignataro challenged the authors to consider a wider scope and applied his substantial expertise with many technologies and their measurement features in his extensive comments. Frank Brockners also shared useful comments and so did Footer Foote. We thank them all!

## Authors' Addresses

J. Ignacio Alvarez-Hamelin
Universidad de Buenos Aires
Av. Paseo Coln 850
C1063ACV Buenos Aires
Argentina
Phone: +54 11 5285-0716
Email: ihameli@cnet.fi.uba.ar
URI: http://cnet.fi.uba.ar/ignacio.alvarez-hamelin/

Al Morton
AT&T Labs
200 Laurel Avenue South
Middletown, NJ 07748
United States of America
Phone: +1 732 420 1571
Email: acm@research.att.com

Joachim Fabini
TU Wien
Gusshausstrasse 25/E389
1040 Vienna
Austria
Phone: +43 1 58801 38813
Email: Joachim.Fabini@tuwien.ac.at
URI: http://www.tc.tuwien.ac.at/about-us/staff/joachim-fabini/

Carlos Pignataro
Cisco Systems, Inc.
7200-11 Kit Creek Road
Research Triangle Park, NC 27709
United States of America
Email: cpignata@cisco.com

Ruediger Geib
Deutsche Telekom
Heinrich Hertz Str. 3-7
64295 Darmstadt
Germany
Phone: +49 6151 5812747

Email: Ruediger.Geib@telekom.de