IPPM Working Group
Internet-Draft
Intended status: Standards Track
Expires: 23 March 2026

G. Mirsky
Ericsson
W. Lingqiang
G. Zhui
ZTE Corporation
H. Song
Futurewei Technologies
P. Thubert
Independent
19 September 2025

Hybrid Two-Step Performance Measurement Method
draft-ietf-ippm-hybrid-two-step-06

Abstract

The ~~development and advancements in~~adoption of network operation automation have brought new measurement methodology requirements. Among them is the ability to collect instant network operational state as the packet being processed by the networking elements along its forwarding path through ~~the~~a network domain. ~~That task can be solved~~An approach to address that requirement is to ~~using~~ use on-path telemetry, also called hybrid measurement. An on-path telemetry method allows operators ~~the~~to collect~~ion of essential~~key information that reflects the operational state and network performance experienced by ~~the~~ a packet through the forwarding path. This document introduces a method complementary to on-path telemetry that causes the generation of network telemetry information. This method, referred to as Hybrid Two-Step (HTS), separates the act of measuring ~~and/~~or calculating the performance metric from collecting and transporting network operational state. ~~The~~ An HTS packet traverses the same set of nodes and links as the trigger packet, thus simplifying the correlation of informational elements originating ~~on~~ from nodes traversed by the trigger packet.

Commenté [MB1]: Use consistent term as you also have «network element» in the main text.

Commenté [MB2]: IOAM is an example of hybrid.

I would reword.

Commenté [MB3]: Covers also both.

Commenté [MB4]: Not introduced.

Commenté [MB5]: Maybe delete and leave these details to the main document.

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 23 March 2026.

Table of Contents

1. Introduction

~~Successful resolution of~~Addressing the challenges of automated network operation, as part of, for example, overall service orchestration or data center operation, relies on a timely collection of accurate information that reflects the state of involved network elements on an unprecedented scale.

Commenté [MB6]: Why this case in particular? Isn't this covered by the service orchestration part as well?

Commenté [MB7]: Do we need this?

Because performing the analysis and acting upon the collected information ~~requires~~require considerable computing and storage resources, the network operational state information ~~is unlikely to~~might not be processed by the network elements themselves but ~~will~~ may be ~~exported~~ offloaded to big data systems for processing and storing. The process of generating, and collecting network operational state information, also referred to in this document as network telemetry, and transporting it for post-processing should work equally well with data flows or injected in ~~the~~ dedicated ~~network~~ test packets. [RFC7799] describes a combination of elements of passive and active measurement as a hybrid measurement.

Several technical methods have been proposed to enable the collection of network operational state information instantaneous to ~~the~~ packets processing, among them [P4.INT] and [RFC9197]. The instantaneous~~,~~ (i.e., in the data packet itself~~,~~) collection of telemetry information simplifies the process of attribution of telemetry information to ~~the~~a ~~particular~~ specific monitored flow. On the other hand, this collection method impacts the data packets, potentially changing their treatment by the network ~~node~~elements. Also, the amount of information that ~~the~~ an instantaneous method collects might be incomplete because of the limited space it can be allotted. Other proposals ~~defined~~ define methods to collect telemetry information in a separate packet from each ~~node~~ network element traversed by ~~the~~ a monitored data flow. [RFC9326] is an example of this approach to collect~~ing~~ telemetry information. These methods allow data collection from any arbitrary forwarding path and avoid directly impacting data packets (that is, packets that carry user data). On the other hand, ~~the correlation~~ unambiguously correlating ~~of~~ data ~~and with~~ the monitored flow requires that each packet with telemetry information also includes characteristic information about the monitored flow.

This document introduces Hybrid Two-Step (HTS) as a new method of

telemetry collection that improves accuracy of a measurement by

separating the act of measuring or calculating the performance metric

from the ~~collecting~~ collection and transport~~ing~~ of this information while

minimizing the overhead of the generated load in a network.

HTS method extends the two-step mode of Residence Time Measurement (RTM) defined in [RFC8169] to on-path network operational state collection and transport. HTS allows the collection of telemetry information

from any arbitrary forwarding path. HTS instruments data packets of the monitored flow or specially constructed test packets that are already

Commenté [MB8]: Implicit correlation may be considered, but it is suboptimal.

Conveying flow info in the packet is superior. The modification is to insist o this.

Commenté [MB9]: As a new paragraoh

a mis en forme : Surlignage

Commenté [MB10]: I'm not sure about this as there is a format to be supported.

a mis en forme : Surlignage

equipped with a shim of on-path telemetry protocol to use as an HTS trigger packet, making the process of attribution of telemetry to the data flow simple.

2. Conventions used in this document

2.1. Acronyms ~~and Terminology~~

~~RTM Residence Time Measurement [RFC8169]~~

~~ECMP Equal Cost Multipath [RFC2992]~~

MTU Maximum Transmission Unit [RFC1191]

HTS Hybrid Two-Step (Section 1)

HMAC Hashed Message Authentication Code [RFC2104]

TLV Type-Length-Value (Figure 4)

~~RTT Round-Trip Time [RFC2681]~~

2.2. Terminology

This document makes use of the following terms:

Characteristic information refers to the ~~the interpretation in this document follows the definition of~~ "Characteristic" in Section 3.2 in [I-D.ietf-nmop-terminology].

HTS domain is an example of a "limited domain" as defined in [RFC8799]. An HTS domain may be identical to an IOAM domain (see Section 3 of [RFC9197]) or a controlled domain where the Alternate-Marking Method is used to measure performance metrics (see Section 7.1 of [RFC9341]).

HTS egress node~~, defined in Section 4.4,~~ is a network element that consumes HTS Trigger packet and terminates an HTS follow-up packet. Refer to Section 4.4.

HTS ingress node~~, defined in Section 4.2,~~ is a~~n HTS system~~ network element that generates an HTS Trigger and HTS follow-up packets. Refer to Section 4.2.

HTS intermediate node~~, defined in Section 4.3,~~ is ~~an HTS system~~a network element that Receives and forwards ~~the~~ HTS Trigger and HTS follow-up packets~~,~~ in an HTS domain.

Network telemetry ~~- the interpretation in this document~~ is the same as in [RFC9232]. Refer to Section 4.3.

Network Operational State ~~- the interpretation in this document~~refers to ~~follows the use of~~ "Operational State" in [RFC8342].

Commenté [MB11]: Used only once

Commenté [MB12]: Used only once

Commenté [MB13]: Used only once.

Commenté [MB14]: I would position it right after ingress entry (logical flow)

Commenté [MB15]: After reading the document, I think that we may simplify here and merge these two entries. «Telemetry» can be used to refer to both.

2.2. 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.

3. Problem Overview

Performance measurements are meant to provide data that characterize conditions experienced by traffic flows in ~~the~~ a network. Such measurements are used for various purposes, including ~~and possibly~~ trigger operational changes (e.g., re-route of flows~~,~~ or changes in resource allocations) to better align with the intended state. Modifications to a network ~~are determined based on~~take into account the performance metric information available when a change is to be made. The correctness of this determination is thus based on the quality of the collected metrics data. The quality of collected measurement data is defined by:

* the resolution and accuracy of each measurement;

* the predictability of both the time at which each measurement is made and the timeliness of measurement collection data delivery for use.

Consider the case of delay measurement that relies on collecting time of packet arrival at the ingress interface and time of the packet transmission at the egress interface. The method includes recording a local clock value on receiving the first ~~octet~~ byte of an affected message at the device ingress, and again recording the clock value on transmitting the first byte of the same message at the device egress. In this ideal case, the difference between the two recorded clock times corresponds to the time that the message spent in traversing the device. In practice, the time recorded can differ from the ideal case by any fixed amount. A correction can be applied to compute the same time difference considering the known fixed time associated with the actual measurement. In this way, the resulting time difference reflects any variable delay associated with queuing.

Depending on the implementation, it may be challenging to compute the difference between message arrival and departure times and - on the fly - add the necessary residence time information to the same message. ~~And t~~That task may become even more challenging if the packet is encrypted. Recording the departure of a packet time in the same packet may be decremental to the accuracy of the measurement because the departure time includes the variable time component (such as that associated with buffering and queuing of the packet). A similar problem may lower the quality of, for example, information that characterizes utilization of the egress interface. If unable to obtain the data consistently, without variable delays for additional processing, information may not accurately reflect the egress

Commenté [MB16]: As there are other criteria

Commenté [MB17]: Not sure to get the point made here.

Commenté [MB18]: Which node? Network?

Commenté [MB19]: Depends if time sync is place.

interface state. To mitigate this problem [RFC8169] ~~defined~~ defines an RTM
two-step mode.

Another challenge associated with methods that collect network operational state information into ~~the~~ actual data packets is the risk to exceed the Maximum Transmission Unit (MTU) size on the path, especially if ~~the packet traverses overlay domains or VPNs~~encapsulation is used. Since ~~the~~ fragmentation ~~is~~ may not be supported~~not~~ ~~available~~ at the ~~transport~~ network, operators may have to reduce MTU size advertised to the client layer or risk missing network operational state data for the part, most probably the latter part, of the path.

Performance measurement methods that instrument data flows inherently collect one-way performance metrics at the egress of the measurement domain. In some networks ~~(, for example,~~e.g., wireless networks that are in the scope of [RFC9450]~~)~~, it is beneficial to collect the telemetry data, including the calculated performance metrics, that reflects conditions experienced by the monitored flow at a network ~~node~~ element other than ~~the~~ an egress network ~~node~~element. For example, a head-end can optimize path selection based on the compounded information that reflects network conditions and resource utilization. This mode is referred to as the upstream collection and the other - downstream collection to differentiate between two modes of telemetry collection.

4. Theory of Operation

The HTS method consists of two phases:

* Performing a measurement and/or obtaining network operational state information on a network ~~node~~element. HTS Trigger is a data or test packet instrumented to trigger the collection of telemetry information on a network ~~node~~element.

* Collecting and transporting the measurement and/or the telemetry information. HTS Follow-up is a packet constructed to transport telemetry information that includes operational state and performance measurements originated on the nodes along the path traversed by the HTS Trigger.

4.1. HTS Packets

4.1.1. HTS Trigger Packets

4.4.1.1 in In-Situ OAM

~~0 1 2 3~~
~~0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1~~
~~+-+~~
~~| Namespace-ID | Flags |Extension-Flags|~~

Commenté [MB20]: ???

Commenté [MB21]: Why egress? Why not any border element?

Commenté [MB22]: Of what?

Commenté [MB23]: Can we please have a figure with a sample topo and then use that figure to illustrate the use of various packet types?

Commenté [MB24]: Point to where this phase is described

Commenté [MB25]: Why those are called out separately here but the follow-up reasons mainly about telemetry that also include measurements, etc.?

Commenté [MB26]: Before diving into protocol-specifics, can we clarify:

•Clarify whether one or more follow-up packets can be triggered?
•If multiple follow-ups are supported, how to indicate «end» of telemetry associated with a trigger packet?

Commenté [MB27]: Point to where this phase is described

Commenté [MB28]: These are only variants of the trigger.

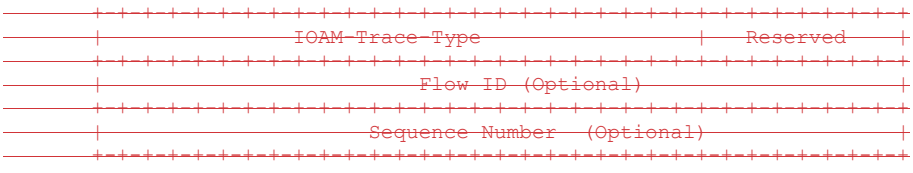

Figure 1: Hybrid Two-Step Trace IOAM Header

An HTS Trigger may be carried in a data packet or a specially constructed test packet. **For example,** an HTS Trigger could be a packet that has IOAM Option-Type set to the "IOAM Hybrid Two-Step Option-Type" value (TBA1) allocated by IANA (see Section 7.1).

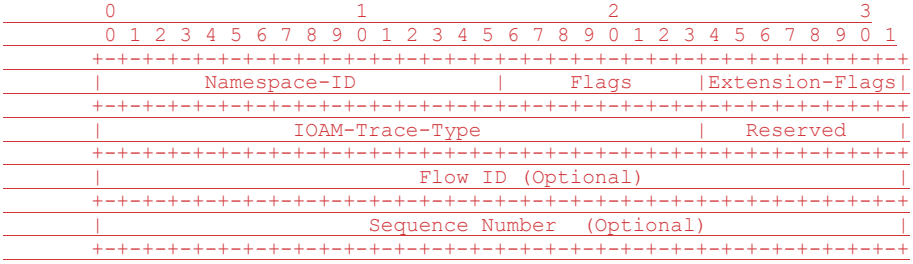

Figure 1: Hybrid Two-Step Trace IOAM Header

The HTS Trigger includes HTS IOAM Header (shown in Figure 1) consists of:

* IOAM Namespace-ID - as defined in Section 5.3 of [RFC9197].

* Flags - as defined in Section 3.2 of [RFC9326].

* Extension-Flags - as defined in Section 3.2 of [RFC9326].

* IOAM-Trace-Type - as defined in Section 5.4 of [RFC9197].

* optional Flow ID (Optional) - as defined in Section 3.2 of [RFC9326].

* optional Sequence Number (Optional) - as defined in Section 3.2 of [RFC9326].

4.1.1.2. HTS Trigger in the Alternate-Marking Method

A packet in the flow to which the Alternate-Marking method, defined in [RFC9341] and [RFC9342], is applied can be used as an HTS Trigger.

The nature of the HTS Trigger is a transport network layer-specific, and its description is outside the scope of this document. The packet that includes the an HTS Trigger in this document is also referred to as the trigger packet.

Commenté [MB29]: We may say that it includes user data.

Commenté [MB30]: At first read, this seems weird to have an example here.

I would add a preamble text that this version of the document specifies the trigger/follow-up with ioam as main implementation, but the considerations discussed here can be applicable to other contexts.

a mis en forme : Police :Gras

a mis en forme : Surlignage

Commenté [MB31]: Why is that mentioned here?

4.1.3. HTS Follow-up Packets

The HTS method uses ~~the~~ HTS Follow-up packets~~,~~ (referred to as the follow-up packet~~,~~) to collect measurement and network operational state data from ~~the~~ network ~~nodes~~elements. The node that creates ~~the~~ an HTS Trigger also generates the relevant HTS Follow-up packet. In some use cases~~,~~ (e.g., when HTS is used to collect the telemetry, including performance metrics, calculated based on a series of measurements), ~~an HTS~~a follow-up packet can be originated without using the HTS Trigger. ~~The~~A follow-up packet contains characteristic information sufficient for participating HTS nodes to associate it with the monitored data flow.

The characteristic information can be obtained using the information of the trigger packet or constructed by a node that originates the follow-up packet. **As the follow-up packet is expected to traverse the same sequence of nodes, one element of the characteristic information is the information that determines the path in the data plane.** For example, in a ~~segment~~ Segment ~~routing~~ Routing (SR) domain [RFC8402], a list of ~~segment~~ Segment identifiers (SIDs) of the trigger packet is applied to the follow-up packet. ~~And i~~In the case of the ~~service~~ Service ~~function~~ Function ~~chain~~ Chain (SFC) based on the Network Service Header (NSH) [RFC8300], the Base Header and Service Path Header of the trigger packet will be applied to the follow-up packet.

Also, when HTS is used to collect the telemetry information in an IOAM domain, the IOAM trace option header [RFC9197] of the trigger packet is applied in the follow-up packet. ~~The~~ A follow-up packet also uses the same network information used to load-balance flows in equal-cost multipath (ECMP) as the trigger packet~~,~~ (e.g., IPv6 Flow Label [RFC6437] or an entropy label [RFC6790]). The exact composition of the characteristic information is specific for each transport network, and its definition is outside the scope of this document.

Only one outstanding follow-up packet MUST be on the node for the given path. That means that if the node receives an HTS Trigger for the flow on which it still waits for the follow-up packet to the previous HTS Trigger, the node will originate the follow-up packet to transport the former set of the network operational state data and transmit it before it sends the follow-up packet with the latest collection of network operational state information.

The following sections describe the operation of HTS nodes in the downstream mode of collecting the telemetry information. In the upstream mode, the behavior of HTS nodes, in general, identical with the exception that **the HTS Trigger packet does not precede the HTS Follow-up packet.**

4.2. Operation of ~~the~~ HTS Ingress Nodes

Commenté [MB32]: How to ensure that follow-up messages are not injected by other nodes?

Commenté [MB33]: Again, rather than repeating this, I suggest that we define what we mean by telemetry in the terminology section. This main text will only refer to telemetry.

Commenté [MB34]: For which case?

Commenté [MB35]: The sentence right before says that trigger may not be used. How to reconcile both?

a mis en forme : Police :Gras

Commenté [MB36]: As several parts of a packet may change on the path, please assess whether there are implications that arer worth to call out.

Please add implications of those in the OPS consideration section.

a mis en forme : Police :Gras

Commenté [MB37]: Please move this to the Operational Considerations section.

The use of these techniques may not be sufficient to ensure path-congruence, btw.

Commenté [MB38]: That is?

Commenté [MB39]: These are operational considerations. Please move and discuss these matters there. Thanks.

Commenté [MB40]: Which node?

a mis en forme : Surlignage

Commenté [MB41]: I have troubles to understand the intended behavior.

Commenté [MB42]: Please add this as an entry in the terminology section.

Commenté [MB43]: Please add this as an entry in the terminology section.

a mis en forme : Police :Gras

Commenté [MB44]: As many SFs may be located on the same node, and these nodes may embed multiple SFs, are there implication on how telemetry is inserted?

A node that originates ~~the~~ an HTS Trigger is referred to as ~~the~~ HTS ingress node.

As stated, the ingress node originates the follow-up packet. The follow-up packet has the transport network encapsulation identical with the trigger packet followed by the HTS shim and one or more telemetry information elements encoded as Type-Length-Value (TLV). Figure 2 displays **an example of the follow-up packet format**.

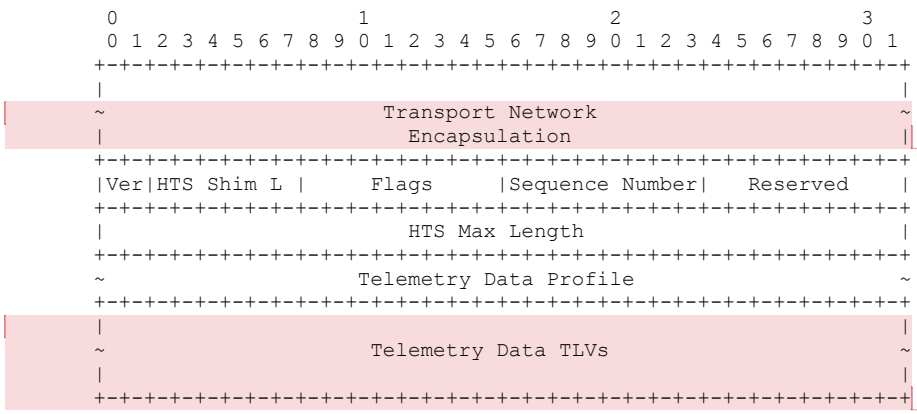

Figure 2: Follow-up Packet Format

The descriptions of the ~~Fields~~ fields of the HTS shim are as follows:

Version (Ver) is ~~the~~ a two-bit~~bits~~ long field. It specifies the version of the HTS shim format. This document defines the format for the 0b00 value of the field.

HTS Shim Length is ~~the~~ a six bits-long field. It defines the length of the HTS shim in octets. The minimal value of the field is eight octets.

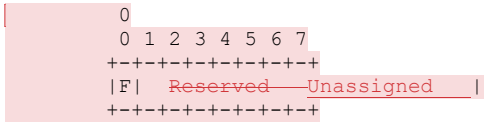

Figure 3: Flags Field Format

Flags is eight-~~bits~~ bit long. The format of the Flags field is ~~displayed~~shown in Figure 3.

- Full (F) flag MUST be set to zero by the node originating the
t to zero by the node originating the

HTS follow-up packet and MUST be set to one by the node that

does not add its telemetry data to avoid exceeding MTU size.

Commenté [MB45]: --How to detect/discard follow-ups generated by misbehaving nodes from within the network?

Commenté [MB46]: The node must keep track of how the trigger was generated/forwarded. I think we need some normative behavior.

a mis en forme : Surlignage

Commenté [MB47]: --If this is only an example, why the description of the corresponding fields uses normative language?

a mis en forme : Police :Gras, Surlignage

a mis en forme : Police :Gras, Surlignage

Commenté [MB48]: To be defined.

Commenté [MB49]: How the length of enclosed TLVs is determined given that profile is also variable length

Commenté [MB50]: Can we also indicate what intermediate nodes are supposed to do? Is it only forwarding packets + MUST NOT alter any part of the packets, else?

Commenté [MB51]: Which parts of the Figure 2 is referred as «HTS shim»?

Commenté [MB52]: Not sure we can define the length if we don't delimit what the shim.

Commenté [MB53]: Are those tracked in a registry?

I guess we need so.

Commenté [MB54]: I'm not sure to understand the intended behavior. Are we referring to an ingress node?

If no telemetry info is included, why is this called follow-up then?

- The node originating the follow-up packet MUST zero the Reserved field and ignore it on the receipt.

Sequence Number is one octet-long field. The zero-based value of the field reflects the place of the HTS follow-up packet in the sequence of the HTS follow-up packets that originated in response to the same HTS trigger. The ingress node MUST set the value of the field to zero.

Reserved is one octet-long field. It MUST be zeroed on transmission and ignored on receipt.

HTS Max Length is four octet-long field. The value of the HTS Max Length field indicates the maximum length of the HTS Follow-up packet in octets. An operator MUST be able to configure the HTS Max Length field's value. The value SHOULD be set equal to the path MTU.

Telemetry Data Profile is ~~the~~ an optional variable-length field of bit-size flags. Each flag indicates the requested type of telemetry data to be collected at each HTS node. The increment of the field is four bytes with a minimum length of zero. For example, IOAM-Trace-Type information defined in [RFC9197], Sequence Number and/or Flow ID (Figure 1) can be used in the Telemetry Data Profile field.

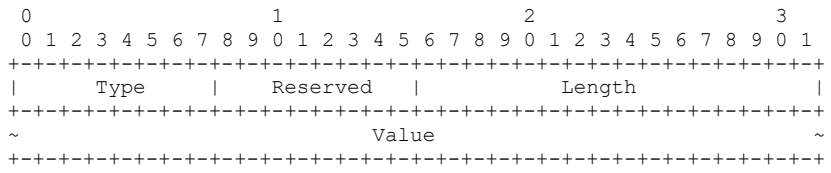

Figure 4: Telemetry Data TLV Format

Telemetry Data TLV is a variable-length field. Multiple TLVs MAY be placed in an HTS packet. Additional TLVs may be enclosed within a given TLV, subject to the semantics of the (outer) TLV in question. Figure 4 presents the format of a Telemetry Data TLV, where fields are defined as the following:

- Type - a one-octet-long field that characterizes the interpretation of the Value field.

- Reserved - one-octet-long field.

- Length - two-octet-long field equal to the length of the Value field in octets.

- Value - a variable-length field. The value of the Type field determines its interpretation and encoding. IOAM data fields, defined in [RFC9197], MAY be carried in the Value field.

All multibyte fields defined in this specification are in network byte order.

Commenté [MB55]: If the sequence number indicates a position, why it has to be set to zero?

Commenté [MB56]: I guess you meant configured the max allowed length of data that can be enclosed.

The length of the field must reflect the actual enclosed data, not the configured one.

BTW, I would move this into a manageability subsection under OPS considerations where these matters are discussed.

Commenté [MB57]: This assumes one single path.

I guess this should be set to the minimum of MTU of available paths.

Commenté [MB58]: Which flag?

Commenté [MB59]: Not sure to get this.

Do you mean that this should be a multiple of 4?

If so, I guess padding will be required in some cases.

Commenté [MB60]: Where these values are tracked?

Commenté [MB61]: I don't think the normative language makes sense here.

If we think these are useful, then it is simple to define types for those.

Commenté [MB62]: Can be moved to be cited early in the document.

4.3. Operation of ~~the~~ HTS Intermediate Nodes

Upon receiving ~~the~~ a trigger packet, ~~the~~ an HTS intermediate node MUST:

* copy the transport information;

* start the HTS Follow-up Timer for the obtained flow;

* transmit the trigger packet.

Upon receiving ~~the~~ a follow-up packet, ~~the~~ an HTS intermediate node MUST:

1. verify that ~~the~~ a matching transport information exists and the Full flag is cleared, then stop the associated HTS Follow-up Timer;

2. otherwise, transmit the received packet. Proceed to Step 8;

3. collect telemetry data requested in the Telemetry Data Profile field or defined by the local HTS policy;

4. if adding the collected telemetry would not exceed HTS Max Length field's value, then append data as a new Telemetry Data TLV and transmit the follow-up packet. Proceed to Step 8;

5. otherwise, set the value of the Full flag to one, copy the transport information from the received follow-up packet and transmit it accordingly;

6. originate the new follow-up packet using the transport information copied from the received follow-up packet. The value of the Sequence Number field in the HTS shim MUST be set to the value of the field in the received follow-up packet incremented by one;

7. copy collected telemetry data into the first Telemetry Data TLV's Value field and then transmit the packet;

8. processing completed.

If the HTS Follow-up Timer expires, ~~the~~ an intermediate node MUST:

* originate ~~the~~ a follow-up packet using transport information associated with the expired timer;

* initialize the HTS shim by setting the Version field's value to 0b00 and Sequence Number field to 0. Values of HTS Shim Length and Telemetry Data Profile fields MAY be set according to the local policy.

* copy telemetry information into Telemetry Data TLV's Value field

Commenté [MB63]: This may required some inspection to detect packet types. May be worth to be mentioned in the OPS section with potential impact on performance.

Commenté [MB64]: Of course, this assumes no validation error is detected. Otherwise, MUST transmit, etc. is problematic.

Commenté [MB65]: Provide more details about this.

At least we should indicate these are not related to info of any user packet.

Commenté [MB66]: Strat a timer for random triggers may be used to overload the node. I guess we need to put rate-limit in place to avoid exhausting these nodes resources.

Commenté [MB67]: What is the default/recommended value? I suggest this to be configurable.

Please group all configurable parameters under a manageability subsection under OPS considerations section.

Commenté [MB68]: What is the role of this timer?

Commenté [MB69]: Is it allowed to alter any part of the trigger?

Commenté [MB70]: First mention of this. Idem as similar parameters, please add a pointer to the section where this discussed.

Commenté [MB71]: Please indicate how the identity of the node that injected a telemetry data is known.

Commenté [MB72]: The transport information may change if the node is collocated with a CGN for example.

Commenté [MB73]: Previous text says that only an ingress injects these packets. I'm not sure to follow here.

and transmit the packet.

If ~~the~~ an intermediate node receives a "late" follow-up packet, i.e., a packet to which the node has no associated HTS Follow-up timer, the node MUST forward the "late" packet.

4.4. Operation of ~~the~~ HTS Egress Nodes

Upon receiving the trigger packet, ~~the~~ an HTS egress node MUST:

* copy the transport information;

* start the HTS Collection timer for the obtained flow.

When the egress node receives ~~the~~ a follow-up packet for the known flow, i.e., the flow to which the Collection timer is running, the node for each of Telemetry Data TLVs MUST:

* if HTS is used in the authenticated mode, verify the authentication of the Telemetry Data TLV using the Authentication sub-TLV (see Section 6);

* ~~copy~~ extract telemetry information from the Value field;

* restart the corresponding Collection timer.

When the Collection timer expires, the egress relays the collected telemetry information for processing and analysis to a local or remote agent.

5. Operational Considerations

Correctly attributing information originated by ~~the particular~~a trigger packet to the ~~proper~~ appropriate HTS Follow-up packet is essential for the HTS protocol. That can be achieved using characteristic information that uniquely identifies the trigger packet within a given HTS domain. For example, a combination of the flow identifier and packet's sequence number within that flow, as Flow ID and Sequence Number in IOAM Direct Export [RFC9326], can be used to correlate between stored telemetry information and the appropriate HTS Follow-up packet. In case the trigger packet doesn't include data that distinguish it from other trigger packets in the HTS domain, then for the particular flow, there MUST be no more than one HTS Trigger, values of HTS timers bounded by the rate of the trigger generation for that flow. In practice, the minimal interval between HTS Trigger packets SHOULD be selected from the range determined by the round-trip time (RTT) between HTS Ingress and HTS Egress nodes as [RTT/2, RTT].

5.1. Deploying HTS in a Multicast Network

Previous sections discussed the operation of HTS in a unicast network. Multicast services are important, and the ability to collect telemetry information is invaluable in delivering a high quality of experience. While the replication of data packets is

Commenté [MB74]: Should remind that telemetry data is stripped for the trigger, etc.?

Commenté [MB75]: How a node knows that it is behaving as egress for this trigger?

Commenté [MB76]: Same comment as for other timers

Commenté [MB77]: Why integrity protection is not covered for ingress and intermediate nodes?

Commenté [MB78]: How the collection is terminated?

Commenté [MB79]: I don't parse this.

Commenté [MB80]: Time synchronization

Commenté [MB81]: Add a subsection to cover this case

Commenté [MB82]: This is underspecified

I tend to suggest to remove the multicast discussion here.

necessary, replication of HTS
Ffollow-up packets is not. Replication
of multicast data packets down a multicast tree ~~may be~~is set based on multicast routing information or explicit information included in the special header~~,~~ as, for example, in Bit-Indexed Explicit Replication [RFC8296]. A replicating node processes the HTS packet as defined below:

* the first transmitted multicast packet MUST be followed by the received corresponding HTS packet as described in Section 4.3;

* each consecutively transmitted copy of the original multicast packet MUST be followed by the new HTS packet originated by the replicating node that acts as an intermediate HTS node when the HTS Follow-up timer expired.

As a result, there are no duplicate copies of Telemetry Data TLV for the same pair of ingress and egress interfaces. At the same time, all ingress/egress pairs traversed by ~~the~~ a given multicast packet reflected in their respective Telemetry Data TLV. Consequently, a centralized controller would reconstruct and analyze the state of the particular multicast distribution tree based on HTS packets collected from egress nodes.

5.X MTU Considerations

6. Authentication in HTS

Telemetry information may be used to drive network operation, closing the control loop for self-driving, self-healing networks. Thus, it is critical to provide a mechanism to protect the telemetry information collected using the HTS method. This document defines an optional authentication of a Telemetry Data TLV that protects the collected information's integrity.

The format of the Authentication sub-TLV is displayed in Figure 5.

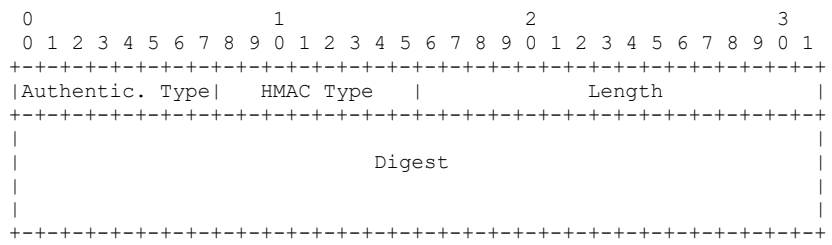

Figure 5: HMAC sub-TLV

where fields are defined as follows:

* Authentication Type - is a one-octet-long field, value 1 is allocated by IANA (Section 7.2).

* Length - two-octet-long field, set equal to the length of the Digest field in octets.

Commenté [MB83]: Both trigger and follow-up?

Commenté [MB84]: Why?

Commenté [MB85]: What does this mean?

Commenté [MB86]: I really don't understand this behavior

Commenté [MB87]: Please consider adding a discussion about MTU.

Commenté [MB88]: Move this one to be listed earlier in the document. This should before describing the various nodes behavior.

a mis en forme : Surlignage

Commenté [MB89]: Shouldn't we also provided integrity protection of the trigger packet? Also, control data of the follow-up?

* HMAC Type - is a one-octet-long field that identifies the type of the HMAC and the length of the digest and the length of the digest according to the HTS HMAC Type sub-registry (~~see~~ Section 7.4).

* Digest - is a variable-length field that carries HMAC digest of the text that includes the encompassing TLV.

This specification defines the use of HMAC-SHA-256 truncated to 128 bits ([RFC4868]) in HTS. Future specifications may define the use in HTS of more advanced cryptographic algorithms or the use of digest of a different length. HMAC is calculated as defined in [RFC2104] over text as the concatenation of the Sequence Number field of the follow-up packet (see Figure 2) and the preceding data collected in the Telemetry Data TLV. The digest then MUST be truncated to 128 bits and written into the Digest field. Distribution and management of shared keys are outside the scope of this document. In the HTS authenticated mode, the Authentication sub-TLV MUST be present in each Telemetry Data TLV. HMAC MUST be verified before using any data in the included Telemetry Data TLV. If HMAC verification fails, the system MUST stop processing corresponding Telemetry Data TLV and notify an operator. Specification of the notification mechanism is outside the scope of this document.

7. IANA Considerations

7.1. IOAM Option-Type for HTS

~~The IOAM Option-Type registry is requested in [RFC9197].~~ IANA is requested to allocate a new code point for the "IOAM Option-Type" registry as listed in Table 1.

Value	Name	Description	Reference
TBA1	IOAM Hybrid Two-Step (HTS) Option-Type	HTS Exporting	This document

Table 1: IOAM Option-Type for HTS

7.2. HTS TLV Registry Group

IANA is requested to create "Hybrid Two-Step" registry group.

IANA is requested to create the "HTS TLV Type" registry ~~in~~ under the "Hybrid Two-Step" registry group. All code points in the range 1 through 175 in this registry shall be allocated according to the "IETF Review" policy (Section 4.5 of ~~procedure specified in~~ [RFC8126]). Code points in the range 176 through 239 in this registry shall be allocated according to the "First Come First Served" ~~procedure specified in~~policy (Section 4.4 of [RFC8126]). The remaining code points are allocated according to Table 2:

Commenté [MB90]: We need a section for expert guidance.

Value	Description	Reference
0	Reserved	This document
1- 175	Unassigned	~~This document~~
176 - 239	Unassigned	~~This document~~
240 - 251	Experimental	This document
252 - 254	Private Use	This document
255	Reserved	This document

Table 2: HTS TLV Type Registry

7.3. HTS Sub-TLV Type ~~Sub-r~~Registry

IANA is requested to create the "HTS sub-TLV Type" ~~sub-~~registry as part of the "HTS TLV Type" registry. All code points in the range 1 through 175 in this registry shall be allocated according to the "IETF Review" ~~procedure~~ policy (Section 4.5 of ~~specified in~~ [RFC8126]). Code points in the range 176 through 239 in this registry shall be allocated according to the "First Come First Served" ~~procedure~~ policy (Section 4.4 of~~specified in~~ [RFC8126]). The remaining code points are allocated according to Table 3:

Commenté [MB91]: We need DE guidance.

Value	Description	TLV Used	Reference
0	Reserved	None	This document
1	HMAC	Any	This document
2 - 175	Unassigned		~~This document~~
176 - 239	Unassigned		~~This document~~
240 - 251	Experimental		This document
252 - 254	Private Use		This document
255	Reserved	None	This document

Table 3: HTS Sub-TLV Type Sub-registry

7.4. HMAC Type Sub-registry

IANA is requested to create the HMAC Type sub-registry as part of the HTS TLV Type registry. All code points in the range 1 through 127 in this registry shall be allocated according to the "IETF Review" procedure specified in [RFC8126]. Code points in the range 128

through 239 in this registry shall be allocated according to the "First Come First Served" procedure specified in [RFC8126]. The remaining code points are allocated according to Table 4:

Value	Description	Reference
0	Reserved	This document
1	HMAC-SHA-256 16 octets long	This document
2 - 127	Unassigned	~~This document~~
128 - 239	Unassigned	~~This document~~
240 - 249	Experimental	This document
250 - 254	Private Use	This document
255	Reserved	This document

Table 4: HMAC Type Sub-registry

8. Security Considerations

Nodes that practice the HTS method are presumed to share a trust model that depends on the existence of a trusted relationship among nodes. This is necessary as these nodes are expected to correctly modify the specific content of the data in the follow-up packet, and the degree to which HTS measurement is useful for network operation depends on this ability. In practice, this means either confidentiality or integrity protection cannot cover those portions of messages that contain the network operational state data. Though there are methods that make it possible in theory to provide either or both such protections and still allow for intermediate nodes to make detectable yet authenticated modifications, such methods do not seem practical at present, particularly for protocols that used to measure latency and/or jitter.

This document defines the use of authentication (Section 6) to protect the integrity of the telemetry information collected using the HTS method. Privacy protection can be achieved by, for example, sharing the IPsec tunnel with a data flow that generates information that is collected using HTS.

While it is possible for a supposed compromised node to intercept and modify the network operational state information in the follow-up packet; this is an issue that exists for nodes in general - for all data that to be carried over the particular networking technology - and is therefore the basis for an additional presumed trust model associated with an existing network.

9. Acknowledgments

Authors express their gratitude and appreciation to Joel Halpern for the most helpful and insightful discussion on the applicability of HTS in a Service Function Chaining domain. Also, the authors thank

Commenté [MB92]: Similar as previous sections

Commenté [MB93]: •IOAM security considerations should be quoted here;
•Insist that deploying this mechanism also assumes that operators are deploying mechanism to monitor and detect misbehaving nodes. Refer to rfc8300 for an example.
•Discuss how follow-ups are not leaked to unauthorized entities/customers.

Commenté [MB94]: Discuss measures to avoid exhausting node resources, detect and prevent replay.

Bjørn Ivar Teigen for the discussion about ensuring proper correlation between generated telemetry information and an HTS Follow-up packet. And a special thank you to Xiao Min for thorough review and thoughtful suggestions that helped in improving the document.

10. References

10.1. Normative References

[RFC2104] Krawczyk, H., Bellare, M., and R. Canetti, "HMAC: Keyed-Hashing for Message Authentication", RFC 2104, DOI 10.17487/RFC2104, February 1997, <https://www.rfc-editor.org/info/rfc2104>.

[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>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[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>.

[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>.

[RFC9326] Song, H., Gafni, B., Brockners, F., Bhandari, S., and T. Mizrahi, "In Situ Operations, Administration, and Maintenance (IOAM) Direct Exporting", RFC 9326, DOI 10.17487/RFC9326, November 2022, <https://www.rfc-editor.org/info/rfc9326>.

10.2. Informative References

[I-D.ietf-nmop-terminology] Davis, N., Farrel, A., Graf, T., Wu, Q., and C. Yu, "Some Key Terms for Network Fault and Problem Management", Work in Progress, Internet-Draft, draft-ietf-nmop-terminology-23, 18 August 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-nmop-terminology-23>.

[P4.INT] "In-band Network Telemetry (INT)", P4.org Specification, November 2020.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, DOI 10.17487/RFC1191, November 1990, <https://www.rfc-editor.org/info/rfc1191>.

[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>.

[RFC2992] Hopps, C., "Analysis of an Equal-Cost Multi-Path Algorithm", RFC 2992, DOI 10.17487/RFC2992, November 2000, <https://www.rfc-editor.org/info/rfc2992>.

[RFC4868] Kelly, S. and S. Frankel, "Using HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 with IPsec", RFC 4868, DOI 10.17487/RFC4868, May 2007, <https://www.rfc-editor.org/info/rfc4868>.

[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>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <https://www.rfc-editor.org/info/rfc6790>.

[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>.

[RFC8169] Mirsky, G., Ruffini, S., Gray, E., Drake, J., Bryant, S., and A. Vainshtein, "Residence Time Measurement in MPLS Networks", RFC 8169, DOI 10.17487/RFC8169, May 2017, <https://www.rfc-editor.org/info/rfc8169>.

[RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January 2018, <https://www.rfc-editor.org/info/rfc8296>.

[RFC8300] Quinn, P., Ed., Elzur, U., Ed., and C. Pignataro, Ed., "Network Service Header (NSH)", RFC 8300, DOI 10.17487/RFC8300, January 2018, <https://www.rfc-editor.org/info/rfc8300>.

[RFC8342] Bjorklund, M., Schoenwaelder, J., Shafer, P., Watsen, K., and R. Wilton, "Network Management Datastore Architecture (NMDA)", RFC 8342, DOI 10.17487/RFC8342, March 2018, <https://www.rfc-editor.org/info/rfc8342>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

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

[RFC9232] Song, H., Qin, F., Martinez-Julia, P., Ciavaglia, L., and A. Wang, "Network Telemetry Framework", RFC 9232,

DOI 10.17487/RFC9232, May 2022,
<https://www.rfc-editor.org/info/rfc9232>.

[RFC9341] Fioccola, G., Ed., Cociglio, M., Mirsky, G., Mizrahi, T., and T. Zhou, "Alternate-Marking Method", RFC 9341, DOI 10.17487/RFC9341, December 2022, <https://www.rfc-editor.org/info/rfc9341>.

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

[RFC9450] Bernardos, CJ., Ed., Papadopoulos, G., Thubert, P., and F. Theoleyre, "Reliable and Available Wireless (RAW) Use Cases", RFC 9450, DOI 10.17487/RFC9450, August 2023, <https://www.rfc-editor.org/info/rfc9450>.

Authors' Addresses

Greg Mirsky
Ericsson
Email: gregimirsky@gmail.com

Wang Lingqiang
ZTE Corporation
No 19 ,East Huayuan Road
Beijing
100191
China
Phone: +86 10 82963945
Email: wang.lingqiang@zte.com.cn

Guo Zhui
ZTE Corporation
No 19 ,East Huayuan Road
Beijing
100191
China
Phone: +86 10 82963945
Email: guo.zhui@zte.com.cn

Haoyu Song
Futurewei Technologies
2330 Central Expressway
Santa Clara,
United States of America
Email: hsong@futurewei.com

Pascal Thubert
Independent
06330 Roquefort-les-Pins
France

Email: pascal.thubert@gmail.com