MPLS Working Group G. Mirsky
Internet-Draft S. Ruffini
Intended status: Standards Track E. Gray
Expires: September 10, 2015 Ericsson
J. Drake
Juniper Networks
S. Bryant
Cisco Systems
A. Vainshtein
ECI Telecom
March 9, 2015

Residence Time Measurement in MPLS network
draft-mirsky-mpls-residence-time-05

Abstract

This document specifies G-ACh based Residence Time Measurement and how it can be used by time synchronization protocols being transported over MPLS domain.

Table of Contents

1. Introduction

Time synchronization protocols, Network Time Protocol version 4 (NTPv4) [RFC5905] and Precision Time Protocol (PTP) Version 2 [IEEE.1588.2008] can be used to synchronize clocks across network domain. In some scenarios calculation of time packet of time

synchronization protocol spends within a node, called Residence Time, can improve accuracy of clock synchronization. This document defines new Generalized Associated Channel (G-ACh) that can be used in Multi-Protocol Label Switching (MPLS) network to measure Residence Time over Label Switched Path (LSP). Mechanisms for transport of time synchronization protocol packets over MPLS are out of scope in this document.

1.1. Conventions used in this document

1.1.1. Terminology

MPLS: Multi-Protocol Label Switching

ACH: Associated Channel

TTL: Time-to-Live

G-ACh: Generic Associated Channel

GAL: Generic Associated Channel Label

NTP: Network Time Protocol

ppm: parts per million

PTP: Precision Time Protocol

LSP: Label Switched Path

LSR: Label Switching Router

OAM: Operations, Administration, and Maintenance

RSO: RTM Set Object

RTM: Residence Time Measurement

IGP: Internal Gateway Protocol

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 [RFC2119].

2. Residence Time Measurement

Packet Loss and Delay Measurement for MPLS Networks [RFC6374] can be used to measure one-way or two-way end-to-end propagation delay over LSP or PW. But these metrics are insufficient for use in some applications, for example, time synchronization across a network as defined in the Precision Time Protocol (PTP). PTPv2 [IEEE.1588.2008] uses "residence time", the time it takes for a PTPv2 event packet to transit a node. Residence times are accumulated in the correctionField of the PTP event messages or of the associated follow-up messages (or Delay_Resp message associated with the Delay_Req message) in case of two-step clocks (detailed discussion in Section 7). The residence time values are specific to each output PTP port and message.

This accumulated residence time MAY then be applied to correct the propagated time for node delays, effectively making these nodes transparent.

This document proposes mechanism to accumulate packet residence time from all LSRs that support the mechanism across a particular LSP.

3. G-ACh for Residence Time Measurement

RFC 5586 [RFC5586] and RFC 6423 [RFC6423] extended applicability of PW Associated Channel (ACH) [RFC5085] to LSPs. G-ACh provides a mechanism to transport OAM and other control messages. Processing by arbitrary transit LSRs can be triggered through controlled use of the Time-to-Live (TTL) value. In a way that is analogous to PTP operations, the packet residence time can be handled by the RTM capable node either as "one-step clock" or as a "two-step clock".

The packet format for Residence Time Measurement (RTM) is presented in Figure 1

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 0 0 1|Version|   Reserved    |           RTM Channel         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                          Scratch Pad                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type              |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              Value                            |
~                                                               ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: G-ACh packet format for Residence Time Measurement

- The Version field is set to 0, as defined in RFC 4385 [RFC4385].

- The Reserved field MUST be set to 0 on transmit and ignored on receipt.

- The RTM G-ACh field, value to be allocated by IANA, identifies the packet as such.

- The Scratch Pad field is 8 octets in length. The first RTM-capable LSR MUST initialize the Scratch Pad field, it SHOULD set it to zero value. The Scratch Pad is used to accumulate the residence time spent in each RTM capable LSR transited by the packet on its path from ingress LSR to egress LSR. Its format is IEEE double precision and its units are nanoseconds. Note: depending on one-step or two-step operation (Section 7), the residence time might be related to the same packet carried in the Value field or to a packet carried in a different RTM packet.

- The Type field identifies the type of Value that the TLV carries. IANA will be asked to create a sub-registry in Generic Associated Channel (G-ACh) Parameters Registry called "MPLS RTM TLV Registry".

- The Length field contains the number of octets of the Value field.

- The optional Value field may be used to carry a packet of a given time synchronization protocol. If packet data is carried in the RTM message, then this is identified by Type accordingly. The data MAY be NTP [RFC5905] or PTP [IEEE.1588.2008]. It is important to note that the packet may be authenticated or encrypted and carried over MPLS LSP edge to edge unchanged while

residence time being accumulated in the Scratch Pad field. Sub-TLVs MAY be included in the Value field.

- o The TLV MUST be included in the RTM message, even if the length of the Value field is zero.

3.1. PTP Packet Sub-TLV

Figure 2 presents format of a PTP sub-TLV that MUST be precede every PTP packet carried in RTM TLV.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type              |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Flags     | Resv  |PTPType|           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Port ID                            |
|                                                               |
|                               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               |           Sequence ID         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: PTP Sub-TLV format

where Flags field has format

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|S|  Reserved   |
+-+-+-+-+-+-+-+-+
```

Figure 3: Flags field format for Residence Time Measurement

- o The Type field identifies PTP sub-TLV defined in the Table 19 Values of messageType field in [IEEE.1588.2008].

- o The Length field of the PTP sub-TLV contains the number of octets of the Value field and MUST be 20.

- o The Flags field currently defines one bit, the S-bit, that defines whether or not the current message has been processed by a 2-step node, where the flag is cleared if the message has been handled exclusively by 1-step nodes and there is no follow-up message, and set if there has been at least one 2-step node and a follow-up message is forthcoming.

- o The PTPType indicates the type of PTP packet carried in the TLV. PTPType is the messageType field of the PTPv2 packet whose values are defined in the Table 19 [IEEE.1588.2008].

- o The 10 octets long Port ID field contains the identity of the source port. The Sequence ID is the sequence ID of the PTP message carried in the Value field of the message.

4. Control Plane Theory of Operation

The operation of RTM depends upon TTL expiry to deliver an RTM packet from one RTM capable interface to the next along the path from ingress LSR to egress LSR, which means that an LSR with RTM capable interfaces needs to be able to compute a TTL which will cause the expiry of an RTM packet at the next LSR with RTM capable interfaces.

However, because of Equal Cost Multipath, labels distributed by LDP do not necessarily instantiate a single path between a given ingress/egress LSR pair but rather MAY create a graph in which different flows will take different paths through this network. This means one doesn't necessarily know the path that RTM packets will take or even if they all take the same path. So, in an environment in which not all interfaces in an IGP domain support RTM, it is effectively impossible to use TTL expiry to deliver RTM packets and hence RTM cannot be used for LSPs instantiated using LDP, if multi-pathing is in use and not all LSRs are RTM-capable. In the special but important case of environment in which all interfaces in an IGP domain support RTM, setting the TTL to 1 will always cause the expiry of an RTM packet on the next RTM capable downstream LSR and hence in such an environment, RTM can be used for LSPs instantiated using LDP.

Also, if it is possible and desirable, multi-path forwarding may be disabled, at least for the set of packets that includes RTM.

Generally speaking, RTM is more useful for an LSP instantiated using RSVP-TE [RFC3209] because the LSP's path can be determined.

4.1. RTM Capability

Note that RTM capability of a node is with respect to the pair of interfaces that will be used to forward an RTM packet. In general, the ingress interface of this pair must be able to capture the arrival time of the packet and encode it in some way such that this information will be available to the egress interface.

The supported modes (1-step verses 2-step) of any pair of interfaces is then determined by the capability of the egress interface. In both cases, the egress interface implementation MUST be able to

determine the precise departure time of the same packet and determine from this, and the arrival time information from the corresponding ingress interface, the difference representing the residence time for the packet.

An interface with the ability to do this and update the associated correctionField in real-time (i.e. ? while the packet is being forwarded) is said to be 1-step capable.

Hence while both ingress and egress interfaces are required to support RTM, for the pair to be ?RTM-capable?, it is the egress interface that determines whether or not the node is 1-step or 2-step capable with respect to the interface-pair.

The RTM capability used in the sub-TLV shown in Figure 4 is thus associated with the egress port of the node making the advertisement, while the ability of any pair of interfaces that includes this egress interface to support any mode of RTM depends on the ability of that interface to record packet arrival time in some way that can be conveyed to and used by that egress interface.

When an IGP is used to carry the above defined RTM capability sub-TLV, the implementation MUST associate the advertisement with the interface that has the ability used to determine its supported RTM capabilities, and MUST NOT propagate this sub-TLV via any interface that does not have the associated ingress ability described in this section.

4.2. RTM Capability Sub-TLV

The format for the RTM Capabilities sub-TLV is presented in Figure 4

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type(TBA5)          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RTM |                        Reserved                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: RTM Capability sub-TLV

- Type value will be assigned by IANA from appropriate registries.

- Length MUST be set to 4.

- RTM is a three-bit long bit-map field with values defined as follows:

   * 0b001 - one-step RTM supported;

   * 0b010 - two-step RTM supported;

   * 0b100 - reserved.

- Reserved field must be set to all zeroes on transmit and ignored on receipt.

[RFC4202] explains that ?the Interface Switching Capability Descriptor describes switching capability of an interface. For bi-directional links, the switching capabilities of an interface are defined to be the same in either direction. I.e., for data entering the node through that interface and for data leaving the node through that interface". That principle SHOULD be applied when a node advertises RTM Capability.

A node that supports RTM MUST be able to act in two-step mode and MAY also support one-step RTM mode. Detailed discussion of one-step and two-step RTM modes in Section 7.

4.3. RTM Capability Advertisement in OSPFv2

The capability to support RTM on a particular link advertised in the OSPFv2 Extended Link Opaque LSA [I-D.ietf-ospf-prefix-link-attr] as RTM Capability sub-TLV, presented in Figure 4, of the OSPFv2 Extended Link TLV.

Type value will be assigned by IANA from the OSPF Extended Link TLV Sub-TLVs registry that will be created per [I-D.ietf-ospf-prefix-link-attr] request.

4.4. RTM Capability Advertisement in OSPFv3

The capability to support RTM on a particular link in OSPFv3 can be advertised by including an RTM Capability sub-TLV defined in Section 4.3 in the following TLVs defined in [I-D.ietf-ospf-ospfv3-lsa-extend] Intra-Area-Prefix TLV, IPv6 Link-Local Address TLV, or IPv4 Link-Local Address TLV when these are included in E-Link-LSA.

4.5. RTM Capability Advertisement in IS-IS

The RTM capability logically belongs to a group of parameters characterized as "generic information not directly related to the operation of the IS-IS protocol" [RFC6823]. Hence the capability to process RTM messages can be advertised by including RTM Capability sub-TLV in GENINFO TLV [RFC6823].

With respect to the Flags field of the GENINFO TLV:

- o The S bit MUST be cleared to prevent the RTM Capability sub-TLV from leaking between levels.

- o The D bit of the Flags field MUST be cleared as well.

- o The I bit and the V bit MUST be set accordingly depending on whether RTM capability being advertised for IPv4 or IPv6 interface of the node.

Application ID (TBA6) will be assigned from the Application Identifiers for TLV 251 IANA registry. The RTM Capability sub-TLV, presented in Figure 4, MUST be included in GENINFO TLV in Application Specific Information.

4.6. RSVP-TE Control Plane Operation to Support RTM

Though RTM capability is per interface throughout this document we will refer to an LSR as RTM capable LSR when:

- o ingress LSR's LSP interface is RTM capable;

- o transit LSR's ingress and egress interfaces for the given LSP are RTM capable;

- o egress LSR's egress interface is RTM capable.

An ingress LSR that wishes to perform RTM along a path through an MPLS network to an egress LSR verifies that the selected egress LSR has an interface that supports RTM via the egress LSR's advertisement of the RTM Capability sub-TLV. In the Path message that the ingress LSR uses to instantiate the LSP to that egress LSR it places initialized Record Route and RTM Set Objects Section 4.7, which tell the egress LSR that RTM is desired for this LSP.

In the Resv message that the egress LSR sends in response to the received Path message, it includes initialized Record Route and RTM Set Objects (RSO). The RTM Set Object contains an ordered list, from egress LSR to ingress LSR, of the RTM capable LSRs along the LSP's path. Each such LSR will use the ID of the first LSR in the RTM Set Object in conjunction with the Record Route Object to compute the hop count to its downstream LSR with reachable RTM capable interface. It will also insert its ID at the beginning of the RTM Set Object before forwarding the Resv upstream.

After the ingress LSR receives the Resv, it MAY begin sending RTM packets to the first RTM capable LSR on the LSP's path. Each RTM

packet has its Scratch Pad field initialized and its TTL set to expire on that first subsequent RTM capable LSR.

It should be noted that RTM can also be used for LSPs instantiated using [RFC3209] in an environment in which all interfaces in an IGP support RTM. In this case the RTM Set Object MAY be omitted.

4.7. RTM_SET Object

RTM capable interfaces can be recorded via RTM_SET object (RSO). The RTM Set Class is TBA7. Currently one C_Type is defined, Type TBA8 RTM Set. The RTM_SET object format presented in Figure 5

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                         Sub-objects                           ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: RTM Set object format

The contents of a RTM_SET object are a series of variable-length data items called sub-objects. The sub-objects are defined in Section 4.7.1 below.

The RSO can be present in both RSVP Path and Resv messages. If a Path message contains multiple RSOs, only the first RSO is meaningful. Subsequent RSOs SHOULD be ignored and SHOULD NOT be propagated. Similarly, if in a Resv message multiple RSOs are encountered following a FILTER_SPEC before another FILTER_SPEC is encountered, only the first RSO is meaningful. Subsequent RSOs SHOULD be ignored and SHOULD NOT be propagated.

4.7.1. RSO Sub-objects

The RTM Set object contains an ordered list, from egress LSR to ingress LSR, of the RTM capable LSRs along the LSP's path.

The contents of a RTM_SET object are a series of variable-length data items called sub-objects. Each sub-object has its own Length field. The length contains the total length of the sub-object in bytes, including the Type and Length fields. The length MUST always be a multiple of 4, and at least 8 (smallest IPv4 sub-object).

Sub-objects are organized as a last-in-first-out stack. The first sub-object relative to the beginning of RSO is considered the top.

The last sub-object is considered the bottom. When a new sub-object is added, it is always added to the top.

Three kinds of sub-objects for RSO are currently defined.

4.7.1.1. IPv4 Sub-object

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Type       |     Length    |             Flags             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       IPv4 address                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: IPv4 sub-object format

Type

0x01 IPv4 address

Length

The Length contains the total length of the sub-object in bytes, including the Type and Length fields. The Length is always 8.

IPv4 address

A 32-bit unicast host address.

Flags

TBD

4.7.1.2. IPv6 Sub-object

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Type       |     Length    |             Flags             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                         IPv6 address                          |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 7: IPv6 sub-object format

Type

0x02 IPv6 address

Length

The Length contains the total length of the sub-object in bytes, including the Type and Length fields. The Length is always 20.

IPv6 address

A 128-bit unicast host address.

Flags

TBD

4.7.1.3. Unnumbered Interface Sub-object

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |     Length    |              Flags            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Router ID                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Interface ID                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 8: IPv4 sub-object format

Type

0x03 Unnumbered interface

Length

The Length contains the total length of the sub-object in bytes, including the Type and Length fields. The Length is always 12.

Router ID

The Router ID interpreted as discussed in the Section 2 of RFC 3447 [RFC3477].

Interface ID

The identifier assigned to the link by the LSR specified by the Router ID.

Flags

TBD

5. Data Plane Theory of Operation

After instantiating an LSP for a path using RSVP-TE [RFC3209] as described in Section 4.6 or if this is the special case of homogeneous RTM-capable IP/MPLS domain discussed in the last paragraph of Section 4, ingress LSR MAY begin sending RTM packets to the first downstream RTM capable LSR on that path. Each RTM packet has its Scratch Pad field initialized and its TTL set to expire on the next downstream RTM-capable LSR. Each RTM-capable LSR on the explicit path receives an RTM packet and records the time at which it receives that packet at its ingress interface as well as the time at which it transmits that packet from its egress interface; this should be done as close to the physical layer as possible to ensure precise accuracy in time determination. The RTM-capable LSR determines the difference between those two times; for 1-step operation, this difference is determined just prior to or while sending the packet, and the RTM-capable egress interface adds it to the value in the Scratch Pad field of the message in progress. Note, for the purpose of calculating a residence time, a common free running clock synchronizing all the involved interfaces may be sufficient, as, for example, 4.6 ppm accuracy leads to 4.6 nanosecond error for residence time on the order of 1 millisecond.

For 2-step operation, the difference between packet arrival time (at an ingress interface) and subsequent departure time (from an egress interface) is determined at some later time prior to sending a subsequent follow-up message, so that this value can be used to update the correctionField in the follow-up message.

See Section 7 for further details on the difference between 1-step and 2-step operation.

The RTM capable LSR also sets the RTM packet's TTL to expire on the next downstream RTM capable LSR.

The last RTM-capable LSR on the LSP MAY then use the value in the Scratch Pad field to perform time correction, if there is no follow-up message. For example, the egress LSR may be a PTP Boundary Clock synchronized to a Master Clock and will use the value in the Scratch Pad Field to update PTP's correctionField.

6. Applicable PTP Scenarios

The proposed approach can be directly integrated in a PTP network based on delay request-response mechanism. The RTM capable LSR nodes act as end-to-end transparent clocks, and typically boundary clocks, at the edges of the MPLS network, use the value in the Scratch Pad field to update the correctionField of the corresponding PTP event packet prior to performing the usual PTP processing.

7. One-step Clock and Two-step Clock Modes

One-step mode refers to the mode of operation where an egress interface updates the correctionField value of an original event message. Two-step mode refers to the mode of operation where this update is made in a subsequent follow-up message.

Processing of the follow-up message, if present, requires the downstream end-point to wait for the arrival of the follow-up message in order to combine correctionField values from both the original (event) message and the subsequent (follow-up) message. In a similar fashion, each 2-step node needs to wait for the correct follow-up message, if there is one, in order to update that follow-up message (as opposed to creating a new one. Hence the first node that uses 2-step mode MUST do two things:

1. Mark the original event message to indicate that a follow-up message will be forthcoming (this is necessary in order to

   Let any subsequent 2-step node know that there is already a follow-up message, and

   Let the end-point know to wait for a follow-up message;

2. Create a follow-up message in which to put the RTM determined as an initial correctionField value.

IEEE 1588v2 [IEEE.1588.2008] defines this behavior for PTP messages.

Thus, for example, with reference to the PTP protocol, the PTPType field identifies whether the message is a Sync message, Follow_up message, Delay_Req message, or Delay_Resp message. The 10 octet long Port ID field contains the identity of the source port, that is, the specific PTP port of the boundary clock connected to the MPLS network. The Sequence ID is the sequence ID of the PTP message carried in the Value field of the message.

PTP messages also include a bit that indicates whether or not a follow-up message will be coming. This bit - once it is set by a

2-step mode device - must stay set accordingly until the original and follow-up message are combined by an end-point (such as a boundary clock).

Thus, an RTM packet, containing residence time information relating to an earlier packet, also contains information identifying that earlier packet.

For compatibility with PTP, RTM (when used for PTP packets) must behave in a similar fashion. To do this, a 2-step RTM capable egress interface will need to examine the S-bit in the Flags field of the PTP sub-TLV (for RTM messages that indicate they are for PTP) and - if it is clear (set to zero), it MUST set it and create a follow-up PTP Type RTM message. If the S bit is already set, then the RTM capable node MUST wait for the RTM message with the PTP type of follow-up and matching originator and sequence number to make the corresponding residence time update to the Scratch Pad field.

In practice an RTM operating according to two-step clock behaves like a two-steps transparent clock.

A 1-step capable RTM node MAY elect to operate in either 1-step mode (by making an update to the Scratch Pad field of the RTM message containing the PTP even message), or in 2-step mode (by making an update to the scratch pad of a follow-up message when its presence is indicated), but MUST NOT do both.

Two main subcases can be identified for an RTM node operating as a two-step clock:

A) If any of the previous RTM capable node or the previous PTP clock (e.g. the BC connected to the first LSR), is a two-step clock, the residence time is added to the RTM packet that has been created to include the associated PTP packet (i.e. follow-up message in the downstream direction), if the local RTM-capable LSR is also operating as a two-step clock. This RTM packet carries the related accumulated residence time and the appropriate values of the Sequence Id and Port Id (the same identifiers carried in the packet processed) and the Two-step Flag set to 1.

Note that the fact that an upstream RTM-capable node operating in the two-step mode has created a follow-up message does not require any subsequent RTM capable LSR to also operate in the 2-step mode, as long as that RTM-capable LSR forwards the follow-up message on the same LSP on which it forwards the corresponding previous message.

A one-step capable RTM node MAY elect to update the RTM follow-up message as if it were operating in two-step mode, however, it MUST NOT update both messages.

A PTP event packet (sync) is carried in the RTM packet in order for an RTM node to identify that residence time measurement must be performed on that specific packet.

To handle the residence time of the Delay request message on the upstream direction, an RTM packet must be created to carry the residence time on the associated downstream Delay Resp message.

The last RTM node of the MPLS network in addition to update the correctionField of the associated PTP packet, must also properly handle the two-step flag of the PTP packets.

B) When the PTP network connected to the MPLS and RTM node, operates in one-step clock mode, the associated RTM packet must be created by the RTM node itself. The associated RTM packet including the PTP event packet needs now to indicate that a "follow up" message will be coming.

The last RTM node of the LSP, modeif it receives an RTM message with a PTP payload indicating a follow-up message will be forthcoming, must generate a follow-up message and properly set the two-step flag of the PTP packets.

8. IANA Considerations

8.1. New RTM G-ACh

IANA is requested to reserve a new G-ACh as follows:

Value	Description	Reference
TBA1	Residence Time Measurement	This document

Table 1: New Residence Time Measurement

8.2. New RTM TLV Registry

IANA is requested to create sub-registry in Generic Associated Channel (G-ACh) Parameters Registry called "MPLS RTM TLV Registry". All code points in the range 0 through 127 in this registry shall be allocated according to the "IETF Review" procedure as specified in [RFC5226] . Remaining code points are allocated according to the

table below. This document defines the following new values RTM TLV type s:

Value	Description	Reference
0	Reserved	This document
1	No payload	This document
2	PTPv2	This document
3	NTP	This document
4-127	Reserved	IETF Consensus
128 - 191	Reserved	First Come First Served
192 - 255	Reserved	Private Use

Table 2: RTM TLV Type

8.3. New RTM Sub-TLV Registry

IANA is requested to create sub-registry in MPLS RTM TLV Registry, requested in Section 8.2, called "MPLS RTM Sub-TLV Registry". All code points in the range 0 through 127 in this registry shall be allocated according to the "IETF Review" procedure as specified in [RFC5226] . Remaining code points are allocated according to the table below. This document defines the following new values RTM sub-TLV types:

Value	Description	Reference
0	Reserved	This document
1	PTP 2-step	This document
2-127	Reserved	IETF Consensus
128 - 191	Reserved	First Come First Served
192 - 255	Reserved	Private Use

Table 3: RTM Sub-TLV Type

8.4. RTM Capability sub-TLV

IANA is requested to assign a new type for RTM Capability sub-TLV from future OSPF Extended Link TLV Sub-TLVs registry as follows:

Value	Description	Reference
TBA2	RTM Capability	This document

Table 4: RTM Capability sub-TLV

8.5. IS-IS RTM Application ID

IANA is requested to assign a new Application ID for RTM from the Application Identifiers for TLV 251 registry as follows:

Value	Description	Reference
TBA3	RTM	This document

Table 5: IS-IS RTM Application ID

8.6. RTM_SET Object RSVP Class Number, Class Type and Sub-object Types

IANA is requested to assign a new Class Number for RTM_SET object as follows:

Value	Description	Reference
TBA4	RTM_SET object	This document

Table 6: RTM_SET object Class

IANA is requested to assign a new Class Type for RTM_SET object as follows:

Value	Description	Reference
TBA5	RTM Set	This document

Table 7: RTM_SET object Class Type

IANA requested to create new sub-registry for sub-object types of RTM_SET object as follows:

Value	Description	Reference
0	Reserved	
1	IPv4 address	This document
2	IPv6 address	This document
3	Unnumbered interface	This document
4-127	Reserved	IETF Consensus
128 - 191	Reserved	First Come First Served
192 - 255	Reserved	Private Use

Table 8: RTM_SET object sub-object types

9. Security Considerations

Routers that support Residence Time Measurement are subject to the same security considerations as defined in [RFC5586] .

In addition - particularly as applied to use related to PTP - there is a presumed trust model that depends on the existence of a trusted relationship of at least all PTP-aware nodes on the path traversed by PTP messages. This is necessary as these nodes are expected to correctly modify specific content of the data in PTP messages and proper operation of the protocol depends on this ability.

As a result, the content of the PTP-related data in RTM messages that will be modified by intermediate nodes cannot be authenticated, and the additional information that must be accessible for proper operation of PTP 1-step and 2-step modes MUST be accessible to intermediate nodes (i.e. - MUST NOT be encrypted in a manner that makes this data inaccessible).

While it is possible for a supposed compromised LSR to intercept and modify the G-ACh content, this is an issue that exists for LSRs in general - for any and all data that may be carried over an LSP - and is therefore the basis for an additional presumed trust model associated with existing LSPs and LSRs.

The ability for potentially authenticating and/or encrypting RTM and PTP data that is not needed by intermediate RTM/PTP-capable nodes is for further study.

Security requirements of time protocols are provided in RFC 7384 [RFC7384].

10. Acknowledgements

TBD

11. References

11.1. Normative References

[I-D.ietf-ospf-ospfv3-lsa-extend]
Lindem, A., Mirtorabi, S., Roy, A., and F. Baker, "OSPFv3 LSA Extendibility", draft-ietf-ospf-ospfv3-lsa-extend-06 (work in progress), February 2015.

[I-D.ietf-ospf-prefix-link-attr]
Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", draft-ietf-ospf-prefix-link-attr-03 (work in progress), February 2015.

[IEEE.1588.2008]
"Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", IEEE Standard 1588, March 2008.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, February 2006.

[RFC5085] Nadeau, T. and C. Pignataro, "Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires", RFC 5085, December 2007.

[RFC5586] Bocci, M., Vigoureux, M., and S. Bryant, "MPLS Generic Associated Channel", RFC 5586, June 2009.

[RFC5905] Mills, D., Martin, J., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, June 2010.

[RFC6423] Li, H., Martini, L., He, J., and F. Huang, "Using the Generic Associated Channel Label for Pseudowire in the MPLS Transport Profile (MPLS-TP)", RFC 6423, November 2011.

[RFC6823] Ginsberg, L., Previdi, S., and M. Shand, "Advertising Generic Information in IS-IS", RFC 6823, December 2012.

11.2. Informative References

[RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, September 2011.

[RFC7384] Mizrahi, T., "Security Requirements of Time Protocols in Packet Switched Networks", RFC 7384, October 2014.

Authors' Addresses

Greg Mirsky
Ericsson

Email: gregory.mirsky@ericsson.com

Stefano Ruffini
Ericsson

Email: stefano.ruffini@ericsson.com

Eric Gray
Ericsson

Email: eric.gray@ericsson.com

John Drake
Juniper Networks

Email: jdrake@juniper.net

Stewart Bryant
Cisco Systems

Email: stbryant@cisco.com

Alexander Vainshtein
ECI Telecom

Email: Alexander.Vainshtein@ecitele.com