MPLS Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 1, 2015

G. Mirsky
S. Ruffini
Ericsson
J. Drake
Juniper Networks
S. Bryant
Cisco Systems
A. Vainshtein
ECI Telecom

June 30, 2014

Residence Time Measurment in MPLS network
draft-mirsky-mpls-residence-time-02

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.

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 http://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 January 1, 2015.

Copyright Notice

Table of Contents

1. Introduction

Time synchronization protocols, Network Time Protocol version 4 (NTPv4) [RFC5905] and Precision Time Protocol (PTP) Version 2, a.k.a. IEEE-1588 v.2, can be used to syncronized clocks across network domain. In some scenarios calculation of time packet of time syncronization protocol spends within a node, called Residence Time, can improve accuracy of clock syncronization. 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). Transport of packets of a time synchronization protocol over MPLS domain is outside of scope of 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: part per million

PTP: Precision Time Protocol

LSP: Label Switched Path

LSR: Label Switched Router

OAM: Operations, Administration, and Maintenance

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 none of these metrics is useful for time syncronization across a network. For example, PTPv2 uses "residence time", time it takes for a PTPv2 event packet to transit a node. The 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. The residence time values are specific to each output PTP port and message.

Note the delay of propagation over a link connected to a port receiving the PTP event message is handled by IEEE 1588 [IEEE.1588.2008] by means of specific messages, Pdelay_Req and Pdelay_Resp,or Delay_Req and Delay_Resp depending on the applicable delay mechanism, peer-to-peer or delay request-response mechanism respectively.

This document proposes mechanism to accumulate packet residence time from all LSRs that support the mechanism across the 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 presents mechanism to transport OAM and other control messages and trigger their processing by arbitrary transient LSRs through controlled use of Time-to-Live (TTL) value.

Packet format for Residence Time Measurement (RTM) presented in Figure 1

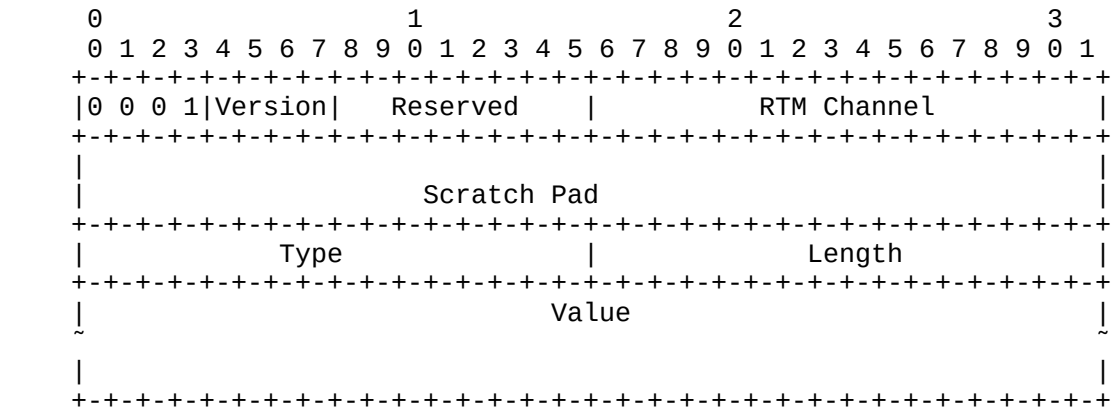

Figure 1: G-ACh packet format for Residence Time Measurement

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 and is used to accumulate the residence time spent in LSRs transited by the packet on its path from ingress LSR to egress LSR. Its format is IEEE double precision and its units are nanoseconds.

The Type field identifies type of Value that the TLV carries. IANA will be asked to create sub-registry in Generic Associated Channel (G-ACh) Parameters Registry called "MPLS RTM TLV Registry". The Length field is 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 the packet carried in the RTM message, then it accordingly identified by distinct Type, and 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. The TLV MUST be included in the RTM.

4. Control Plane Theory of Operation

A router will announce its support for RTM in a new sub-TLV, the RTM Capable TLV which will be defined in a subsequent version of this document, for the router capabilities TLV defined in RFC 4970 (OSPF) [RFC4970] and RFC 4971 (IS-IS) [RFC4971].

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

However, because of Equal Cost Multipath, labels distributed by LDP do not instantiate a single path between a given ingress/egress LSR pair but rather a graph and different flows will take different paths through this graph. This means one doesn’t 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 routers 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. In the special but important case of environment in which all routers 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.

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

4.1. RSVP-TE Control Plane Operation to Support RTM

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 supports RTM via the egress LSR’s advertisement of the RTM Capable 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 (see below) Objects, 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. The latter object will be defined in a subsequent version of this document and it 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 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 RTM capable LSR. 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 will 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 LSR.

It should be noted that RTM can also be used for LSPs instantiated using [RFC3209] in an environment in which all routers in an IGP support RTM. In this case the RTM Set Object is not used.

5. Data Plane Theory of Operation

After instantiating an LSP for a path using RSVP-TE [RFC3209] as described in Section 4.1 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 RTM capable downstream LSR on that path. Each RTM packet has its Scratch Pad field initialized and its TTL set to expire on the next downstream LSR. Each RTM capable LSR that receives an RTM packet records the time at which it receives that packet as well as the time at which it transmits that packet; this should be done as close to the physical layer as possible. Just prior to sending that packet, it takes the difference between those two times and adds it to the value in the Scratch Pad field. Note, for the purpose of calculating a residence time, a free running clock may be sufficient, as, for example, 4.6 ppm accuracy leads to 4,6 ns error for residence time in the order of 1 ms.

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

The egress LSR may then use the value in the Scratch Pad field to perform time correction. 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 Correction Field.

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.

Under certain assumptions the proposed solution in a network where peer delay mechanism is used is also possible. The solution in this case requires the definition of a specific protocol to be used to calculate the link delays according to a peer delay link measurement approach. This is not described in this version of the draft.

7. IANA Considerations

7.1. New RTM G-ACh

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

Value	Description	Reference
X	Residence Time Measurement	This document

Table 1: New Residence Time Measurement

7.2. 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 within this registry shall be allocated according to the "IETF Review" procedure as specified in [RFC5226] This document defines the following new values RTM TLV type

Value	Description	Reference
0	Reserved	This document
TBD1	No payload	This document
TBD2	PTPv2	This document
TBD3	NTP	This document

Table 2: RTM TLV Type

8. Security Considerations

Routers that support Residence Time Measurement are subject to the same security considerations as defined in [RFC5586] and [RFC6423].

9. Acknowledgements

TBD

10. References

10.1. Normative References

[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.

[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.

[RFC4970] Lindem, A., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 4970, July 2007.

[RFC4971] Vasseur, JP., Shen, N., and R. Aggarwal, "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information", RFC 4971, July 2007.

[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.

10.2. Informative References

[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.

Authors' Addresses

Greg Mirsky
Ericsson

Email: gregory.mirsky@ericsson.com

Stefano Ruffini
Ericsson

Email: stefano.ruffini@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