Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 2, 2014

M. Wasserman
Painless Security
D. Eastlake
D. Zhang
Huawei Technologies

January 31, 2014

Transparent Interconnection of Lots of Links (TRILL) over IP
draft-mrw-trill-over-ip-04.txt

Abstract

The Transparent Interconnection of Lots of Links (TRILL) protocol is implemented by devices called TRILL Switches or RBridges (Routing Bridges). TRILL supports both point-to-point and multi-access links and is designed so that a variety of link protocols can be used between TRILL switch ports. This document standardizes methods for encapsulating TRILL in IP(v4 or v6) to provide a unified TRILL campus.

Status of This Memo

Copyright Notice

Table of Contents

1. Requirements Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Introduction

TRILL switches (RBridges) are devices that implement the IETF TRILL protocol [RFC6325] [I-D.eastlake-isis-rfc6326bis] [I-D.ietf-trill-rfc6327bis].

RBridges provide transparent forwarding of frames within an arbitrary network topology, using least cost paths for unicast traffic. They support not only VLANs and Fine Grained Labels [I-D.ietf-trill-fine-labeling] but also multipathing of unicast and multi-destination traffic. They use IS-IS link state routing and encapsulation with a hop count. They are compatible with IEEE 802.1 customer bridges, and can incrementally replace them.

Ports on different RBridges can communicate with each other over various link types, such as Ethernet [RFC6325] or PPP [RFC6361].

This document defines a method for RBridges to communicate over UDP/IP(v4 or v6). TRILL over IP will allow remote, Internet-connected RBridges to form a single RBridge campus, or multiple TRILL over IP networks within a campus to be connected as a single TRILL campus via a TRILL over IP backbone.

TRILL over IP connects RBridge ports using IPv4 or IPv6 as a transport in such a way that the ports appear to TRILL to be connected by a single multi-access link. Therefore, if more than two RBridge ports are connected via a single TRILL over IP link, any pair of them can communicate.

To support the scenarios where RBridges are connected via links (such as the public Internet) that are not under the same administrative control as the TRILL campus, this document specifies the use of Datagram Transport Layer Security (DTLS) [RFC6347] to secure the communications between RBridges running TRILL over IP.

3. Use Cases for TRILL over IP

This section introduces two application scenarios (a remote office scenario and an IP backbone scenario) which cover the most typical of situations where network administrators may choose to use TRILL over an IP network.

3.1. Remote Office Scenario

In the Remote Office Scenario, a remote TRILL network is connected to a TRILL campus across a multihop IP network, such as the public Internet. The TRILL network in the remote office becomes a logical part of TRILL campus, and nodes in the remote office can be attached

to the same VLANs or Fine Grained Labels[I-D.ietf-trill-fine-labeling] as local campus nodes. In many cases, a remote office may be attached to the TRILL campus by a single pair of RBridges, one on the campus end, and the other in the remote office. In this use case, the TRILL over IP link will often cross logical and physical IP networks that do not support TRILL, and are not under the same administrative control as the TRILL campus.

3.2. IP Backbone Scenario

In the IP Backbone Scenario, TRILL over IP is used to connect a number of TRILL networks to form a single TRILL campus. For example, a TRILL over IP backbone could be used to connect multiple TRILL networks on different floors of a large building, or to connect TRILL networks in separate buildings of a multi-building site. In this use case, there may often be several TRILL switches on a single TRILL over IP link, and the IP link(s) used by TRILL over IP are typically under the same administrative control as the rest of the TRILL campus.

3.3. Important Properties of the Scenarios

There are a number of differences between the above two application scenarios, some of which drive features of this specification. These differences are especially pertinent to the security requirements of the solution, how multicast data frames are handled, and how the TRILL switch ports discover each other.

3.3.1. Security Requirements

In the IP Backbone Scenario, TRILL over IP is used between a number of RBridge ports, on a network link that is in the same administrative control as the remainder of the TRILL campus. While it is desirable in this scenario to prevent the association of rogue RBridges, this can be accomplished using existing IS-IS security mechanisms. There may be no need to protect the data traffic, beyond any protections that are already in place on the local network.

In the Remote Office Scenario, TRILL over IP may run over a network that is not under the same administrative control as the TRILL network. Nodes on the network may think that they are sending traffic locally, while that traffic is actually being sent, in a UDP/IP tunnel, over the public Internet. It is necessary in this scenario to protect the integrity and confidentiality of user traffic, as well as ensuring that no unauthorized RBridges can gain access to the RBridge campus. The issues of protecting integrity and confidentiality of user traffic are addressed by using DTLS for both IS-IS frames and data frames between RBridges in this scenario.

3.3.2. Multicast Handling

In the IP Backbone scenario, native multicast may be supported on the TRILL over IP link. If so, it can be used to send TRILL IS-IS and multicast data packets, as discussed later in this document. Alternatively, multi-destination packets can be transmitted serially.

In the Remote Office Scenario there will often be only one pair of RBridges connecting a given site and, even when multiple RBridges are used to connect a Remote Office to the TRILL campus, the intervening network may not provide reliable (or any) multicast connectivity. The issues such as complex key management also makes it difficult to provide strong data integrity and confidentiality protections for multicast traffic. For all of these reasons, the connections between local and remote RBridges will be treated like point-to-point links, and all TRILL IS-IS control messages and multicast data packets that are transmitted between the Remote Office and the TRILL campus will be serially transmitted, as discussed later in this document.

3.3.3. RBridge Neighbor Discovery

In the IP Backbone Scenario, RBridges that use TRILL over IP will use the normal TRILL IS-IS Hello mechanisms to discover the existence of other RBridges on the link [I-D.ietf-trill-rfc6327bis], and to establish authenticated communication with those RBridges.

In the Remote Office Scenario, a DTLS session will need to be established between RBridges before TRILL IS-IS traffic can be exchanged, as discussed below. In this case, one of the RBridges will need to be configured to establish a DTLS session with the other RBridge. This will typically be accomplished by configuring the RBridge at a Remote Office to initiate a DTLS session, and subsequent TRILL exchanges, with a TRILL over IP-enabled RBridge attached to the TRILL campus.

4. TRILL Packet Formats

To support the TRILL base protocol standard [RFC6325]. , two types of packets will be transmitted between RBridges: TRILL Data frames and TRILL IS-IS packets.

4.1. TRILL Data Packet

The on-the-wire form of a TRILL Data packet in transit between two neighboring RBridges is as shown below:

```
      +-------------+----------+--------------+----------+
      | TRILL Data  |  TRILL   | Native Frame |   Link   |
      | Link Header |  Header  |   Payload    |  Trailer |
      +-------------+----------+--------------+----------+
```

Where the Encapsulated Native Frame is similar to Ethernet frame format with a VLAN tag or Fine Grained Label [I-D.ietf-trill-fine-labeling] but with no trailing Frame Check Sequence (FCS).

4.2. TRILL IS-IS Packet

TRILL IS-IS packets are formatted on-the-wire as follows:

```
      +-------------+--------------+----------+
      | TRILL IS-IS |  TRILL IS-IS |   Link   |
      | Link Header |    Payload   |  Trailer |
      +-------------+--------------+----------+
```

The Link Header and Link Trailer in these formats depend on the specific link technology. The Link Header usually contains one or more fields that distinguish TRILL Data from TRILL IS-IS. For example, over Ethernet, the TRILL Data Link Header ends with the TRILL Ethertype while the TRILL IS-IS Link Header ends with the L2 -IS-IS Ethertype; on the other hand, over PPP, there are no Ethertypes but PPP protocol code points are included that distinguish TRILL Data from TRILL IS-IS.

In TRILL over IP, we will use UDP/IP (v4 or v6) as the link header, and the TRILL packet type will be determined based on the UDP destination port number. In TRILL over IP, no Link Trailer is specified, although one may be added when the resulting IP packets are encapsulated for transmission on a network (e.g. Ethernet).

5. Link Protocol Specifics

TRILL Data packets can be unicast to a specific RBridge or multicast to all RBridges on the link. TRILL IS-IS packets are always multicast to all other RBridge on the link (except for MTU PDUs, which may be unicast). On Ethernet links, the Ethernet multicast address All-RBridges is used for TRILL Data and All-IS-IS-RBridges for TRILL IS-IS.

To properly handle TRILL base protocol packets on a TRILL over IP link, either native multicast mode must be enabled on that link, or multicast must be simulated using serial unicast, as discussed below.

In TRILL Hello PDUs used on TRILL IP links, the IP addresses of the connected IP ports are their real SNPA (SubNetwork Point of Attachment) addresses and, for IPv6, the 16-byte IPv6 address is used; however, for easy of code re-use designed for common 48-bit SNPAs, for TRILL over IPv4, a 48-bit synthetic SNPA that looks like a unicast MAC address is constructed for use in the SNPA field of TRILL Neighbor TLVs [I-D.eastlake-isis-rfc6326bis][I-D.ietf-trill-rfc6327bis] on the link. This synthetic SNPA is as follows:

```
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  0xFE          |  0x00          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  IPv4 upper half                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  IPv4 lower half                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This synthetic SNPA/MAC address has the local (0x02) bit on in the first byte and so cannot conflict with any globally unique 48-bit Ethernet MAC. However, at the IP level, where TRILL operates on an IP link, there are only IP stations, not MAC stations, so conflict on the link with a real MAC address would be impossible in any case.

6. Port Configuration

Each RBridge physical port used for a TRILL over IP link MUST have at least one IP (v4 or v6) address. Implementations MAY allow a single physical port to operate as multiple IPv4 and/or IPv6 logical ports. Each IP address constitutes a different logical port and the RBridge with those ports MUST associate a different Port ID with each logical port.

TBD: MUST be able to configure a list of IP addresses for serial unicast. MUST be able to configure a non-standard IP multi-cast address if native multicast is being used.

7. TRILL over UDP/IP Format

The general format of a TRILL over UDP/IP packet is shown below.

```
      +----------+--------+----------------------+
      | IP       | UDP    |  TRILL               |
      | Header   | Header |  Payload             |
      +----------+--------+----------------------+
```

Where the UDP Header is as follows:

```
      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
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |    Source Port = Entropy      |      Destination Port         |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |           UDP Length          |        UDP Checksum           |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |  TRILL Payload ...
```

Source Port - see Section 10.2

Destination Port - indicates TRILL Data or IS-IS, see Section 14

UDP Length - as specified in [RFC768]

UDP Checksum - as specified in [RFC768]

The TRILL Payload starts with the TRILL Header (not including the TRILL Ethertype) for TRILL Data packets and starts with the 0x83 Intradomain Routeing Protocol Discriminator byte (thus not including the L2-IS-IS Ethertype) for TRILL IS-IS packets.

8. Handling Multicast

By default, both TRILL IS-IS packets and multi-destination TRILL Data packets are sent to an All-RBridges IPv4 or IPv6 multicast Address as appropriate (see Section 14); however, a TRILL over IP port may be configured to use serial unicast with a list of unicast addresses of other stations to which multi-destination packets are sent.

TBD

9. Use of DTLS

All RBridges that support TRILL over IP MUST implement DTLS and support the use of DTLS to secure both TRILL IS-IS and TRILL data packets. When DTLS is used to secure a TRILL over IP link, the DTLS session MUST be fully established before any TRILL IS-IS or data frames are exchanged.

RBridges that implement TRILL over IP SHOULD support the use of certificates for DTLS and, if they support certificates, MUST support the following algorithm:

- TLS_RSA_WITH_AES_128_CBC_SHA [RFC5246]

RBridges that support TRILL over IP MUST support the use of pre-shared keys for DTLS. If the communicating RBridges have IS-IS Hello authentication enabled with a pre-shared key, then, by default a key derived from that TRILL Hello pre-shared key is used for DTLS unless some other pre-shared key is configured. The following cryptographic algorithms MUST be supported for use with pre-shared keys:

- TLS_PSK_WITH_AES_128_CBC_SHA [RFC5246]

If the derived default preshared key is used, it is derived as follows:

```
HMAC-SHA256 ("TRILL IP", IS-IS-shared key )
```

In the above "|" indicates concatenation, HMAC-SHA256 is as described in [FIPS180] [RFC6234] and "TRILL IP" is the eight byte US ASCII [ASCII] string indicated.

10. Transport Considerations

10.1. Recursive Ingress

TRILL is designed to transport end station traffic to and from IEEE 802.1Q conformant end stations and IP is frequently transported over IEEE 802.3 or similar protocols supporting 802.1Q conformant end stations. Thus, an end station data frame EF might get TRILL ingressed to TRILL(EF) which was then sent on a TRILL over IP over an 802.3 link resulting in an 802.3 frame of the form 802.3(IP(TRILL(EF))). There is a risk of such a packet being re-ingressed by the same TRILL campus, due to physical or logical misconfiguration, looping round, being further re-ingressed, etc. The packet might get discarded if it got too large but if fragmentation is enabled, it would just keep getting split into fragments that would continue to loop and grow and re-fragment until the path was saturated with junk and packets were being discarded due to queue overflow. The TRILL Header TTL would provide no protection because each TRILL ingress adds a new Header and TTL.

To protect against this scenario, TRILL over IP output ports MUST be able to test whether a TRILL packet they are above to send is, in fact a TRILL ingress of a TRILL over IP over 802.3 or the like packets. That is, is it of the form TRILL(802.3(IP(TRILL(...)))? If

so, the default action of the TRILL over IP output port is to discard the packet. However, there are cases where some level of nested ingress is desired so it MUST be possible to configure the port to allow such packets.

10.2. Fat Flows

For the purpose of load balancing, it could be worthwhile to consider how to transport the TRILL packets over the Equal Cost Multiple Paths (ECMPs) existing in the IP path.

The ECMP election for the IP traffics could be based, at least for IPv4, on the quintuple of the outer IP header { Source IP, Destination IP, Source Port, Destination Port, and IP protocol }. Such tuples, however, can be exactly the same for all TRILL Data packets between two RBridge ports, even if there is a huge amount of data being sent. Therefore, in order to support ECMP, a RBridge SHOULD set the Source Port as an entropy field for ECMP decisions. This idea is also introduced in [I-D.yong-tsvwg-gre-in-udp-encap].

10.3. Congestion Considerations

TRILL can carry many different protocols as a payload. When a TRILL over IP flow carries primarily IP-based traffic, the aggregate traffic is assumed to be TCP friendly due to the congestion control mechanisms used by the payload traffic. Packet loss will trigger the necessary reduction in offered load, and no additional congestion avoidance action is necessary. When a TRILL over IP flow carries payload traffic that is not known to be TCP friendly and the flow runs across a path that could potentially become congested, additional mechanisms MUST be employed to ensure that the offered load on the TRILL link over IP is reduced appropriately during periods of congestion. This is not necessary in the case of a TRILL link over IP through an over- provisioned network, where the potential for congestion is avoided through the over-provisioning of the network.

11. MTU Considerations

In TRILL each RBridge advertises the largest LSP frame it can accept (but not less than 1,470 bytes) on any of its interfaces (at least those interfaces with adjacencies to other RBridges in the campus) in its LSP number zero through the originatingLSPBufferSize TLV [RFC6325] [I-D.eastlake-isis-rfc6326bis]. The campus minimum MTU, denoted Sz, is then established by taking the minimum of this advertised MTU for all RBridges in the campus. Links that do not meet the Sz MTU are not included in the routing topology. This

protects the operation of IS-IS from links that would be unable to accommodate some LSPs.

A method of determining originatingLSPBufferSize for an RBridge with one or more TRILL over IP portsis described in [I-D.ietf-trill-clear-correct]. However, if an IP link either can accommodate jumbo frames or is a link on which IP fragmentation is enabled and acceptable, then it is unlikely that the IP link will be a constraint on the RBridge's originatingLSPBufferSize. On the other hand, if the IP link can only handle smaller frames and fragmentation is to be avoided when possible, a TRILL over IP port might constrain the RBridge's originatingLSPBufferSize. Because TRILL sets the minimum values of Sz at 1,470 bytes, there may be links that meet the minimum MTU for the IP protocol (1,280 bytes for IPv6, theoretically 68 bytes for IPv4) on which it would be necessary to enable fragmentation for TRILL use.

The optional use of TRILL IS-IS MTU PDUs, as specified in [RFC6325] and [I-D.ietf-trill-rfc6327bis] can provide added assurance of the actual MTU of a link.

12. Middlebox Considerations

TBD

13. Security Considerations

TRILL over IP is subject to all of the security considerations for the base TRILL protocol [RFC6325]. In addition, there are specific security requirements for different TRILL deployment scenarios, as discussed in the "Use Cases for TRILL over IP" section above.

This document specifies that all RBridges that support TRILL over IP MUST implement DTLS, and makes it clear that it is both wise and good to use DTLS in all cases where a TRILL over IP link will traverse a network that is not under the same administrative control as the rest of the TRILL campus. DTLS is necessary, in these cases to protect the privacy and integrity of data traffic.

TRILL over IP is completely compatible with the use of IS-IS security, which can be used to authenticate RBridges before allowing them to join a TRILL campus. This is sufficient to protect against rogue RBridges, but is not sufficient to protect data packets that may be sent, in UDP/IP tunnels, outside of the local network, or even across the public Internet. To protect the privacy and integrity of that traffic, use DTLS.

In cases were DTLS is used, the use of IS-IS security may not be necessary, but there is nothing about this specification that would prevent using both DTLS and IS-IS security together. In cases where both types of security are enabled, by default, a key derived from the IS-IS key will be used for DTLS.

14. IANA Considerations

IANA has allocated the following destination UDP Ports for the TRILL IS-IS and Data channels:

UDP Port	Protocol
(TBD)	TRILL IS-IS Channel
(TBD)	TRILL Data Channel

IANA has allocated one IPv4 and one IPv6 multicast address, as shown below, which correspond to the All-RBridges and All-IS-IS-RBridges multicast MAC addresses that the IEEE Registration Authority has assigned for TRILL. Because the low level hardware MAC address dispatch considerations for TRILL over Ethernet do not apply to TRILL over IP, one IP multicast address for each version of IP is sufficient.

[Values recommended to IANA:]

Name	IPv4	IPv6
All-RBridges	233.252.14.0	FF0X:0:0:0:0:0:0:205

Note: when these IPv4 and IPv6 multicast addresses are used and the resulting IP frame is sent over Ethernet, the usual IP derived MAC address is used.

[Need to discuss scopes for IPv6 multicast (the "X" in the addresses) somewhere. Default to "site" scope but MUST be configurable?]

15. Acknowledgements

This document was written using the xml2rfc tool described in RFC 2629 [RFC2629].

The following people have provided useful feedback on the contents of this document: Sam Hartman, Adrian Farrel.

Some material has been derived from draft-ietf-mpls-in-udp by Xiaohu Xu, Nischal Sheth, Lucy Yong, Carlos Pignataro, and Yongbing Fan.

16. References

16.1. Normative References

[ASCII] "American National Standards Institute (formerly United States of America Standards Institute), "USA Code for Information Interchange", ANSI X3.4-1968, ANSI X3.4-1968 has been replaced by newer versions with slight modifications, but the 1968 version remains definitive for the Internet.", 1968.

[FIPS180] ""Secure Hash Standard (SHS)", United States of American, National Institute of Science and Technology, Federal Information Processing Standard (FIPS) 180-4", March 2012.

[I-D.eastlake-isis-rfc6326bis]
Eastlake, D., Senevirathne, T., Ghanwani, A., Dutt, D., and A. Banerjee, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", draft-eastlake-isis-rfc6326bis-09 (work in progress), August 2012.

[I-D.ietf-trill-clear-correct]
Eastlake, D., Zhang, M., Ghanwani, A., Manral, V., and A. Banerjee, "TRILL: Clarifications, Corrections, and Updates", draft-ietf-trill-clear-correct-06 (work in progress), July 2012.

[I-D.ietf-trill-rfc6327bis]
Eastlake, D., Perlman, R., Ghanwani, A., Yang, H., and V. Manral, "TRILL: Adjacency", draft-ietf-trill-rfc6327bis-03 (work in progress), January 2014.

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

[RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, August 2008.

[RFC6325] Perlman, R., Eastlake, D., Dutt, D., Gai, S., and A. Ghanwani, "Routing Bridges (RBridges): Base Protocol Specification", RFC 6325, July 2011.

16.2. Informative References

[I-D.ietf-trill-fine-labeling]
Eastlake, D., Zhang, M., Agarwal, P., Perlman, R., and D. Dutt, "TRILL (Transparent Interconnection of Lots of Links): Fine-Grained Labeling", draft-ietf-trill-fine-labeling-07 (work in progress), May 2013.

[I-D.yong-tsvwg-gre-in-udp-encap]
Crabbe, E., Yong, L., and K. Building, "Generic UDP Encapsulation for IP Tunneling", draft-yong-tsvwg-gre-in-udp-encap-02 (work in progress), October 2013.

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

[RFC6234] Eastlake, D. and T. Hansen, "US Secure Hash Algorithms (SHA and SHA-based HMAC and HKDF)", RFC 6234, May 2011.

[RFC6347] Rescorla, E. and N. Modadugu, "Datagram Transport Layer Security Version 1.2", RFC 6347, January 2012.

[RFC6361] Carlson, J. and D. Eastlake, "PPP Transparent Interconnection of Lots of Links (TRILL) Protocol Control Protocol", RFC 6361, August 2011.

Authors' Addresses

Margaret Wasserman
Painless Security
356 Abbott Street
North Andover, MA 01845
USA

Phone: +1 781 405-7464
Email: mrw@painless-security.com
URI: http://www.painless-security.com

Donald Eastlake
Huawei Technologies
155 Beaver Street
Milford, MA 01757
USA

Phone: +1 508 333-2270
Email: d3e3e3@gmail.com

Dacheng Zhang
Huawei Technologies
Q14, Huawei Campus
No.156 Beiqing Rd.
Beijing, Hai-Dian District 100095
P.R. China

Email: zhangdacheng@huawei.com