Independent Submission
Request for Comments: 6847
Category: Informational
ISSN: 2070-1721

D. Melman
T. Mizrahi
Marvell
D. Eastlake 3rd
Huawei

January 2013

Fibre Channel over Ethernet (FCoE) over Transparent Interconnection of Lots of Links (TRILL)

Abstract

Fibre Channel over Ethernet (FCoE) and Transparent Interconnection of Lots of Links (TRILL) are two emerging standards in the data center environment. While these two protocols are seemingly unrelated, they have a very similar behavior in the forwarding plane, as both perform hop-by-hop forwarding over Ethernet, modifying the packet's Media Access Control (MAC) addresses at each hop. This document describes an architecture for the integrated deployment of these two protocols.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This is a contribution to the RFC Series, independently of any other RFC stream. The RFC Editor has chosen to publish this document at its discretion and makes no statement about its value for implementation or deployment. Documents approved for publication by the RFC Editor are not a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6847.

Table of Contents

1. Introduction

Data center networks are rapidly evolving towards a consolidated approach, in which Ethernet is used as the common infrastructure for all types of traffic. Storage traffic was traditionally dominated by the Fibre Channel (FC) protocol suite. At the intersection between these two technologies a new technology was born, Fibre Channel over Ethernet (FCoE), in which native FC packets are encapsulated with an FCoE encapsulation over an Ethernet header. FCoE is specified in [FC-BB-5]. (A future version of FCoE is under development and is expected to be specified in a document to be referred to as FC-BB-6; however, this is a work in progress and is beyond the scope of this document.)

Traffic between two FCoE end nodes (ENodes) is forwarded through one or more FCoE Forwarders (FCFs). An FCF takes a forwarding decision based on the Fibre Channel destination ID (D_ID), and enforces security policies between ENodes, also known as zoning. Once an FCF takes a forwarding decision, it modifies the source and destination MAC addresses of the packet, to reflect the path to the next-hop FCF or ENode. An FCoE virtual link is an Ethernet link between an ENode and an FCF, or between two FCFs. An FCoE virtual link may traverse one or more Layer 2 bridges. FCFs use a routing protocol called Fabric Shortest Path First (FSPF) to find the optimal path to each

destination. An FCF typically has one or more native Fibre Channel interfaces, allowing it to communicate with native Fibre Channel devices, e.g., storage arrays.

TRILL [TRILL] is a protocol for transparent least-cost routing, where Routing Bridges (RBridges) forward traffic to their destination based on a least-cost route, using a TRILL encapsulation header. RBridges route TRILL-encapsulated packets based on the egress RBridge nickname in the TRILL header. An RBridge routes a TRILL-encapsulated packet after modifying its MAC addresses to reflect the path to the next-hop RBridge and decrementing a Hop Count field.

TRILL and FCoE bear a strong resemblance in their forwarding planes. Both protocols take a routing decision based on protocol addresses above Layer 2, and both modify the Ethernet MAC addresses on a per-hop basis. Each of the protocols uses its own routing protocol rather than using any type of bridging protocol, such as the spanning tree protocol [802.1Q] or the Shortest Path Bridging protocol [802.1Q].

FCoE and TRILL are both targeted at the data center environment, and their concurrent deployment is self-evident. This document describes an architecture for the integrated deployment of these two protocols.

2. Abbreviations

DCB	Data Center Bridging
ENode	FCoE Node such as server or storage array
EoR	End of Row
FC	Fibre Channel
FCF	FCoE Forwarder
FCoE	Fibre Channel over Ethernet
FCRB	FCF over RBridge
FIP	FCoE Initialization Protocol
FSPF	Fabric Shortest Path First
LAN	Local Area Network
RBridge	Routing Bridge

SAN Storage Area Network

ToR Top of Rack

TRILL Transparent Interconnection of Lots of Links

WAN Wide Area Network

3. FCoE over TRILL

3.1. FCoE over a TRILL Cloud

The simplest approach for running FCoE traffic over a TRILL network is presented in Figure 1. The figure illustrates a TRILL-enabled network, in which FCoE traffic is transparently forwarded over the TRILL cloud. The figure illustrates two ENodes, a Server and an FCoE Storage Array, an FCF, and a native Fibre Channel SAN connected to the FCF.

FCoE traffic between the two ENodes is sent from the first ENode over the TRILL cloud to the FCF, and then back through the TRILL cloud to the second ENode.

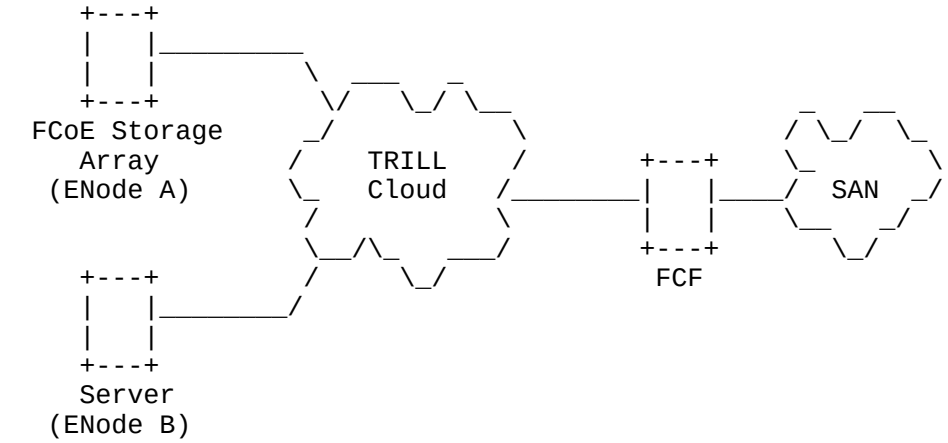

Figure 1. The "Separate Cloud" Approach

The configuration in Figure 1 separates the TRILL cloud(s) and the FCoE cloud(s). The TRILL cloud routes FCoE traffic as standard Ethernet traffic, and appears to the ENodes and FCF as an Ethernet LAN. FCoE traffic routed over the TRILL cloud includes FCoE data frames, as well as FCoE control traffic, including FCoE

Initialization Protocol (FIP) frames. To eliminate frame loss due to queue overflow, the switches in any TRILL Cloud used with FCoE would likely implement and use the relevant DCB protocols [TRILLPFC] [TRILLCN].

The main drawbacks of the Separate Cloud approach are that RBridges and FCFs are separate nodes in the network, resulting in more cabling and boxes, and that communication between ENodes usually requires traversing the TRILL cloud twice, so there are twice as many hops. As mentioned above, data center networking is converging towards a consolidated and cost-effective approach, where the same infrastructure and equipment are used for both data and storage traffic, and where high efficiency and minimal number of hops are important factors when designing the network topology.

The Separate Cloud approach is presented as background to clarify the motivation to develop an alternative approach with a higher level of integration.

3.2. FCoE over an RBridge

3.2.1. FCRB

Rather than using the Separate Cloud approach discussed in Section 3.1, an alternate approach is presented, where each switch incorporates both an FCF entity and an RBridge entity. This consolidated entity is referred to as FCoE-forwarder-over-RBridge (FCRB).

Figure 2 illustrates an FCRB and its main building blocks. An FCRB can be functionally viewed as two independent entities:

o An FCoE Forwarder (FCF) entity.

o An RBridge entity.

The FCF entity is connected to one of the ports of the RBridge, and appears to the RBridge as a native Ethernet host. A detailed description of the interaction between the layers is presented in Section 3.2.3.

Note: In this document, the term "FCF" is used slightly differently than defined in [FC-BB-5] to emphasize the concept that an FCRB is logically similar to an RBridge cascaded to an FCF. In the terminology defined in [FC-BB-5], an FCRB would be referred to as an FCF, and the FCF building block in Figure 2 would be referred to as an FC switching element.

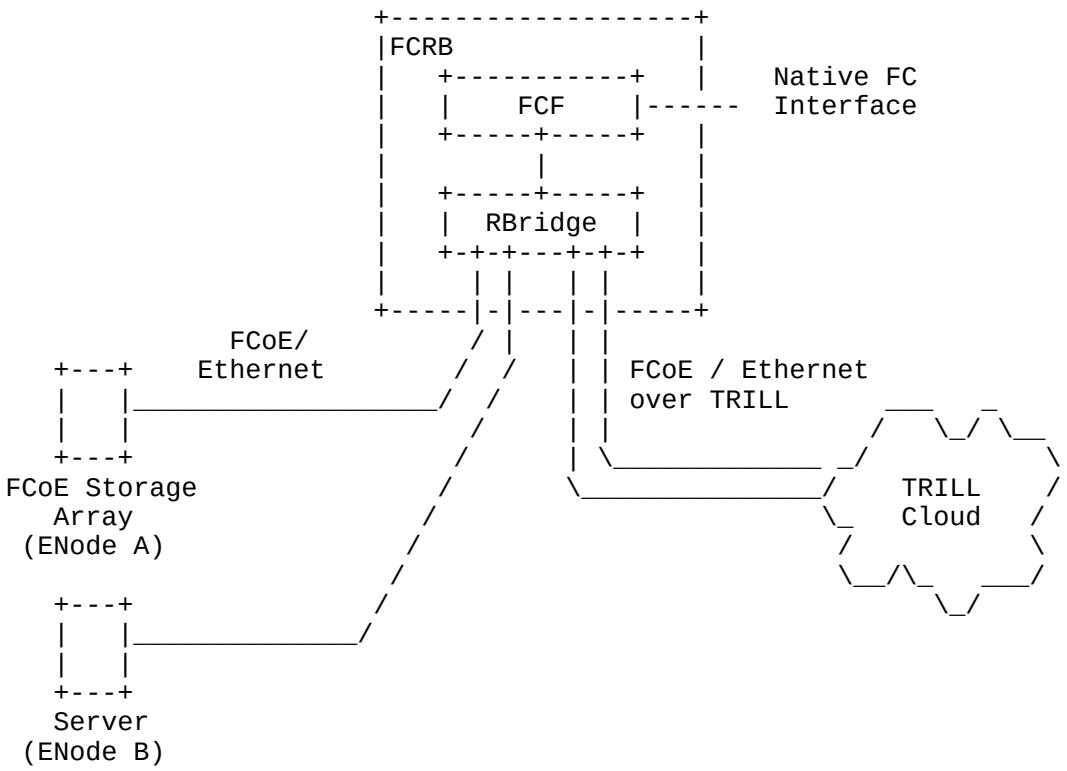

Figure 2. FCRB Entity in the Network

The FCRB entity maintains layer independence between the TRILL and FCoE protocols, while enabling both protocols on the same network.

Note that FCoE traffic is always forwarded through an FCF and cannot be forwarded directly between two ENodes. Thus, FCoE traffic between ENodes A and B in the topology in Figure 1 is forwarded through the path

ENode A-->TRILL cloud-->FCF-->TRILL cloud-->ENode B

As opposed to the topology in Figure 1, the FCF in Figure 2 is adjacent to ENodes A and B. In Figure 2, the FCRB is connected to ENodes A and B, and functions as the edge RBridge that connects these two nodes to the TRILL cloud, as well as the FCF that forwards traffic between these two nodes. Thus, traffic between ENodes A and B in the topology in Figure 2 is forwarded through the path

ENode A-->FCRB-->ENode B

Hence, the usage of FCRB entities allows TRILL and FCoE to use common infrastructure and equipment, as opposed to requiring separate infrastructure as shown in the Separate Cloud topology presented in Figure 1.

3.2.2. Topology

The network configuration illustrated in Figure 3 shows a typical topology of a data center network. Servers are hierarchically connected through Top-of-Rack (ToR) switches, also known as access switches, and each set of racks is aggregated through an End-of-Row (EoR) switch. The EoR switches are aggregated to the core switches, which may be connected to other clouds, such as an external WAN or a native FC SAN.

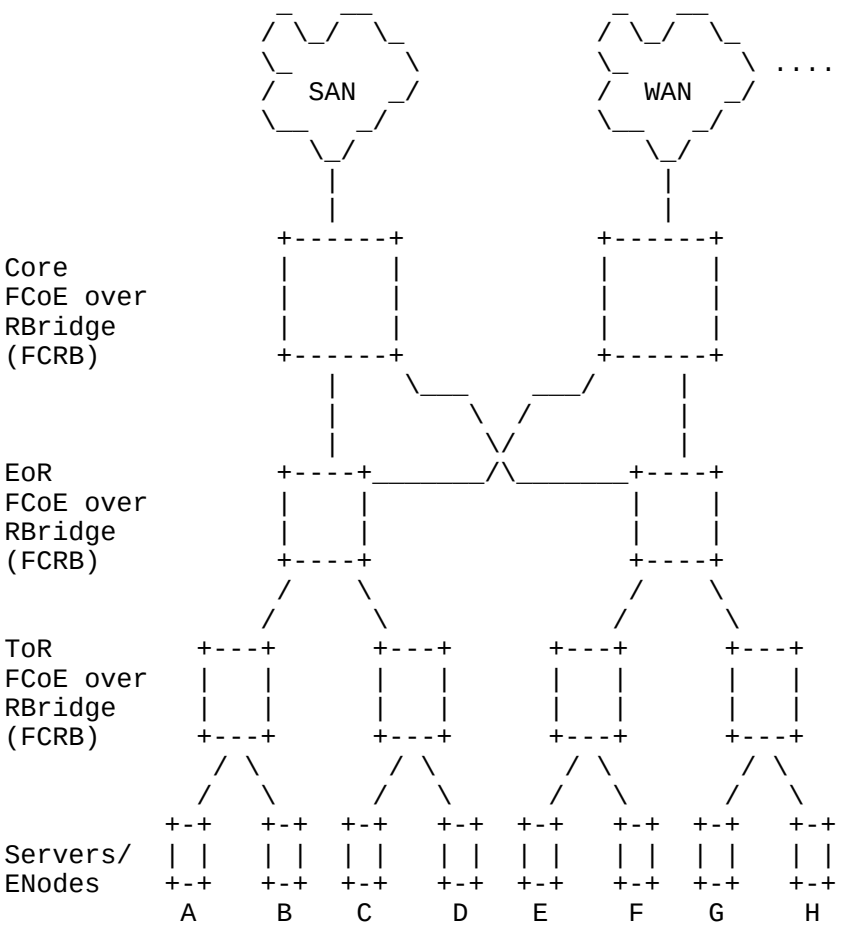

Figure 3. FCoE over RBridge Topology

Note that in the example in Figure 3, all the ToR, EoR, and core switches are FCRB entities, but it is also possible for some of the network nodes to be pure RBridges, creating a topology in which FCRBs are interconnected through TRILL clouds.

3.2.3. The FCRB Flow

3.2.3.1. Example - ENode to ENode

FCoE traffic sent between the two ENodes A and B in Figure 3 is transmitted through the ToR FCRB, since A and B are connected to the same ToR. Traffic between ENodes A and C must be forwarded through the EoR FCRB.

The FCoE jargon distinguishes between two deployment modes:

- Sparse mode: an FCoE packet sent between two FCFs may be forwarded over several hops of a Layer 2 network, allowing the underlying Layer 2 network to determine the path between the two FCFs.

- Dense mode: each node along the path between two FCFs is also an FCF, and the network is configured such that the forwarding decision at each hop is taken at the FCF layer, allowing the path between the two FCFs to be based on the FSPF routing protocol.

Figure 4 illustrates the traffic between ENodes A and C, which are not connected to the same ToR. The following two subsections describe the forwarding procedure in the Dense mode and in the Sparse mode, respectively.

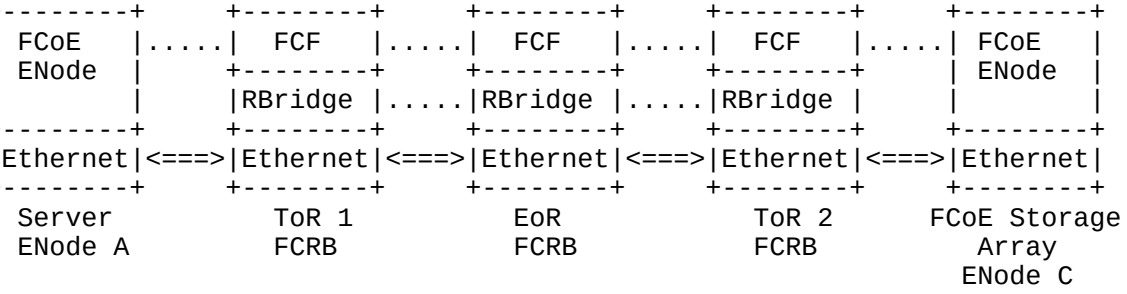

Figure 4. Traffic between two ENodes - Example

3.2.3.1.1. Forwarding from A to C in Dense Mode

- FCoE traffic from A is sent to ToR 1 over the Ethernet interface. The destination MAC address is the address of the FCF entity at ToR 1.
- ToR 1:
  - The packet is forwarded to the FCF entity at the ToR. Thus, forwarding between ENode A and the FCF at the ToR is analogous to forwarding between two Ethernet hosts.
  - The FCF entity at the ToR takes a forwarding decision based on the FC addresses. This decision is based on the FSPF routing protocol at the FCF layer. The FCF entity at the ToR forwards the packet to the FCF entity in the EoR.
  - The FCF then updates the destination MAC address of the packet to the address of the EoR FCF.
  - The packet is forwarded to the RBridge entity, where it is encapsulated in a TRILL header, and sent to the RBridge at the EoR over a single hop of the TRILL network.
- The RBridge entity in the EoR FCRB, acting as the egress RBridge, decapsulates the TRILL header and forwards the FCoE packet to the FCF entity. From this point, the forwarding process is similar to the one described above for the ToR.
- A similar forwarding process takes place at the next-hop ToR FCRB, where the FCRB finally forwards the FCoE packet to the target, ENode C.

3.2.3.1.2. Forwarding from A to C in Sparse Mode

- Traffic is forwarded to ToR 1, as described in Section 3.2.3.1.1.
- The FCF in ToR 1, based on an FSPF forwarding decision, forwards the packet to the FCF in ToR 2. The destination MAC address of the FCoE packet is updated, reflecting the FCF in ToR 2. The RBridge entity in ToR 2 adds a TRILL encapsulation, with an egress RBridge nickname representing ToR 2.
- The packet reaches the EoR. The RBridge entity in the EoR routes the packet to the RBridge entity in ToR 2.
- The packet reaches ToR 2. From this point on, the process is identical to the one described in Section 3.2.3.1.1.

3.2.3.2. Example - ENode to Native FC Node

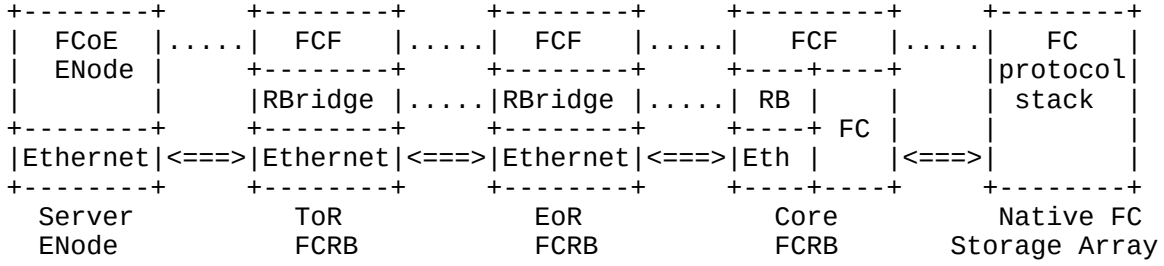

Figure 5. Example of Traffic between an ENode and a Native FC Storage Array

Figure 5 illustrates a second example, where traffic is sent between an ENode and an FC Storage Array, based on the network topology in Figure 3.

- FCoE traffic from the ENode is sent to the ToR over the Ethernet interface. The forwarding process through the ToR FCRB and through the EoR is similar to the corresponding steps in Section 3.2.3.1.

- When the packet reaches the core FCRB, the egress RBridge entity decapsulates the TRILL header and forwards the FCoE packet to the FCF entity. The packet is then forwarded as a native FC packet through the FC interface to the native FC node.

3.2.3.3. Example - ENode to ENode with Non-FCRB EoR

The example illustrated in Figure 6 is similar to the one shown in Figure 4, except that the EoR is an RBridge rather than an FCRB.

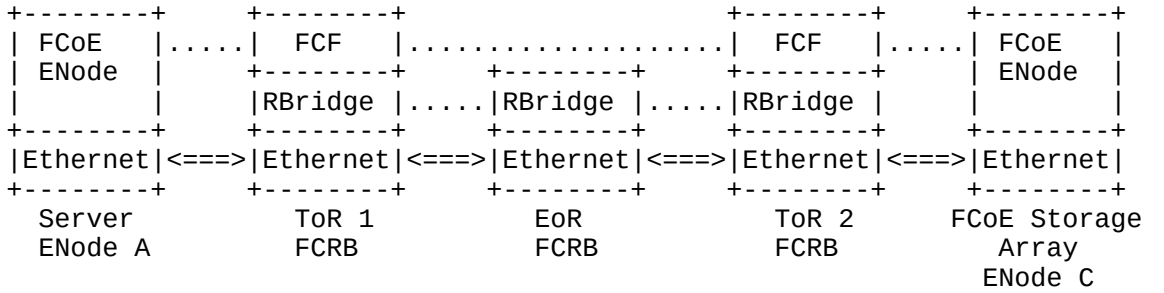

Figure 6. Example of Traffic between Two ENodes

An FCoE packet sent from ENode A to C is forwarded as follows:

- The packet is sent to the FCF in ToR 1, as in the previous example.
- The FCF in ToR 1 takes a forwarding decision based on the FC addresses and forwards the packet to the next-hop FCF, which resides in ToR 2. This forwarding decision is taken at the FCF layer and is based on the FSPF routing protocol.
- The packet is then forwarded to the RBridge entity in ToR 1, where it is encapsulated in a TRILL encapsulation, and forwarded to the RBridge at ToR 2. The packet is routed over the TRILL cloud through the RBridge at the EoR. The path through the TRILL cloud is determined by TRILL's IS-IS routing protocol.
- Once the packet reaches ToR 2, it is forwarded in a similar manner to the description in Section 3.2.3.1.

This example demonstrates that it is possible to have a hybrid network, in which some of the nodes are FCRBs and some of the nodes are RBridges. The forwarding procedure in this example is somewhat similar to the sparse-mode forwarding described in Section 3.2.3.1.2.

3.2.3.4. Example - FCoE Control Traffic through an FCRB

The previous subsections focused on the data plane, i.e., storage data exchanges transported over an FCoE encapsulation. FCoE also requires control and management traffic that is used for initializing sessions (i.e., FIP), distributing routing information (i.e., FSPF), and administering and managing fabric.

The FCoE Initialization Protocol (FIP) uses Ethernet frames with a dedicated Ethertype, allowing the FCF to distinguish these frames from other traffic. FIP uses both unicast and multicast traffic. The following example describes the forwarding scheme of a multicast FIP packet sent through the network depicted in Figure 4:

- ENode A generates a multicast frame to a multicast MAC address that represents all the FCFs (All-FCF-MAC).
- The packet is forwarded to the ToR FCRB node. The RBridge entity forwards a copy of the packet to its FCF entity, and also sends the packet through the TRILL cloud as a multicast TRILL encapsulated packet.

- o Each of the FCRBs then receives the packet, forwards a copy to its FCF entity, and forwards the packet through the TRILL network, allowing all the FCFs to receive the packet.

While FIP packets have a dedicated Ethertype and frame format, other types of FCoE control and management frames use the same FCoE encapsulation as FCoE data traffic. Thus, the forwarding scheme for such control traffic is similar to the examples described in the previous subsections, with the exception that these frames can be sent between ENodes, between FCFs, or between ENodes and FCFs.

4. Security Considerations

For general TRILL security considerations, see [TRILL].

For general FCoE security considerations, see Annex D of [FC-BB-5].

There are no additional security implications imposed by this document.

5. Acknowledgments

The authors gratefully acknowledge Ayandeh Siamack and David Black for their helpful comments. The authors also thank the T11 committee for reviewing the document, and in particular Pat Thaler and Joe White for their useful input.

6. References

6.1. Normative References

[TRILL] Perlman, R., Eastlake 3rd, D., Dutt, D., Gai, S., and A. Ghanwani, "Routing Bridges (RBridges): Base Protocol Specification", RFC 6325, July 2011.

[FC-BB-5] ANSI INCITS 462: "Information Technology - Fibre Channel - Backbone - 5 (FC-BB-5)", May 2010.

6.2. Informative References

[802.1Q] "IEEE Standard for Local and metropolitan area networks - Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks", IEEE Std 802.1Q(tm), 2012 Edition, October 2012.

[TRILLPFC] Eastlake 3rd, D., Wadekar, M., Ghanwani, A., Agarwal, P., and T. Mizrahi, "TRILL: Support of IEEE 802.1 Priority-based Flow Control and Enhanced Transmission Selection", Work in Progress, January 2013.

[TRILLCN] Eastlake 3rd, D., Wadekar, M., Ghanwani, A., Agarwal, P., and T. Mizrahi, "TRILL: Support of IEEE 802.1 Congestion Notification", Work in Progress, January 2013.

Authors' Addresses

David Melman
Marvell
6 Hamada St.
Yokneam, 20692 Israel

EMail: davidme@marvell.com

Tal Mizrahi
Marvell
6 Hamada St.
Yokneam, 20692 Israel

EMail: talmi@marvell.com

Donald Eastlake 3rd
Huawei USA R&D
155 Beaver Street
Milford, MA 01757 USA

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