

INTERNET-DRAFT  
Intended Status: Proposed Standard

Mohammed Umair  
Kingston Smiler S  
Shaji Ravindranathan  
IP Infusion  
Lucy Yong  
Donald Eastlake 3rd  
Huawei Technologies  
November 02, 2015

Expires: May 05, 2016

**Date Center Interconnect using TRILL**  
**<[draft-muks-trill-dci-00.txt](#)>**

**Abstract**

This document describes a TRILL based DCI solution using VTSD. VTSD (Virtual TRILL Service/Switch Domain) is specified in [[draft-VTSD](#)]. This draft describes the advantages provided by a TRILL based DCI solution over an existing MPLS L2VPN solution, advantages such as bandwidth scaling and providing multiple active pseudowires.

**Status of this Memo**

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at  
<http://www.ietf.org/1id-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at  
<http://www.ietf.org/shadow.html>

**Copyright and License Notice**

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

<a href="#">1</a>	Introduction . . . . .	<a href="#">3</a>
<a href="#">1.1</a>	Terminology . . . . .	<a href="#">4</a>
<a href="#">2</a>	Date Center Topology . . . . .	<a href="#">6</a>
<a href="#">2</a>	Appointed Forwarders . . . . .	<a href="#">7</a>
<a href="#">3</a>	Multiple Parallel pseudowires. . . . .	<a href="#">8</a>
<a href="#">4</a>	Active-Active Pseudowire . . . . .	<a href="#">9</a>
<a href="#">4.1</a>	Port-based AC operations. . . . .	<a href="#">10</a>
<a href="#">4.2</a>	VLAN-based AC operations. . . . .	<a href="#">10</a>
<a href="#">5</a>	MPLS encapsulation and Loop free provider PSN/MPLS . . . . .	<a href="#">10</a>
<a href="#">6</a>	Frame processing . . . . .	<a href="#">10</a>
<a href="#">6.1</a>	Frame processing between data center T2 switch and TIR. . . . .	<a href="#">10</a>
<a href="#">6.2</a>	Frame processing between TIR's . . . . .	<a href="#">11</a>
<a href="#">7</a>	MAC Address learning and withdrawal . . . . .	<a href="#">12</a>
<a href="#">8</a>	Active-Active Access with VTSD . . . . .	<a href="#">12</a>
<a href="#">9</a>	ARP/ND proxy . . . . .	<a href="#">12</a>
<a href="#">10</a>	MAC mass-withdrawal . . . . .	<a href="#">12</a>
<a href="#">11</a>	Security Considerations . . . . .	<a href="#">13</a>
<a href="#">12</a>	IANA Considerations . . . . .	<a href="#">13</a>
<a href="#">13</a>	References . . . . .	<a href="#">13</a>
<a href="#">13.1</a>	Normative References . . . . .	<a href="#">13</a>
<a href="#">10.2</a>	Informative References . . . . .	<a href="#">13</a>
	Authors' Addresses . . . . .	<a href="#">13</a>



## **1 Introduction**

Pseudo Wire Emulation Edge-to-Edge (PWE3) is a mechanism that emulates the essential attributes of a service such as Ethernet over a Packet Switched Network (PSN). The required functions of PWS include encapsulating service-specific PDUs arriving at an ingress port, and carrying them across a path or tunnel, managing their timing and order, and any other operations required to emulate the behavior and characteristics of the service as faithfully as possible.

The IETF Transparent Interconnection of Lots of Links (TRILL) protocol [[RFC6325](#)] [[RFC7177](#)] [[rfc7180bis](#)] provides transparent forwarding in multi-hop networks with arbitrary topology and link technologies using a header with a hop count and link-state routing. TRILL provides optimal pair-wise forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic. Intermediate Systems (ISs) implementing TRILL are called Routing Bridges(RBridges)or TRILL Switches.

The [[draft-VTSD](#)] introduces a new terminology called VTSD. VTSD is a logical RBridge resides inside TIR (TRILL Intermediate Router) that should be capable of performing all the operations that a standard TRILL switch can do, along with IP and MPLS functions. A TIR is a Provider Edge (PE) device where VTSD resides and provides TRILL DCI solution. VTSD is connected to the Layer2 interface towards the DC and PW interface towards the MPLS core

TRILL as a protocol enables optimal use of the links in a layer2 network and running TRILL inside the TIR or VTSD provides a way for optimally utilizing the following:

1. The PWE3 mesh connectivity in the MPLS core using parallel pseudowires.
2. The PWE3 attachment circuit interface, when there are more than one attachment circuit interfaces using active-active pseudowires.
3. Providing a RING based DCI solution along with traditional mesh / hub-spoke topology.
4. Optimally re-route the traffic from one pseudowire to another pseudowire when there is a failure. This is possible as VTSD doesn't follow split-horizon for loop free topology.

When there is a requirement to increase the bandwidth of a particular



DCI link, with TRILL DCI, new pseudowires could be created with the same endpoints. These pseudowires are termed as parallel pseudowires. As these pseudowires are attached to VTSD (which is a TRILL RBridge), the TRILL protocol takes care of optimally load sharing the traffic across these parallel pseudowires.

Similarly when there is a requirement to increase the bandwidth of customer facing interface (attachment circuit), this can be achieved effectively by adding new attachment circuit interfaces and attaching them to the same VTSD.

The objective of a pseudowire (PW) connected in parallel or mesh or ring is to maintain connectivity across the packet switched network (PSN) used by the emulated service. In this model all pseudowires that are part of a service domain will carry data traffic without making any of the pseudowire go in to standby mode.

## 1.1 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](#)].

Acronyms used in this document include the following:

AC	- Attachment Circuit [ <a href="#">RFC4664</a> ]
Access Port	- A TRILL switch port configured with the "end station service enable" bit on, as described in <a href="#">Section 4.9.1 of [RFC6325]</a> . All AC's, VTSD ports connected to CE's, should configured as TRILL Access port.
AF	- Appointed Forwarder [ <a href="#">RFC6325</a> ], [ <a href="#">RFC6439</a> ] and [RFC6439bis].
Data Label	- VLAN or FGL
ECMP	- Equal Cost Multi Pathing
FGL	- Fine-Grained Labeling [ <a href="#">RFC7172</a> ]
IS-IS	- Intermediate System to Intermediate System [ <a href="#">IS-IS</a> ]



LAN	- Local Area Network
Link	- The means by which adjacent TRILL switches or VTSD is connected. May be a bridged LAN
MLAG	- Multi-Chassis Link Aggregation
MPLS	- Multi-Protocol Label Switching
PE	- Provider Edge Device
PSN	- Packet Switched Network
PW	- Pseudowire [ <a href="#">RFC4664</a> ]
RBridge	- An alternative name for TRILL Switch
TIR	- TRILL Intermediate Router (Devices where Pseudowire starts and Terminates)
TRILL	- Transparent Interconnection of Lots of Links OR Tunneled Routing in the Link Layer
TRILL Site	- A part of a TRILL campus that contains at least one RBridge.
TRILL switch	- A device implementing the TRILL protocol. An alternative name for an RBridge.
Trunk port	- A TRILL switch port configured with the "end station service disable" bit on, as described in <a href="#">Section 4.9.1 of [RFC6325]</a> . All pseudowires should be configured as TRILL Trunk port.
VLAN	- Virtual Local Area Network
VPLS	- Virtual Private LAN Service
VPTS	- Virtual Private TRILL Service
VSI	- Virtual Service Instance [ <a href="#">RFC4664</a> ]
VTSI	- Virtual TRILL Service Instance





- VTSD                      - Virtual TRILL Switch Domain OR  
                              Virtual TRILL Service Domain  
                              A Virtual RBridge that segregates  
                              one tenant's TRILL database as well  
                              as traffic from the other.
- VTSD-AP                  - A VTSD TRILL Access port can be a  
                              AC or a logical port connected with  
                              CE's. it can be a combination of  
                              physical port and Data Label.  
                              OR just Physical port connected to  
                              CE's

## 2. Data Center Topology

The reference topology that will be used for our discussion is a 3 tier traditional topology. Although other topologies may be utilized within the data center, most of such L2 based data centers may be modeled as a 3 tier traditional topology. The reference topology is illustrated in Figure 1. To keep terminologies simple and uniform, in this document these layers will be referred to as Tier-1, Tier-2 and Tier-3 "tiers", and the switches in these layers will be termed as T1SW, T2SW etc. For simplicity reasons, the entire DC topology will not be mentioned in the further sections. Only the relevant nodes will be shown with the above mentioned node nomenclature.

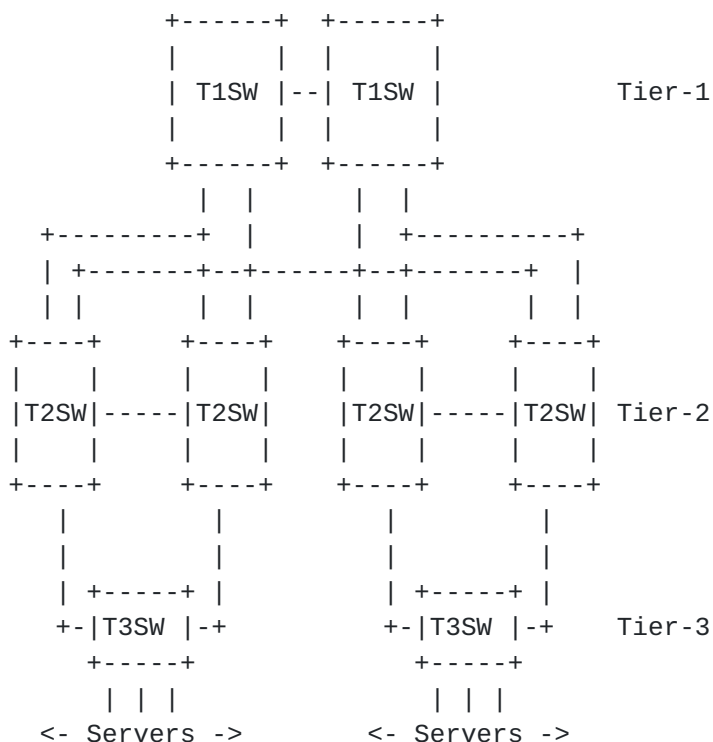




Figure 1: Typical DC network topology

## 2. Appointed Forwarders

TRILL supports multi-access LAN (Local Area Network) links that can have multiple end stations and R Bridges attached. Where multiple R Bridges are attached to a link, native traffic to and from end stations on that link is handled by a subset of those R Bridges called "Appointed Forwarders" [[rfc6439bis](#)], with the intent that native traffic in each VLAN be handled by at most one R Bridge. An R Bridge can be Appointed Forwarder for many VLANs.

The Appointed Forwarder mechanism is irrelevant to any link on which end station service is not offered. This includes links configured as point-to-point IS-IS links and any link with all R Bridge ports on that link configured as trunk ports. (In TRILL, configuration of a port as a "trunk port" just means that no end station service will be provided. It does not imply that all VLANs are enabled on that port). Furthermore, Appointed Forwarder status has no effect on the forwarding of TRILL Data frames. It only affects the handling of native frames.

By default, the DRB (Designated R Bridge) on a link is in-charge of native traffic for all VLANs on the link. The DRB may, if it wishes, act as Appointed Forwarder for any VLAN and it may appoint other R Bridges that have ports on the link as Appointed Forwarder for one or more VLANs.

The DRB may appoint other R Bridges on the link with any one of the mechanism described in [[rfc6439bis](#)].

A R Bridge on a multi-access link forms adjacency [[RFC7177](#)] with other R Bridge if the VLAN's configured/enabled between them are common. For example there are four R Bridges attached to multi-access link, say RB1, RB2, RB3 and RB4. RB1 and RB2 are configured with single VLAN "VLAN 2", whereas RB3 and RB4 are configured with "VLAN 3". Assume that there are no Native VLAN's present on any of the R Bridges connected to multi-access link. Since TRILL Hellos are sent with VLAN Tag enabled on the interface, RB3 and RB4 drops the hellos of RB1 and RB2 (since they are not configured for VLAN 2). Similarly RB1 and RB2 drops the Hellos of RB3 and RB4. This results in RB1 and RB2 not forming adjacency with RB3 and RB4. RB1 and RB2 after electing DRB and forming adjacency between them, will decide about VLAN 2 AF. Similarly RB3 and RB4 decide about the VLAN 3 AF.

As VTSD should be capable of performing all the operations a standard TRILL Switch should do, it should also be capable of performing



Appointed Forwarder selection. A group of VTSD that are configured for same service's (VLAN's in our case) on different TIR's will form adjacencies, whereas VTSD which are enabled for different VTSDI will never form adjacencies.

### 3. Multiple Parallel pseudowires.

TRILL supports multiple parallel adjacencies between neighbor RBridges. [Appendix C of \[RFC6325\]](#) and [section 3.5 of \[RFC7177\]](#) describes this in detail. Multipathing across such parallel connections can be done for unicast TRILL Data traffic on a per-flow basis, but is restricted for multi-destination traffic. VTSD should also support this functionality.

TRILL DCI Pseudowires which belong to same VTSD instance in a TIR and connected to same remote TIR are referred to as parallel pseudowires. These parallel pseudowires corresponds to a single link inside VTSD.

Here all pseudowires should be capable of carrying traffic.

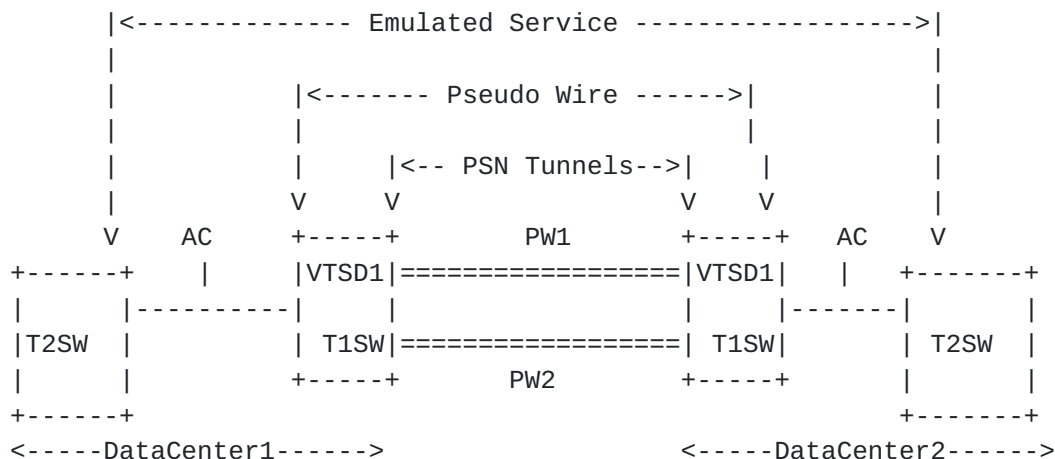


Figure 2: Parallel pseudowires with TRILL DCI

In above Figure 2, PW1 and PW2 are parallel pseudowires, as these pseudowires belongs to same VTSD and provides a connectivity across same TIRs.



This mechanism provides a way for actively increasing and optimally utilizing the bandwidth in the service provider network without affecting the existing traffic.

#### 4. Active-Active Pseudowire

[RFC6718] describes pseudowire Redundancy mechanism, wherein among the pair of pseudowires, one pseudowire will be selected as a active pseudowire and the other will be selected as a standby pseudowire. The standby pseudowire will not forward any user traffic under normal circumstances. The introduction of VTSD in TRILL DCI provides a very simple mechanism for providing multiple active pseudowires.

Pseudowires which belongs to the same VTSD instance inside the same TIR or between TIR's will be in active-active state. These pseudowires are able to carry data-traffic without making any one of pseudowire to go in standby mode.

To distribute traffic between pseudowires, TRILL protocol will be used.

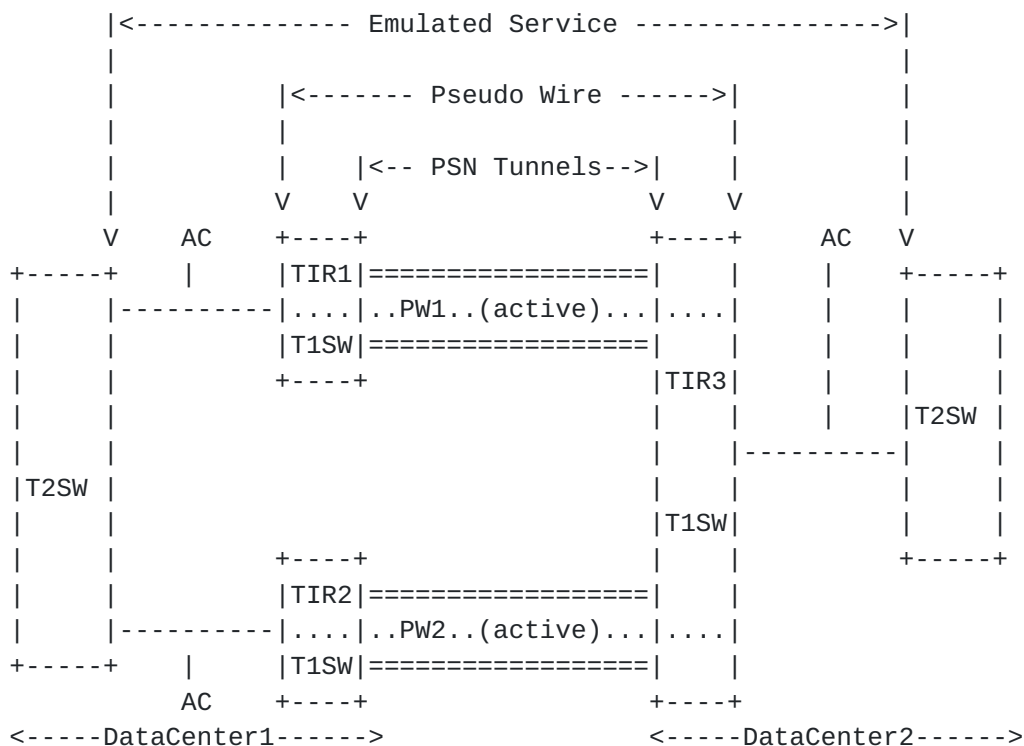


Figure 3: Dual-Home AC with Active-Active PW's





In the above Figure 3, pseudowires PW1 and PW2 are in active state and will be capable of carrying user traffic without making anyone of the pseudowire go in standby mode. The above Figure illustrates an application of multiple active pseudowires, where DC1's T2 switch (T2SW) is dual-homed with the TIR switch. This scenario is designed to actively load share the emulated service among the two TIRs attached to the multi-homed switch.

The attachment circuit can be of either Port-based Attachment Circuit or VLAN-based Attachment Circuit.

#### **4.1. Port-based AC operations.**

In this case, the VTSDs in TIR1 and TIR2 will form TRILL adjacency via AC ports. If the attachment circuit port can carry N number of end-station service VLANs, then TIR1 and TIR2's VTSDs can equally distribute them using AF Mechanism of TRILL.

#### **4.2. VLAN-based AC operations.**

Likewise in Port-based AC, in this case also the VTSDs in TIR1 and TIR2 will form TRILL adjacency via AC ports. Since only one VLAN end-station service is enabled, only one TIR's VTSD can become AF for that VLAN. Hence native traffic can be processed by any one of the AC.

### **5. MPLS encapsulation and Loop free provider PSN/MPLS**

TRILL with MPLS encapsulation over pseudowire is specified in [\[RFC7173\]](#), and requires no changes in the frame format.

TRILL DCI doesn't require to employ Split Horizon mechanism in the provider PSN network, as TRILL takes care of Loop free topology using Distribution Trees. Any multi-destination frame will traverse a distribution tree path. All distribution trees are calculated based on TRILL base protocol standard [\[RFC6325\]](#) as updated by [\[RFC7180bis\]](#).

### **6. Frame processing**

This section specifies frame processing from data center T2 switch and TIR's

#### **6.1. Frame processing between data center T2 switch and TIR.**



In a multi-homed CE topology where in a data center switch is connected to two PEs / TIRs, AF mechanism described in [section 2](#) will be used to decide which TIR/VTSD will carry the traffic for a particular VLAN. This is applicable to the case wherein the data center switch is connected to a PE/TIR device via multiple layer 2 interfaces to increase the bandwidth.

As a frame gets ingressed into a TIR (or any one of the TIR, when the tier2 switches are connected to multiple TIR's) after having AF check, the TIR encapsulates the frame with TRILL and MPLS headers and forwards the frame on a pseudowire. If parallel pseudowires are present, the TRILL protocol running in VTSD will select any one of the pseudowire and forward the TRILL Data packet. Multi-destination packets will be forwarded on Distribution tree's path [[rfc7180bis](#)]

The advantage of using TRILL for distribution of frames is, even if any of the paths or links fails between DC switch and TIR's or between TIR's, frames can be always be forwarded to any of available UP links or paths through other links/pseudowires.

If multiple equal paths are available, TRILL will distribute traffic among all the paths.

Also VTSD doesn't depend on the routing or signaling protocol that is running between TIRs, provided there is a tunnel available with proper encapsulation mechanism.

Any multi-destination frames when ingressed to TIR's will traverse one of the Distribution-Trees, with strong RFC Checks. Hop count field in TRILL Header will avoid loops or duplication of Traffic.

## **[6.2](#). Frame processing between TIR's**

When a frame gets ingressed into a VTSD inside TIR, the TRILL protocol will forward the frames to the proper pseudowire. When multiple paths / pseudowires are available between the TIR's then shortest path, calculated through TRILL protocol, will be used. If multiple paths are of equal cost, then TRILL protocol will do ECMP load spreading. If any multi-destination frame gets received by the VTSD through a pseudowire, TRILL will do an RPF check and will take proper action.

Once a frame gets to the VTSD through pseudowire, MPLS header will be de-capsulated, further action will be taken depending on the egress nickname field of TRILL header. If egress nickname is the nickname of this VTSD, MAC address table and AF lookup will be performed and the



frame will be forwarded by decapsulating the TRILL header. If egress nickname belongs to some other VTSD, frame will be forwarded on a pseudowire connected to that VTSD by encapsulating with an MPLS header.

## **7. MAC Address learning and withdrawal**

MAC address learning and withdrawal mechanism on a RBridge is specified in [section 4.8. of \[RFC6325\]](#), this document requires no changes for MAC address learning and its withdrawal.

## **8. Active-Active Access with VTSD**

TBD

## **9. ARP/ND proxy**

TBD

## **10. MAC mass-withdrawal**

TBD



## **11. Security Considerations**

TBD

## **12. IANA Considerations**

TBD

## **13. References**

### **13.1. Normative References**

[IS-IS]    "Intermediate system to Intermediate system routing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002, 2002".

[rfc7180bis] Eastlake, D., et al, "TRILL: Clarifications, Corrections, and Updates", [draft-ietf-trill-rfc7180bis](#), work in progress.,.

[[draft-VTSD](#)] Umair, M., Smiler, K., Eastlake, D., Yong, L., "TRILL Transparent Transport over MPLS" [draft-muks-trill-transport-over-mpls](#), work in progress.,.

[rfc6439bis] Eastlake, D., et al., "TRILL: Appointed Forwarders", [draft-eastlake-trill-rfc6439bis](#), work in progress.,.

### **10.2. Informative References**

Authors' Addresses

Mohammed Umair  
IP Infusion  
RMZ Centennial





Mahadevapura Post  
Bangalore - 560048 India  
EMail: mohammed.umair2@gmail.com

Kingston Smiler S  
IP Infusion  
RMZ Centennial  
Mahadevapura Post  
Bangalore - 560048 India  
EMail: kingstonsmiler@gmail.com

Shaji Ravindranathan  
IP Infusion  
3965 Freedom Circle, Suite 200  
Santa Clara, CA 95054 USA  
EMail: srnathan2014@gmail.com

Lucy Yong  
Huawei Technologies  
5340 Legacy Drive  
Plano, TX 75024  
USA  
Phone: +1-469-227-5837  
EMail: lucy.yong@huawei.com

Donald E. Eastlake 3rd  
Huawei Technologies  
155 Beaver Street  
Milford, MA 01757  
USA

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

