Provider Provisioned VPN Working Group
Internet Draft
Expiration Date: June 2002

Sunil Khandekar
Vach Kompella
Joe Regan
Nick Tingle
TiMetra Inc

Tom Soon
SBC Communications

Pascal Menezes
Terabeam Networks

Giles Heron
PacketExchange Ltd.

Marc Lassere
Riverstone Networks

Ron Haberman
Rick Wilder
Masergy Communications

Kireeti Kompella
Juniper Networks

Marty Borden
Atrica Inc

Hierarchical Virtual Private LAN Service
draft-khandekar-ppvpn-hvpls-mpls-00.txt

Status of this Memo

Abstract

This document proposes scalable extensions to the Virtual Private LAN Segment (VPLS) solution described in [VPLS], by introducing hierarchical connectivity. This document also describes support

for participation of non-bridging PE devices in a VPLS solution.

1. Placement of this Memo in Sub-IP Area

RELATED DOCUMENTS

draft-vkompella-ppvpn-vpsn-mpls-00.txt
draft-heron-ppvpn-vpsn-reqmts-00.txt
draft-lasserre-tls-mpls-00.txt

WHERE DOES THIS FIT IN THE PICTURE OF THE SUB-IP WORK

This fits in the PPVPN box.

WHY IS IT TARGETED AT THIS WG

This work fits in the PPVPN working group charter. It describes scalable extensions to a service that uses an emulation of a Layer 2 medium to create a provider provisioned virtual private network 0, specifically, a Transparent LAN service.

JUSTIFICATION

We believe the WG should consider this draft because it specifies extensions for a class of layer 2 VPN

2. Introduction

The solution described in [VPLS] requires a full mesh of tunnel LSPs between all the PE routers that participate in the VPLS service. For each VPLS service, n*(n-1) VCs must be setup between the PE routers. While this creates signaling overhead, the real detriment to large-scale deployment is the packet replication requirements for each provisioned VCs on a PE router. Hierarchical connectivity, described in this document reduces signaling and replication overhead to allow large-scale deployment.

In many cases, service providers place smaller edge devices in multi-tenant buildings and aggregate them into a PE device in a large Central Office (CO) facility. In some instances, standard IEEE 802.1q (Dot 1Q) tagging techniques may be used to facilitate mapping CE interfaces to PE VPLS access points. When this is done, a hierarchical architecture is created outside the context of VPLS; no service level signaling is present between the PE router and the MTU bridge.

To avoid issues with Spanning Tree Protocol (STP), 'VLAN' tag provisioning and non-Ethernet access networks, it is beneficial to

extend VPLS service tunneling techniques into the MTU domain. This

can be accomplished by treating the MTU device as a PE device and provision VCs between it and every other edge, as described in [VPLS]. An alternative is to utilize [MARTINI-ENCAP] VCs between the MTU and selected VPLS enabled PE routers. This document focuses on this alternative approach. The [VPLS] mesh core tier VCs (Hub) are augmented with access tier VCs (Spoke) to form a two tier hierarchical VPLS (H-VPLS).

Spoke VCs may be expanded to include any L2 tunneling mechanism, expanding the scope of the first tier to include non-bridging VPLS PE routers. The non-bridging PE router would extend a Spoke VC from a Layer-2 switch or Router that connects to it, through the service core network, to a bridging VPLS PE router supporting Hub VCs.

This document also describes support for participation of non-bridging devices (routers) in a VPLS solution.

3. Hierarchical connectivity

This section describes the hub and spoke connectivity model and describes the requirements of the bridging capable and non-bridging devices for supporting the spoke connections.

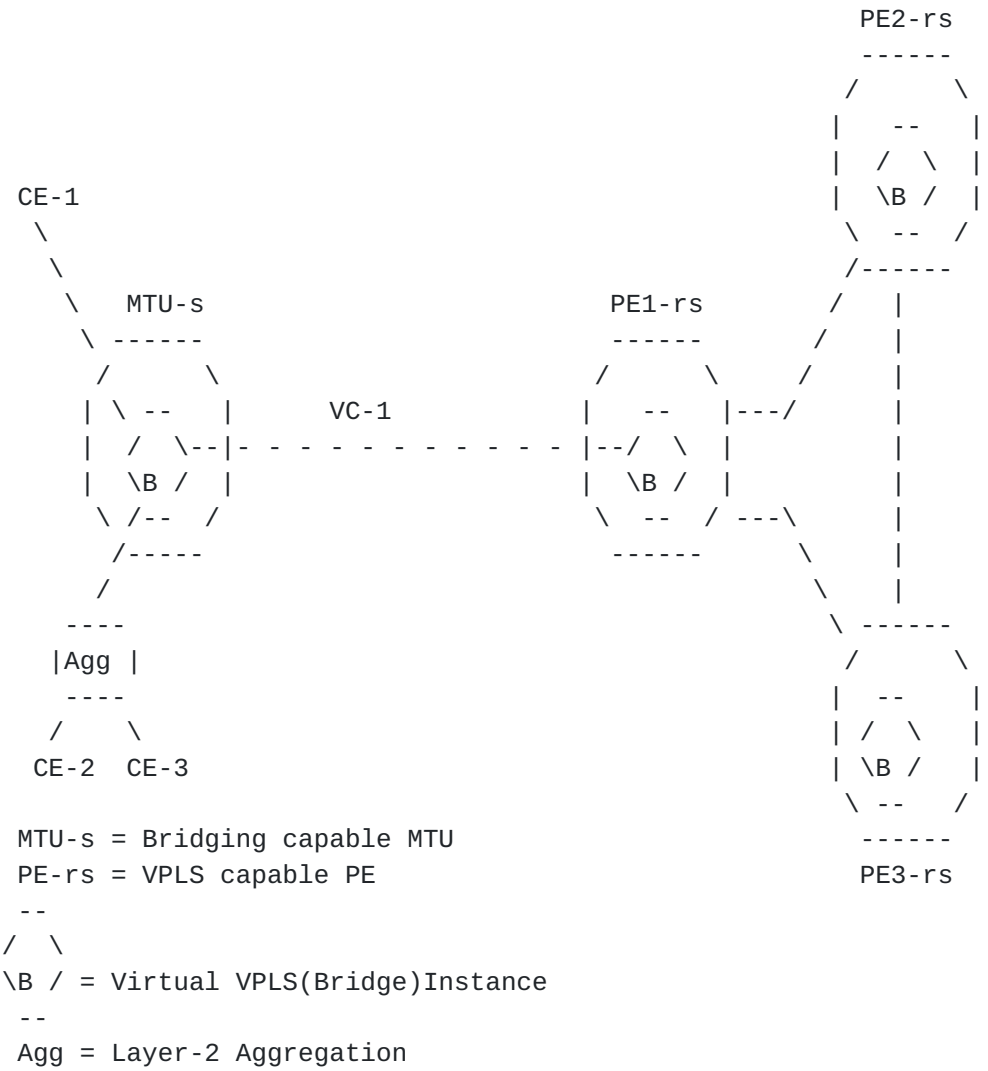
For rest of this discussion we will refer to a bridging capable MTU device as MTU-s and a non-bridging capable PE device as PE-r. A routing and bridging capable device will be referred to as PE-rs.

3.1. Spoke connectivity for bridging-capable devices

As shown in figure-1, consider the case where an MTU-s device has a single connection to the PE-rs device placed in the CO. The PE-rs devices are connected in a full mesh. To participate in the VPLS service, MTU-s device creates a single point-to-point tunnel LSP to the PE-rs device in the CO. We will call this the spoke connection. For each VPLS service, a single spoke VC is setup between an MTU-s and the PE-rs based on [MARTINI-SIG] and [MARTINI-ENCAP]. Unlike traditional [MARTINI-ENCAP] VCs that terminate on a physical port at each end, the spoke VC terminates on a virtual bridge instance on the MTU-s and the PE-rs devices. The MTU-s device and the PE-rs device treat each spoke connection like an access port of the VPLS service. On access ports, the combination of the physical port and the vlan tag is used to associate the traffic to a VPLS instance while the combination of vc-id and vc-labels (ingress and egress) are used to associate the traffic on the virtual spoke port with a VPLS instance.

The signaling and association of the spoke connection to the VPLS

service may be done by introducing extensions to the LDP signaling as specified in [MARTINI-SIG]. This will be specified in a future version of this document.



3.1.1. MTU-s Operation

The MTU-s device is defined as a device that supports layer-2 switching functionality and does all the normal bridging functions of learning and replication on all its ports, including the virtual spoke port. Packets to unknown destination are replicated to all ports in the service including the virtual spoke port. Once the MAC address is learned, traffic between CE1 and CE2 will be switched locally by the MTU-s device conserving the link capacity of the connection to the PE-rs. Similarly traffic between CE1 or CE2 and any remote destination is switched directly on to the spoke connection and sent to the PE-rs over the point-to-point VC LSP.

Since the MTU-s is bridging capable, only a single VC is required per VPLS instance for any number of access connections in the same VPLS service. This further reduces the signaling overhead between

the MTU-s and PE-rs.

3.1.2. PE-rs Operation

The PE-rs device is a device that supports all the bridging functions for VPLS service and supports the routing and MPLS encapsulation, i.e. it supports all the functions described in [VPLS]. The operation on the PE-rs node is identical to that described in [VPLS] with one addition. A point-to-point VC associated with the VPLS is regarded as a virtual port. The operation on the virtual spoke port is identical to the operation on an access port as described in the earlier section. As shown in figure-1, each PE-rs device switches traffic between aggregated [MARTINI-ENCAP] VCs that look like virtual ports and the network side VPLS VCs.

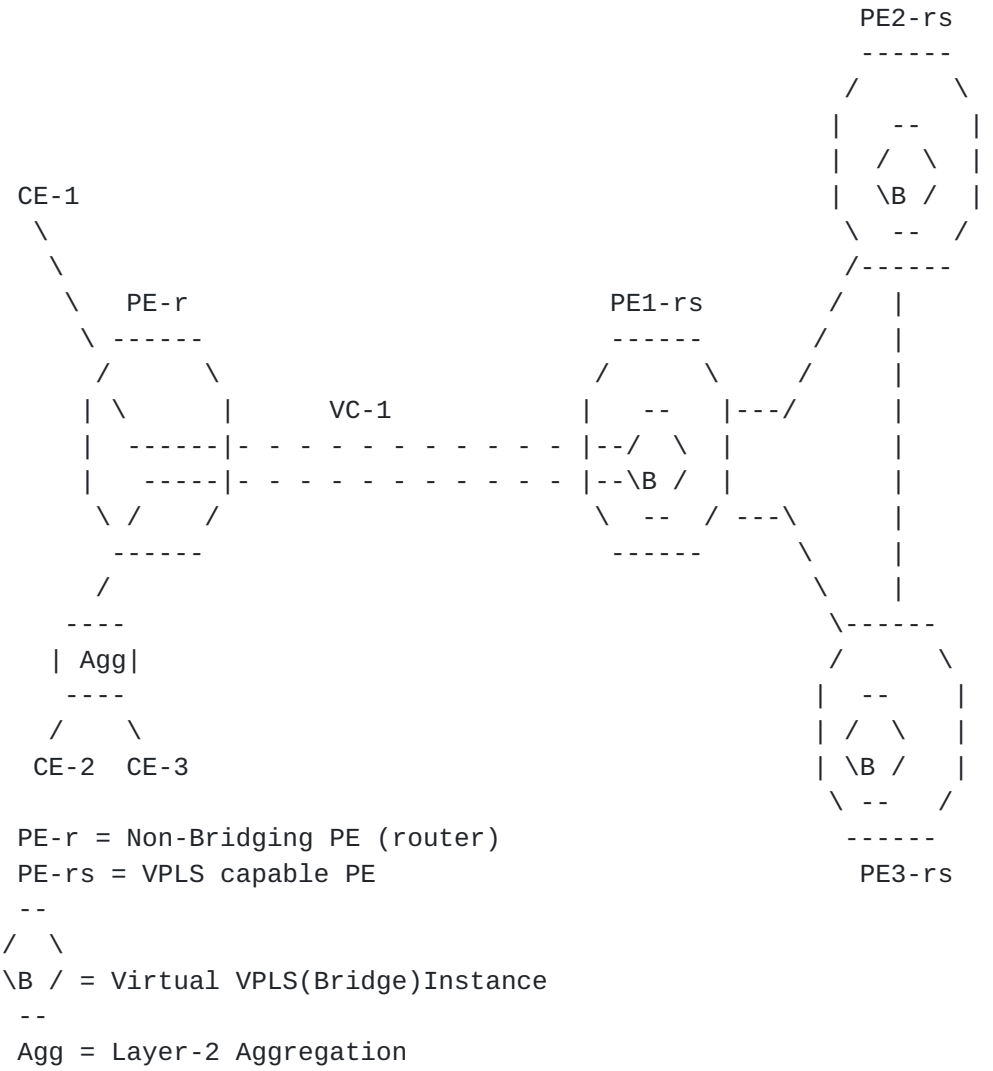
3.2. Advantages of spoke connectivity

Spoke connectivity offers several scaling and operational advantages for creating large scale VPLS implementations, while retaining the ability to offer all the functionality of the VPLS service.

- Eliminates the need for a full mesh of tunnels and full mesh of VCs per service between all devices participating in the VPLS service.
- Minimizes signaling overhead since fewer VC-LSPs are required for the VPLS service.
- Segments VPLS nodal discovery. MTU-s needs to be aware of only the PE-rs node although it is participating in the VPLS service that spans multiple devices. On the other hand, every VPLS PE-rs must be aware of every other VPLS PE-rs device and all of it's locally connected MTU-s and PE-r devices.
- Addition of other sites requires configuration of the new MTU-s device but does not require any provisioning of the existing MTU-s devices on that service.
- Hierarchical connections can be used to create VPLS service that spans multiple service provider domains. This is explained in a later section.

3.3. Spoke connectivity for non-bridging devices

In some cases, a bridging PE-rs device may not be deployed in some CO while a PE-r might already be deployed. If there is a need to provide VPLS service from the CO where the PE-rs device is not available, the service provider may prefer to use the PE-r device in the interim. In this section, we explain how a PE-r device that does not support any of the bridging functionality as described in [VPLS] can participate in the VPLS service.



As shown in figure-2, the PE-r device creates a point-to-point tunnel LSP to a PE-rs device. Then for every access port that needs to participate in a VPLS service, the PE-r device creates a point-to-point [MARTINI-SIG] VC that terminates on the physical port at the PE-r and terminates on the virtual bridge instance of the VPLS service at the PE-rs.

3.3.1. PE-r Operation

The PE-r device is defined as a device that supports routing but does not support any bridging functions. However, it is capable of setting up [MARTINI-SIG] VCs between itself and the PE-rs. For every port that is supported in the VPLS service, a [MARTINI-SIG] VC is setup from the PE-r to the PE-rs. Once the VCs are setup, there is no learning or replication function required on part of the PE-r. All traffic received on an access port is transmitted on the VC associated with that access port. Similarly all traffic

received on a VC is transmitted to the access port where the VC terminates. Thus traffic from CE1 destined for CE2 is switched at PE-rs and not at PE-r.

This approach adds more overhead than the bridging capable (MTU-s) spoke approach since a VC is required for every access port that participates in the service versus a single VC required per service (regardless of access ports) when a MTU-s type device is used. However, this approach offers the advantage of offering a VPLS service in conjunction with a routed internet service from CO where the PE-rs device is not yet deployed while the PE-r device is deployed.

3.3.2. PE-rs Operation

The operation of PE-rs is independent of the type of device at the other end of the spoke connection. Whether there is a bridging capable device (MTU-s) at the other end of the spoke connection or there is a non-bridging device (PE-r) at the other end of the spoke connection, the operation of PE-rs is exactly the same. Thus, the spoke connection from the PE-r is treated as a virtual port and the PE-rs device switches traffic between the virtual port, access ports and the network side VPLS VCs once it has learned the MAC addresses.

4. Redundant Spoke Connections

An obvious weakness of the hub and spoke approach described thus far is that the MTU device has a single connection to the PE-rs device. In case of failure of the connection or the PE-rs device, the MTU device suffers total loss of connectivity.

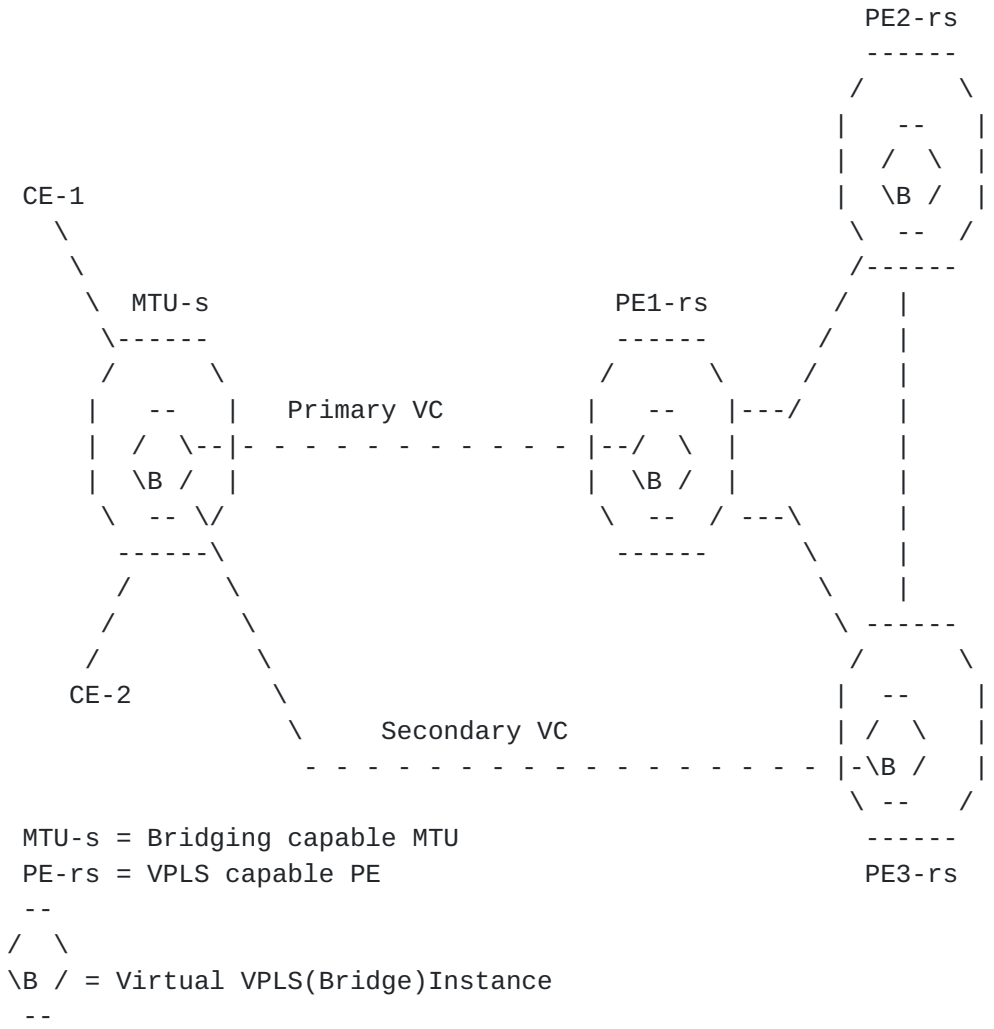
In this section, we describe how redundant connections can be provided to avoid total loss of connectivity from the MTU device. The mechanism described is identical for both, MTU-s and PE-r type of devices

4.1. Dual-homed MTU-s OR PE-r device

To protect from connection failure of the VC or the failure of the PE-rs device, the MTU-s device or the PE-r device is dual-homed into two PE-rs devices, as shown in figure-3. The PE-rs devices must be part of the same VPLS service instance.

An MTU-s device will setup two [MARTINI-ENCAP] VCs (one each to PE1-rs and PE3-rs) for each VPLS instances. One of the two VC is designated as primary and is the one that is actively used under normal conditions, while the second VC is designated as secondary and is held in a standby state. The MTU-s device or the PE-r device negotiates the VC-labels for both the primary and secondary VC, but does not use the secondary VC unless the primary VC fails. Since only one link is active at a given time, a loop does not

exist and hence 802.1D spanning tree is not required.



4.2. Failure detection and recovery

The MTU-s device controls the usage of the VC links to the PE-rs nodes. Since LDP signaling is used to negotiate the VC-labels, the keepalive messages used for the LDP session are used to detect failure of the primary VC.

Upon failure of the primary VC, the MTU-s device immediately switches to the secondary VC. At this point, the PE3-rs device that terminates the secondary VC starts learning MAC addresses on the spoke VC. All other PE-rs nodes in the network think that CE-1 and CE-2 are behind PE1-rs and may continue to send traffic to PE1-rs until they learn that the devices are now behind PE3-rs. The relearning process can take a long time and may adversely affect the connectivity of higher level protocols from CE1 and CE2. To enable faster convergence, the PE1-rs device where the primary VC

failed SHOULD send out a flush message, using the MAC TLV as defined in [VPLS], to all other PE-rs devices participating in the VPLS service. Upon receiving the message, all PE-rs flush the MAC addresses learned from PE1-rs. This creates more traffic

temporarily, since the remote PE-rs device that is transmitting to the CE1 and CE2 must replicate the traffic until it learns that the devices are now behind PE3-rs. This approach, however, speeds upon the convergence times when devices move from PE-rs to PE-rs.

5. Multi-domain VPLS service

Hierarchy can also be used to create a large scale VPLS service within a single domain or a service that spans multiple domains without requiring full mesh connectivity between all VPLS capable devices. Two fully meshed VPLS networks are connected together using a single LSP tunnel between the VPLS gateway devices. A single VC is setup per VPLS service to connect the two domains together. The VPLS gateway device joins two VPLS services together to form a single multi-domain VPLS service. The requirements and functionality required from a VPLS gateway device will be explained in a future version of this document.

6. Security Considerations

No new security issues result from this draft. It is recommended in that LDP security (authentication) methods be applied. This would prevent unauthorized participation by a PE in a VPLS. Traffic separation for VPLS is maintained using VC labels or IEEE 802.1q VLAN tags. However, for additional levels of security, the customer MAY deploy end-to-end security, which is out of the scope of this draft.

7. References

IETF Drafts

[VPSN-REQ] Heron et al, "Requirements for Virtual Private Switched Networks", (draft-heron-ppvpn-vpsn-reqmts-00.txt), work in progress, July 2001.

[PPVPN-REQ] "Service Requirements for Provider Provisioned Virtual Private Networks", M. Carugi, et al. August 2001. draft-ietf-ppvpn-requirements-02.txt. Work in progress.

[VPSN] VKompella et al, "Virtual Private Switched Network Services over an MPLS Network" , (draft-vkompella-ppvpn-vpsn-mpls-00.txt), work in progress, July 2001.

[TLS] Lasserre et al, "Transparent VLAN Services over MPLS". (draft-lasserre-tls-mpls-00.txt), work in progress, August 2001.

[VPLS] Vkompella Lasserre et al, "Signaling Virtual Private LAN Segments", work in progress, November 2001.

[MARTINI-ENCAP] Martini et al, "Encapsulation Methods for Transport of Layer 2 Frames Over IP and MPLS Networks", (draft-martini-l2circuit-encap-mpls-04txt), work in progress, November 2001.

[MARTINI-SIG] Martini et al, "Transport of Layer 2 Frames Over MPLS", (draft-martini-l2circuit-trans-mpls-08.txt), work in progress, November 2001.

[LDP] "LDP Specification", L. Andersson, et al. RFC 3036.January 2001.

8. Authors' Addresses

10. Author Information

Sunil Khandekar
TiMetra Networks
274 Ferguson Dr.
Mountain View, CA 94043
Tel.: +1 (650) 237-5105
Email: sunil@timetra.com

Vach Kompella
TiMetra Networks
274 Ferguson Dr.
Mountain View, CA 94043
Tel.: +1 (650) 237-5152
Email: vkompella@timetra.com

Joe Regan
TiMetra Networks
274 Ferguson Dr.
Mountain View, CA 94043
Tel.: +1 (650) 237-5103
Email: jregan@timetra.com

Nick Tingle
TiMetra Networks

274 Ferguson Dr.
Mountain View, CA 94043
Tel.: +1 (650) 237-5105

Email: sunil@timetra.com

Pascal Menezes
Terabeam
14833 NE 87th St.
Redmond, WA, USA
Email: Pascal.Menezes@Terabeam.com
Phone: +1 (206) 686-2001

Tom S. C. Soon
SBC Technology Resources Inc.
4698 Willow Road
Pleasanton, CA 94588
Tel.: +1 (925) 598-1227
Email: sxsoon@tri.sbc.com

Marc Lasserre
Riverstone Networks
5200 Great America Pkwy
Santa Clara, CA 95054
Tel.: +1 (408) 878-6550
Email: marc@riverstonenet.com

Giles Heron
PacketExchange Ltd.
The Truman Brewery
91 Brick Lane
LONDON E1 6QL
United Kingdom
Tel.: +44 7880 506185
Email: giles@packetexchange.net

Ron Haberman
Masergy Communications
2901 Telestar Ct.
Falls Church, VA 22042
Tel.: +1 (703) 846-0159
Email: ronh@masergy.com

Rick Wilder
Masergy Communications
2901 Telestar Ct.
Falls Church, VA 22042
Tel.: +1 (703) 846-0529
Email: rwilder@masergy.com

Kireeti Kompella
Juniper Networks

1194 N. Mathilda Ave
Sunnyvale, CA 94089
kireeti@juniper.net

Marty Borden
Atrica, Inc.
30 Shaker Lane
Littleton, MA 01460
mborden@atrica.com