

L2VPN

Weiguo Hao
Liang Xia
Shunwan Zhuang
Huawei
Vic Liu
China Mobile
July 4, 2014

Internet Draft

Intended status: Informational
Expires: January 2015

Inter-AS Option B between NV03 and MPLS EVPN network
draft-hao-l2vpn-inter-nvo3-evpn-00.txt

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/ietf/lid-abstracts.txt>.

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

This Internet-Draft will expire on January 4, 2015.

Copyright Notice

Copyright (c) 2013 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

Internet-Draft

EVPN Inter-As Option-B

July 2014

carefully, as they describe your rights and restrictions with respect to this document.

Abstract

This draft describes option-B inter-as connection between NV03 network and MPLS EVPN network. Comparing to traditional MPLS EVPN Option-B inter-as connection, this draft provides enhancement for heterogeneous network multi-as connection, the control plane and data plane procedures in NV03 network are described.

Table of Contents

1.	Introduction	2
2.	Conventions used in this document.....	3
3.	Reference model	5
4.	Option-A inter-as solution overview.....	6
5.	Inter-as option-B routing distribution process.....	7
	5.1. Ethernet Tag ID conversion on ASBR.....	7
	5.2. Ethernet Auto-Discovery Route process.....	8
	5.2.1. Optimized MPLS Label solution on ASBR.....	9
	5.3. Ethernet Segment Route process.....	10
	5.4. Inclusive Multicast Ethernet Tag Route process.....	10
	5.5. MAC/IP advertisement route process	10
6.	Inter-as option-B data plane procedures	12
	6.1. Internal DC to external DC direction	12
	6.2. External DC to internal DC direction	12
7.	Inter-as option-B solution between PBB-EVPN network and NV03 network	13
8.	Security Considerations.....	13
9.	IANA Considerations	13
10.	References	13
	10.1. Normative References.....	13
	10.2. Informative References.....	13
11.	Acknowledgments	14

[1.](#) Introduction

In cloud computing era, multi-tenancy has become a core requirement for data centers. Since NV03 can satisfy multi-tenancy key

requirements, this technology is being deployed in an increasing number of cloud data center network. NV03 focuses on the construction of overlay networks that operate over an IP (L3) underlay transport network. It can provide layer 2 bridging and

layer 3 IP service for each tenant. VXLAN and NVGRE are two typical NV03 technologies. NV03 overlay network can be controlled through centralized NVE-NVA architecture or through distributed BGP VPN protocol.

NV03 has good scaling properties from relatively small networks to networks with several million tenant systems (TSs) and hundreds of thousands of virtual networks within a single administrative domain. In NV03 network, 24-bit VN ID is used to identify different virtual networks, theoretically 16M virtual networks can be supported in a data center. In a data center network, each tenant may include one or more layer 2 virtual network and in normal cases each tenant corresponds to one routing domain (RD). Normally each layer 2 virtual network corresponds to one or more subnets.

To provide cloud service to external data center client, data center networks should be connected with WAN networks. BGP MPLS based Ethernet VPNs(EVPN)[EVPN] is being deployed in an increasing number of WAN networks. If EVPN CEs in external DC and TSs in internal DC belong to same subnet of same tenant, they are in same broadcast domain and can freely layer 2 communicate with each other in the broadcast domain.

Normally internal data center and external EVPN network belongs to different autonomous system(AS). This requires the setting up of inter-as connections at Autonomous System Border Routers(ASBRs) between NV03 network and external EVPN network.

Currently, a typical connection mechanism between a data center network and an MPLS EVPN network is similar to Inter-AS Option-A of [RFC4364](#), but it has scalability issue if there is huge number of tenants in data center networks. To overcome the issue, inter-as Option-B between NV03 network and BGP MPLS EVPN network is proposed in this draft.

[2.](#) Conventions used in this document

EVI - An EVPN instance spanning across the PEs participating in that

EVPN.

MAC-VRF - A Virtual Routing and Forwarding table for MAC addresses on a PE for an EVI.

Network Virtualization Edge (NVE) - An NVE is the network entity that sits at the edge of an underlay network and implements network virtualization functions.

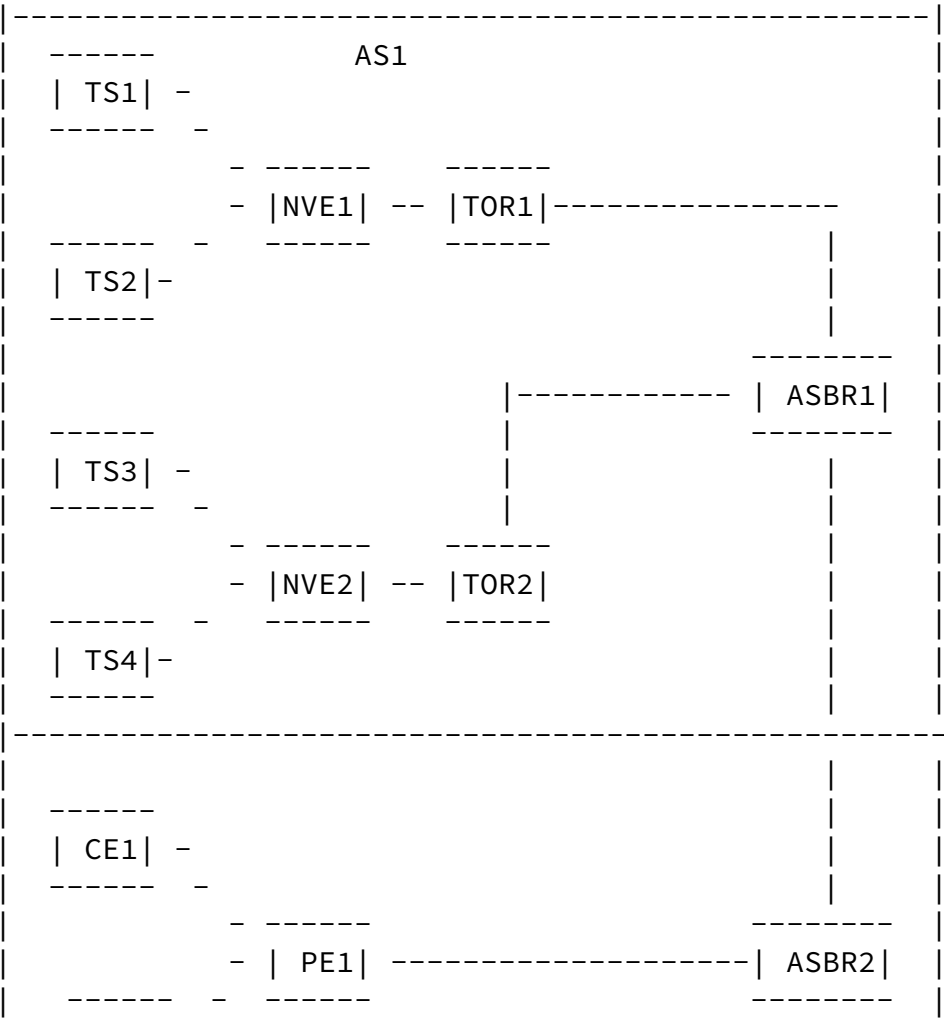
Tenant System - A physical or virtual system that can play the role of a host, or a forwarding element such as a router, switch, firewall, etc. It belongs to a single tenant and connects to one or more VNs of that tenant.

VN - A VN is a logical abstraction of a physical network that provides L2 network services to a set of Tenant Systems.

RD - Route Distinguisher. RDs are used to maintain uniqueness among identical routes in different MAC-VRFs, The route distinguisher is an 8-octet field prefixed to the customer's MAC address. The resulting 12-octet field is a unique "VPN-MAC" address.

RT - Route targets. It is used to control the import and export of routes between different MAC-VRFs.

3. Reference model



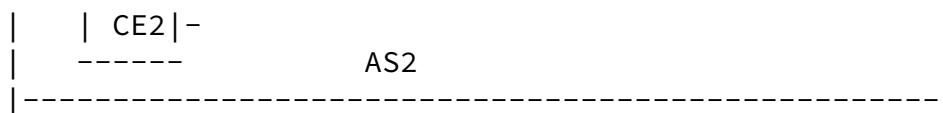


Figure 1 Reference model

Figure 1 shows an arbitrary Multi-AS VPN interconnectivity scenario between NV03 network and MPLS EVPN network. NVE1, NVE2, and ASBR1 forms NV03 overlay network in internal DC. TS1 and TS2 connect to NVE1, TS3 and TS4 connect to NVE2. PE1 and ASBR2 forms MPLS EVPN network in external DC. CE1 and CE2 connect to PE1. The NV03 network belongs to AS 1, the MPLS EVPN network belongs to AS 2.

There are two tenants of tenant 1 and tenant 2 in NV03 network. MAC-VRF 1 and MAC-VRF 2 are created on NVE1 and NVE2 for these two tenants. TSs(TS1 and TS3) in tenant 1 and CE(CE1) in VPN-Red belongs to same subnet and broadcast domain, TSs(TS2 and TS4) in tenant 2

and CE(CE2) in VPN-Green belongs to same subnet and broadcast domain. The TSs and CEs in same broadcast domain can freely layer 2 communicate with each other.

The TS and CE information in above figure 1 are as follows:

TS	Tenant	IP Address	MAC	VN ID
TS1	1	10.1.1.2	MAC1	10
TS2	2	20.1.1.2	MAC2	20
TS3	1	10.1.1.3	MAC3	10
TS4	2	20.1.1.3	MAC4	20

Table 1 TS information in NV03 network

CE	Route Distinguisher	Route Target	IP Address	MAC
CE1	VPN-Red1	1:1	10.1.1.4	MAC5

CE2	VPN-Green1	2:2	20.1.1.4	MAC6
+-----+	-----+	+-----+	-----+	+-----+

Table 2 CE information in MPLS/IP VPN network

4. Option-A inter-as solution overview

In Option-A inter-as solution, peering ASBRs are connected by multiple sub-interfaces, each ASBR acts as a PE, and thinks that the other ASBR is a CE. EVPN instances are configured at AS border routers (ASBR1 and ASBR2) so that each ASBRs associate each such sub-interface with a EVPN instance. It requires MAC look up on ASBRs. MAC address propagation for each EVPN instance between ASBR1 and ASBR2 relies on data plane learning mechanism. In the data-plane, VLANs are used for VPN traffic separation.

Option-A inter-as solution has following issues:

1. Up to 16 million (16M) sub-interfaces need to exist between the ASBRs, if there are 16M VN in NV03 network.
2. UP to 16M EVPN instances need to be supported on border routers.

3. Several million MAC routing entries need to be supported on border routers.

Inter-as option B between NV03 network and MPLS EVPN network can be used to address these issues. Due to it is for multi-as interconnection between heterogeneous networks, so there are some differences from traditional homogenous EVPN Inter-AS Option-B.

5. Inter-as option-B routing distribution process

In option-B inter-as solution, an EBGp session is used to distribute labeled EVPN NLRI between the ASBRs. The advantage of this option is that it's more scalable, as there is no need to have one sub-interface per VPN between ASBRs.

There are four Route Types in EVPN NLRI, they are Ethernet Segment Route, Ethernet Auto-Discovery Route, Inclusive Multicast Ethernet Tag Route and MAC/IP advertisement route which are used for Designated Forwarder Election, fast Convergence and aliasing or

backup-path, multicast traffic handling and unicast traffic handling respectively.

For inter-as option-B interconnection between EVPN and NV03 network, When a ASBR receives BGP update message carrying the routes from peer PEs or NVEs in local AS, it should re-constructs these message and advertises it to peer ASBR, the Next Hop field of the MP_REACH_NLRI attribute should be set to a routable IP address of the ASBR. When the peer ASBR receives the message, the ASBR also should re-construct the message and advertise it to peer PEs or NVEs in its local AS, the Next Hop field of the MP_REACH_NLRI attribute also should be set to a routable IP address of the ASBR.

In NV03 network, there are two options for mapping the VNI to an EVI [EVPN-OVERLAY], one is single subnet per EVI, and another one is multiple subnets per EVI.

In the following subsection, a detail explanation will be given on how to re-construct EVPN update message and how to generate incoming and outgoing forwarding table on ASBR.

[5.1.](#) Ethernet Tag ID conversion on ASBR

A broadcast domain can be identified by different Ethernet Tag ID in NV03 and MPLS EVPN network. The Ethernet Tag ID mapping relationship between NV03 and MPLS EVPN network should be configured on each ASBR in beforehand. For example, VLAN 10 in EVPN network and VN 100 in

NV03 network belong to same broadcast domain, NV03 network uses 100 as Ethernet Tag ID, EVPN network uses 10 as Ethernet Tag ID.

When a ASBR receives BGP update message carrying EVPN NLRI from peer ASBR, it should replace Ethernet Tag ID field with local corresponding value and then advertise the message to peer PEs or NVEs in its local AS.

[5.2.](#) Ethernet Auto-Discovery Route process

There are two Ethernet A-D route types, one is per ES route, and another one is per EVI. The "ESI Label Extended Community" MUST be included in the route, it is to indicate ES's redundancy mode and to advertise ESI Label for split-horizon filtering.

When an ASBR in NV03 network receives Ethernet A-D per ES route, the ASBR learns a ES and multi-homed NVEs correspondence, the ES's redundancy mode. If "Single-Active" flag in "ESI Label Extended Community" is set, the ES is operating in Single-Active redundancy mode. Otherwise, it is operating in All-Active redundancy mode. The Ethernet A-D per EVI route can be used for Aliasing and Backup-Path, aliasing is used for all-active mode, backup-path is for single-active mode.

When an ASBR in NV03 network receives Ethernet A-D per EVI route, the ASBR should allocate new MPLS Label and advertises it to all peer ASBRs in Ethernet Auto-Discovery Route MPLS Label field. In NV03 network, the MPLS Label allocation principle is: If ESI is 0, MPLS label is allocated per NVE per VN(This is single-homed case). Otherwise, MPLS label is allocated per ESI per VN((This is multi-homed case). MPLS VPN Label and <remote NVE,VN ID> correspondence is used to generate incoming forwarding table on each ASBR, traffic forwarding from external to internal DC direction relies on the incoming forwarding table.

In multi-homed scenario, when an ESI occurs link failure and lost connection with a NVE, the NVE should trigger ASBR in its local AS to mass update its local forwarding table by Ethernet A-D per ES route. This is called fast convergence procedure. The ASBR doesn't need re-allocate MPLS Label for each VN on the ESI and advertise to peer AS, i.e., fast convergence process is restricted to local AS, the Ethernet A-D route per ES doesn't need to be transmitted to peer AS.

For aliasing and backup-path procedures, these procedures also don't need cross different AS domain, they are only restricted in local AS,

each ASBR in local AS needs to process Ethernet A-D per EVI route from PEs or NVEs in local AS for these procedures.

In aliasing case, when a ASBR in NV03 network receives traffic data from external DC to external DC, the traffic will be forwarded to all-active remote NVEs in load balancing mode. For each aliasing ES and VN, there is a corresponding incoming forwarding table item which includes one MPLS Label and multiple <NVE,VN ID> pairs on ASBR, the NVE is a member of remote multi-homed NVEs attaching the

aliasing ES.

In backup-path case, for each backup-path ES and VN, there is a corresponding incoming forwarding table item which includes one MPLS Label and one <NVE,VN ID> on ASBR, the NVE is primary NVE that advertises the MAC/IP advertisement route in the VN. When a ASBR receives first MAC/IP advertisement route from remote primary NVE, it will know the primary NVE and generate the incoming forwarding table item.

5.2.1. Optimized MPLS Label solution on ASBR

MPLS Label consumption on ASBR is high through the above per ESI per VN solution, the optimized allocation solution is provided as follows:

If multiple ESIs are operating in all-active mode and attached to the exact same set of NVEs, then these ESIs can share same MPLS Label for same VN to save MPLS Label space on ASBR.

If multiple ESIs are operating in single-active mode and attached to the exact same set of NVEs, primary NVEs for these ESI are same NVE, then the VNs on these ESIs can share same MPLS Label for same VN to save MPLS VPN Label space on ASBR.

In this case, if a ESI occurs link failure and lost connection with a NVE, the NVE advertises Ethernet Auto-Discovery Route per ES to each ASBR in its local AS, the ASBRs knows that the ESI is attached to a different set of NVEs, it should re-allocate new MPLS Labels for each VN on the ESI, mass update its incoming forwarding table, then advertise these MPLS Labels using Ethernet Auto-Discovery route per EVI to peer ASBR.

When peer ASBR receives the Ethernet Auto-Discovery route per EVI, it allocates new MPLS Label and replaces the value in Ethernet Auto-Discovery Route MPLS Label field, then advertises it to all peer PEs.

Remote PEs in peer AS should update all its MAC entries with the new MPLS Label associated with the ESI and EVI.

[5.3.](#) Ethernet Segment Route process

Due to this route is used for DF election and multi-homed PE or NVE devices won't straddle between MPLS EVPN and NV03 network, so when a ASBR receives BGP update message carrying the route from peer PE or NVE in its own AS, it just removes it from the message, the route don't need to be transmitted to peer AS.

[5.4.](#) Inclusive Multicast Ethernet Tag Route process

Similar to regular EVPN inter-as solution, when a ASBR receives from one of its IBGP neighbors a BGP Update message that carries the route, it re-advertises it to peer ASBRs and these peer ASBRs re-advertise it to peer PEs or NVEs in its local AS. The re-advertised routes MUST be the same as the original ones, except for the PMSI Tunnel attribute in Inclusive Multicast Ethernet Tag Route and Ethernet Tag ID. If ingress replication is used between ASBRs, the Tunnel Type in PMSI Tunnel attribute is set to Ingress Replication, the Leaf Information Required flag is set to 1, the PMSI Tunnel attribute carries no MPLS labels.

[5.5.](#) MAC/IP advertisement route process

Because the ASBR in NV03 network has already assigned MPLS Label for each ESI(or NVE in single-homed case) and each VN when it received Ethernet Auto-Discovery Route from remote NVEs in its local AS, so the ASBR receives first MAC/IP advertisement route from a <ES,VN>, it will search the already assigned MPLS Label for the <ES,VN>, generate a incoming forwarding item, fuel MPLS Label field in the MAC/IP advertisement route, and then send it to peer ASBR. The incoming forwarding table is used for traffic forwarding from external DC to internal DC direction.

In above figure 1, all TSs are single-homed to a NVE, MPLS VPN Label is assigned per NVE per VN, the incoming forwarding table on ASBR in NV03 network is as follows:

MPLS VPN Label	NVE + VN ID
1000	NVE1 + 10
2000	NVE1 + 20
1001	NVE2 + 10
2001	NVE2 + 20

Incoming forwarding table

When ASBR1 in NV03 network receives from EBGp neighbors ASBR2 a BGP Update message that carries MAC/IP advertisement route, it should allocate VN ID per MPLS VPN Label, generate outgoing forwarding table, and then advertises it to peer NVEs in its local AS.

In above figure 1, ASBR1 allocates VN ID for each VPN Label receiving from ASBR2, and then ASBR2 advertises the MAC/IP advertisement route with new allocated VN ID to each NVE (NVE1 and NVE2). The role of the VN ID is similar to the role of In VPN Label in ASBR1, it has local significance on ASBR1, each VN ID corresponds to each MPLS VPN Label on ASBR2; The VN ID space should be assigned in beforehand and should be orthogonal to the VN ID space for tenant identification(for example, assuming ASBR1 has local connecting TSS of tenant 1 to 100, VN ID 1 to 100 are allocated for these tenants, other VN ID other than 1 to 100 can be allocated for outgoing forwarding table purpose). The allocated VN ID and its corresponding out VPN Label forms an outgoing forwarding table which is used to forward NV03 traffic from internal DC to external DC. The outgoing forwarding table on ASBR1 is as follows:

VN ID	Out VPN Label
10000	3000
10001	4000

Outgoing forwarding table

6. Inter-as option-B data plane procedures

This section describes the step by step procedures of data forward for either: a) internal DC to external DC data flows, or b) the external DC to internal DC data flows.

6.1. Internal DC to external DC direction

1. TS1 sends traffic to NVE1, the destination MAC is CE1's MAC address of MAC5.
2. NVE1 looks up MAC-VRF 1's MAC forwarding table, then it gets NV03 tunnel encapsulation information. The destination outer address is ASBR1's IP address, VN ID is 10000. NVE1 performs NV03 encapsulation and sends the traffic to ASBR1.
3. ASBR1 decapsulates NV03 encapsulation and gets VN ID 10000. Then it looks up outgoing forwarding table based on the VN ID and gets MPLS VPN label 3000. Finally it pushes MPLS VPN label for the IP traffic and sends it to ASBR2.
4. ASBR2 swaps VPN Label, then sends the traffic to PE1 through IGP tunnel.
5. PE1 terminates IGP tunnel, pops MPLS VPN label 3000, looks up local MAC-VRF 1, and then forwards the traffic to CE1.

6.2. External DC to internal DC direction

1. CE1 sends traffic to PE1, destination MAC is TS1's MAC address of MAC1.
2. PE1 looks up local MAC forwarding table in VPN-RED, pushes MPLS VPN label 1000, then searches IGP tunnel, then the PE sends the traffic to ASBR2 through IGP tunnel.
3. ASBR2 terminates IGP tunnel, swaps MPLS VPN label, then sends the traffic to ASBR1.
4. ASBR1 looks up incoming forwarding table and gets NV03 encapsulation, then performs NV03 encapsulation and sends the traffic to NVE1. The destination outer IP is NVE1's IP, VN ID is 10.
5. NVE1 decapsulates NV03 encapsulation, gets local EVPN instance 1 relying on VN ID 10, looks up local MAC-VRF 1, then sends the traffic to TS1.

Internet-Draft

EVPN Inter-As Option-B

July 2014

[7.](#) Inter-as option-B solution between PBB-EVPN network and NV03 network

For the further study.

[8.](#) Security Considerations

Similar to the security considerations for inter-as Option-B in [\[RFC4364\]](#) the appropriate trust relationship must exist between NV03 network and MPLS EVPN network. EVPN routes in NV03 network should neither be distributed to nor accepted from the public Internet, or from any BGP peers that are not trusted. For other general VPN Security Considerations, see [\[RFC4364\]](#).

[9.](#) IANA Considerations

This document requires no IANA actions. RFC Editor: Please remove this section before publication.

[10.](#) References

[10.1.](#) Normative References

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

[10.2.](#) Informative References

- [1] [\[RFC4364\]](#) E. Rosen, Y. Rekhter, " BGP/MPLS IP Virtual Private Networks (VPNs)", [RFC 4364](#), February 2006.
- [2] [EVPN] Sajassi et al., "BGP MPLS Based Ethernet VPN", [draft-ietf-l2vpn-evpn-02.txt](#), work in progress, February, 2012.
- [3] [NVA] D.Black, etc, "An Architecture for Overlay Networks (NV03)", [draft-ietf-nvo3-arch-01](#), February 14, 2014
- [4] [EVPN-OVERLAY] A. Sajassi,etc, "'A Network Virtualization Overlay Solution using EVPN'", [draft-sd-l2vpn-evpn-overlay-03](#), June, 2014
- [5] [NOV3-FRWK] Lasserre et al., "Framework for DC Network Virtualization", [draft-ietf-nvo3-framework-01.txt](#), work in

progress, October 2012.

- [6] [NVGRE] Sridhavan, M., et al., "NVGRE: Network Virtualization using Generic Routing Encapsulation", [draft-sridharan-virtualization-nvgre-01.txt](#), July 8, 2012.

Hao & et,al

Expires January 4, 2015

[Page 13]

Internet-Draft

EVPN Inter-As Option-B

July 2014

- [7] [VXLAN] Dutt, D., et al, "VXLAN: A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", [draftmahalingam-dutt-dcops-vxlan-02.txt](#), August 22, 2012.

[11](#). Acknowledgments

Authors like to thank Junlin Zhang for his valuable inputs.

Authors' Addresses

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China
Email: haoweiguo@huawei.com

Liang Xia (Frank)
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China
Email: frank.xialiang@huawei.com

Shunwan Zhuang
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China
Email: zhuangshunwan@huawei.com

Vic Liu

China Mobile
32 Xuanwumen West Ave, Beijing, China
Email: liuzhiheng@chinamobile.com

Hao & et,al

Expires January 4, 2015

[Page 14]