

NV03  
Internet-Draft  
Intended status: Informational  
Expires: January 5, 2013

B. Carpenter  
Univ. of Auckland  
S. Jiang  
Huawei Technologies Co., Ltd  
July 4, 2012

Layer 3 Addressing Considerations for Network Virtualization Overlays  
draft-carpenter-nvo3-addressing-00

## Abstract

This document discusses network layer addressing issues for virtual network overlays in large scale data centres hosting many virtual servers for multiple customers.

## Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on January 5, 2013.

## Copyright Notice

Copyright (c) 2012 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.

Internet-Draft

NV03 Addresssing

July 2012

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Aspects of addressing in virtual overlay networks . . . . .	<a href="#">3</a>
<a href="#">2.1.</a>	Address Independence and Isolation . . . . .	<a href="#">3</a>
<a href="#">2.2.</a>	Multiple Data Centres . . . . .	<a href="#">4</a>
<a href="#">2.3.</a>	Address mapping . . . . .	<a href="#">4</a>
<a href="#">2.4.</a>	Address migration . . . . .	<a href="#">5</a>
<a href="#">2.5.</a>	DNS . . . . .	<a href="#">5</a>
<a href="#">2.6.</a>	Dual Stack Operation . . . . .	<a href="#">6</a>
<a href="#">3.</a>	Consequences for IPv4 address management . . . . .	<a href="#">6</a>
<a href="#">4.</a>	Consequences for IPv6 address management . . . . .	<a href="#">7</a>
<a href="#">5.</a>	Security Considerations . . . . .	<a href="#">7</a>
<a href="#">6.</a>	IANA Considerations . . . . .	<a href="#">7</a>
<a href="#">7.</a>	Acknowledgements . . . . .	<a href="#">8</a>
<a href="#">8.</a>	Change log [RFC Editor: Please remove] . . . . .	<a href="#">8</a>
<a href="#">9.</a>	References . . . . .	<a href="#">8</a>
<a href="#">9.1.</a>	Normative References . . . . .	<a href="#">8</a>
<a href="#">9.2.</a>	Informative References . . . . .	<a href="#">8</a>
	Authors' Addresses . . . . .	<a href="#">9</a>

Internet-Draft

NV03 Addresssing

July 2012

## 1. Introduction

A common technique in large data centres hosting servers and services for many customers is to use virtual layer 3 network overlays (NV03) to organise, manage and separate the virtual servers used by individual customers. The related problems are discussed in [[I-D.narten-nvo3-overlay-problem-statement](#)], and a framework for the main components of a solution is described in [[I-D.lasserre-nvo3-framework](#)].

[Note: In this draft we do not yet give detailed references to the various NV03 drafts, none of which discuss addressing issues in detail.]

When emulating a large number of virtual hosts on whatever physical network topology is used, probably involving multiple LAN segments, virtual LANs at layer2, and both routers and switches, the question of the IP addressing scheme for the virtual hosts is not trivial. The intention of the present document is to describe the resulting consequences for IP address management in such an environment. Firstly the general aspects are discussed, and then the consequences for both IPv4 and IPv6 addressing schemes are described.

## 2. Aspects of addressing in virtual overlay networks

### 2.1. Address Independence and Isolation

In a typical hosting centre, there will be a number of customers, who are quite possibly mutual competitors who happen to use the same hosting centre. It is essential that virtual hosts assigned to one customer cannot communicate directly with, or even be aware of, virtual hosts assigned to any other customer. It is also essential that virtual networks are operationally independent to the maximum possible extent.

Therefore, to simplify operations by clearly separating independent virtual networks (VNs) from one another, and to enhance both real and perceived confidentiality of each network, it is desirable for the addresses used by each virtual layer 3 network to be allocated and managed independently of all the others. A consequence of this is that it becomes reasonably straightforward to configure layer 3 routing such that traffic from one VN can never unintentionally enter another one, because each network has its own well defined range of addresses. Similarly, it is also reasonably straightforward to configure firewalls or filters to detect and block any unwanted traffic between VNs, even if there is a routing misconfiguration.

For these reasons alone, it is necessary for each VN to have its own well-defined layer 3 address space that is managed independently of all other VNs. This requirement is independent of whether the physical hosts that contain the virtual hosts are on one or more physical or bridged LANs or VLANs. In other words, once the layer 3 topology is virtualised, layer 2 address independence and isolation is neither necessary nor sufficient to guarantee layer 3 address independence and isolation.

It should be understood that independent address allocation does not imply unique addresses. In the IPv4 context, as further discussed below, it is very likely that multiple VNs will use the same ambiguous address space, running over the same physical network infrastructure.

## [2.2.](#) Multiple Data Centres

A given customer might have virtual hosts spread across multiple data centres (DCs). Furthermore, those data centres might be owned and operated by competing enterprises. The only safe assumption is that a single address range cannot span multiple DCs, and that a virtual host being relocated to another DC might need to be renumbered. The addressing scheme for virtual hosts must be compatible with such a situation. Most likely this means extending the requirement for address independence and isolation to cover separate parts of a given customer's total set of virtual servers. In other words the usage scenario for any given customer must be able to deal with virtual hosts in multiple independent and mutually isolated layer 3 address spaces, and with the risk of occasional virtual host renumbering.

This complicates the issues of routing configuration and address filtering. If a VN extends over multiple DCs, VN routing across DC borders must be supported for the address ranges concerned, and address filtering must also be applied in a consistent way at each DC hosting part of a given VN.

### [2.3.](#) Address mapping

Several of the NV03 documents state the need for an address mapping scheme. Some aspects of this are discussed in [\[I-D.kreeger-nvo3-overlay-cp\]](#). It is generally assumed that an NV03 system will be built using tunnels, and the required mapping is between virtual host addresses and tunnel end points. The addressing scheme for virtual hosts needs to be consistent with the mapping system adopted and whatever dynamic update protocol is used for that mapping.

In the case of a VN that covers multiple DCs, the mapping scheme must

also support multiple DCs. The mapping update protocol will need to exchange mapping information between tunnel endpoints at all DCs involved. This information needs to be specified in some detail, and it must be decided whether this protocol needs to be run per VN or per DC, how this protocol decides which DCs it should talk with, etc.

### [2.4.](#) Address migration

As discussed in [\[I-D.narten-nvo3-overlay-problem-statement\]](#), current virtual host mechanisms assume that a host's IP address is fixed. If a workload is migrated from one physical host to another, the migration mechanisms assume that existing transport layer associations such as TCP sessions stay alive, and the succesful migration of a job in progress relies on this.

As workload conditions change in a large data centre, virtual hosts may need to be migrated from one physical host to another, and quite possibly this will mean moving to a different physical LAN. However, the virtual host address itself should remain constant as just mentioned. The addressing scheme adopted needs to be consistent with this requirement. Another way to view this is as an inverted form of renumbering - instead of the address of a given host changing, a

given address is reassigned to a different physical host, thereby representing the move of a given virtual host.

When such a move occurs, there will normally be changes in both the layer 2/layer 3 mapping (given by ARP, Neighbour Discovery, etc.) and the virtual host address to tunnel mapping mentioned above. However, an address which is moved in this way should still remain part of the same aggregate for routing purposes. Otherwise, an immediate change to the routing configuration will be needed as well.

As mentioned above, if a virtual host needs to be migrated between DCs, it might be unavoidable for its virtual address to change. In this case an application layer mechanism will be needed to recover from the resulting loss of transport layer sessions.

## [2.5.](#) DNS

A Domain Name Service, which resolves queries for hostnames into IP addresses, can reduce the direct dependence of customer applications on IP addresses. If a virtual host is always connected using its hostname, the renumbering issue during inter-DC migrations, mentioned in previous section, would be significantly mitigated. However, this would imply a need for rapid DNS updates.

## [2.6.](#) Dual Stack Operation

In the case of a dual stack deployment, where each virtual host has both an IPv4 and an IPv6 address, there will presumably be some sort of interdependency of the two addressing schemes. At least, the virtual subnet topologies would usually be the same for the two addressing schemes, and virtual hosts would need to migrate simultaneously for IPv4 and IPv6 purposes. If this was not the case, scenarios requiring IPv4/IPv6 interworking might arise unexpectedly, which would be inconvenient and inefficient. A particular case would be the migration of a virtual host between two DCs, one of which supports dual stack and the other of which supports only a single stack.

In general, these situations would require a layer 3 IPv4/IPv6

translator within a VN. This solution should be avoided if possible.

### 3. Consequences for IPv4 address management

In IPv4, it must be assumed that in many if not most cases, the virtual hosts will be numbered out of ambiguous private address space [[RFC1918](#)]. The only safe assumption for a general model is that any individual VN may use the same address space as any other. This increases the importance of the requirements for address isolation, independence and mapping. In fact, without address mapping (and in some scenarios network address translation) a large scale IPv4 NV03 system could not be made to work.

Since the IPv4 addresses in use will be ambiguous, management tools must be carefully designed so that operators will never need to rely on addresses alone to identify individual servers. For example, when an address is presented to an operator for any reason, it should always be tagged with some sort of VN identifier. The same goes for any place that an address is logged or stored for any other reason. Legacy software and tools that do not do this should be avoided as much as possible.

An interesting aspect of using, say, Net 10 for every VN instance is that the number of virtual hosts can be quite large, up to  $2^{24}$  (in excess of 16 million). This frees the designer from traditional limits on the size of an IPv4 subnet.

An unavoidable consequence of using [RFC1918](#) addresses is that the virtual hosts, if accessed by outside users, will be hidden behind either an application layer proxy or a NAT. In both cases these might be part of a load balancing system.

### 4. Consequences for IPv6 address management

In IPv6, there is no concept of ambiguous private space. Each VN can have its own global-scope address prefix. This removes the operational problems casued by ambiguous addresses in IPv4.

Even a basic /64 prefix would allow for more virtual host addresses than would ever be possible in IPv4, so again the designer is not

restricted by any absolute limit on subnet size. Nevertheless, it is advisable to use a shorter prefix such as /48 or /56 for each VN, so that a VN can span more than one LAN using standard IPv6 routing without difficulty. It is unclear that tunnels and address mapping are needed for IPv6-based VNs, due to the absence of ambiguous addresses.

A choice can be made between a regular IPv6 prefix from the customer's own IPv6 space or from ISP-assigned space, and a Unique Local Address prefix [[RFC4193](#)], [[I-D.liu-v6ops-ula-usage-analysis](#)]. The latter has the advantage of needing no administrative procedure before assigning it, and it is also routinely blocked by site and ISP border routers, like an [RFC1918](#) prefix. However, apart from that the choice of IPv6 prefix has little external importance and is mainly a matter of convenience.

There is no requirement for IPv6 prefix translation [[RFC6296](#)] between the virtual hosts and any outside users. However, the presence of such translation, or of some form of load balancing, cannot be excluded.

## [5.](#) Security Considerations

Routing configurations and filters in firewalls and routers should be constructed such that, by default, packets from virtual hosts in one VN cannot be forwarded into another VN. Traffic to and from a given VN should only be allowed for the designated users of that VN, and for the VN management and operations tools.

An independent and isolated addressing scheme is not by itself a security solution. While it might avoid the most trivial and straightforward penetration attempts, it is in no way a substitute for a security solution that responds to specific threat models in the NV03 situation.

## [6.](#) IANA Considerations

This document requests no action by IANA.

## [7.](#) Acknowledgements



Valuable comments and contributions were made by ... and others.

This document was produced using the xml2rfc tool [[RFC2629](#)].

## [8.](#) Change log [RFC Editor: Please remove]

[draft-carpenter-nvo3-addressing-00](#): original version, 2012-07-04.

## [9.](#) References

### [9.1.](#) Normative References

- [RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", [BCP 5](#), [RFC 1918](#), February 1996.
- [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", [RFC 2460](#), December 1998.
- [RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", [RFC 4193](#), October 2005.

### [9.2.](#) Informative References

- [I-D.kreeger-nvo3-overlay-cp]  
Black, D., Dutt, D., Kreeger, L., Sridhavan, M., and T. Narten, "Network Virtualization Overlay Control Protocol Requirements", [draft-kreeger-nvo3-overlay-cp-00](#) (work in progress), January 2012.
- [I-D.lasserre-nvo3-framework]  
Lasserre, M., Balus, F., Morin, T., Bitar, N., and Y. Rekhter, "Framework for DC Network Virtualization", [draft-lasserre-nvo3-framework-02](#) (work in progress), June 2012.
- [I-D.liu-v6ops-ula-usage-analysis]  
Liu, B., Jiang, S., and C. Byrne, "Analysis and recommendation for the ULA usage", [draft-liu-v6ops-ula-usage-analysis-02](#) (work in progress), March 2012.
- [I-D.narten-nvo3-overlay-problem-statement]  
Narten, T., Sridhavan, M., Dutt, D., Black, D., and L.

Kreeger, "Problem Statement: Overlays for Network Virtualization",  
[draft-narten-nvo3-overlay-problem-statement-02](#) (work in progress), June 2012.

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

[RFC6296] Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Prefix Translation", [RFC 6296](#), June 2011.

#### Authors' Addresses

Brian Carpenter  
Department of Computer Science  
University of Auckland  
PB 92019  
Auckland, 1142  
New Zealand

Email: [brian.e.carpenter@gmail.com](mailto:brian.e.carpenter@gmail.com)

Sheng Jiang  
Huawei Technologies Co., Ltd  
Q14, Huawei Campus  
No.156 Beiqing Road  
Hai-Dian District, Beijing 100095  
P.R. China

Email: [jiangsheng@huawei.com](mailto:jiangsheng@huawei.com)

