Internet Engineering Task Force
Internet Draft
Intended status: Informational
Expires: May 2013

Nabil Bitar
Verizon

Marc Lasserre
Florin Balus
Alcatel-Lucent

Thomas Morin
France Telecom Orange

Lizhong Jin
Bhumip Khasnabish
ZTE

November 28, 2012

# NVO3 Data Plane Requirements
# draft-bl-nvo3-dataplane-requirements-03.txt

## 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), 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/1id-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 May 28, 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.

Abstract

Several IETF drafts relate to the use of overlay networks to support large scale virtual data centers. This draft provides a list of data plane requirements for Network Virtualization over L3 (NVO3) that have to be addressed in solutions documents.

Table of Contents

# 1. Introduction

## 1.1. Conventions used in this document

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].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

## 1.2. General terminology

The terminology defined in [NVO3-framework] is used throughout this document. Terminology specific to this memo is defined here and is introduced as needed in later sections.

DC: Data Center

BUM: Broadcast, Unknown Unicast, Multicast traffic

TS: Tenant System

VAP: Virtual Access Point

VNI: Virtual Network Instance

VNID: VNI ID





## 2. Data Path Overview

The NVO3 framework [NVO3-framework] defines the generic NVE model depicted in Figure 1:

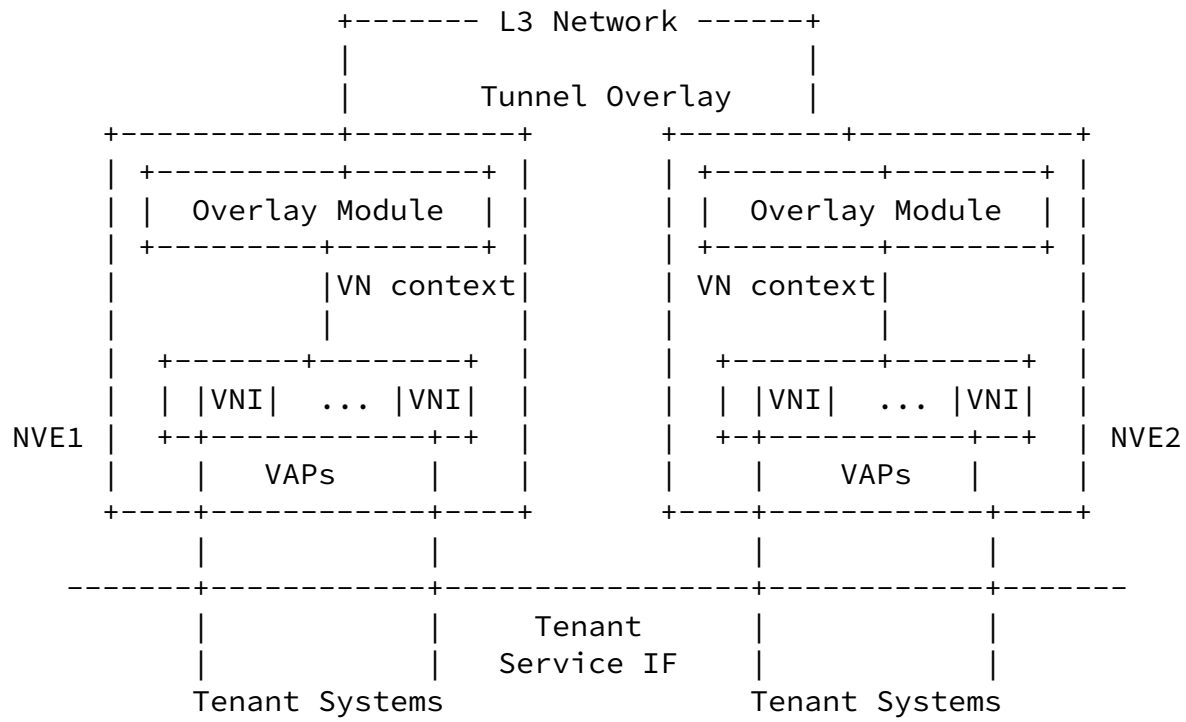


Figure 1 : Generic reference model for NV Edge

When a frame is received by an ingress NVE from a Tenant System over a local VAP, it needs to be parsed in order to identify which virtual network instance it belongs to. The parsing function can examine various fields in the data frame (e.g., VLANID) and/or

associated interface/port the frame came from.

Once a corresponding VNI is identified, a lookup is performed to determine where the frame needs to be sent. This lookup can be based on any combinations of various fields in the data frame (e.g., destination MAC addresses and/or destination IP addresses). Note that additional criteria such as 802.1p and/or DSCP markings might be used to select an appropriate tunnel or local VAP destination.

Lookup tables can be populated using different techniques: data plane learning, management plane configuration, or a distributed control plane. Management and control planes are not in the scope of





this document. The data plane based solution is described in this document as it has implications on the data plane processing function.

The result of this lookup yields the corresponding information needed to build the overlay header, as described in section 3.3. This information includes the destination L3 address of the egress NVE. Note that this lookup might yield a list of tunnels such as when ingress replication is used for BUM traffic.

The overlay header MUST include a context identifier which the egress NVE will use to identify which VNI this frame belongs to.

The egress NVE checks the context identifier and removes the encapsulation header and then forwards the original frame towards the appropriate recipient, usually a local VAP.

## 3. Data Plane Requirements

### 3.1. Virtual Access Points (VAPs)

The NVE forwarding plane MUST support VAP identification through the following mechanisms:

- Using the local interface on which the frames are received, where the local interface may be an internal, virtual port in a VSwitch or a physical port on the ToR
- Using the local interface and some fields in the frame header, e.g. one or multiple VLANs or the source MAC

### 3.2. Virtual Network Instance (VNI)

VAPs are associated with a specific VNI at service instantiation time.

A VNI identifies a per-tenant private context, i.e. per-tenant policies and a FIB table to allow overlapping address space between tenants.

There are different VNI types differentiated by the virtual network service they provide to Tenant Systems. Network virtualization can be provided by L2 and/or L3 VNIs.

#### 3.2.1. L2 VNI

An L2 VNI MUST provide an emulated Ethernet multipoint service as if Tenant Systems are interconnected by a bridge (but instead by using





a set of NVO3 tunnels). The emulated bridge MAY be 802.1Q enabled (allowing use of VLAN tags as a VAP). An L2 VNI provides per tenant virtual switching instance with MAC addressing isolation and L3 tunneling. Loop avoidance capability MUST be provided.

Forwarding table entries provide mapping information between MAC addresses and L3 tunnel destination addresses. Such entries MAY be populated by a control or management plane, or via data plane.

In the absence of a management or control plane, data plane learning MUST be used to populate forwarding tables. As frames arrive from VAPs or from overlay tunnels, standard MAC learning procedures are used: The source MAC address is learned against the VAP or the NVO3 tunnel on which the frame arrived. This implies that unknown unicast traffic be flooded i.e. broadcast.

When flooding is required, either to deliver unknown unicast, or broadcast or multicast traffic, the NVE MUST either support ingress replication or multicast. In this latter case, the NVE MUST be able to build at least a default flooding tree per VNI. In such cases, multiple VNIs MAY share the same default flooding tree. The flooding tree is equivalent with a multicast (*,G) construct where all the NVEs for which the corresponding VNI is instantiated are members. The multicast tree MAY be established automatically via routing and signaling or pre-provisioned.

When tenant multicast is supported, it SHOULD also be possible to select whether the NVE provides optimized multicast trees inside the VNI for individual tenant multicast groups or whether the default VNI flooding tree is used. If the former option is selected the VNI SHOULD be able to snoop IGMP/MLD messages in order to efficiently join/prune Tenant System from multicast trees.

### 3.2.2. L3 VNI

L3 VNIs MUST provide virtualized IP routing and forwarding. L3 VNIs MUST support per-tenant forwarding instance with IP addressing isolation and L3 tunneling for interconnecting instances of the same VNI on NVEs.

In the case of L3 VNI, the inner TTL field MUST be decremented by (at least) 1 as if the NVO3 egress NVE was one (or more) hop(s) away. The TTL field in the outer IP header MUST be set to a value appropriate for delivery of the encapsulated frame to the tunnel exit point. Thus, the default behavior MUST be the TTL pipe model where the overlay network looks like one hop to the sending NVE. Configuration of a "uniform" TTL model where the outer tunnel TTL is





set equal to the inner TTL on ingress NVE and the inner TTL is set to the outer TTL value on egress MAY be supported.

L2 and L3 VNIs can be deployed in isolation or in combination to optimize traffic flows per tenant across the overlay network. For example, an L2 VNI may be configured across a number of NVEs to offer L2 multi-point service connectivity while a L3 VNI can be co-located to offer local routing capabilities and gateway functionality. In addition, integrated routing and bridging per tenant MAY be supported on an NVE. An instantiation of such service may be realized by interconnecting an L2 VNI as access to an L3 VNI on the NVE.

The L3 VNI does not require support for Broadcast and Unknown Unicast traffic. The L3 VNI MAY provide support for customer multicast groups. When multicast is supported, it SHOULD be possible to select whether the NVE provides optimized multicast trees inside the VNI for individual tenant multicast groups or whether a default VNI multicasting tree, where all the NVEs of the corresponding VNI are members, is used.

## 3.3. Overlay Module

The overlay module performs a number of functions related to NVO3 header and tunnel processing.

The following figure shows a generic NVO3 encapsulated frame:

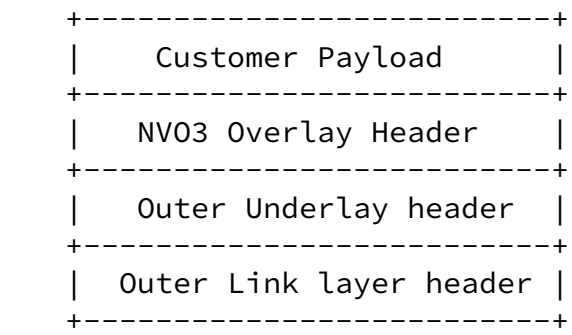


Figure 2 : NVO3 encapsulated frame

where

. Customer payload: Ethernet or IP based upon the VNI type





. NVO3 overlay header: Header containing VNI context information and other optional fields that can be used for processing this packet.

. Outer underlay header: Can be either IP or MPLS

. Outer link layer header: Header specific to the physical transmission link used

### 3.3.1. NVO3 overlay header

An NVO3 overlay header MUST be included after the underlay tunnel header when forwarding tenant traffic. Note that this information can be carried within existing protocol headers (when overloading of specific fields is possible) or within a separate header.

#### 3.3.1.1. Virtual Network Context Identification

The overlay encapsulation header MUST contain a field which allows the encapsulated frame to be delivered to the appropriate virtual network endpoint by the egress NVE. The egress NVE uses this field to determine the appropriate virtual network context in which to process the packet. This field MAY be an explicit, unique (to the administrative domain) virtual network identifier (VNID) or MAY express the necessary context information in other ways (e.g. a locally significant identifier).

It SHOULD be aligned on a 32-bit boundary so as to make it efficiently processable by the data path. It MUST be distributable by a control-plane or configured via a management plane.

In the case of a global identifier, this field MUST be large enough to scale to 100's of thousands of virtual networks. Note that there is no such constraint when using a local identifier.

#### 3.3.1.2. Service QoS identifier

Traffic flows originating from different applications could rely on differentiated forwarding treatment to meet end-to-end availability and performance objectives. Such applications may span across one or more overlay networks. To enable such treatment, support for multiple Classes of Service across or between overlay networks MAY be required.





To effectively enforce CoS across or between overlay networks, NVEs MAY be able to map CoS markings between networking layers, e.g., Tenant Systems, Overlays, and/or Underlay, enabling each networking layer to independently enforce its own CoS policies. For example:

- TS (e.g. VM) CoS

  - Tenant CoS policies MAY be defined by Tenant administrators

  - QoS fields (e.g. IP DSCP and/or Ethernet 802.1p) in the tenant frame are used to indicate application level CoS requirements

- NVE CoS

  - NVE MAY classify packets based on Tenant CoS markings or other mechanisms (eg. DPI) to identify the proper service CoS to be applied across the overlay network

  - NVE service CoS levels are normalized to a common set (for example 8 levels) across multiple tenants; NVE uses per tenant policies to map Tenant CoS to the normalized service CoS fields in the NVO3 header

- Underlay CoS

  - The underlay/core network MAY use a different CoS set (for example 4 levels) than the NVE CoS as the core devices MAY have different QoS capabilities compared with NVEs.

  - The Underlay CoS MAY also change as the NVO3 tunnels pass between different domains.

Support for NVE Service CoS MAY be provided through a QoS field, inside the NVO3 overlay header. Examples of service CoS provided part of the service tag are 802.1p and DE bits in the VLAN and PBB ISID tags and MPLS TC bits in the VPN labels.

### 3.3.2. Tunneling function

This section describes the underlay tunneling requirements. From an encapsulation perspective, IPv4 or IPv6 MUST be supported, both IPv4 and IPv6 SHOULD be supported, MPLS tunneling MAY be supported.





#### 3.3.2.1. LAG and ECMP

For performance reasons, multipath over LAG and ECMP paths SHOULD be supported.

LAG (Link Aggregation Group) [IEEE 802.1AX-2008] and ECMP (Equal Cost Multi Path) are commonly used techniques to perform load-

balancing of microflows over a set of a parallel links either at Layer-2 (LAG) or Layer-3 (ECMP). Existing deployed hardware implementations of LAG and ECMP uses a hash of various fields in the encapsulation (outermost) header(s) (e.g. source and destination MAC addresses for non-IP traffic, source and destination IP addresses, L4 protocol, L4 source and destination port numbers, etc). Furthermore, hardware deployed for the underlay network(s) will be most often unaware of the carried, innermost L2 frames or L3 packets transmitted by the TS. Thus, in order to perform fine-grained load-balancing over LAG and ECMP paths in the underlying network, the encapsulation MUST result in sufficient entropy to exercise all paths through several LAG/ECMP hops. The entropy information MAY be inferred from the NVO3 overlay header or underlay header.

All packets that belong to a specific flow MUST follow the same path in order to prevent packet re-ordering. This is typically achieved by ensuring that the fields used for hashing are identical for a given flow.

All paths available to the overlay network SHOULD be used efficiently. Different flows SHOULD be distributed as evenly as possible across multiple underlay network paths. For instance, this can be achieved by ensuring that some fields used for hashing are randomly generated.

#### 3.3.2.2. DiffServ and ECN marking

When traffic is encapsulated in a tunnel header, there are numerous options as to how the Diffserv Code-Point (DSCP) and Explicit Congestion Notification (ECN) markings are set in the outer header and propagated to the inner header on decapsulation.

[RFC2983] defines two modes for mapping the DSCP markings from inner to outer headers and vice versa. The Uniform model copies the inner DSCP marking to the outer header on tunnel ingress, and copies that outer header value back to the inner header at tunnel egress. The Pipe model sets the DSCP value to some value based on local policy at ingress and does not modify the inner header on egress. Both models SHOULD be supported.





ECN marking MUST be performed according to [RFC6040] which describes the correct ECN behavior for IP tunnels.

#### 3.3.2.3. Handling of BUM traffic

NVO3 data plane support for either ingress replication or point-to-multipoint tunnels is required to send traffic destined to multiple locations on a per-VNI basis (e.g. L2/L3 multicast traffic, L2 broadcast and unknown unicast traffic). It is possible that both methods be used simultaneously.

There is a bandwidth vs state trade-off between the two approaches. User-definable knobs MUST be provided to select which method(s) gets used based upon the amount of replication required (i.e. the number of hosts per group), the amount of multicast state to maintain, the duration of multicast flows and the scalability of multicast protocols.

When ingress replication is used, NVEs MUST track for each VNI the related tunnel endpoints to which it needs to replicate the frame.

For point-to-multipoint tunnels, the bandwidth efficiency is increased at the cost of more state in the Core nodes. The ability to auto-discover or pre-provision the mapping between VNI multicast trees to related tunnel endpoints at the NVE and/or throughout the core SHOULD be supported.

### 3.4. External NVO3 connectivity

NVO3 services MUST interoperate with current VPN and Internet services. This may happen inside one DC during a migration phase or as NVO3 services are delivered to the outside world via Internet or VPN gateways.

Moreover the compute and storage services delivered by a NVO3 domain may span multiple DCs requiring Inter-DC connectivity. From a DC perspective a set of gateway devices are required in all of these cases albeit with different functionalities influenced by the overlay type across the WAN, the service type and the DC network technologies used at each DC site.

A GW handling the connectivity between NVO3 and external domains represents a single point of failure that may affect multiple tenant services. Redundancy between NVO3 and external domains MUST be supported.

### 3.4.1. GW Types

#### 3.4.1.1. VPN and Internet GWs

Tenant sites may be already interconnected using one of the existing VPN services and technologies (VPLS or IP VPN). If a new NVO3 encapsulation is used, a VPN GW is required to forward traffic between NVO3 and VPN domains. Translation of encapsulations MAY be required. Internet connected Tenants require translation from NVO3 encapsulation to IP in the NVO3 gateway. The translation function SHOULD NOT require provisioning touches and SHOULD NOT use intermediate hand-offs, for example VLANs.

#### 3.4.1.2. Inter-DC GW

Inter-DC connectivity MAY be required to provide support for features like disaster prevention or compute load re-distribution. This MAY be provided via a set of gateways interconnected through a WAN. This type of connectivity MAY be provided either through extension of the NVO3 tunneling domain or via VPN GWs.

#### 3.4.1.3. Intra-DC gateways

Even within one DC there may be End Devices that do not support NVO3 encapsulation, for example bare metal servers, hardware appliances and storage. A gateway device, e.g. a ToR, is required to translate the NVO3 to Ethernet VLAN encapsulation.

### 3.4.2. Path optimality between NVEs and Gateways

Within the NVO3 overlay, a default assumption is that NVO3 traffic will be equally load-balanced across the underlying network consisting of LAG and/or ECMP paths. This assumption is valid only as long as: a) all traffic is load-balanced equally among each of the component-links and paths; and, b) each of the component-links/paths is of identical capacity. During the course of normal operation of the underlying network, it is possible that one, or more, of the component-links/paths of a LAG may be taken out-of-service in order to be repaired, e.g.: due to hardware failure of cabling, optics, etc. In such cases, the administrator should configure the underlying network such that an entire LAG bundle in the underlying network will be reported as operationally down if there is a failure of any single component-link member of the LAG bundle, (e.g.: N = M configuration of the LAG bundle), and, thus, they know that traffic will be carried sufficiently by alternate, available (potentially ECMP) paths in the underlying network. This is a likely an adequate assumption for Intra-DC traffic where

presumably the costs for additional, protection capacity along alternate paths is not cost-prohibitive. Thus, there are likely no additional requirements on NVO3 solutions to accommodate this type of underlying network configuration and administration.

There is a similar case with ECMP, used Intra-DC, where failure of a single component-path of an ECMP group would result in traffic shifting onto the surviving members of the ECMP group. Unfortunately, there are no automatic recovery methods in IP routing protocols to detect a simultaneous failure of more than one component-path in a ECMP group, operationally disable the entire ECMP group and allow traffic to shift onto alternative paths. This is problem is attributable to the underlying network and, thus, out-of-scope of any NVO3 solutions.

On the other hand, for Inter-DC and DC to External Network cases that use a WAN, the costs of the underlying network and/or service (e.g.: IPVPN service) are more expensive; therefore, there is a requirement on administrators to both: a) ensure high availability (active-backup failover or active-active load-balancing); and, b) maintaining substantial utilization of the WAN transport capacity at nearly all times, particularly in the case of active-active load-balancing. With respect to the dataplane requirements of NVO3 solutions, in the case of active-backup fail-over, all of the ingress NVE's MUST dynamically adapt to the failure of an active NVE GW when the backup NVE GW announces itself into the NVO3 overlay immediately following a failure of the previously active NVE GW and update their forwarding tables accordingly, (e.g.: perhaps through dataplane learning and/or translation of a gratuitous ARP, IPv6 Router Advertisement, etc.) Note that active-backup fail-over could be used to accomplish a crude form of load-balancing by, for example, manually configuring each tenant to use a different NVE GW, in a round-robin fashion. On the other hand, with respect to active-active load-balancing across physically separate NVE GW's (e.g.: two, separate chassis) an NVO3 solution SHOULD support forwarding tables that can simultaneously map a single egress NVE to more than one NVO3 tunnels. The granularity of such mappings, in both active-backup and active-active, MUST be unique to each tenant.

#### 3.4.2.1. Triangular Routing Issues,a.k.a.: Traffic Tromboning

L2/ELAN over NVO3 service may span multiple racks distributed across

different DC regions. Multiple ELANs belonging to one tenant may be interconnected or connected to the outside world through multiple Router/VRF gateways distributed throughout the DC regions. In this scenario, without aid from an NVO3 or other type of solution, traffic from an ingress NVE destined to External gateways will take





a non-optimal path that will result in higher latency and costs, (since it is using more expensive resources of a WAN). In the case of traffic from an IP/MPLS network destined toward the entrance to an NVO3 overlay, well-known IP routing techniques MAY be used to optimize traffic into the NVO3 overlay, (at the expense of additional routes in the IP/MPLS network). In summary, these issues are well known as triangular routing.

Procedures for gateway selection to avoid triangular routing issues SHOULD be provided. The details of such procedures are, most likely, part of the NVO3 Management and/or Control Plane requirements and, thus, out of scope of this document. However, a key requirement on the dataplane of any NVO3 solution to avoid triangular routing is stated above, in Section 3.4.2, with respect to active-active load-balancing. More specifically, an NVO3 solution SHOULD support forwarding tables that can simultaneously map a single egress NVE to more than one NVO3 tunnels. The expectation is that, through the Control and/or Management Planes, this mapping information MAY be dynamically manipulated to, for example, provide the closest geographic and/or topological exit point (egress NVE) for each ingress NVE.

## 3.5. Path MTU

The tunnel overlay header can cause the MTU of the path to the egress tunnel endpoint to be exceeded.

IP fragmentation SHOULD be avoided for performance reasons.

The interface MTU as seen by a Tenant System SHOULD be adjusted such that no fragmentation is needed. This can be achieved by configuration or be discovered dynamically.

Either of the following options MUST be supported:

- Classical ICMP-based MTU Path Discovery [RFC1191] [RFC1981] or Extended MTU Path Discovery techniques such as defined in

[RFC4821]

- Segmentation and reassembly support from the overlay layer operations without relying on the Tenant Systems to know about the end-to-end MTU

- The underlay network MAY be designed in such a way that the MTU can accommodate the extra tunnel overhead.





### 3.6. Hierarchical NVE

It might be desirable to support the concept of hierarchical NVEs, such as spoke NVEs and hub NVEs, in order to address possible NVE performance limitations and service connectivity optimizations.

For instance, spoke NVE functionality MAY be used when processing capabilities are limited. A hub NVE would provide additional data processing capabilities such as packet replication.

NVEs can be either connected in an any-to-any or hub and spoke topology on a per VNI basis.

### 3.7. NVE Multi-Homing Requirements

Multi-homing techniques SHOULD be used to increase the reliability of an nvo3 network. It is also important to ensure that physical diversity in an nvo3 network is taken into account to avoid single points of failure.

Multi-homing can be enabled in various nodes, from tenant systems into TORs, TORs into core switches/routers, and core nodes into DC GWs.

Tenant systems can either be L2 or L3 nodes. In the former case (L2), techniques such as LAG or STP for instance MAY be used. In the latter case (L3), it is possible that no dynamic routing protocol is enabled. Tenant systems can be multi-homed into remote NVE using several interfaces (physical NICS or vNICS) with an IP address per interface either to the same nvo3 network or into different nvo3 networks. When one of the links fails, the corresponding IP is not reachable but the other interfaces can still be used. When a tenant

system is co-located with an NVE, IP routing can be relied upon to handle routing over diverse links to TORs.

External connectivity MAY be handled by two or more nvo3 gateways. Each gateway is connected to a different domain (e.g. ISP) and runs BGP multi-homing. They serve as an access point to external networks such as VPNs or the Internet. When a connection to an upstream router is lost, the alternative connection is used and the failed route withdrawn.





## 3.8. OAM

NVE MAY be able to originate/terminate OAM messages for connectivity verification, performance monitoring, statistic gathering and fault isolation. Depending on configuration, NVEs SHOULD be able to process or transparently tunnel OAM messages, as well as supporting alarm propagation capabilities.

Given the critical requirement to load-balance NVO3 encapsulated packets over LAG and ECMP paths, it will be equally critical to ensure existing and/or new OAM tools allow NVE administrators to proactively and/or reactively monitor the health of various component-links that comprise both LAG and ECMP paths carrying NVO3 encapsulated packets. For example, it will be important that such OAM tools allow NVE administrators to reveal the set of underlying network hops (topology) in order that the underlying network administrators can use this information to quickly perform fault isolation and restore the underlying network.

The NVE MUST provide the ability to reveal the set of ECMP and/or LAG paths used by NVO3 encapsulated packets in the underlying network from an ingress NVE to egress NVE. The NVE MUST provide the ability to provide a "ping"-like functionality that can be used to determine the health (liveness) of remote NVE's or their VNI's. The NVE SHOULD provide a "ping"-like functionality to more expeditiously aid in troubleshooting performance problems, i.e.: blackholing or other types of congestion occurring in the underlying network, for

NVO3 encapsulated packets carried over LAG and/or ECMP paths.

## 3.9. Other considerations

### 3.9.1. Data Plane Optimizations

Data plane forwarding and encapsulation choices SHOULD consider the limitation of possible NVE implementations, specifically in software based implementations (e.g. servers running VSwitches)

NVE SHOULD provide efficient processing of traffic. For instance, packet alignment, the use of offsets to minimize header parsing, padding techniques SHOULD be considered when designing NVO3 encapsulation types.

The NV03 encapsulation/decapsulation processing in software-based NVEs SHOULD make use of hardware assist provided by NICs in order to speed up packet processing.





### 3.9.2. NVE location trade-offs

In the case of DC traffic, traffic originated from a VM is native Ethernet traffic. This traffic can be switched by a local VM switch or ToR switch and then by a DC gateway. The NVE function can be embedded within any of these elements.

The NVE function can be supported in various DC network elements such as a VM, VM switch, ToR switch or DC GW.

The following criteria SHOULD be considered when deciding where the NVE processing boundary happens:

- Processing and memory requirements
  - Datapath (e.g. lookups, filtering, encapsulation/decapsulation)
  - Control plane processing (e.g. routing, signaling, OAM)
- FIB/RIB size

- Multicast support
    - Routing protocols
    - Packet replication capability
- Fragmentation support
- QoS transparency
- Resiliency

## 4. Security Considerations

This requirements document does not raise in itself any specific security issues.

## 5. IANA Considerations

IANA does not need to take any action for this draft.





## 6. References

### 6.1. Normative References

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

### 6.2. Informative References

[NVOPS] Narten, T. et al, "Problem Statement: Overlays for Network Virtualization", draft-narten-nvo3-overlay-problem-statement (work in progress)

[NVO3-framework] Lasserre, M. et al, "Framework for DC Network Virtualization", draft-lasserre-nvo3-framework (work in progress)

[OVCPREQ] Kreeger, L. et al, "Network Virtualization Overlay Control Protocol Requirements", draft-kreeger-nvo3-overlay-cp (work in progress)

[FLOYD] Sally Floyd, Allyn Romanow, "Dynamics of TCP Traffic over ATM Networks", IEEE JSAC, V. 13 N. 4, May 1995

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC1191] Mogul, J. "Path MTU Discovery", RFC1191, November 1990

[RFC1981] McCann, J. et al, "Path MTU Discovery for IPv6", RFC1981, August 1996

[RFC4821] Mathis, M. et al, "Packetization Layer Path MTU Discovery", RFC4821, March 2007

[RFC2983] Black, D. "Diffserv and tunnels", RFC2983, Cotober 2000

[RFC6040] Briscoe, B. "Tunnelling of Explicit Congestion Notification", RFC6040, November 2010

[RFC6438] Carpenter, B. et al, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", RFC6438, November 2011

[RFC6391] Bryant, S. et al, "Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network", RFC6391, November 2011





## 7. Acknowledgments

In addition to the authors the following people have contributed to this document:

Shane Amante, Level3

Dimitrios Stiliadis, Rotem Salomonovitch, Alcatel-Lucent

Larry Kreeger, Cisco

This document was prepared using 2-Word-v2.0.template.dot.

## Authors' Addresses


Nabil Bitar
Verizon
40 Sylvan Road
Waltham, MA 02145
Email: nabil.bitar@verizon.com

Marc Lasserre
Alcatel-Lucent
Email: marc.lasserre@alcatel-lucent.com

Florin Balus
Alcatel-Lucent
777 E. Middlefield Road
Mountain View, CA, USA 94043
Email: florin.balus@alcatel-lucent.com

Thomas Morin
France Telecom Orange
Email: thomas.morin@orange.com

Lizhong Jin
ZTE
Email : lizhong.jin@zte.com.cn

Bhumip Khasnabish
ZTE
Email : Bhumip.khasnabish@zteusa.com