Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: August 18, 2014

E. Gray, Ed.
Ericsson
N. Bitar
Verizon
X. Chen
Huawei Technologies
M. Lasserre
Alcatel-Lucent
T. Tsou
Huawei Technologies (USA)
February 14, 2014

NVO3 Gap Analysis - Requirements Versus Available Technology Choices
draft-ietf-nvo3-gap-analysis-01

Abstract

This document evaluates candidate protocols against the NVO3 requirements. Gaps are identified and further work recommended.

Table of Contents

1. Introduction

The initial charter of the NVO3 Working Group requires it to identify any gaps between the requirements identified and available technoloogy solutions as a prerequisite to rechartering or concluding the Working Group (if no gaps exist). This document is intended to provide the required gap analysis.

This document provides a tabulation of candidate solutions and their ability to satisfy each requirement identified by the Working Group.

Areas of work are identified where further work is required to ensure that the requirements are met.

The major areas covered in this document include:

- o Operational Requirements [I-D.ashwood-nvo3-operational-requirement]

- o Management Requirements (TBD)

- o Control (Plane) Requirements [I-D.ietf-nvo3-nve-nva-cp-req]

- o Dataplane Requirements [I-D.ietf-nvo3-dataplane-requirements]

Since the Working Group has yet to complete (and in some cases adopt) documents describing requirements for some of these areas, not all areas are complete in the present version of this document.

The initial candidate technologies are:

- o NVGRE [I-D.sridharan-virtualization-nvgre],

- o VxLAN [I-D.mahalingam-dutt-dcops-vxlan],

- o L2VPN: VPLS [RFC4761][RFC4762] and EVPN [I-D.ietf-l2vpn-evpn], and

- o L3VPN [RFC4365].

2. Terminology and Conventions

2.1. Requirements Language

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].

2.2. Conventions

In sections providing analysis of requirements defined in referenced documents, section numbers from each referenced document are used as they were listed in that document.

In order to avoid confusing those section numbers with the section numbering in this document, the included numbering is parenthesized.

L2VPN is represented (in tables and analysis, as a technology) by the two differing approaches: VPLS and EVPN.

2.3. Terms and Abbreviations

This document uses terms and acronyms defined in [RFC3168], [I-D.ietf-nvo3-framework], [I-D.ietf-nvo3-dataplane-requirements], [I-D.kreeger-nvo3-hypervisor-nve-cp] and [I-D.ietf-nvo3-nve-nva-cp-req]. Acronyms are included here for convenience but are meant to remain aligned with definitions in the references included.

ECN: Explicit Congestion Notification [RFC3168]

NVA: Network Virtualization Authority [I-D.ietf-nvo3-nve-nva-cp-req]

NVE: Network Virtualization Edge [I-D.ietf-nvo3-framework]

VAP: Virtual Access Point [I-D.ietf-nvo3-dataplane-requirements]

VNI: Virtual Network Instance [I-D.ietf-nvo3-framework]

VNIC: Virtual Network Interface Card (NIC) [I-D.kreeger-nvo3-hypervisor-nve-cp]

VNID: Virtual Network Identifier [I-D.ietf-nvo3-nve-nva-cp-req]

This document uses the following additional general terms and abbreviations:

DSCP: Differentiated Services Code-Point

ECMP: Equal Cost Multi-Path

L2VPN: Layer 2 Virtual Private Network

L3VPN: Layer 3 Virtual Private Network

NVO3: Network Virtualization Overlay over L3

VM: Virtual Machine

VN: Virtual Network

3. Operational Requirements

TBD

4. Management Requirements

TBD

5. Control Plane Requirements

The NVO3 Problem Statement [I-D.ietf-nvo3-overlay-problem-statement], describes 3 categories of control functions:

1. Control functions associated with implementing the Network Virtualization Authority (e.g. - signaling and control required for interactions between multiple NVA devices).

2. Control functions associated with interactions between an NVA and a Network Virtualization Edge (NVE).

3. Control functions associated with attaching and detaching a Virtual Machine (VM) from a particular Virtual Network Instance (VNI).

As sometimes happens, there is not a 1:1 mapping of the work areas defined in [I-D.ietf-nvo3-overlay-problem-statement] and requirements documents intended to address the problems that have been identified there.

Current control-plane requirement documents include the following:

- NVE-NVA control-plane requirements [I-D.ietf-nvo3-nve-nva-cp-req]

- Control-plane requirements specific to VM-to-NVE interactions [I-D.kreeger-nvo3-hypervisor-nve-cp]

In the following subsections, we consider the data-plane candidate solutions and proposed or existing control plane solutions that may apply to each.

In each case, the control-plane solutions can be divided into support for Layer-2 and Layer-3 services, and for each of these cases, the data-plane solutions considered will be limited to those services and solutions that make sense for that case.

Tables are thus organized into separate tables for both L2 and L3 data/control service options. It may turn out that - for all potential control-plane solutions - each solution applies equally to all data-plane solutions considered for the layer applicable.

If this turns out to be the case, then the tables may be further simplified - possibly by reducing each pair of L2/L3 tables to a single table where the columns are simply "Layer-2" and "Layer-3."

The intent is to show potential mapping of data-plane to applicable control-plane alternatives and evaluate each applicable control-plane against defined control-plane requirements.

The way this document attempts to do this is to list the control planes that may be applicable to each of the candidate data-planes in table footnotes and then stating in table footnotes the extent to which candidate control plane technologies satisfy each requirement.

As with tables in other sections of this draft, the rows in each table list the applicable requirements found in analogous sections of applicable requirements documents.

5.1. NVE-NVA Control-Plane Requirements

In this section, numbering of requirement headings corresponds to section numbering in [I-D.ietf-nvo3-nve-nva-cp-req].

(3.1) Inner to Outer Address Mapping

The requirements document [I-D.ietf-nvo3-nve-nva-cp-req] states that avoiding the need to "flood" traffic to support learning of mapping information from the data-plane is a goal of NVO3 candidate technological approaches.

For each candidate technology, (how) is the mapping of header information present in tenant traffic mapped to corresponding header information to be used in overlay encapsulation (this includes addresses, context identification, etc.) determined?

Supported Approach	VxLAN	VPLS	EVPN
Control Protocol Mapping Acquisition? - - - Data-Plane Learning?	- - -	- - -	- -

Table 1: Inner:Outer Address Mapping (L2)

Supported Approach	NVGRE	L3VPN
Control Protocol Mapping Acquisition? - - - Data-Plane Learning?	- - -	- - -

Table 2: Inner:Outer Address Mapping (L3)

(3.2) Underlying Network Multi-Destination Address(es)

The requirements document [I-D.ietf-nvo3-nve-nva-cp-req] lists 3 approaches that may be used to deliver traffic to multiple destinations in an overlay virtual network:

1. Use the capabilities of the underlay network.

2. Require a sending NVE to replicate traffic.

3. Use a replication service provided within the overlay network.

For each delivery approach, it may be necessary to map specific multipoint (e.g. - broadcast, unknown destination or multicast) traffic to (for instance) addresses used to deliver this traffic via the underlay network.

For each technological approach, which delivery approaches are supported and does the technology provide a method by which an NVE needing to send multi-destination traffic can determine to what address, or addresses to which to send this traffic?

Supported Approach	VxLAN	VPLS	EVPN
Underlay Network Capability			
- - -	- - -	- - -	- - -
NVE Sender Replication			
- - -	- - -	- - -	- - -
Replication Service			

Table 3: Multi-Destination Delivery (L2)

Supported Approach	NVGRE	L3VPN
Underlay Network Capability		
- - -	- - -	- - -
NVE Sender Replication		
- - -	- - -	- - -
Replication Service		

Table 4: Multi-Destination Delivery (L3)

(3.3) VN Connect/Disconnect Notification

The requirements document [I-D.ietf-nvo3-nve-nva-cp-req] states as an assumption that a mechanism exists in the overlay technology by which an NVE is notified of Tenant Systems attaching and detaching from a specific Virtual Network (VN).

For each candidate technology, does the technology currently support these functions?

Requirement	VxLAN	VPLS	EVPN
Connect Notification - - - Disconnect Notification	- - -	- - -	- - -

Table 5: Connect/Disconnect Notification (L2)

Requirement	NVGRE	L3VPN
Connect Notification - - - Disconnect Notification	- - -	- - -

Table 6: Connect/Disconnect Notification (L3)

(3.4) VN Name to VNID Mapping

The requirements document [I-D.ietf-nvo3-nve-nva-cp-req] concludes that having a means to map for a "VN Name to a "VN ID" may be useful.

For each technological approach we are considering, is this function currently available?

Function	VxLAN	VPLS	EVPN
VN-Name:VN-ID Mapping			

Table 7: VN Name to VN ID Mapping (L2)

Function	NVGRE	L3VPN
VN-Name:VN-ID Mapping		

Table 8: VN Name to VN ID Mapping (L3)

5.2. VM-to-NVE Specific Control-Plane Requirements

In this section, numbering of requirement headings corresponds to section numbering in [I-D.kreeger-nvo3-hypervisor-nve-cp].

(4.1) VN Connect/Disconnect

The requirements document [I-D.kreeger-nvo3-hypervisor-nve-cp] states as a requirement that a mechanism must exist by which an NVE is notified when an end device requires a connection, or no longer requires a connection, to a specific Virtual Network (VN).

The requirements document further states as a requirement that the mechanism(s) used in a candidate technological approach must provide a local indicator (e.g. - 802.1Q tag) that the end device will use in sending traffic to, or receiving traffic from, the NVE (where that traffic is associated with the connected VN).

As an additional related requirement, the requirements document states that the NVE - once notified of a connection to a VN (by VN Name), needs to have a means for getting associated VN context information from the NVA.

For each candidate technology, does the technology currently support these functions?

Requirement	VxLAN	VPLS	EVPN
Connect Notification			
- - -	- - -	- - -	- - -
Local VN Indicator			
- - -	- - -	- - -	- - -
VN Name to VN Context Mapping			
- - -	- - -	- - -	- - -
Disconnect Notification			

Table 9: VN Connect/Disconnect (L2)

Requirement	NVGRE	L3VPN
Connect Notification		
- - -	- - -	- - -
Local VN Indicator		
- - -	- - -	- - -
VN Name to VN Context Mapping		
- - -	- - -	- - -
Disconnect Notification		

Table 10: VN Connect/Disconnect (L3)

(4.2) VNIC Address Association

The requirements document [I-D.kreeger-nvo3-hypervisor-nve-cp] lists two approaches for acquiring VNIC address association information:

1. Data Plane Learning (i.e. - by inspecting source addresses in traffic received from an end device).

2. Explicit signaling from the end device when a specific VNIC address is to be associated with a tenant system.

Supported Approaches	VxLAN	VPLS	EVPN
Data Plane Learning			
- - -	- - -	- - -	- - -
Explicit Signaling			

Table 11: VNIC Address Association (L2)

Supported Approaches	NVGRE	L3VPN
Data Plane Learning		
- - -	- - -	- - -
Explicit Signaling		

Table 12: VNIC Address Association (L3)

(4.3) VNIC Address Disassociation

TBD

(4.4) VNIC Shutdown/Startup/Migration

TBD

(4.5) VN Profile

TBD

6. Data Plane Requirements

In this section, numbering of requirement headings corresponds to section numbering in [I-D.ietf-nvo3-dataplane-requirements].

(3.1) Virtual Access Points (VAPs)

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
MUST support VAP identification					
- - -	- - -	- - -	- -	- -	- - -
1) Local interface	YES				
- - -	- - -	- - -	- -	- -	- - -
2) Local interface + fields in frame header	YES				

Table 13: VAP Identification Requirements

(3.2) Virtual Network Instance (VNI)

Network virtualization can be provided by L2 and/or L3 VNIs.

(3.2.1) L2 VNI

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
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).					
- - -	- - -	- - -	- -	- -	- - -
Loop avoidance capability MUST be provided.					
- - -	- - -	- - -	- -	- -	- - -
Data plane learning MUST be supported as the default means to populate forwarding tables.					
- - -	- - -	- - -	- -	- -	- - -
When flooding is required for delivery of broadcast, unknown unicast or multicast (BUM) traffic, the NVE MUST either support ingress replication or multicast.					
- - -	- - -	- - -	- -	- -	- - -
If using multicast, the NVE MUST be able to build at least one default flooding tree for use by local VNIs for flooding to NVEs belonging to the same VN.					

Table 14: L2 VNI Service

(3.2.2) L3 VNI

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
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.					
- - -	- - -	- - -	- -	- -	- - -
For L3 VNI, the inner TTL field MUST be decremented by at least 1 (as if the NVO3 egress was at least 1 hop away).					
- - -	- - -	- - -	- -	- -	- - -
TTL in the outer IP header MUST be set to a value appropriate for delivery of the encapsulated packet to the tunnel exit point.					
- - -	- - -	- - -	- -	- -	- - -
The default behavior for TTL MUST use the "pipe" model.					

Table 15: L3 VNI Service

(3.3.1) NVO3 overlay header

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
An NVO3 overlay header MUST be included after the underlay tunnel header when forwarding tenant traffic.	YES	YES	YES	YES	YES

Table 16: Overlay Header

(3.3.1.1) Virtual Network Context Identification

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
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. .	YES	YES	YES	YES	YES
- - -	- - -	- - -	- -	- -	- - -
If Global Identifiers are used, the identifier field MUST be large enough to scale to hundreds of thousands of VNs.					

Table 17: Virtual Network Context Identification

(3.3.2.1) LAG and ECMP

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
In order to perform fine-grained load-balancing, the data-plane encapsulation MUST result in sufficient entropy to exercise all paths through several LAG/ECMP hops.					
- - -	- - -	- - -	- -	- -	- - -
All packets belonging to any specific flow MUST follow the same path in order to prevent packet re-ordering.	NO				

Table 18: Multipath Support

(3.3.2.2) DiffServ and ECN marking

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
[RFC2983] defines two modes for mapping the DSCP markings from inner to outer headers and vice versa. Both models SHOULD be supported.	NO				

Table 19: DSCP and ECN Marking

(3.3.2.3) Handling of broadcast, unknown unicast, and multicast 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.

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
User-configurable knobs MUST be provided to select which method(s) are used based upon the amount of replication required.					
- - -	- - -	- - -	- -	- -	- - -
When ingress replication is used, NVEs MUST track maintain (for each VNI) the related tunnel endpoints to which it needs to replicate the frame.					

Table 20: Handling of Broadcast, Unknown Unicast, and Multicast Traffic

(3.4) External NVO3 connectivity

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
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.	YES				
Redundancy between NVO3 and external domains MUST be supported.					

Table 21: Interoperation

(3.4.2.1) Load-balancing

When using 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.

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
The granularity of such mappings, in both active-backup and active-active, MUST be specific to each tenant.					

Table 22: Gateway Load-balancing

(3.5) Path MTU

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
Classical ICMP-based MTU Path Discovery ([RFC1191], [RFC1981]) or Extended MTU Path Discovery techniques such as defined in [RFC4821].	NO				
- - -	- - -	- - -	- -	- -	- - -
Fragmentation and reassembly support from the overlay layer operations without relying on the Tenant Systems to know about the end-to-end MTU.	YES				

Table 23: Path MTU

(3.7) NVE Multi-Homing Requirements

Requirement	NVGRE	VxLAN	VPLS	EVPN	L3VPN
Multi-homing techniques SHOULD be used to increase the reliability of an NVO3 network.	NO				

Table 24: Multihoming

7. Summary and Conclusions

TBD

8. Acknowledgements

The Authors would like to acknowledge the technical contributions of Florin Balus, Luyuan Fang, Sue Hares, Wim Henderickx, Yves Hertoghs, Yuichi Ikejiri, Rangaraju Iyengar, Mircea Pisica, Evelyn Roch, Ali Sajassi, Peter Ashwood-Smith and Lucy Yong as well as the initial help in editing the XML source for the document from Tom Taylor.

9. IANA Considerations

This memo includes no request to IANA.

10. Security Considerations

Security considerations of the requirements documents referenced by this analysis document apply.

11. References

11.1. Normative References

[I-D.ashwood-nvo3-operational-requirement]
Ashwood-Smith, P., Iyengar, R., Tsou, T., Sajassi, A., Boucadair, M., Jacquenet, C., and M. Daikoku, "NVO3 Operational Requirements", draft-ashwood-nvo3-operational-requirement-03 (work in progress), July 2013.

[I-D.hertoghs-nvo3-lisp-controlplane-unified]
Hertoghs, Y., Maino, F., Moreno, V., Smith, M., Farinacci, D., and L. Iannone, "A Unified LISP Mapping Database for L2 and L3 Network Virtualization Overlays", draft-hertoghs-nvo3-lisp-controlplane-unified-01 (work in progress), February 2014.

[I-D.ietf-l2vpn-evpn]
Sajassi, A., Aggarwal, R., Henderickx, W., Isaac, A., and J. Uttaro, "BGP MPLS Based Ethernet VPN", draft-ietf-l2vpn-evpn-05 (work in progress), February 2014.

[I-D.ietf-nvo3-dataplane-requirements]
Bitar, N., Lasserre, M., Balus, F., Morin, T., Jin, L., and B. Khasnabish, "NVO3 Data Plane Requirements", draft-ietf-nvo3-dataplane-requirements-02 (work in progress), November 2013.

[I-D.ietf-nvo3-framework]
Lasserre, M., Balus, F., Morin, T., Bitar, N., and Y. Rekhter, "Framework for DC Network Virtualization", draft-ietf-nvo3-framework-05 (work in progress), January 2014.

[I-D.ietf-nvo3-nve-nva-cp-req]
Kreeger, L., Dutt, D., Narten, T., and D. Black, "Network Virtualization NVE to NVA Control Protocol Requirements", draft-ietf-nvo3-nve-nva-cp-req-01 (work in progress), October 2013.

[I-D.ietf-nvo3-overlay-problem-statement]
Narten, T., Gray, E., Black, D., Fang, L., Kreeger, L., and M. Napierala, "Problem Statement: Overlays for Network Virtualization", draft-ietf-nvo3-overlay-problem-statement-04 (work in progress), July 2013.

[I-D.kreeger-nvo3-hypervisor-nve-cp]
Kreeger, L., Narten, T., and D. Black, "Network Virtualization Hypervisor-to-NVE Overlay Control Protocol Requirements", draft-kreeger-nvo3-hypervisor-nve-cp-01 (work in progress), February 2013.

[I-D.mahalingam-dutt-dcops-vxlan]
Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "VXLAN: A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", draft-mahalingam-dutt-dcops-vxlan-08 (work in progress), February 2014.

[I-D.sridharan-virtualization-nvgre]
Sridharan, M., Greenberg, A., Wang, Y., Garg, P., Venkataramiah, N., Duda, K., Ganga, I., Lin, G., Pearson, M., Thaler, P., and C. Tumuluri, "NVGRE: Network Virtualization using Generic Routing Encapsulation", draft-sridharan-virtualization-nvgre-04 (work in progress), February 2014.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, November 1990.

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.

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

[RFC2983] Black, D., "Differentiated Services and Tunnels", RFC 2983, October 2000.

[RFC4365] Rosen, E., "Applicability Statement for BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4365, February 2006.

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

[RFC4821] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", RFC 4821, March 2007.

[RFC6040] Briscoe, B., "Tunnelling of Explicit Congestion Notification", RFC 6040, November 2010.

11.2. Informative References

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, September 2001.

Authors' Addresses

Eric Gray (editor)
Ericsson
120 Morris Avenue
Pitman, New Jersey 08071
USA

Email: eric.gray@ericsson.com

Nabil Bitar
Verizon
40 Sylvan Road
Waltham, Massachusetts 02145
USA

Email: nabil.bitar@verizon.com

Xiaoming Chen
Huawei Technologies

Email: ming.chen@huawei.com

Marc Lasserre
Alcatel-Lucent

Email: marc.lasserre@alcatel-lucent.com

Tina Tsou
Huawei Technologies (USA)
2330 Central Expressway
Santa Clara, California 95050
USA

Phone: +1 408 330 4424
Email: Tina.Tsou.Zouting@huawei.com
URI: http://tinatsou.weebly.com/contact.html