Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: January 16, 2014

P. Ashwood-Smith
Huawei Technologies
R. Iyengar
T. Tsou
Huawei Technologies USA
A. Sajassi
Cisco Technologies
M. Boucadair
C. Jacquenet
France Telecom
M. Daikoku
KDDI corporation

July 15, 2013

NVO3 Operational Requirements
draft-ashwood-nvo3-operational-requirement-03

Abstract

This document provides framework and requirements for Network Virtualization over Layer 3 (NVO3) Operations, Administration, and Maintenance (OAM). This document for the most part gathers requirements from existing IETF drafts and RFCs which have already extensively studied this subject for different data planes and layering. As a result this draft is high level and broad. We begin to ask which are truly required for NVO3 and expect the list to be narrowed by the working group as subsequent versions of this draft are created.

Table of Contents

1. Introduction

This document provides framework and requirements for Network virtualization over Layer 3(NVO3) Operation, Administration, and Maintenance (OAM). Given that this OAM subject is far from new and has been under extensive investigation by various IETF working groups (and several other standards bodies) for many years, this document draws from existing work, starting with [RFC6136]. As a result, sections of [RFC6136] have been reused with minor changes with the permission of the authors.

NVO3 OAM requirements are expected to be a subset of IETF/IEEE etc. work done so far; however, we begin with a full set of requirements and expect to prune them through several iterations of this document.

1.1. OSI Definitions of OAM

The scope of OAM for any service and/or transport/network infrastructure technologies can be very broad in nature. OSI has defined the following five generic functional areas commonly abbreviated as "FCAPS" [NM-Standards]:

- Fault Management,
- Configuration Management,
- Accounting Management,
- Performance Management, and
- Security Management.

This document focuses on the Fault, Performance and to a limited extent the Configuration Management aspects. Other functional aspects of FCAPS and their relevance (or not) to NVO3 are for further study.

Fault Management can typically be viewed in terms of the following categories:

- Fault Detection;

- Fault Verification;
- Fault Isolation;
- Fault Notification and Alarm Suppression;
- Fault Recovery.

Fault detection deals with mechanism(s) that can detect both hard failures such as link and device failures, and soft failures, such as software failure, memory corruption, misconfiguration, etc. Fault detection relies upon a set of mechanisms that first allow the observation of an event, then the use of a protocol to dynamically notify a network/system operator (or management system) about the event occurrence, then the use of diagnostic tools to assess the nature and severity of the fault.

After verifying that a fault has occurred along the data path, it is important to be able to isolate the fault to the level of a given device or link. Therefore, a fault isolation mechanism is needed in Fault Management. A fault notification mechanism should be used in conjunction with a fault detection mechanism to notify the devices upstream and downstream to the fault detection point. The fault notification mechanism should also notify NMS systems.

> The terms "upstream" and "backward" are used here to denote the direction(s) from which data traffic is flowing. The terms "downstream" and "forward" denote the direction(s) to which data traffic is forwarded.

For example, when there is a client/server relationship between two layered networks (e.g., the NVO3 layer is a client of the outer IP server layer, while the inner IP layer is a client of the NVO3 server layer 2), fault detection at the server layer may result in the following fault notifications:

- Sending a forward fault notification from the server layer to the client layer network(s) using the fault notification format appropriate to the client layer.
- Sending a backward fault notification to the server layer, if applicable, in the reverse direction.
- Sending a backward fault notification to the client layer, if applicable, in the reverse direction.

Finally, fault recovery deals with recovering from the detected failure by switching to an alternate available data path (depending

on the nature of the fault) using alternate devices or links. In fact, the controller can provision another virtual network, thus automatically resolving the reported problem.

The controller may also directly monitor the status of virtual network components such as Network Virtualization Edge elements (NVEs) [NVO3-framework] in order to respond to their failures. In addition to forward and backward fault notifications, the controller may deliver notifications to a higher level orchestration component, e.g., one responsible for Virtual Machine (VM) provisioning and management.

Note, given that the IP network on which NVO3 resides is usually self healing, it is expected that recovery by the NVO3 layer would not normally be required, although there may be a requirement for that layer to log that the problem has been detected and resolved. The special cases of a static IP overlay network, or possibly of a centrally controlled IP overlay network, may, however, require NVO3 involvement in fault recovery.

Performance Management deals with mechanism(s) that allow determining and measuring the performance of the network/services under consideration. Performance Management can be used to verify the compliance to both the service-level and network-level metric objectives/specifications. Performance Management typically consists of measuring performance metrics, e.g., Frame Loss, Frame Delay, Frame Delay Variation (aka Jitter), Frame throughput, Frame discard, etc., across managed entities when the managed entities are in available state. Performance Management is suspended across unavailable managed entities.

1.2. 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 [RFC2119].

1.3. Relationship with Other OAM Work

This document leverages requirements that originate with other OAM work, specifically the following:

- [RFC6136] provides a template and some of the high level requirements and introductory wording.
- [IEEE802.1ag] is expected to provide a subset of the requirements for NVO3 both at the Tenant level and also within the L3 Overlay network.

- o [Y.1731] is expected to provide a subset of the requirements for NVO3 at the Tenant level.

- o Section 3.8 of [NVO3-DP-Reqs] lists several requirements specifically concerning ECMP/LAG.

2. Terminology

The terminology defined in [NVO3-framework] and [NVO3-DP-Reqs] is used throughout this document. We introduce no new terminology.

3. NVO3 Reference Model

Figure 1 below reproduces the generic NVO3 reference model as per [NVO3-framework].

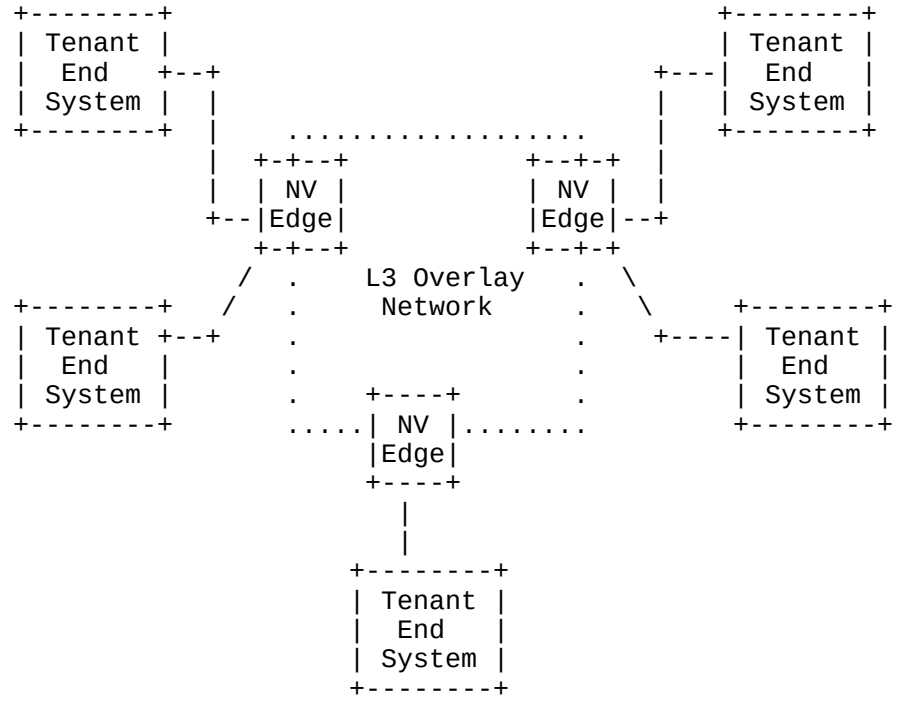

Figure 1: Generic reference model for DC network virtualization over a Layer3 infrastructure

Figure 2 below, reproduces the Generic reference model for the NV Edge (NVE) as per [NVO3-DP-Reqs].

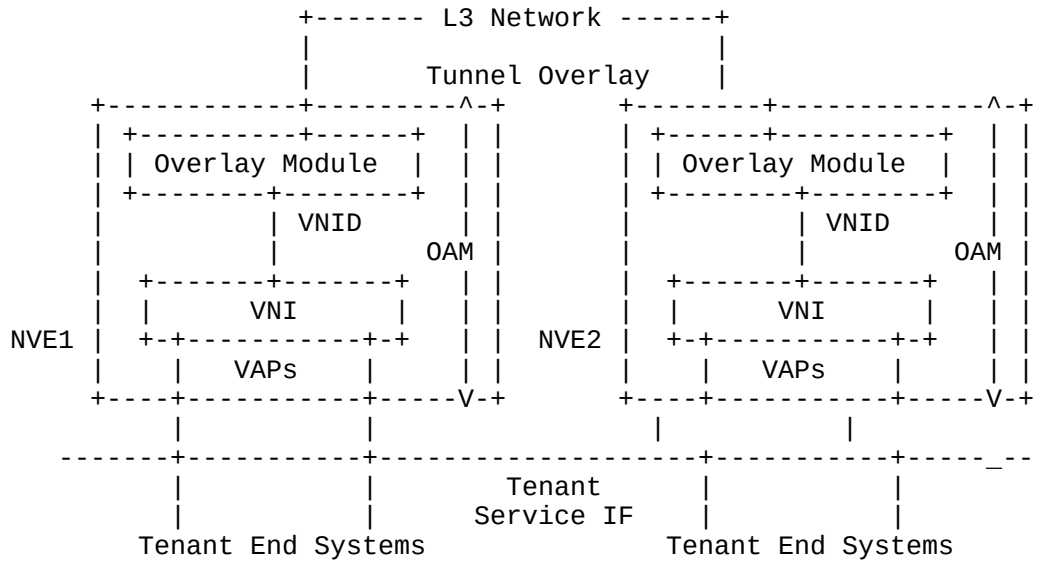

Figure 2: Generic reference model for NV Edge

4. OAM Framework for NVO3

Figure 1 showed the generic reference model for a DC network virtualization over an L3 (or L3VPN) infrastructure while Figure 2 showed the generic reference model for the Network Virtualization (NV) Edge.

L3 network(s) or L3 VPN networks (either IPv6 or IPv4, or a combination thereof), provide transport for an emulated layer 2 created by NV Edge devices. Unicast and multicast tunneling methods (de-multiplexed by Virtual Network Identifier (VNID)) are used to provide connectivity between the NV Edge devices. The NV Edge devices then present an emulated layer 2 network to the Tenant End Systems at a Virtual Network Interface (VNI) through Virtual Access Points (VAPs). The NV Edge devices map layer 2 unicast to layer 3 unicast point-to-point tunnels and may either map layer 2 multicast to layer 3 multicast tunnels or may replicate packets onto multiple layer 3 unicast tunnels.

4.1. OAM Layering

The emulated layer 2 network is provided by the NV Edge devices to which the Tenant End Systems are connected. This network of NV Edges can be operated by a single service provider or can span across multiple administrative domains. Likewise, the L3 Overlay Network can be operated by a single service provider or span across multiple administrative domains.

While each of the layers is responsible for its own OAM, each layer may consist of several different administrative domains. Figure 3 shows an example.

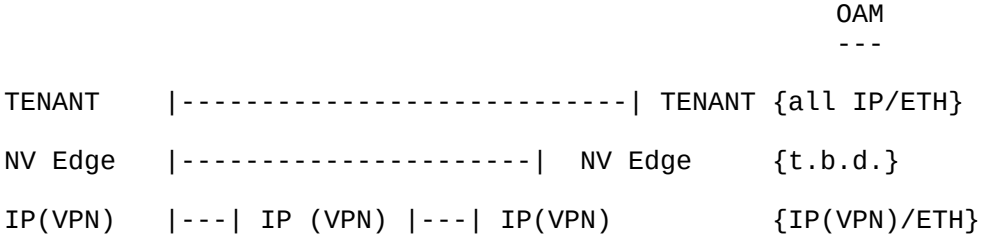

Figure 3: OAM layers in an NVO3 network

For example, at the bottom, at the L3 IP overlay network layer IP(VPN) and/or Ethernet OAM mechanisms are used to probe link by link, node to node etc. OAM addressing here means physical node loopback or interface addresses.

Further up, at the NV Edge layer, NVO3 OAM messages are used to probe the NV Edge to NV Edge tunnels and NV Edge entity status. OAM addressing here likely means the physical node loopback together with the VNI (to de-multiplex the tunnels).

Finally, at the Tenant layer, the IP and/or Ethernet OAM mechanisms are again used but here they are operating over the logical L2/L3 provided by the NV-Edge through the VAP. OAM addressing at this layer deals with the logical interfaces on Vswitches and Virtual Machines.

4.2. OAM Domains

Complex OAM relationships exist as a result of the hierarchical layering of responsibility and of breaking up of end-to-end responsibility.

The OAM domain above NVO3, is expected to be supported by existing IP and L2 OAM methods and tools.

The OAM domain below NVO3, is expected to be supported by existing IP /L2 and MPLS OAM methods and tools. Where this layer is actually multiple domains spliced together, the existing methods to deal with these boundaries are unchanged. Note however that exposing LAG/ECMP detailed behavior may result in additional requirements to this domain, the details of which will be specified in the future versions of this draft.

When we refer to an OAM domain in this document, or just 'domain', we therefore refer to a closed set of NV Edges and the tunnels which interconnect them. Inter-domain OAM considerations will be specified in the future versions of this draft.

5. NVO3 OAM Requirements

The following numbered requirements originate from [RFC6136]. All are included however where they seem obviously not relevant (to the present authors) an explanation as to why is included.

5.1. Discovery

R1) NVO3 OAM MUST allow an NV Edge device to dynamically discover other NV Edge devices that share the same VNI within a given NVO3 domain. This may be based on a discovery mechanism used to set up data path forwarding between NVEs.

5.2. Connectivity Fault Management

5.2.1. Connectivity Fault Detection

R2) NVO3 OAM MUST allow proactive connectivity monitoring between two or more NV Edge devices that support the same VNIs within a given NVO3 domain. NVO3 OAM MAY act as a protection trigger. That is, automatic recovery from transmission facility failure by switchover to a redundant replacement facility may be triggered by notifications from NVO3 OAM.

R3) NVO3 OAM MUST allow monitoring/tracing of all possible paths in the underlay network between a specified set of two or more NV Edge devices. Using this feature, equal cost paths that traverse LAG and/ or ECMP may be differentiated.

5.2.2. Connectivity Fault Verification

R4) NVO3 OAM MUST allow connectivity fault verification between two or more NV Edge devices that support the same VNI within a given NVO3 domain.

5.2.3. Connectivity Fault localization

R5) NVO3 OAM MUST allow connectivity fault localization between two or more NV Edge devices that support the same VNI within a given NVO3 domain.

5.2.4. Connectivity Fault Notification and Alarm Suppression

R6) NVO3 OAM MUST support fault notification to be triggered as a result of the faults occurring in the underneath network infrastructure. This fault notification SHOULD be used for the suppression of redundant service-level alarms.

5.3. Frame Loss

R7) NVO3 OAM MUST support measurement of per VNI frame loss between two NV Edge devices that support the same VNI within a given NVO3 domain.

5.4. Frame Delay

R8) NVO3 OAM MUST support measurement of per VNI two-way frame delay between two NV edge devices that support the same VNI within a given NVO3 domain.

R9) NVO3 OAM MUST support measurement of per VNI one-way frame delay between two NV Edge devices that support the same VNI within a given NVO3 domain.

5.5. Frame Delay Variation

R10) NVO3 OAM MUST support measurement of per VNI frame delay variation between two NV Edge devices that support the same VNI within a given NVO3 domain.

5.6. Frame Throughput

R11) NVO3 OAM MAY [*** Should this be stronger? ***] support measurement of per VNI frame throughput throughput (in frames and bytes) between two NV Edge devices that support the same VNI within a given NVO3 domain. This feature could be an effective way to confirm whether or not assigned path bandwidth conforms to service level agreement before providing the path between two NV Edge devices.

5.7. Frame Discard

R12) NVO3 OAM MAY support measurement of per VNI frame discard between two NV Edge devices that support the same VNI within a given

NVO3 domain. This feature MAY be effective to monitor bursty traffic between two NV Edge devices.

5.8. Availability

A service may be considered unavailable if the service frames/packets do not reach their intended destination (e.g., connectivity is down) or the service is degraded (e.g., frame loss and/or frame delay and/or delay variation threshold is exceeded). Entry and exit conditions may be defined for the unavailable state. Availability itself may be defined in the context of a service type. Since availability measurement may be associated with connectivity, frame loss, frame delay, and frame delay variation measurements, no additional requirements are specified currently.

5.9. Data Path Forwarding

R13) NVO3 OAM frames MUST be forwarded along the same path (i.e., links (including LAG members) and nodes) as the NVO3 data frames.

R14) NVO3 OAM frames MUST provide a mechanism to exercise/trace all data paths that result due to ECMP/LAG hops in the underlay network.

5.10. Scalability

R15) NVO3 OAM MUST be scalable such that an NV edge device can support proactive OAM for each VNI that is supported by the device. (Note - Likely very hard to achieve with hash based ECMP/LAG).

5.11. Extensibility

R16) NVO3 OAM should be extensible such that new functionality and information elements related to this functionality can be introduced in the future.

R17) NVO3 OAM MUST be defined such that devices not supporting the OAM are able to forward the OAM frames in a similar fashion as the regular NVO3 data frames/packets.

5.12. Security

R18) NVO3 OAM frames MUST be prevented from leaking outside their NVO3 domain.

R19) NVO3 OAM frames from outside an NVO3 domain MUST be prevented from entering the said NVO3 domain when such OAM frames belong to the same level or to a lower-level OAM. (Trivially met because hierarchical domains are independent technologies.)

R20) NVO3 OAM frames from outside an NVO3 domain MUST be transported transparently inside the NVO3 domain when such OAM frames belong to a higher-level NVO3 domain. (Trivially met because hierarchical domains are independent technologies).

5.13. Transport Independence

Similar to transport requirement from [RFC6136], we expect NVO3 OAM will leverage the OAM capabilities of the transport layer (e.g., IP underlay).

R21) NVO3 OAM MAY allow adaptation/interworking with its IP underlay OAM functions. For example, this would be useful to allow fault notifications from the IP layer to be sent to the NVO3 layer and likewise exposure of LAG / ECMP will require such non-independence.

5.14. Application Independence

R22) NVO3 OAM MUST [*** discuss -- is this too strong? ***] be independent of the application technologies and specific application OAM capabilities.

[Comment -- ECM: Noticed Nicira implementation has a dedicated NVP manager node to play the role of FCAPS here. It is both application layer and OAM layer. May not meet this requirement. In reality, due to the nature of overlay network, very often, vendors are going to make everything all together to a dedicated manager node.]

5.15. Prioritization

R23) NVO3 OAM messages MUST be preferentially treated in NVE and between NVEs, since NVO3 OAM MAY be used to trigger protection switching. As noted above (R2), protection switching is the automatic replacement of a failed transmission facility with a working one providing equal or greater capacity, typically within a few tens of milliseconds from fault detection.

[Comment -- ECM: giving NVO3 OAM messages priority treatment may interfere with measurements of frame delay and jitter.]

6. Items for Further Discussion

This section identifies a set of operational items which may be elaborated further if these items fall within the scope of the NVO3.

- VNID renumbering support

    * Means to change the VNID assigned to a given instance MUST [*** discuss: is this too strong? ***] be supported.

    * System convergence subsequent to VNID renumbering MUST NOT take longer than a few seconds, to minimize impact on the tenant systems.

    * A VNE MUST be able to map a VNID with a virtual network context.

- VNI migration and management operations

    * Means to delete an existing VNI MUST be supported.

    * Means to add a new VNI MUST be supported.

    * Means to merge several VNIs MAY be supported.

    * Means to retrieve reporting data per VNI MUST be supported.

    * Means to monitor the network resources per VNI MUST be supported.

- Support of planned maintenance operations on the NVO3 infrastructure

    * Graceful procedure to allow for planned maintenance operation on NVE MUST be supported. This includes undoing any configuration changes made for maintenance purposes after completion of the maintenance.

- Support for communication among virtual networks

    * For global reachability purposes, communication among virtual networks MUST be supported. This can be enforced using a NAT function.

- Activation of new network-related services to the NVO3

    * Means to assist in activating new network services (e.g., multicast) without impacting running service should be supported.

- Inter-operator NVO3 considerations

    * As NVO3 may be deployed over inter-operator infrastructure, coordinating OAM actions in each individual domain are required to ensure an end-to-end OAM. In particular, this assumes

existence of agreements on the measurement and monitoring methods, fault detection and repair actions, extending QoS classes (e.g., DSCP mapping policies), etc.

[[DISCUSSION NOTE: Should inter-operator issues be declared out of scope?]]

7. IANA Considerations

This memo includes no request to IANA.

8. Security Considerations

TBD

9. Acknowledgements

The authors are grateful for the contributions of David Black, Dennis Qin, Erik Smith and Ziye Yang to this latest version.

10. References

10.1. Normative References

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

10.2. Informative References

[IEEE802.1ag]
IEEE, "IEEE Standard for Local and metropolitan area networks - Virtual Bridged Local Area Networks, Amendment 5: Connectivity Fault Management", 2007.

[IEEE802.1ah]
IEEE, "IEEE Standard for Local and metropolitan area networks - Virtual Bridged Local Area Networks, Amendment 6: Provider Backbone Bridges", 2008.

[NM-Standards]
ITU-T, "ITU-T Recommendation M.3400 (02/2000) - TMN Management Functions", February 2000.

[NVO3-DP-Reqs]
Bitar, N., Lasserre, M., Balus, F., Morin, T., Jin, L., and B. Khasnabish, "NVO3 Data Plane Requirements", October 2012.

[NVO3-framework] Lasserre, M., Balus, F., Morin, T., Bitar, N., and Y. Rekhter, "Framework for DC Network Virtualization", July 2012.

[RFC6136] Sajassi, A. and D. Mohan, "Layer 2 Virtual Private Network (L2VPN) Operations, Administration, and Maintenance (OAM) Requirements and Framework", RFC 6136, March 2011.

[Y.1731] ITU-T, "ITU-T Recommendation Y.1731 (02/08) - OAM functions and mechanisms for Ethernet based networks", February 2008.

Authors' Addresses

Peter Ashwood-Smith
Huawei Technologies
303 Terry Fox Drive, Suite 400
Kanata, Ontario K2K 3J1
Canada

Phone: +1 613 595-1900
Email: Peter.AshwoodSmith@huawei.com

Ranga Iyengar
Huawei Technologies USA
2330 Central Expy
Santa Clara, CA 95050
USA

Email: ranga.Iyengar@huawei.com

Tina Tsou
Huawei Technologies USA
2330 Central Expy
Santa Clara, CA 95050
USA

Email: Tina.Tsou.Zouting@huawei.com

Ali Sajassi
Cisco Technologies
170 West Tasman Drive
San Jose, CA 95134
USA

Email: sajassi@cisco.com

Mohamed Boucadair
France Telecom
Rennes 35000
France

Email: mohamed.boucadair@orange.com

Christian Jacquenet
France Telecom
Rennes 35000
France

Email: christian.jacquenet@orange.com

Masahiro Daikoku
KDDI corporation
3-10-10, Iidabashi, Chiyoda-ku
Tokyo 1028460
Japan

Email: ms-daikoku@kddi.com