Abstract

High-Availability (HA) is a very important requirement throughout the history of carrier network, many technologies have emerged for it. With the trend of Network Function Virtualization (NFV), network function are migrated from dedicated hardware to software running over COTS servers, the same SLA of HA should be provided depending on network service itself. But some new challenges are brought by NFV, one example is Virtualized Network Function (VNF) cluster caused by the HA and performance limitation of individual VNF instance. For the VNF cluster, some gaps exist between with the available HA technologies on issues of multi-homing, state synchronization, share-risk prevention and HA role election, especially in the network which has a large scale deployment of NFV.

This document firstly identifies the challenges emerged within NFV deployed networks. Then, available HA technologies are reviewed and the detailed gap analysis between them with the new challenges is discussed in depth. At last, the summary of these gaps is presented.

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 July 3, 2014.

Copyright Notice
Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents 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.

Table of Contents

1. Introduction ................................................. 3
2. Terminology .................................................. 4
3. New challenges and requirements to NFV HA solution ........ 5
   3.1. New features of NFV .................................... 5
   3.2. VNF cluster ........................................... 6
   3.3. Failure detection and handling ......................... 6
   3.4. Consideration of central management server .......... 7
   3.5. Policy enforcement .................................... 7
   3.6. NFV HA solution requirements ......................... 7
4. Gaps between available HA solutions and NFV’s challenges ... 9
   4.1. VRRP .................................................. 9
   4.2. BFD .................................................. 10
   4.3. APS .................................................. 10
   4.4. NSR, NSF, SSO and GR ................................ 11
   4.5. STP .................................................. 11
   4.6. FRR .................................................. 12
   4.7. OAM .................................................. 13
5. Summary ...................................................... 14
6. IANA Considerations ........................................ 16
7. Security Considerations .................................... 17
8. Informative References ..................................... 18
Author’s Address ............................................... 19
1. Introduction

For the benefits of reduced operational and capital costs, automated deployment, and enhanced elasticity, Network Function (e.g. Firewall, IPS, Load Balancer, etc) Virtualization technology would be widely supported by DC and Carrier network. But, one key issue, the HA requirement of NFV should be considered over again to provide the same SLA of HA depending on the network services itself.

Considering the reliability requirements, NFV HA architecture should support several key points detailed below:

- Redundancy mechanism: LACP, VRRP, ECMP, etc;
- Failure detection: IEEE 802.1ag, ITU-T Y.1731, BFD, MPLS(-TP) OAM, etc;
- Failure notification: APS, etc;
- State synchronization: NSR, NSF, SSO, GR, etc;
- Failure handling (switchover and failover): STP, EAPS, MPLS FRR, etc.

[VNFP-PS] has provided the problem statement and working scope analysis focusing on the VNF reliability and high availability issues. One accompany draft [VNFP-UC] provides an overview of VNF HA use cases.

In this document, we reviewed the challenges of NFV and traditional HA solution, and summarize the gap between them.
2. Terminology

NFV  Network Function Virtualization
VRRP  Virtual Router Redundancy Protocol
STP  Spanning Tree Protocol
FRR  Fast Reroute
BFD  Bidirectional Forwarding Detection
APS  Automatic Protection Switching
HA  High Availability
3. New challenges and requirements to NFV HA solution

As described in [VNFP-PS] and [VNFP-UC], NFV brings some new features and requirements to current network, which also results in new challenges to NFV HA solution in the following aspects:

3.1. New features of NFV

NFV network should support high scalability, cost reduction, high flexibility, high automation.

- High scalability: NFV HA solution should support the large scale network. To support the HA of large number of VNFs in virtual network, the highly efficient and minimum impact to network HA solution is surely needed. NFV HA solution should also support an automatic and fast mechanism to discover and add/delete new VNFs into/out of VNF cluster;

- Cost reduction: One of the intentions of deploying NFV is to reduce the cost, thus, whether for the VNF cluster or working/protection group deployment, making full use of its members by active/active mode is the preferred choice. And, 1:N redundancy mechanism should be the basic requirement rather than 1:1, 1+1 or 1+N. But, multiple active VNFs maybe result in a performance problem. Because every active VNF needs to synchronize its state to every other VNF in the same VNF cluster. The full-mesh connectivity for state synchronization is very complex and will consume too much bandwidth and system resource to bring additional delay;

- High flexibility: NFV brings high flexibility to network to make it easy for operator to scale up or down the virtualized network, and migrate VNF instances to other location on demand. Thus, VNF instances can be highly distributed in DC networks, Carrier networks and even customer premises, and they can be migrated dynamically. It is hard to located VNF VMs to fixed hosts. If the VNF instances in the same VNF cluster are located in the same host or the chains of VNF are crossed, it will reduce the SLA of HA due to share-risk problem. Another problem is related to service chain. For the consideration of HA, the working and backup service chains should not be crossed in the same location (e.g. host, hypervisor, VM, VNF), otherwise it will lead to a single point of failure;

- High automation: NFV deployed network can be a large scale, highly distributed, and highly dynamic network. Which means a large amount of operations and functions must be automatically performed. Otherwise, too much manual configurations and
interventions are not affordable. For example, the deployment and configuration of redundancy mechanisms, the failure detection and switchover, the adaptation to dynamic changes of VNFs should all be performed automatically and quickly.

3.2. VNF cluster

NFV has transferred the running network functions from physical platform to virtualized platform, always result in the limitation of individual VNF instance’s performance and reliability. So, NFV normally uses cluster with more members to meet the performance and reliability requirement. With the VNF numbers in VNF cluster increasing, some aspects should be considered:

- State synchronization: State synchronization is the essential function for NFV, especially for the clustering mechanism. Otherwise, VNF cluster cannot work in several scenarios. One example is when forward flow and return flow hit different VNFs. The other example is switchover case. With the number of VNFs in cluster increased, state synchronization between them with full mesh becomes more complex and resource consuming. A general and efficient technology of state synchronization for various kinds of VNFs is needed;

- Dynamic change of VNF members’ role: A VNF member’s role (active, especially standby) in VNF cluster can dynamically change in the run time due to switchover, NFV HA solution must adapt to it.

3.3. Failure detection and handling

- Failure detection: In NFV infrastructure, failure can happen in multiple layers: hardware, hypervisor, VM, or VNF instance. So, the new failure detection mechanism should have the capability to detect the failure in all these layers. This mechanism should be simple, uniform, protocol independent, and software driven. One candidate is BFD;

- Failure handling approaches: NFV technology brings new approaches to handle failures rather than traditional network. For instance, scaling up can resolve the problem of over-load. Restarting VNF instance or VNF dynamic relocation can all be used when VNF fails. The new HA solutions should make full use of these new NFV features.
3.4. Consideration of central management server

The north bound interfaces of VNF can be connected to the central management server to report VNF's running status. By this way, central management server can decide the roles of VNFs in VNF cluster and be used for the switchover mechanism. It can also be used for the state synchronization.

3.5. Policy enforcement

There could be some policies reflecting the different reliability class of the service and hence affecting the selection of VNF instances. Examples would include isolation policies requiring that VNF instances be placed on separate physical servers or separate DC sites. Another example is to place some VNF instances in topologically closed locations. Other policies and related HA issues are TBD.

3.6. NFV HA solution requirements

For coping with the above new challenges, the NFV HA solution should provide the functionalities as following:

- Redundancy mechanism
  - The function to support VNF cluster with active/active mode
  - The function to support 1:N redundancy of VNF cluster;
  - The function to support VNFs to join/leave or scale up/down efficiently and automatically;
  - The function of share-risk prevention.

- Failure detection
  - The simple, uniform, protocol independent, and software driven function to detect the failure of multiple layers of NFV infrastructure.

- Failure notification
  - The function to notify central management server about VNF failure;
  - The function to notify other interested or influenced VNF about VNF failure.
- State synchronization
  * The function to support the dynamic change of VNF member’s role in VNF cluster due to switchover by tracking the binding relationship between active VNFs with standby VNFs;
  * The function to support VNF keep-alive monitoring and efficient state synchronization to avoid full mesh connectivity by using central controlled technology.

- Failure handling (switchover and failover)
  * The function to support quick convergence for large VNF cluster;
  * The function to support central management server;
  * The function to utilize new NFV features, for example scaling up/down, restarting and dynamic relocation of VNF to overcome VNF failure or over-load.

- Compatibility
  * Supporting different provider’s VNF;
  * The compatibility to existed hardware network elements.
4. Gaps between available HA solutions and NFV’s challenges

The available HA solutions are the combination of several technologies deployed in different points of network. The overview is as followed:

- VRRP is a common redundancy mechanisms in Layer 3 and supports failure handling;
- BFD is a common technology for failure detection;
- APS protocol is mainly used for failure notification which naturally leads to failure handling process;
- NSR, NSF, SSO and GR can support state synchronization between neighboring devices or standby RP during the device restarts;
- STP and MPLS FRR are deployed as failure handling mechanisms in Layer 2 or 3.

4.1. VRRP

The Virtual Router Redundancy Protocol (VRRP) is designed to eliminate the single point of failure inherent in the static default routed environment over Layer 3. The nodes or ports in the same group share the same virtual IP address with different MAC addresses. The elements in VRRP group elect the active role using pre-set priority. This priority/role changes only if switchover or fallback occurred. The active VRRP element notifies hosts in the subnet using gratuitous ARP. If the active VRRP element fails, the standby elements in the VRRP group would select a new active element. The new active element will sent gratuitous ARP to notify the corresponding hosts to update their MAC table. Consequently, the flows with the corresponding virtual IP as destination IP will be lead to the new active element.

VRRP works as an integration solution of redundancy and failure handling mechanism.

For requirements of redundancy mechanism:

- VRRP supports cluster. However, there is only one active element per VRRP group. It does not support multiple active elements in the same group. Multiple active elements means multiple different VRRP groups distinguished by different virtual IP. The related hosts’ gateway configuration should be planned carefully to these IFs, which is very inflexible;
o VRRP can only support most 255 elements in one group, this maybe a drawback for NFV deployment in cross DC site scenarios in the future;

o Only active element can send VRRP ADVERTISEMENT message to notify standby elements in same group, however standby elements can not announce their presence. Thus the automatic discovery of new elements is a problem for state synchronization;

o VRRP does not support share-risk prevention.

For requirements of failure handling:

o VRRP group is an autonomy system without the support of central management server. Thus, it cannot take advantage of it to make an optimized choice between the VRRP elements flexibly according to their respective performance;

o The standby elements take over the load of active element when it fails. While, if the active one fails due to overload, this is not effective. VNF can scale up to solve this problem, while VRRP cannot.

4.2. BFD

BFD is a lightweight hello protocol designed to run over multiple transport protocols (e.g. IPv4, IPv6, MPLS, etc) used for Layer 3 failure detection. Any interested client (e.g. OSPF, BFP, HSRP, etc) registers with BFD and is notified as soon as BFD detects a neighbor loss. BFD establishes monitoring sessions between two neighbors and detects link or node failures if no BFD packet is received for a period.

IBFD is a good candidate of failure detection solution for the NFV network due to its features of simplicity, efficiency, uniformity, protocol independence, software driven. But, BFD can only detect the failures of Layer 3 and does not support Layer 2. How it supports the failure detection for VNF cluster and multiple layer NFV infrastructures also needs more study.

4.3. APS

APS (Automatic Protection Switching) protocol is a mature and proven mechanism specified for bidirectional protection switching which needs the coordination of the two endpoints of the transport entity in SONET/SDH networks. It can also be used for Ethernet [ITU-T G.8031] and MPLS [ITU-T Y.1720] network now.
One endpoint transmits a new APS packet immediately to inform the far
end endpoint for the coordination of protection switching when a
change in the transmitted status occurs (e.g. link/node failure,
force switch, signal fail, etc).

APS protocol is mainly used for failure notification which will
naturally leads to failure handling process.

For requirements of redundancy mechanism:

- APS does not support VNF cluster with active/active mode;
- Usually, APS provide 1:1 or 1+1 backup, and no more than 1:14
  backup will limit the scale of NFV network;
- APS does not support dynamic election mechanism of path role. The
  role of path is pre-configured;
- APS does not support share-risk prevention.

For requirements of failure notification:

- APS does not provide north bound interface to central management
  server.

For requirements of failure handling:

- APS does not support the new NFV failure handling features.

4.4. NSR, NSF, SSO and GR

TBD.

4.5. STP

The Spanning Tree Protocol (STP) is a network protocol that ensures a
loop-free topology for any bridged Ethernet local area network. The
basic function of STP is to prevent bridge loops and the broadcast
storm that results from them. Spanning tree also allows a network
design to include spare (redundant) links to provide automatic backup
paths if an active link fails, without the danger of bridge loops, or
the need for manual enabling/disabling of these backup links.

STP works as an integration solution of redundancy and failure
handling mechanism.

For requirements of redundancy mechanism:
o STP does not support VNF cluster with active/active mode;

o STP do not provide the elect mechanism for dynamically electing active and standby elements in elements groups, the roles is determined by configured network layout and pre-assigned priority and port ID etc. After convergence, the role of port can only be changed by manual configuration or at next time of convergence when some ports fail;

o STP does not provide the share-risk prevention mechanism.

For requirements of failure handling:

o STP does not provide north bound interface to central management server;

o STP does not support new NFV failure handling features.

4.6. FRR

MPLS Fast Reroute is a local restoration network resiliency mechanism. It is actually a feature of resource reservation protocol (RSVP) traffic engineering (RSVP-TE). In MPLS local protection each label switched path (LSP) passing through a facility is protected by a backup path which originates at the node immediately upstream to that facility.

FRR works as an integration solution of redundancy and failure handling mechanism.

For requirements of redundancy mechanism:

o FRR does not support VNF cluster with active/active mode;

o The FRR paths cannot dynamically elect active or standby paths, it is manual configured or layout by aptotic algorithms like LFA(loop free alternate)[RFC 5286];

o FRR does not support share-risk prevention;

o Usually, 2 or a little more paths can be protected in FRR, the large number of VNFs in a cluster will make network layout very complex.

For requirements of failure handling:

o FRR does not provide north bound interfaces to central management server;
FRR also has the similar problem with VRRP in the deployment of new NFV failure handling features.

4.7. OAM

TBD
5. Summary

In conclusion, there is a gap between the available HA technologies and the new challenges of NFV.

<table>
<thead>
<tr>
<th></th>
<th>VRRP</th>
<th>APS</th>
<th>STP</th>
<th>FRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support active/active cluster</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Support 1:N backup</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Automatic scalability</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Share-risk prevention</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Figure 1: Gap Analysis Table of Redundancy Mechanism

Note:

1. For NFV, VNF cluster with active/active mode is one of the basic requirements;

2. 1:N backup means 1 standby element for N active elements in one VNF cluster.

<table>
<thead>
<tr>
<th></th>
<th>BFD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support failure detection of VNF cluster and multiple layers</td>
<td>TBD</td>
</tr>
</tbody>
</table>

Figure 2: Gap Analysis Table of failure Detection
Figure 3: Gap Analysis Table of failure Notification

<table>
<thead>
<tr>
<th></th>
<th>APS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notify central network</td>
<td>no</td>
</tr>
<tr>
<td>server</td>
<td></td>
</tr>
<tr>
<td>Notify related VNF</td>
<td>yes</td>
</tr>
</tbody>
</table>

Figure 4: Gap Analysis Table of State Synchronization

<table>
<thead>
<tr>
<th></th>
<th>VRRP</th>
<th>APS</th>
<th>STP</th>
<th>FRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>North bound API</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Support new NFV approaches</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Figure 5: Gap Analysis Table of failure handling
6. IANA Considerations

This document has no actions for IANA.
7. Security Considerations

TBD.
8. Informative References

ID draft-zong-vnfpool-problem-statement-01, September 2013.

[VNF-UC] Xia, L., "Use cases and Requirements for Virtual Service Node Pool Management",
Author’s Address

Yang Wang
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu  210012
China

Email: alex.wangyang@huawei.com