



NFV 可靠性探讨

杨旭, 肖子玉, 邵永平, 宋小明

(中国移动通信集团设计院有限公司, 北京 100080)

摘要: 由 NFV 引入带来的硬件可靠性下降, 引出电信网引入 NFV 的可靠性问题。通过对 NFV 可靠性和可用性进行分析, 提出 NFV 架构下网元的典型故障场景、故障检测机制和流程以及 NFV 独有的跨层检测机制和部署建议。最后提出 NFV 可靠性要求及设计建议。

关键词: NFV; 可靠性; 可用性; 核心网

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017086

Discussion on the reliability of NFV

YANG Xu, XIAO Ziyu, SHAO Yongping, SONG Xiaoming

China Mobile Design Institute Co., Ltd., Beijing 100080, China

Abstract: Through the reduction of the hardware reliability which was caused by the introduction of NFV, the reliability issue of introducing NFV into telecommunication network was introduced. Through the analysis of reliability and availability of NFV, the typical fault scenarios, fault detection mechanism and processes of NFV architecture, and the NFV's unique cross-layer detection mechanism and deployment advice was proposed. Finally, the NFV reliability requirements and design recommendations were proposed.

Key words: network function virtualization, reliability, availability, core network

1 引言

网络功能虚拟化 (network function virtualization, NFV) 打破了传统电信设备的竖井式体系, 其核心是网元的分层解耦, 并引入新的 MANO (management and orchestration, 管理和编排) 体系实现全生命周期管理。NFV 架构由 ETSI (European Telecommunications Standards Institute, 欧洲电信标准化协会) 定义, 包括 3 个域: NFV 基础设施 (network function virtuali-

zation infrastructure, NFVI)、虚拟网络功能 (virtualized network function, VNF)、NFV 管理和编排。其中, NFVI 包括虚拟化软件和硬件资源 (计算、存储、网络资源), MANO 由 NFVO (NFV orchestration)、VNFM (VNF manager)、VIM (virtual infrastructure manager) 组成。NFVO 负责网络服务的全生命周期管理, VNFM 负责 VNF 的全生命周期管理, VIM 主要负责基础设施层硬件资源及虚拟化资源的管理、监控和故障上报。

传统核心网设备采用 ATCA (advanced telecom computing architecture, 先进的电信计算平台) 专有硬件, 通过硬件传感器检测及内部的软硬件耦合通知机制, 保证故障检测的实时性和准确性 (毫秒级)。NFV 云化网元通过软件机制实现故障检测 (秒级)。故障处理机制比较如图 1 所示。

引入 NFV 后, VNF 的可靠性应不低于现有传统物理模块 (功能) (physical network function, PNF) 的可靠性。NFV 基于通用硬件设备和开放的云操作系统, 如何在“3个9”的基础设施上构筑“5个9”的业务可靠性是本文关注的重点。可靠性挑战总结如下。

(1) 基础设施可靠性低

IT 通用硬件故障率高, 开源虚拟化软件产生新的故障模式。

(2) 分层解耦带来新的故障点

NFV 整体业务端到端可靠性受限于每一层的可靠性, 且易产生单点故障。

(3) 基础设施资源共享

计算、存储、网络资源共享, 易产生资源抢占; 分布式、弹性容易导致故障扩散。

(4) 软硬件解耦

网络软件化, 基础设施故障信息无法及时上报给 VNF, VNF 只能靠软件手段检测故障; 垂直厂商环境下集成交付, 可靠性是一大挑战。

2 NFV 可靠性模型分析

可靠性 (reliability) (%) = $e^{-t/MTBF}$, t 为观测的时间间隔。可用性 (availability) (%) = $MTBF/(MTBF+MTTR)$, 其中 $MTBF$ 为平均故障间隔时间, $MTTR$ 为平均故障恢复时间。系统的可用性与 $MTBF$ 和 $MTTR$ 有关, 提升可用性可以通过增加 $MTBF$ 或者减小 $MTTR$, 而可靠性仅与 $MTBF$ 有关。电信级可靠性代表系统正常运行时, 具有 99.999% 的电信服务能力, 意味着一年的业务中断时间不超过 5.26 min, 落实到系统上即可用性。

ETSI REL003 中给出的端到端业务可靠性模型示意如图 2 所示。

由图 2 可以看出, 端到端的网络由并联系统和串联系统共同构成, 端到端系统可靠性随冗余并联配置增加而增大, 随串联配置节点数增加而减小。端到端路径的可靠性受限于最低的可靠性组件, 冗余方式的广泛使用以改善系统的可靠性。系统可用性依赖于 $MTTR$ 值, 包括故障检测时间、修复时间, 通过减小 $MTTR$ 增加可用性, 当 $MTTR$ 趋近于 0 时, 系统可用性 A 趋近于 1。

如果所涉及的故障自动恢复对业务的接通率、掉话率、接通时间、计费等有影响, 故障自动恢复的时间应作为故障时间计算。如果业务恢复时间少于业务允许的容忍中断时间, 那么业务被视为连续而未中断。然而一个单一的业务恢复时

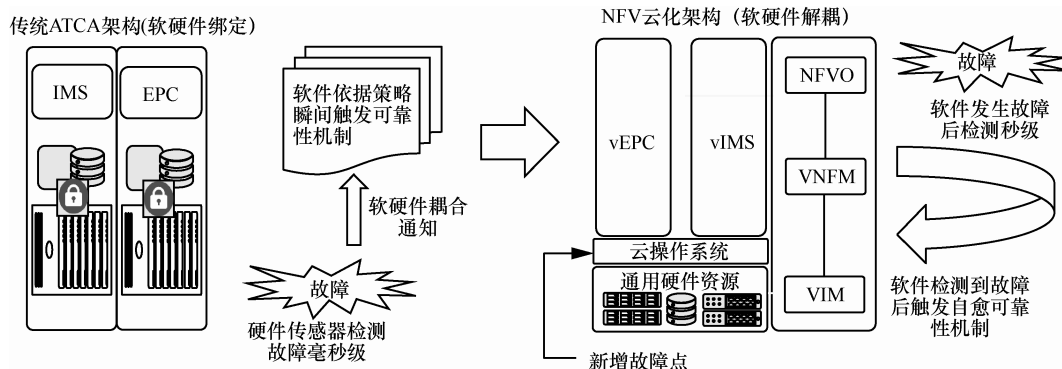


图1 传统网络与NFV网络故障处理机制对比

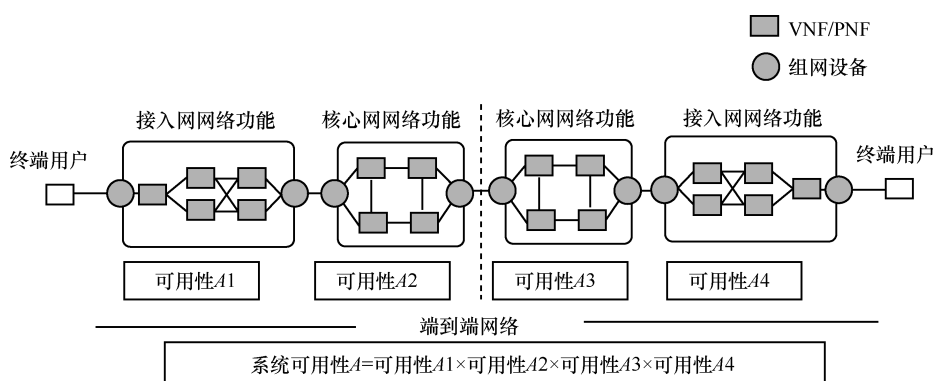


图2 端到端网络服务可靠性模型

表1 服务可用性级别参考

服务可用级别	服务恢复时间门限	服务类型	注
SLA1	5~6 s	网络运营商控制流量 政府、监管紧急业务	可以再划分为几个子级别，如 1A：控制；2A：实时；3A：数据； 可能需要 1+1 冗余以及快速恢复
SLA2	10~15 s	网络运营商务务流量 大型企业、校园客户等	可以再划分为几个子级别，如 1A：VPN；2A：实时；3A：数据； 可能需要 1+1 冗余以及快速恢复
SLA3	20~25 s	一般企业客户	典型的尽力而为的场景； 可能需要 M+1 冗余以及快速恢复

间不适用于所有业务，需要支持多个服务可用级别（service level agreement, SLA）。运营商可以选择结合 SLA1 和 SLA2 可用性作为高可用性级别，并可以进一步细分为几个典型的子级别。TL-9000 论坛规定 15 s 的业务中断时间作为传统电信业务的断供时间。服务可用性级别参考见表 1。

3 NFV 可靠性机制及设计建议

从第 2 节的分析可知，可靠性的关键性指标为故障的快速检测和恢复能力，这是以时间参数来考量的，检测方法和流程与指标强相关。而要实现快速恢复，系统的冗余配置、网元模块的状态、网络的部署都是可靠性的决定性因素。因此本节将对以上问题逐条分析，并在第 3.4 节给出总结。

3.1 VNF 典型故障场景及流程分析

NFV 分层解耦与传统 ATCA 架构相比，引入了新的故障模式，故障模式主要包括：硬件故障

（计算、存储、网络等）、虚拟层故障（虚拟机、虚拟化软件等）、VNF 层故障（VNFC 等）、MANO 故障（NFVO、VNFM、VIM 等）。本文以下将重点以典型的 VNFC 故障以及 NFVI 层故障引起的 VNF 故障为例进行分析。

根据 VNF 有状态、无状态类型，VNF 是否有冗余，VNF 执行故障检测还是 VNF 和 NFVI 均执行故障检测，可以将 VNF 分为表 2 所示的 8 种典型故障场景。

表2 VNF 故障场景分析

状态	VNF 是否有冗余	故障检测	场景
有状态	是	VNF	1
		VNF&NFVI	2
	否	VNF	3
		VNF&NFVI	4
无状态	是	VNF	5
		VNF&NFVI	6
	否	VNF	7
		VNF&NFVI	8

VNFC 无状态：是指无需保持会话参数而提供的应答服务，只要提供冗余实例和负载均衡即可实现。一般采用 $N+M$ 的 active-active 方式，通过负载均衡进行检测。

VNFC 有状态：是指会话过程中依赖前一请求中携带的参数的服务。一般采用 $2N$ 主备冗余方式，状态信息需要实时同步。

下面以场景 1 和场景 2 为例，分析 VNF 故障、故障检测和故障修复流程。场景 1 是 VNFC 故障触发的流程，如图 3 所示。

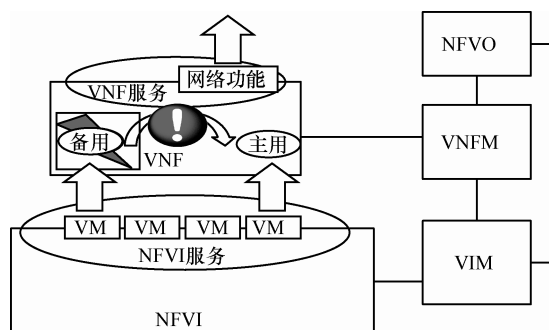


图3 VNF有状态、有冗余、VNFC故障（场景1）

该场景具体流程如下：

(1) VNFC 故障，导致网络功能（如 MME 网元功能）故障；

(2) VNF 检测到 VNFC 故障，VNFC 故障检测使用心跳机制；

(3) VNF 隔离故障 VNFC，然后 VNF 进行主用到备用 VNFC 的故障倒换，网络功能恢复；

(4) VNF 修复备用 VNFC。

其中 (1) ~ (3) 的故障检测和修复时间，在对应的业务恢复时间门限内，则不影响业务可用性。(4) 的 VNFC 修复时间不影响业务可用性。HA (high availability, 高可用性) 机制可以嵌入 VNF 中，也可以由 VNFM 负责，由 VNFM 负责时，VNFM 需要负责故障检测。

场景 2 是 NFVI 层 VM 故障触发的流程，具体如图 4 所示。

该场景具体流程如下：

(1) VM 故障导致 VNFC、网络功能故障；

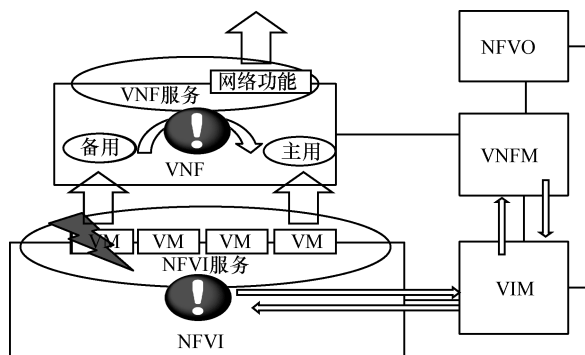


图4 VNF有状态、有冗余、VM故障（场景2）

(2) VNF 通过心跳机制检测到 VNFC 故障；

(3) VNF 隔离故障 VNFC，然后 VNF 进行主备份 VNFC 的故障倒换，网络功能恢复；

(4) NFVI 检测到 VM 故障，Hypervisor 层通常使用软件看门狗机制对 VM 进行故障检测；

(5) NFVI 将 VM 故障信息上报给 VIM，VIM 再上报给 VNFM；

(6) VIM 修复 VM，并上报给 VNFM，VNF 修复备用 VNFC。

该场景在 NFVI 和 VNF 层同时作故障检测和处理，VNF 层故障检测和故障倒换不依赖于 NFVI 层，VM 故障由 NFVI 层检测并在 VIM 和 VNFM 之间协作处理，以修复备用 VM 和备用 VNFC。其中 (1) ~ (3) 的故障检测和修复时间在对应的业务恢复时间门限内，不影响业务可用性。(5)、(6) 的 VM 故障检测和修复以及备用 VNFC 的修复时间不影响业务可用性。

当检测到 VM 故障，首选在该单板上重启该 VM，重启失败后在不同的单板上进行虚拟机重生，时间为分钟级，所以该机制下将有较长一段时间网元组件处于无备份状态。

3.2 分层故障检测及跨层故障关联机制

(1) VNF 层（心跳机制）

在 VNF 内部，一般在负载均衡器与 VNFC 之间或者在 VNFC 之间进行心跳检测和回复。对于一个高可用的 VNF，有必要及时检测和恢复故障，



VNF 专有故障最好由 VNF 本身检测, 当 VNF 不能从 VNF 具体的故障中检测和恢复, 可能需要 VNFM 检测和重启 VNF, 或从 NFVI 请求替换或者新增资源。心跳检测机制频率是关键, 过快可能无法保证检测的准确度, 导致业务频繁切换, 过慢则可能引起较大的业务损失。

目前各厂商的心跳检测时间有所差异 (单次在 100 ms~1 s 内), 在 3~7 次之间均无回复, 则进行主备故障倒换。

(2) 虚拟层 (看门狗软件)

Hypervisor 层提供软件狗功能, 通过虚拟机内部部署的监控模块定期向软件狗“喂狗” (发送心跳), 如果心跳超过阈值, Hypervisor 向 VIM 发送故障通知, VIM 向该虚拟机发重置命令。

目前各厂商检测时间有所差异, 如某厂商单次检测时间为 1 s, 如超时认定 VM 故障, 上报 VIM, 重置 VM。实现原理如图 5 所示。

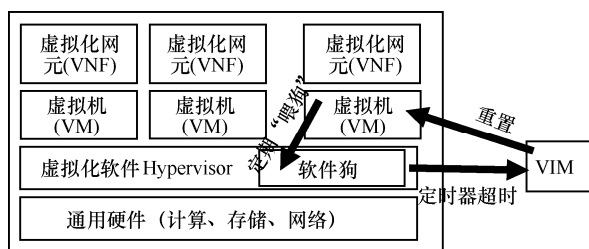


图 5 虚拟层看门狗软件故障检测机制实现原理

(3) 硬件层

硬件故障仍然通过传统的方案解决。NFV 场景下, 硬件故障可以引起业务和虚拟基础设施故障, 因此需要硬件的冗余架构和自动的故障倒换机制。硬件故障上报给 Hypervisor 和 VIM, 以实现 HA。硬件故障不像传统的 ATCA 架构直接上报给 VNF (特殊场景下会有硬件到一些专有的 VNF 的直接通知)。所有关键的硬件故障应在 1 s 内上报给 VIM。可周期性更新硬件运行情况给 VIM, 如每 60 s 更新一次。

参照 ETSI NFV-REL 001 中给出的故障模式及故障检测时间要求见表 3。

表 3 ETSI 给出的故障检测时间要求

故障模式	时间
硬件故障检测	<1 s
虚拟层故障检测	<1 s
告警阈值检测	<1 min
SLA1 故障检测	<1 s
SLA1 故障恢复	5~6 s
SLA2 故障检测	<5 s
SLA2 故障恢复	10~15 s
SLA3 故障检测	<10 s
SLA3 故障恢复	20~25 s

(4) 跨层故障关联机制

NFV 系统中的错误可能导致在不同层次的错误通知, 这些错误可能有因果关系。这些通知不应该被独立处理, 可能导致不一致的恢复行动。另一方面, 如果所有信息传递给集中的故障管理系统, 如 VNF-EMS、VNFM 或者 NFVO 等, 该实体可能不能及时处理所有故障。为了维持弹性和可用性, 在最小的人工干预下, 帮助快速和动态地找到上报故障的根源。

NFV 架构利用分布式故障上报框架, 并允许在多个级别故障关联。本地的故障关联运行在系统的不同层次, 收集发生在该层的不同组件故障信息。故障关联可以在 VNF-EMS、MANO 等。

3.3 区域划分原则及建议

region (域)、cell (小区)、availability zone (AZ, 可行区域)、host aggregate (主机集群) 是 OpenStack 用于划分云资源池的方法, 具体如图 6 所示。

(1) 域

更像一个地理概念, 每个域有自己独立的 API 端点, 域之间完全隔离, 但是多个域之间共享同一个 keystone (认证服务) 和 Horizon (统一界面)。

(2) 小区

主要用来解决 OpenStack 扩展性和规模瓶颈, 每个小区有自己独立的数据库和消息队列。

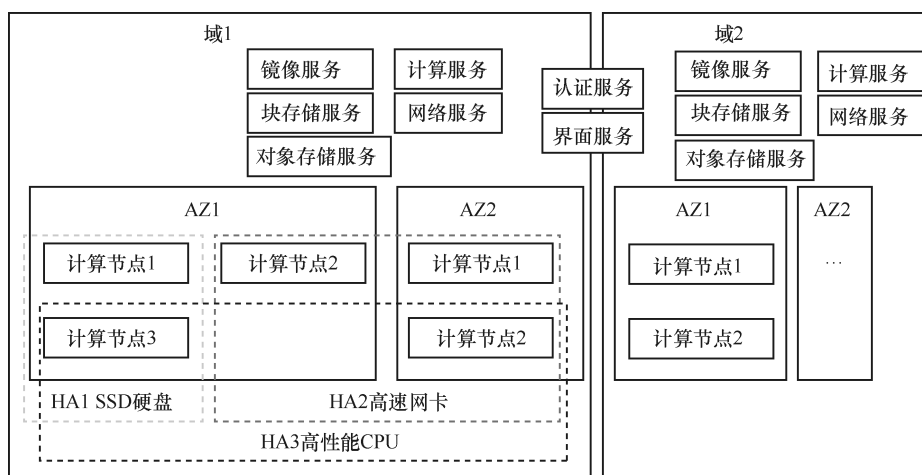


图 6 区域划分示意

（3）可行区域

可以理解为一组计算节点的集合，这组节点具有独立的电力供应设备，比如独立供电的机房、独立供电的机架。可行区域是用户可见的概念，用户在创建实例的时候可以选择创建到哪些可行区域中，以保证实例的可用性。

（4）主机集群

也是一组计算节点的集合，但强调这组节点具有共同的属性，如高性能 CPU 或配置高速网卡等，其主要功能就是实现根据某一属性来划分物理机。主机集群是用户不可见的概念，只对管理员可见，主要用来给 Nova 调度器通过某一属性来进行 VM 的调度。

区域划分建议如下。

（1）域

不同地域的 DC，如果不需要统一资源调度，则可以划分到不同域进行管理。

（2）小区

根据 DC 规模，一个 DC 站点可以划分为一个或多个小区。

（3）可行区域

建议根据独立供电系统，将一个 DC 划分多个可行区域，VNF 可以显式指定部署的可行区域。VNF 池内的网元部署到同 DC 不同可行区域中，提高可靠性。

（4）主机集群

VNF 可以指定 VM 所需的特征属性，调度器会根据属性将 VM 分配到合适的主机集群上，如将配置高速网卡的节点规划为一个主机集群。

3.4 NFV 可靠性要求及设计建议

3.4.1 NFV 可靠性要求

（1）保证业务连续性

NFV 可靠性的主要目标是保证端到端的业务连续性。

（2）故障自动恢复

NFV 应支持自动化的故障倒换机制。故障自动恢复包括：物理主机故障，在物理主机出现故障后，在其他健康主机上重建故障主机上部署的虚拟机及业务；虚拟机故障，在虚拟机出现故障后，尝试本地恢复，如无法恢复，在其他空闲主机上重建虚拟机及业务。

（3）防止单点故障

系统中不应存在单点故障，需要分别在硬件资源层、虚拟资源层、VNF 层、MANO 层考虑冗余模型，满足 SLA 要求。可通过分层架构的弹性措施、标准 NFVI 资源的动态分配以及反亲和性部署原则等方式避免单点故障。

（4）限制故障区域

应该为系统的所有网元定义故障影响区



域。在检测到网元故障后，首先进行故障隔离，然后触发恢复机制。如果故障隔离不成功，需要尽快尝试在更高的级别进行隔离，以阻止错误扩散。应可以结合隔离和告警产生策略，定义不同级别的故障影响区域。如果 VNFC 或者 VNF 频繁发生故障，即使故障恢复时间满足要求，仍然需要触发故障预防流程，交换不稳定的硬件或软件。

3.4.2 分层的可靠性设计建议

(1) VNF 层

高可靠的软件架构：向无状态化、公用数据库（业务处理和数据分离）、微服务方向发展，便于弹性扩缩容、提升服务可用性。

组网及部署：网元内部，VNFC 故障单元反亲和性部署；网元之间，pool 组网（VNF 独立组 pool 或与 PNF 混合组 pool）；站点之间，跨 DC 站点容灾（建议每省至少布局 3 个省级 DC，保障站点容灾同时具备较高的设备利用率）。

故障检测和恢复：应用层不依赖底层构筑高可靠性，秒级检测及业务恢复，利用冗余实现故障快速切换。

VNF 设计需要能够定义对 NFVI 的要求，如地理要求、弹性要求等，使得 VNF 的具体需求能够匹配到 NFVI 的具体能力。

(2) 虚拟化层

VM 采用反亲和性原则，部署在不同的硬件单元上，避免单点故障；支持虚拟机高可用性、热迁移、故障隔离、NUMA 亲和性等特性，关闭 CPU/内存超分配；NFVI 应维持提供给 VNF 足够的 VM，以应对故障；Hypervisor 系统能够及时发现通用设备、虚拟机、虚拟网络和 Hypervisor 自身的故障，并向 VIM 上报告警，需支持软件狗机制；为了防止所有的共享虚拟网络 I/O 和虚拟存储 I/O 资源全部被一个故障 VM 消耗，应能通过配置的共享资源策略为每个虚拟机分配最大和最小共享范围。

(3) 硬件层

硬件层可靠性可以通过一些传统方式解决，如硬件的冗余配置和故障自动倒换。如对于服务器可靠性，需支持电源 1+1 备份和模块热插拔等；对于网络可靠性，需支持网络路径冗余、分平面组网、网卡绑定等；对于存储可靠性，需支持存储多路径等。

服务器支持带外故障检测及上报，包括：温度异常、电源故障、风扇故障等。

根据 HA（主机集群）、AZ（可行区域）等区域规划，对故障单元反亲和性部署；硬件冗余配置（如每 10 台服务器配置 1 台冗余），并通过硬件例行维护提升硬件可用性。

硬件故障应该上报给 Hypervisor 和 VIM。硬件故障不应该向传统的 ATCA 架构一样直接上报给 VNF，周期性更新硬件运行情况给 NFVI 和 VIM。

(4) MANO 层

NFVO 在网络中所处的位置类似于现网 OSS，参照 OSS 容灾要求进行部署。

VNFM 按照 VNF 厂商集中部署，在网络中的位置类似于各厂商 OMC，采用主备份方式部署。

根据 VIM 管理厂商硬件范围及计算节点数量，建议每 DC 部署一套或多套。VIM 以集群方式部署，所有有状态的服务支持 *N-way*、*N+M active-standby* 冗余，无状态的服务应该以 *all active* 的模式部署。

利用 MANO 提供弹性伸缩能力，提升系统可用性。定义不同的业务可用性级别，通过 VNFD 和 NSD 定义 VNF 对 NFVI 的要求。

VIM 应该为 VNF 提供反亲和性方案，在不同的主机集群部署冗余业务。

跨层故障关联，NFV 架构利用分布式故障上报框架，并允许在多个级别进行故障关联，如故障关联点可以分别在 EMS 及 NFVO 进行故障关联。

NFV-MANO 功能将监视提供给 VNF 的资源，

一旦 NFVI 层出现故障, 将采取必要的行动确保受影响的业务的连续性, 如在 VNF 故障后自动重建 VNF, 实现故障快速自动恢复。

4 结束语

引入 NFV 后, 在传统网络中使用的可靠性机制可能仍然适用, 如硬件板卡主备份、负载均衡冗余配置方式、网元组 pool 等。由于实现了资源池化, 使得资源使用效率得到提升, 在硬件资源层面, 利用冗余资源可以达到可靠性目标。同时, 通过 MANO 能够对网络服务的可靠性进行全面监控及主被动管理, 并利用 NFV 快速提供资源的能力(如弹性伸缩), 提升系统故障恢复的灵活性及提高系统的可靠性。

应注意的是, NFV 强调业务可用性, 硬件可用性不是 NFVI 中最关键的, 硬件通常被视为资源池, 如果一些组件不可使用, VNF 会自动从相同的池中重新分配不同的硬件资源, 因此硬件修复成为一个预定的维持活动而不是紧急的行动。VNF 通常通过心跳机制实现故障检测, 心跳的频率至关重要, 如亚健康场景则无法快速切换。NFVI 资源冗余是提升系统的可靠性的主要手段, 在设计过程中需要在弹性、性能和成本之间权衡。

参考文献:

- [1] ETSI. Network functions virtualisation (NFV); resiliency requirements: GS NFV-REL001[S]. 2015.
- [2] ETSI. Network functions virtualisation (NFV); reliability; report on models and features for end-to-end reliability: GS

NFV-REL003[S]. 2016.

- [3] OPENNFV. High availability for OPNFV[R]. 2015.
- [4] ETSI. Network functions virtualisation (NFV); virtual network functions architecture: GS NFV-SWA001[S]. 2014.
- [5] ETSI. Network functions virtualisation (NFV); management and orchestration: GS NFV-MAN001[S]. 2014.

[作者简介]



杨旭(1985-), 男, 中国移动通信集团设计院有限公司工程师, 主要研究方向为 NFV/SDN、5G 架构演进、移动通信核心网和物联网等。



肖子玉(1969-), 女, 中国移动通信集团设计院有限公司教授级高级工程师, 中国通信学会高级会员, 美国北卡罗莱纳州立大学访问学者, 主要研究方向为移动通信核心网、NFV/SDN 5G、IMS、网络和信息安全等。



邵永平(1978-), 男, 中国移动通信集团设计院有限公司高级工程师, 主要研究方向为 NFV/SDN、移动通信核心网、网络和信息安全和物联网等。



宋小明(1985-), 女, 中国移动通信集团设计院有限公司工程师, 主要研究方向为移动通信核心网、信令网和 NFV 等。