International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.3527

(09/2021)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

Cloud Computing

Cloud computing – End-to-end fault and performance management framework of network services in inter-cloud

Recommendation ITU-T Y.3527

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

GLOBAL INFORMATION INFRASTRUCTURE	
General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899
INTERNET PROTOCOL ASPECTS	
General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	Y.1500–Y.1599
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
IPTV over NGN	Y.1900–Y.1999
NEXT GENERATION NETWORKS	
Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Enhancements to NGN	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Computing power networks	Y.2500–Y.2599
Packet-based Networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999
FUTURE NETWORKS	Y.3000–Y.3499
CLOUD COMPUTING	**Y.3500–Y.3599**
BIG DATA	Y.3600–Y.3799
QUANTUM KEY DISTRIBUTION NETWORKS	Y.3800–Y.3999
INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES	
General	Y.4000–Y.4049
Definitions and terminologies	Y.4050–Y.4099
Requirements and use cases	Y.4100–Y.4249
Infrastructure, connectivity and networks	Y.4250–Y.4399
Frameworks, architectures and protocols	Y.4400–Y.4549
Services, applications, computation and data processing	Y.4550–Y.4699
Management, control and performance	Y.4700–Y.4799
Identification and security	Y.4800–Y.4899
Evaluation and assessment	Y.4900–Y.4999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.3527

Cloud computing – End-to-end fault and performance management framework of network services in inter-cloud

Summary

Recommendation ITU-T Y.3527 provides framework and functional requirements of end-to-end (E2E) fault and performance management of network services (NSs) in inter-cloud. The functional requirements are derived from the corresponding typical use cases. In particular, a predictive model for fault and performance issue detection and localization is presented.

History

Edition	Recommendation	Approval	Study Group	Unique ID[*]
1.0	ITU-T Y.3527	2021-09-13	13	11.1002/1000/14760

Keywords

End-to-end, fault, framework, inter-cloud, management, network service, performance.

[*] To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T Y.3527

Cloud computing – End-to-end fault and performance management framework of network services in inter-cloud

1 Scope

This Recommendation specifies framework and functional requirements for end-to-end (E2E) fault and performance management of network services (NSs) in inter-cloud. The scope of this Recommendation includes:

– an overview of E2E fault and performance management of NSs in inter-cloud;

– functional requirements for E2E fault and performance management of NSs in inter-cloud;

– typical use cases for E2E fault and performance management of NSs in inter-cloud;

– a predictive model for fault and performance issues detection and localization.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T M.3010] Recommendation ITU-T M.3010 (2000), *Principles for a telecommunications management network*.

[ITU-T X.711] Recommendation ITU-T X.711 (1997), *Information technology – Open Systems Interconnection – Common management information protocol: Specification*.

[ITU-T X.1601] Recommendation ITU-T X.1601 (2015), *Security framework for cloud computing*.

[ITU-T Y.3502] Recommendation ITU-T Y.3502 (2014) | ISO/IEC 17789:2014, *Information technology – Cloud computing – Reference architecture*.

[ITU-T Y.3515] Recommendation ITU-T Y.3515 (2017), *Cloud computing – Functional architecture of Network as a Service*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 cloud service customer [b-ITU-T Y.3500]: Party which is in a business relationship for the purpose of using cloud services.

3.1.2 cloud service provider [b-ITU-T Y.3500]: Party which makes cloud services available.

3.1.3 inter-cloud computing [b-ITU-T Y.3511]: The paradigm for enabling the interworking between two or more cloud service providers.

3.1.4 network as a service (NaaS) [b-ITU-T Y.3500]: Cloud service category in which the capability provided to the cloud service customer is transport connectivity and related network capabilities.

NOTE – NaaS can provide any of the three cloud capabilities types.

3.1.5 network function [ITU-T Y.3515]: A function of a network infrastructure whose external interfaces and functional behaviour are well specified.

NOTE – Examples of network functions include network switches and network routers.

3.1.6 network service [ITU-T Y.3515]: A collection of network functions with a well specified behaviour.

NOTE – Examples of network services include content delivery networks (CDNs) and IP multimedia subsystem (IMS).

3.1.7 physical network function [ITU-T Y.3515]: A network function implemented via a tightly coupled software and hardware system.

3.1.8 service function chain [b-ITU-T Y.Suppl. 41]: A chain that defines an ordered set of abstract service functions and ordering constraints that must be applied to packets and/or frames and/or flows selected as a result of classification and/or policy.

3.1.9 virtualized network function [ITU-T Y.3515]: A network function that can be deployed as a software on a NaaS cloud service provider infrastructure.

NOTE – Examples of virtualized network functions include virtual switches and virtual routers.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 cloud service: One or more capabilities offered via cloud computing invoked using a defined interface. It may comprise the hardware and hypervisor layers delivering individual servers, border routers, firewalls, load balancers & switches.

NOTE – Expanded version of definition in [b-ITU-T Y.3500].

3.2.2 network function virtualization (NFV): Principle of separating network functions from the hardware they run on by using virtual hardware abstraction.

NOTE – Based on [b-ETSI GR NFV 003].

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

CCS	Cloud Compute and Storage
CDN	Content Delivery Network
CMIP	Common Management Information Protocol
CSC	Cloud Service Customer
CSP	Cloud Service Provider
E2E	End to End
FCAPS	Fault, Configuration, Accounting, Performance and Security
IMS	IP Multimedia Subsystem
IP	Internet Protocol
NaaS	Network as a Service

NC	Network Connectivity
NF	Network Function
NFV	Network Function Virtualization
NS	Network Service
OSS	Operations Support System
PNF	Physical Network Function
PM	Physical Machine
PR	Physical Resources
QoS	Quality of Service
SFC	Service Function Chain
SLA	Service Level Agreement
TMN	Telecommunication Management Network
VM	Virtual Machine
VNF	Virtualized Network Function

5 Conventions

In this Recommendation:

The phrase "**is required to**" indicates a requirement that must be strictly followed and from which no deviation is permitted if conformity to this Recommendation is to be claimed.

The phrase "**is recommended**" indicates a requirement that is recommended but which is not absolutely required. Thus this requirement need not be present to claim conformity.

6 Overview of E2E fault and performance management of NSs in inter-cloud

6.1 NSs in the virtualized environment

According to [ITU-T Y.3515], an NS is a collection of network functions (NFs) with a well specified behaviour, examples of NSs include CDNs and IMSs. An NS can also be made up of NF(s) or component NS(s), characterized by its functional and behavioural specification (see [b-ETSI GR NFV 003]). The NS contributes to the behaviour of the higher layer service, which is characterized by at least performance, dependability and security specifications. E2E NS behaviour is the result of the combination of the behaviours of individual NFs, as well as the network infrastructure composition mechanism.

When supporting NaaS connectivity services, an NS can be described as an abstracted transport connectivity between two endpoints in a virtualized environment where the endpoints may be located in one or more clouds. The virtualization hierarchy of NS is shown in Figure 6-1.

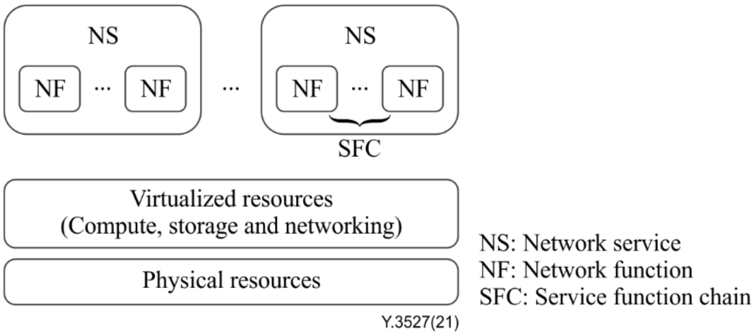


Figure 6-1 – Virtualization hierarchy of NS

An NS can be described as an E2E implementation using a service function chain (SFC), interconnecting the virtual network resources. An SFC is an ordered set of virtualized network functions (VNFs) in the virtualized environment that represents functions like routers and broadband network gateways or middle-boxes like load balancers and firewalls, which act on the traffic in the sequence they appear in the chain. Such VNFs are hosted on virtual machines (VMs) instantiated over physical data centre and network resources. An example of an E2E NS in inter-cloud is shown in Figure 6-2. This NS is composed of VNF1 to VNF5, which belong to different cloud service providers (CSPs).

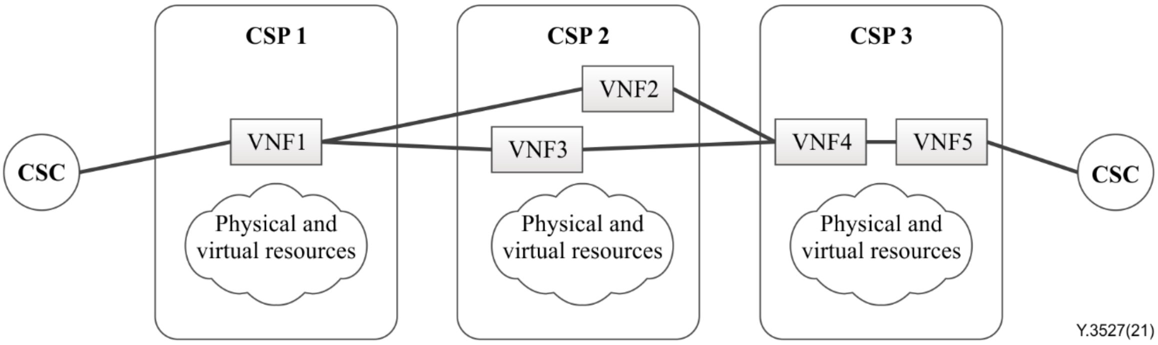

CSP 1
CSP 2
CSP 3
CSC
VNF1
VNF2
VNF3
VNF4
VNF5
CSC
Physical and virtual resources
Physical and virtual resources
Physical and virtual resources
Y.3527(21)

Figure 6-2 – An E2E NS in inter-cloud

6.2 Background

The traditional telecommunication network deployment largely involves use of physical network appliances like routers, switches, broadband remote access servers, as well as middle-boxes like firewalls, deep packet inspectors or load balancers. These integrated hardware and software solutions are normally closed and proprietary leading to vendor lock-in, thereby making expansions and deployment of new services difficult and time consuming. Such equipment is also not amenable to easy scaling or redeployment of resources. The power and space requirements, as well as the total cost of operation, are higher in physical element-based networks.

In traditional networks, time-tested standards relating to fault, configuration, accounting, performance and security (FCAPS) are embodied in the common management information protocol (CMIP) specified in [ITU-T X.711] and telecommunication management network (TMN) specified in [ITU-T M.3010]. Network management based on relevant Recommendations provides five nines (99.999%) availability and carrier grade reliability.

Inter-cloud computing, coupled with the deployment of NFV, provides numerous advantages to CSPs including ease of deployment, ease of scaling, ease of introducing and switching off services, and reduced cost of operation. Therefore, NSs in inter-cloud need a strong fault and performance management system to replace traditional networks. For carrier grade availability and reliability of up to 99.999% for NS inter-cloud, there is a need for standardization of the framework for fault and performance management to deal with complexity in such networks, as the anomalous behaviour could be in the hardware, VMs, VNFs, SFCs or at the service levels.

6.3 Challenges of E2E fault and performance management for NSs in inter-cloud

Telecommunication networks have been traditionally designed to provide high availability and standards-based quality of service (QoS). In inter-cloud, NS deployment over multiple clouds identifies new challenges to equip inter-cloud management systems to deal with management issues. Especially, those NSs relay over underlying both physical and software infrastructure. Therefore, E2E management is related to the physical, virtual layer or the VNFs of inter-cloud environments where VMs are instantiated, on which particular VNFs are hosted.

For NSs in inter-cloud, faults may occur for many more reasons compared to traditional physical telecommunication networks. The virtual resources are created on shared physical resources like servers or network equipment, using virtualization software. One reason why virtual resources may fail is because of the failure of physical resources. Even if the physical resources are healthy, virtual resources may fail. Furthermore, even if both physical and virtual resources are healthy, the VNFs instantiated on these virtual resources may have problems causing NSs to malfunction or totally break down. The myriad levels of malfunctions make handling of fault and performance issues in inter-cloud more complex.

Some of the key challenges for E2E fault and performance management for NSs in inter-cloud are as follows:

- absence of a standardized framework;
- non-applicability of traditional rule-based techniques when used in inter-cloud;
- multiple layers of implementation including physical infrastructure, virtual resource, NFs and NSs;
- massive distribution of NFs and underlying resources over different clouds.

6.4 Goals of E2E fault and performance management of NSs in inter-cloud

The goals of E2E fault and performance management of NSs in inter-cloud can be summarized as follows.

- Detection of any condition that has already led to or could lead to degraded performance or failure, which could be caused by manifested faults, hidden faults or inconspicuous deviations. The goal of fault and performance issue detection is to sense and notify impending or actual fault and performance issues.
- Identification and localization of manifested and impending faults. The goal of fault and performance issue localization is to determine the root cause of the problem by identifying the resources that are malfunctioning or the severity with which the resources may malfunction in the future.

6.5 Framework of E2E fault and performance management of NSs in inter-cloud

The fault and performance management of NSs in inter-cloud is a collaborative process among the elements constituting the service and the management functional components involved. The fault and performance management-related responsibilities are jointly implemented by some functional components of the operations support systems (OSSs) defined in the multi-layer functions of cloud-computing reference architecture [ITU-T Y.3502]. These functional components include OSS-NS, OSS-NF, OSS-cloud compute and storage (OSS-CCS), OSS-network connectivity (OSS NC) and OSS- OSS-physical resources (PR). Their interrelationship in the context of NSs is illustrated in Figure 6-3. For more information about these functional components, please see clauses 7.8 and 8.3 of [ITU-T Y.3515].

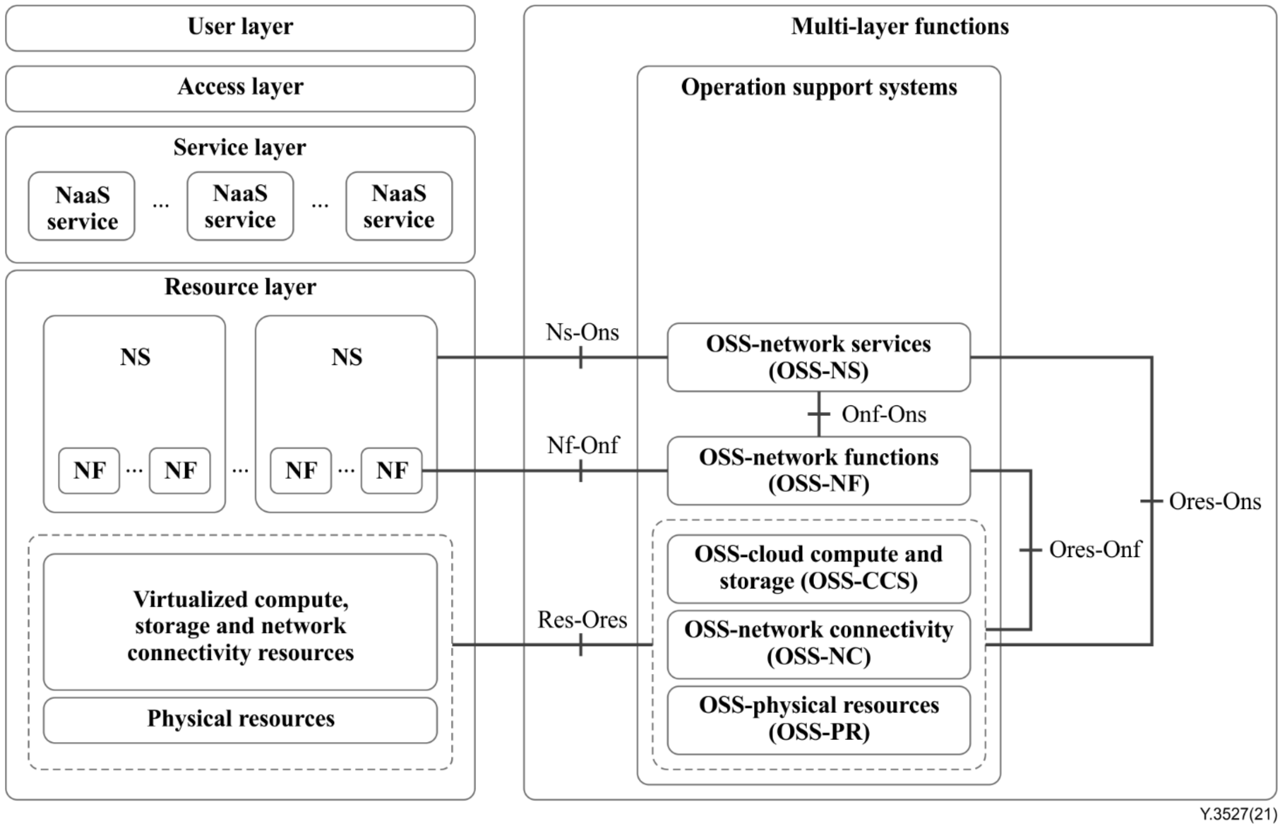

User layer
Access layer
Service layer
NaaS service
...
NaaS service
...
NaaS service
Resource layer
NS
NS
NF
...
NF
...
NF
...
NF
Virtualized compute, storage and network connectivity resources
Physical resources
Multi-layer functions
Operation support systems
Ns-Ons
OSS-network services (OSS-NS)
Onf-Ons
Nf-Onf
OSS-network functions (OSS-NF)
Ores-Ons
Ores-Onf
OSS-cloud compute and storage (OSS-CCS)
Res-Ores
OSS-network connectivity (OSS-NC)
OSS-physical resources (OSS-PR)
Y.3527(21)

Figure 6-3 – OSS functionalities and reference points for fault and performance management of NSs

The responsibilities of functional components related to fault and performance management of NSs are as follows.

- **OSS-NS**: which is responsible for managing the lifecycle of NSs and using available resources or requesting additional resources to maintain the required performance. For handling fault and performance issues, it monitors NSs and resources, and detects anomalous conditions. It gets NF level alarms from OSS-NF and resource level alarms from OSS-CCS, OSS-NC and OSS-PR. It correlates alarms from various sources to localize faults and performance conditions.
- **OSS-NF**: which is responsible for managing the lifecycle of NFs. For handling fault and performance issues, it interacts with NF instances to obtain NF-related fault and performance information. It also collects NF instance-related resource information and sends it to OSS-NS for fault detection and localization.
- **OSS-CCS, OSS-NC and OSS-PR**: which are responsible for collecting alarms related to physical and virtual resources. They forward fault and performance alarms to OSS-NS and OSS-NF for broader correlation and root cause analysis. The fault information may include VM crashes, virtual port malfunction, storage failure and resource unavailability.

The reference points related to fault and performance management of NSs are as follows.

- **Res-Ores**: This reference point covers the interactions between the virtual and physical resources, and the functional components about virtual and physical resource management, i.e., OSS-CCS, OSS-NC and OSS-PR. It includes the interaction related to reporting resource level fault and performance issues.
- **Nf-Onf**: This reference point covers the interactions between the NFs and OSS-NF. It includes the interaction related to reporting NF level fault and performance issues.

- **Ns-Ons**: This reference point covers the interactions between the NSs and OSS-NS. It includes the interaction related to reporting NS level fault and performance issues.
- **Onf-Ons**: This reference point covers the interactions between the OSS-NF and OSS-NS. It includes the interaction related to exchanging information about the creation and modification of NFs, and forwarding fault and performance issues related to NFs.
- **Ores-Onf**: This reference point covers the interactions between the OSS-NF and the functional components for virtual and physical resource management. It includes the interaction related to exchanging information about resource level fault and performance issues.
- **Ores-Ons**: This reference point covers the interactions between the OSS-NS and the functional components for virtual and physical resource management. It includes the interaction related to exchanging information about resource level fault and performance issues.

Figure 6-3 presents the OSS functionalities and reference points for fault and performance management of NSs based on the cloud-computing reference architecture [ITU-T Y.3502] in one CSP. In an inter-cloud environment, the framework of E2E fault and performance management of NSs is presented as Figure 6-4. It is based on fault and performance management chains including the cloud service customer (CSC) and one or more CSPs.

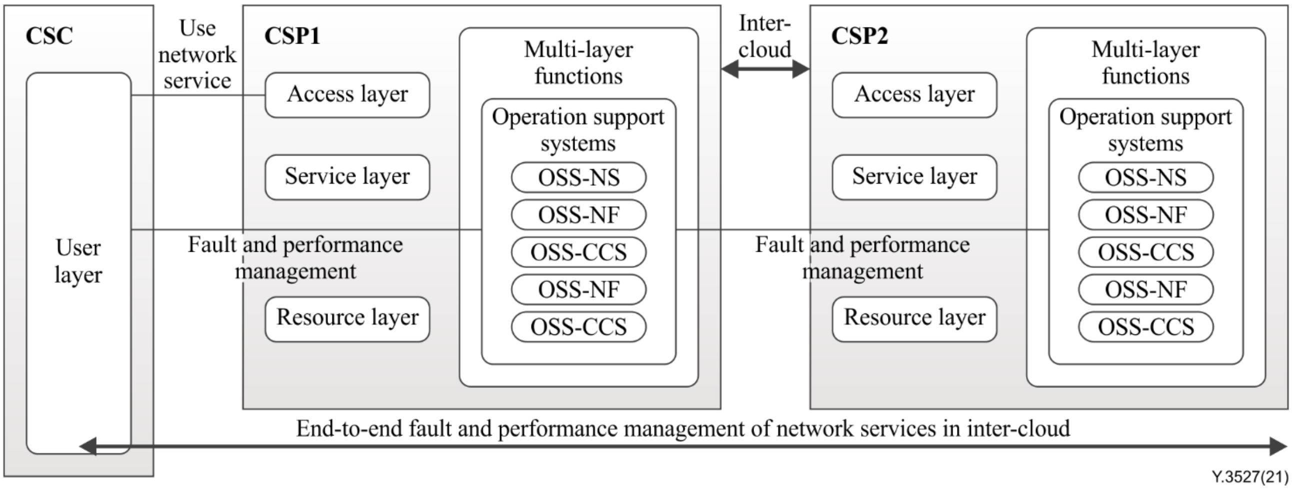


Figure 6-4 – Framework of E2E fault and performance management of NSs in inter-cloud

7 Functional requirements for E2E fault and performance management of NSs in inter-cloud

This clause identifies functional requirements applicable for E2E fault and performance management of NSs in inter-cloud.

7.1 Fault and performance data collection

It is required that a CSP support collection of fault and performance data from all resources even in different CSPs that support the implementation of the NS.

7.2 Fault and performance problem detection

It is required that a CSP support detection of the unavailability and failures of physical and virtual resources that might cause fault or performance problems in NFs running on top of them.

It is required that a CSP support filtering out dependent and routine operational events.

It is recommended that a CSP support classification of detected faults into manifested or impending so that further action can accordingly be taken.

7.3 Fault and performance problem localization

It is required that a CSP support determination of the root cause of a manifest fault by identification of the resources that are malfunctioning.

It is recommended that a CSP support determination of the severity of an impending fault with which the resources may malfunction in the future.

7.4 Fault and performance problem correlation

It is required that a CSP support mapping each fault to the impacted components of the NS.

It is required that a CSP identify unavailability of virtual resources that are or will be affected by failures on the physical resources under them, or identifies VNF instances that are or will be affected by failures of the virtual resources.

7.5 Fault handling in inter-cloud

It is recommended that a CSP support recovery of a manifest fault using an active-standby configuration.

It is recommended that a CSP support handling of an impending fault using an active-standby configuration or VM migration.

8 Security consideration

Security aspects for consideration within the cloud-computing environment are described in [ITU-T X.1601], which analyses security threats and challenges and describes security capabilities that could mitigate these threats and meet security challenges.

Appendix I

Use case of end-to-end fault and performance management of network services in inter-cloud

(This appendix does not form an integral part of this Recommendation.)

I.1 Use case template

The use cases developed in this appendix should adopt the unified format in Table I.1 for better readability and convenient material organization.

Table I.1 – Unified format

Title	The title of the use case
Description	Scenario description of the use case
Roles	Roles involved in the use case
Figure (optional)	Figure to explain the use case, not mandatory
Pre-conditions (optional)	The necessary pre-conditions that should be achieved before starting the use case.
Post-conditions (optional)	The post-condition that will be carried out after the termination of current use case.
Derived requirements	Requirements derived from the use cases, whose detailed description is presented in the relevant clause

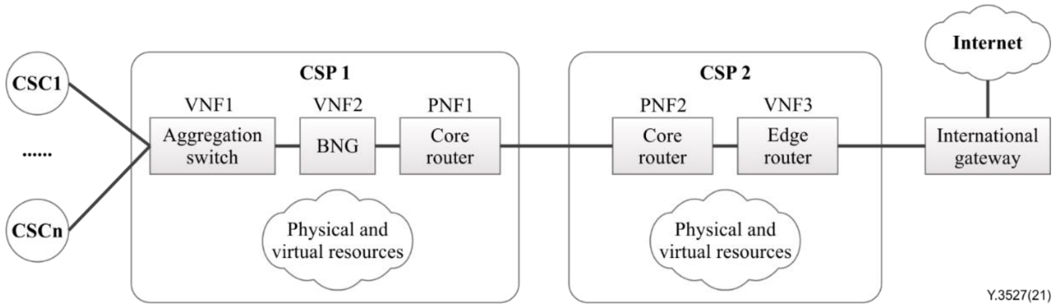
I.2 Virtual broadband service

This use case illustrates fault and performance management for a virtual broadband service in inter-cloud.

Table I.2 – Virtual broadband service

Title	Virtual broadband service
Description	CSP1 provides broadband service to CSC. The broadband service is an NS that is composed of VNFs realized as VNF1 to VNF3 and physical network functions (PNFs) realized as PNF1 to PNF2. VNF1, VNF2 and PNF1 are deployed in CSP1, VNF3 and PNF2 are deployed in CSP2. CSP1 is the primary CSP in the inter-cloud intermediary pattern. As the operator of this NS, CSP1 should have the comprehensive topology of the NS and the relationship between of the elements, e.g., the virtual resources and their supporting physical resources, and the VNF instances and their supporting virtual resources. During the operation of this NS, large volumes of high dimensional data in the form of markers like alarms, notifications, warnings and measurement of performance indicators are produced from both CSP1 and CSP2. As the operator of this NS, CSP1 has the responsibility for managing the fault and performance issues to meet service level agreements (SLAs). To achieve this, CSP1 should collect all markers from CSP1 and CSP2 related to the fault and performance problems of the NS. Based on these data, CSP1 should detect any condition that has already led to or could lead to degraded performance or failure, and identify and localize manifested and impending faults.
Roles	CSC, CSP

Table I.2 – Virtual broadband service

Figure (optional)	 Internet CSC1 ...... CSCn CSP 1 VNF1 VNF2 PNF1 Aggregation switch BNG Core router Physical and virtual resources CSP 2 PNF2 VNF3 Core router Edge router Physical and virtual resources International gateway Y.3527(21)
Pre-conditions (optional)	
Post-conditions (optional)	
Derived requirements	– Fault and performance data collection (see clause 7.1) – Fault and performance problem detection (see clause 7.2) – Fault and performance problem localization (see clause 7.3)

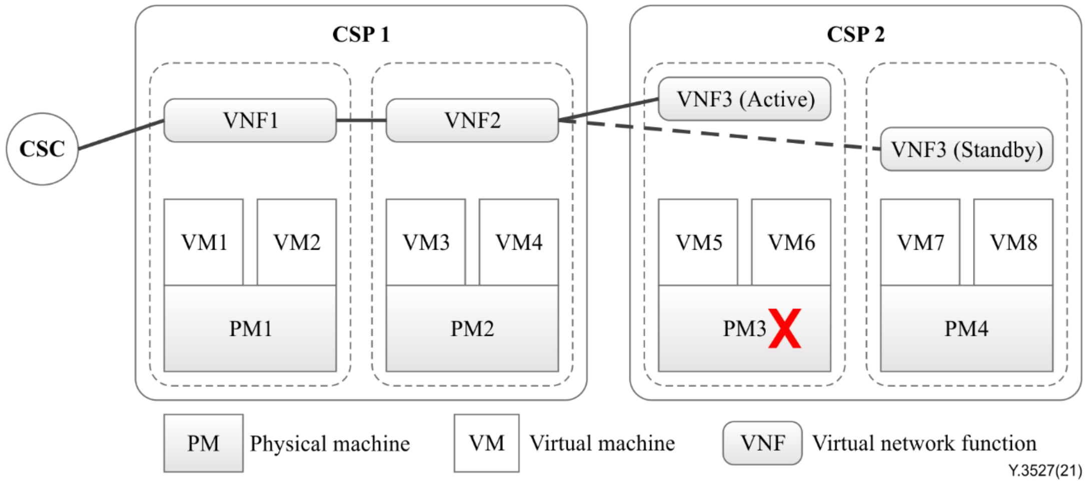
I.3 Fault recovery in inter-cloud

This use case illustrates the process of recovering a fault of an NS in inter-cloud.

Table I.3 – Fault recovery in inter-cloud

Title	Fault recovery in inter-cloud
Description	CSP1 provides an NS to CSC. The NS is composed of VNFs realized as VNF1 to VNF3. VNF1 and VNF2 are deployed in CSP1, VNF3 is deployed in CSP2 with an active-standby mode. VNF3 (active) is deployed on VM5 that launched in physical machine 3 (PM3), VNF3 (standby) is deployed on VM8 that launched in PM4. CSP1 is the primary CSP in the inter-cloud intermediary pattern. As the operator of this NS, CSP1 needs to detect faults in physical resources that can affect proper functioning of virtual resources. Once such a fault is detected, CSP1 shall find out which virtual resources and VNFs are affected by this fault, and decide how to handle it. In this use case, VM5 and VM6 are affected by a fault in PM3. VNF3 (active) is out of function because it is deployed on VM5. To keep the availability of the NS, CSP1 shall switch the VNF3 (standby) to active state and change the access path of the NS. Then CSP1 needs to find another VM that is not hosted in PM4 to deploy VNF3 as the new standby node.
Roles	CSC, CSP

Table I.3 – Fault recovery in inter-cloud

Figure (optional)	 CSP 1 CSP 2 CSC VNF1 VNF2 VNF3 (Active) VNF3 (Standby) VM1 VM2 VM3 VM4 VM5 VM6 VM7 VM8 PM1 PM2 PM3 PM4 PM Physical machine VM Virtual machine VNF Virtual network function Y.3527(21)
Pre-conditions (optional)	A VNF is deployed in active-standby mode and the active node is affected by a hardware fault.
Post-conditions (optional)	The standby node is switched to active node.
Derived requirements	– Fault and performance problems correlation (see clause 7.4) – Fault handling in inter-cloud (see clause 7.5)

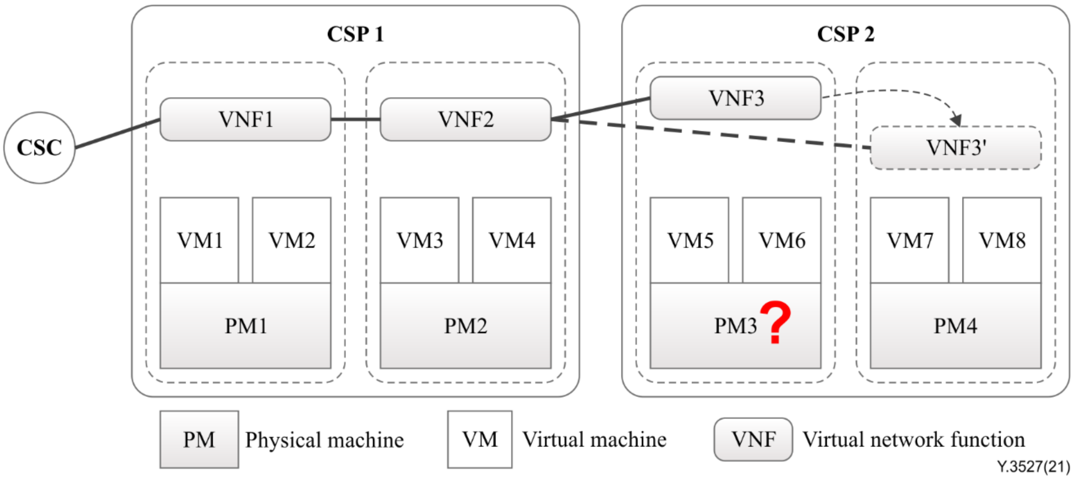
I.4 Fault prediction and handling in inter-cloud

This use case illustrates the process of predicting and handling an impending fault of an NS in inter-cloud.

Table I.4 – Fault prediction and handling in inter-cloud

Title	Fault prediction and handling in inter-cloud
Description	CSP1 provides an NS to CSC. The NS is composed of VNFs realized as VNF1 to VNF3. VNF1 and VNF2 are deployed in CSP1, VNF3 is deployed in CSP2. CSP1 is the primary CSP in the inter-cloud intermediary pattern. As the operator of this NS, CSP1 could predict some impending faults based on the data collected from the physical resources. For example, the CSP1 may find that the central processing unit temperature of PM3 is rising, then exceeding a threshold value, which may trigger a restart of the PM. In this case, CSP1 can predict an impending fault for PM3. As VNF3 is deployed on VM5 that launched in PM3, this impending fault will affect VNF3. To keep the availability of the NS, CSP1 shall handle this impending fault. If VNF3 is configured with active-standby mode as in the use case in clause I.3, CSP1 will switch the standby node to active state and change the access path of the NS. If VNF3 is not configured in active-standby mode, CSP1 will select a healthy PM (in this use case PM4 is the healthy PM) to migrate the VNF3 to a new VM in that PM and change the access path of the NS.
Roles	CSC, CSP

Table I.4 – Fault prediction and handling in inter-cloud

Figure (optional)	 CSP 1 CSP 2 CSC VNF1 VNF2 VNF3 VNF3' VM1 VM2 VM3 VM4 VM5 VM6 VM7 VM8 PM1 PM2 PM3 ? PM4 PM Physical machine VM Virtual machine VNF Virtual network function Y.3527(21)
Pre-conditions (optional)	
Post-conditions (optional)	
Derived requirements	– Fault and performance problems correlation (see clause 7.4) – Fault handling in inter-cloud (see clause 7.5)

Appendix II

A predictive model for fault and performance issues detection and localization

(This appendix does not form an integral part of this Recommendation.)

II.1 Problem statement

For NSs in inter-cloud, fault and performance issues can have complex origins within virtual resources, including computation, storage, networking, and VNFs. In such case, it would be very difficult to capture the intricate relationships among the features (e.g., the location of the fault, resources involved, and markers produced) and the corresponding labels (fault, no fault, impending fault, manifested fault, fault-severity, etc.) through traditional deterministic methods.

Traditional failure detection methods depend on probing or running tests on hardware, which are not accessible to the NSs deployed on virtual resources. Too much probing or software testing may overload the VMs that have been optimized for the NF hosted on them. Attempts to apply other traditional methods, like rule-based approaches involving direct correlation of the markers with the faults, get mired in complexity and prove to be inadequate.

The traditional deterministic methods fail to deliver in virtualized environments in which virtual resources can be dynamically scaled, migrated or destroyed. Predictive techniques are needed to identify and resolve management issues before or after they have occurred.

This Recommendation provides a model based on a judicious combination of shallow as well as deep structures in machine learning to ensure carrier grade availability and reliability.

II.2 A model for fault and performance issues detection and localization

The model approach has predictive and deductive properties to meet fault and performance management requirements. Run time monitoring and measurements, alarms, notifications and warnings, configuration changes, measurements and environmental factors are all used along with models trained with historical data to draw inferences about manifest fault and performance issues. Additionally, decisions about impending faults are taken using these inputs and the predictive properties of machine-learning models.

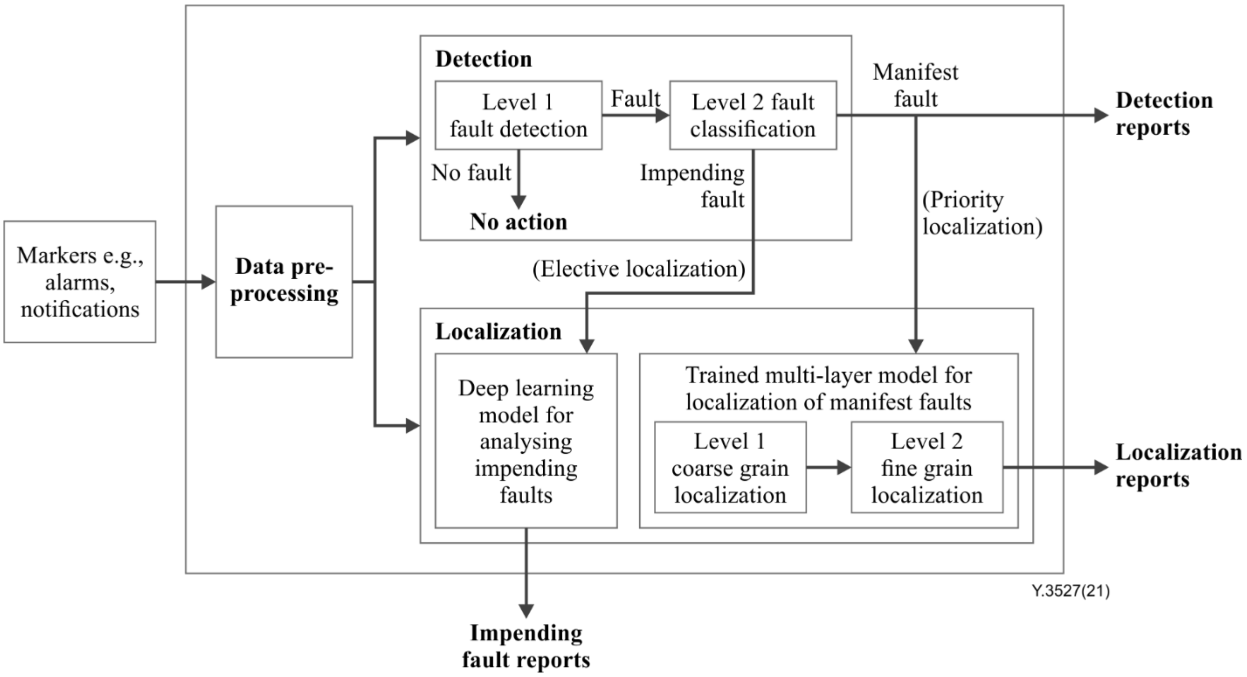

Markers e.g., alarms, notifications
Data pre-processing
Detection
Level 1 fault detection
Fault
Level 2 fault classification
No fault
No action
Impending fault
Manifest fault
Detection reports
(Priority localization)
(Elective localization)
Localization
Deep learning model for analysing impending faults
Trained multi-layer model for localization of manifest faults
Level 1 coarse grain localization
Level 2 fine grain localization
Localization reports
Impending fault reports
Y.3527(21)

Figure II.1 – Model for fault and performance issues detection and localization

The model shown in Figure II.1 consists of three main sub-systems: data pre-processing, detection and localization. Data pre-processing involves collation and normalization of the dataset to remove biases. The pre-processing policy may also involve the reduction of features based on some criterion like correlation with the labels.

The detection sub-system decides whether there is a manifest or an impending fault or a performance issue. Detection is essentially a two-stage binary classification. Firstly, it classifies the outcome into "normal performance" and "abnormal performance" or "fault" and "no fault" categories. Then for the "fault" or "abnormal performance" cases, it decides whether the problem is manifest, i.e., it has already occurred somewhere in the network in some form, or impending, i.e., it might happen in the near future.

The localization sub-system fulfils the localization of the detected faults. Localization of manifested faults is taken up on priority while for the impending faults it is elective, nevertheless important. For the manifested faults, the model uses a multi-layered localization strategy using machine-learning classification models. At localization layer 1, the broad category of the manifested fault is determined, e.g., network performance problem. At localization layer 2, the system makes a finer identification of the problem to assist in the identification of its root cause, i.e., malfunctioning resources or resources subject to performance degradation. For the impending faults, a deep learning strategy uses the markers to predict the severity and location of faults.

II.3 Markers and metrics for fault and performance issue detection and localization

During the operation, CSP networks produce large volumes of high dimensional data in the form of markers like alarms, notifications, observed behaviour, warnings, counter values and measurement of performance indicators. The markers used by CSPs are predominantly at the service and NF level.

There are a large number of markers that are directly or indirectly related to the occurrence of a fault and performance issue. Events produce the markers related to communication, QoS, processing, equipment and environment. Not only do each fault and performance issue usually have multiple markers, but also many markers appear in more than one type of issue. This means that when using machine learning for fault detection and localization, feature engineering, i.e., selection of appropriate markers is required to get better results.

The metrics used by CSPs to measure the health of the network provide important information about fault and performance problems at the macro level. Use of these as features in the training dataset will help learning algorithms to narrow down the scope of the localization effort.

Bibliography

[b-ITU-T Y.3500] Recommendation ITU-T Y.3500 (2014) | ISO/IEC 17788:2014, *Information technology – Cloud computing – Overview and vocabulary*.

[b-ITU-T Y.3511] Recommendation ITU-T Y.3511 (2014), *Framework of inter-cloud computing*.

[b-ITU-T Y.Suppl. 41] Recommendation ITU-T Y.Suppl. 41 (2016), *ITU-T Y.2200-series - Deployment models of service function chaining*.

[b-ETSI GR NFV 003] ETSI Group Report NFV 003 V1.5.1 (2020), *Network functions virtualisation (NFV); Terminology for main concepts in NFV*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	**Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities**
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2021