ETSI GR NFV-EVE 017 V1.1.1 (2020-08)

GROUP REPORT

Network Functions Virtualisation (NFV); Management and Orchestration; Report on the support of real-time/ultra-low latency aspects in NFV related to service and network handling

Reference

DGR/NFV-EVE017

Keywords

MANO, NFV, real time, service

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Group Report (GR) has been produced by ETSI Industry Specification Group (ISG) Network Functions Virtualisation (NFV).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

The present document analyses the impact on management and orchestration of Network Service (NS) instance(s) supporting low latency services from the perspective of the NFV-MANO architectural framework. The following topics are handled:

- Definition of relevant NFV-MANO use cases.
- Analysis of the use-cases and deriving potential requirements.
- Providing relevant recommendations.

The content of the present document is informative.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI GR NFV-IFA 012 (V3.1.1): "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Report on Os-Ma-Nfvo reference point - Application and Service Management Use Cases and Recommendations", October 2018.

[i.2] ETSI GR NFV-IFA 028 (V3.1.1): "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Report on architecture options to support multiple administrative domains".

[i.3] ETSI GS NFV-IFA 032 (V3.2.1): "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Interface and Information Model Specification for Multi-Site Connectivity Services".

[i.4] ETSI GS NFV-IFA 010 (V3.3.1): "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Functional requirements specification".

[i.5] ETSI GS NFV-IFA 011 (V3.2.1): "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; VNF Descriptor and Packaging Specification".

[i.6] ETSI GS NFV-IFA 014 (V3.3.1): "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Network Service Templates Specification".

[i.7] ETSI GS NFV-IFA 031 (V3.3.1): "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Requirements and interfaces specification for management of NFV-MANO".

[i.8] ETSI GS NFV-IFA 027 (V2.4.1): "Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Performance Measurements Specification".

[i.9] ETSI GS NFV-TST 008 (V3.2.1): "Network Functions Virtualisation (NFV) Release 3; Testing; NFVI Compute and Network Metrics Specification".

[i.10] ETSI GR NFV 003 (V1.5.1): "Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV".

[i.11] ETSI TS 122 261 (V15.8.0): "5G; Service requirements for next generation new services and markets (3GPP TS 22.261 version 15.8.0 Release 15)".

[i.12] ETSI GS NFV-IFA 013 (V3.3.1): "Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Os-Ma-Nfvo reference point - Interface and Information Model Specification".

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms given in ETSI GR NFV 003 [i.10] apply.

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in ETSI GR NFV 003 [i.10] and the following apply:

AF	Application Function
MLPOC	Multiple Logical Points Of Contacts
NFVIaaS-C	NFVIaaS Consumer
NFVIaaS-P	NFVIaaS Provider
SLPOC	Single Logical Points Of Contacts

4 Use Cases

4.1 Introduction

In ETSI TS 122 261 [i.11] various services requiring low latency guarantees have been identified. These use cases include urgent healthcare and emergency services, which require latency guarantees from 1 ms to 10 ms. A second class of uses cases includes smart factories and tactile interaction applications, which require latency to stay between 0,5 ms and 1 ms. The requirements towards the network of these use cases are summarized in Table 4.1-1. Thus, it has to be emphasized that these latency requirements are strictly on an end-to-end service level. Consequently, any operations impacting the actual service deployment and run-time operation should guarantee that those upper bounds are not exceeded.

The present document investigates the gaps in NFV-MANO specifications to support low latency services that are provisioned over NFV Network Services (NS). Thus generic use cases are presented which are later analysed from the perspective of NFV-MANO system in supporting and managing such services within strict end-to-end delay bound. The NFV-MANO system will be analysed in order to highlight the various aspects that can potentially impact the latency bounds of an active service over the NFVI. With respect to the analysis, necessary recommendations will be provided.

Table 4.1-1: Performance requirements for low-latency and high reliability scenarios [i.11]

Scenario	Max. allowed end-to-end latency (note 2)	Survival time	Communication service availability (note 3)	Reliability (note 3)	User experienced data rate	Payload size (note 4)	Traffic density (note 5)	Connection density (note 6)	Service area dimension (note 7)
Discrete automation	10 ms	0 ms	99,99%	99,99%	10 Mbps	Small to big	1 Tbps/km^2	100 000/km^2	1000 x 1000 x 30 m
Process automation – remote control	60 ms	100 ms	99,9999%	99,999%	1 Mbps up to 100 Mbps	Small to big	100 Gbps/km^2	1 000/km^2	300 x 300 x 50 m
Process automation – monitoring	60 ms	100 ms	99,9%	99,9%	1 Mbps	Small	10 Gbps/km^2	10 000/km^2	300 x 300 x 50
Electricity distribution – medium voltage	40 ms	25 ms	99,9%	99,9%	10 Mbps	Small to big	10 Gbps/km^2	1 000/km^2	100 km along power line
Electricity distribution – high voltage (note 1)	5 ms	10 ms	99,9999%	99,999%	10 Mbps	Small	100 Gbps/km^2	1 000/km^2 (note 8)	200 km along power line
Intelligent transport systems – infrastructure backhaul	30 ms	100 ms	99,9999%	99,999%	10 Mbps	Small to big	10 Gbps/km^2	1 000/km^2	2 km along a road

NOTE 1: Currently realised via wired communication lines.
NOTE 2: This is the maximum end-to-end latency allowed for the 5G system to deliver the service in the case the end-to-end latency is completely allocated to the 5G system from the UE to the Interface to Data Network.
NOTE 3: Communication service availability relates to the service interfaces, and reliability relates to a given system entity. One or more retransmissions of network layer packets may take place in order to satisfy the reliability requirement.
NOTE 4: Small: payload typically ≤ 256 bytes
NOTE 5: Based on the assumption that all connected applications within the service volume require the user experienced data rate.
NOTE 6: Under the assumption of 100% 5G penetration.
NOTE 7: Estimates of maximum dimensions; the last figure is the vertical dimension.
NOTE 8: In dense urban areas.
NOTE 9: All the values in this table are example values and not strict requirements. Deployment configurations should be taken into account when considering service offerings that meet the targets.

4.2 Use Case 1: Re-routing a low latency network service

This use case considers the situation when a low-latency service may have to be re-routed e.g. due to a network element failing, the topology change inside the NFVI, congestion event.

The failing network elements can be located inside a NFVI-PoP or on elements that are used to ensure interconnectivity between different NFV-PoPs. Apart from network resources the use case also considers the failure of VNFs. This could potentially involve the relocation of a failed VNF, which then would imply changes to the underlying network infrastructure connecting the affected VNF to the other elements of the NS.

The investigation will take into account that a VNF itself could be managing a part of the low latency service itself, e.g. by monitoring redundant links with which it is connected to the other NS elements. The use case should investigate if and how these kinds of VNF could interact with the NFV-MANO system.

The use case does not assume any specific procedures for re-routing of NS, but the NS rerouting process should take into account the latency bounds for the low-latency service. If multiples routes are available; those fulfilling the latency bounds best should be preferred. This would potentially involve monitoring latency bounds for routes (e.g. application latency monitoring).

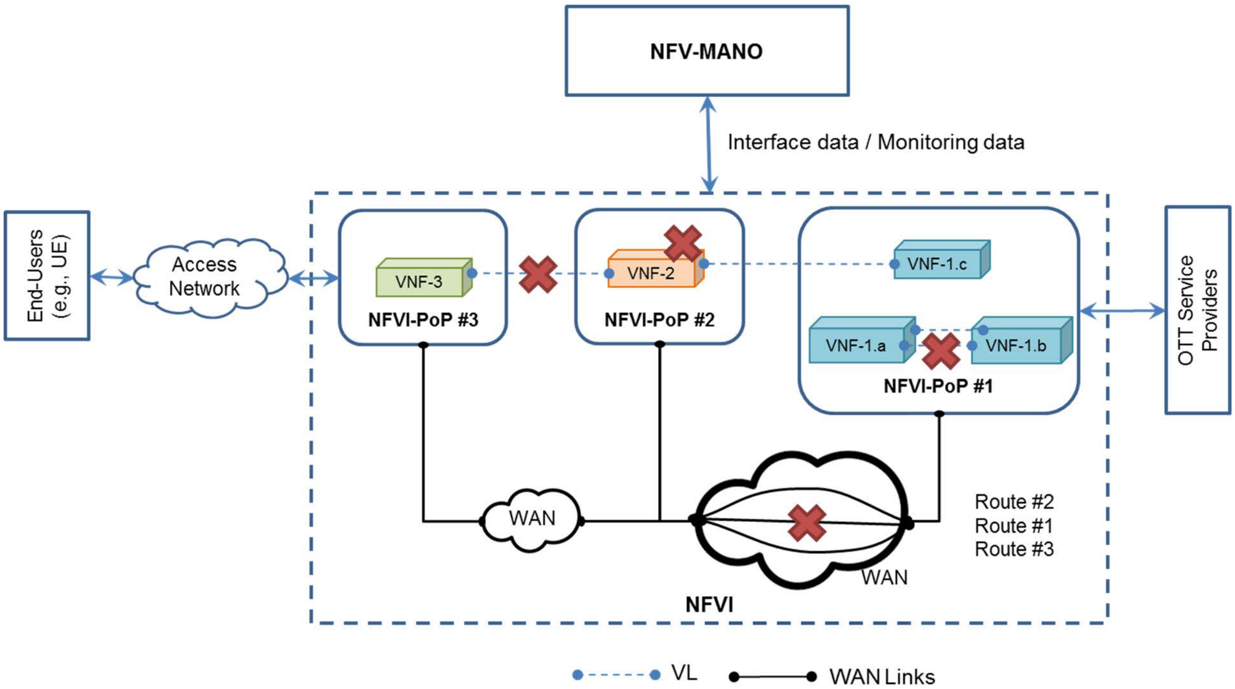

Figure 4.2-1: Overview of the use case 1 scenario

Figure 4.2-1 is presenting a network scenario consisting of 3 NFVI-PoPs interconnected over WAN links. There are two NS instances, a multisite NS and an intra-site NS instance. The multisite NS instance is composed of VNF instances, which are instantiated in different NFVI-PoPs, whereas the intra-site NS is composed of VNF instances that are instantiated within the same NFVI-PoP (i.e. VNF-1.a and VNF-1.b in NFVI-PoP #1 in Figure 4.2-1). A more detailed view on the WAN is showing different routes between NFVI-PoP #1 and NFVI-PoP #2. The following failure possibilities are assumed:

1) The current route #1 is failing and that there are 2 alternatives, one fulfilling the latency bounds of the low-latency service and one not fulfilling them.

2) Failure of a VL on a NS, within a single NFVI-PoP or across multiple NFVI-PoPs.

3) Failure of VNF itself that is part of a NS.

Each failure event will require the NFV-MANO system to trigger recovery actions to ensure time integrity of the affected NS.

4.3 Use Case 2: Mobility for a low latency network service

This use case considers the situation when a low-latency service is established and the client receiving the low-latency service is mobile. In such a situation, the client could change the access point it is connected to while receiving the service. The new access point could be connected to the same NFVI-PoP or even in a different NFVI-PoP. Usually the network connections are preconfigured and the NFV system does not even notice when clients move around. When dealing with low-latency services this stays basically the same but the network conditions might change when many clients receiving low-latency services move. When many clients move to the same new access point the network could get loaded in such a way that the low-latency conditions could no longer be guaranteed. In such a situation, the NFV-MANO system may have to detect and react to re-enforce the latency bounds given for the low-latency service.

In such a situation it is thus required that the NFV-MANO system is able to detect, or get notification of, degradation of low-latency services due to mobile users changing access points. When a service degradation is detected, the NFV-MANO system may have to derive and trigger suitable actions to restore the low-latency characteristics of the degraded service.

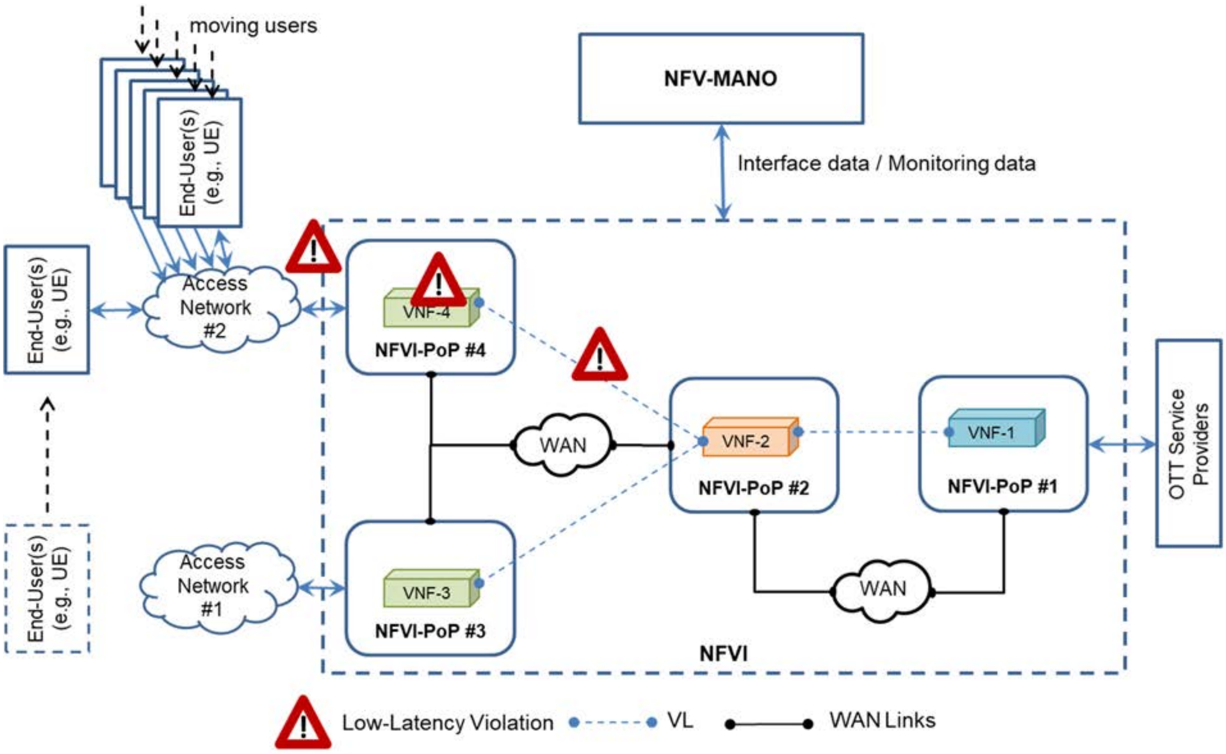

Figure 4.3-1: Overview of the use case 2 scenario

Figure 4.3-1 is presenting a network consisting of 4 NFVI-PoPs hosting different VNFs providing an NS and interconnecting WAN links. NFVI-PoP#3 and NFVI-PoP#4 are ingress for mobile end-users. Figure 4.3-1 also shows many end-users currently moving in to the access network connected to NFVI-PoP#4. Different locations are assumed where low-latency guarantees could be violated due to the mass of end-users moving in.

Each location that causes the violation of these guarantees will require the NFV-MANO system to trigger recovery actions to ensure low latency characteristics of the affected NS.

The investigation will analyse if the requirements for a low-latency NS will introduce additional requirements on NS monitoring to detect degradation of NS low-latency guarantees. In addition it will identify which management elements of the whole NFV-MANO system might be affected and may have to act to restore the low-latency characteristic of the degraded NS.

4.4 Use Case 3: Supporting Low latency Application Function Overlaying the Network Service

In ETSI GR NFV-IFA 012 [i.1] Application Functions (AFs) are introduced, where the AFs rely on the functional/operational characteristics of the NS or one or more of its constituent VNFs that are parts of NS(s) to deliver application services. On the other hand, the VNF(s) and/or the NS(s) may also utilize the functions provided by the AFs for their own operational/functional support, e.g. by analysing their KPIs to improve service delivery.

This use case considers NFV-MANO supporting AFs responsible for delivering low latency services, such as live streaming of multimedia content over NS. This is relevant because the AF instance could be the cause of delay and/or the underlying NS or one or more of its components can impact the latency bounds of the application service, and thus it is important for NFV-MANO to be aware of it. The application that is providing/serving the low-latency service could provide latency information about the end-to-end network path since it usually receives feedback from the clients and can retrieve traffic information from the underlying NS/VNF instance(s) as laid out in ETSI GR NFV-IFA 012 [i.1]. This latency information could then be provided to the management system of the AF.

ETSI GR NFV-IFA 012 [i.1] foresees relying on the interfaces on the Os-Ma-Nfvo reference point for an AF to interact with the NFV-MANO system. In this use case, the relevant notifications from the AF can be used by the OSS/BSS to trigger the NFV-MANO system about degraded conditions to help identify and improve the degraded latency conditions as far as the network path of an NS is concerned that the AF is using. The AF may have to provide information that enables the NFV-MANO system to map the application latency requirements to a NS being managed by NFV-MANO. To fulfil this task the AF might have to use other interfaces apart from the Os-Ma-Nfvo ones. The information provided by the AF can be manifold: it could be a simple Boolean value informing about bad conditions or might contain more fine grained information exchanged between the AF and the VNFs that couple it with the NS.

The use case will consider different causes for latency degradation such as load on the AF, load on the NS itself or users changing access points, etc. In such a situation it is beneficial that the NFV-MANO system gets help in detecting the degradation of low-latency services through notifications triggered by the AF. When a service degradation is detected and identified, the NFV-MANO system may have to derive and enforce suitable measures to resurrect the low-latency guarantees for the degraded service. These measures can be executed by different functional blocks of the NFV-MANO system depending on the cause of latency degradation.

The investigation of this use case will analyse how and to which functional blocks in NFV-MANO system AF specific performance and latency information may have to be provided.

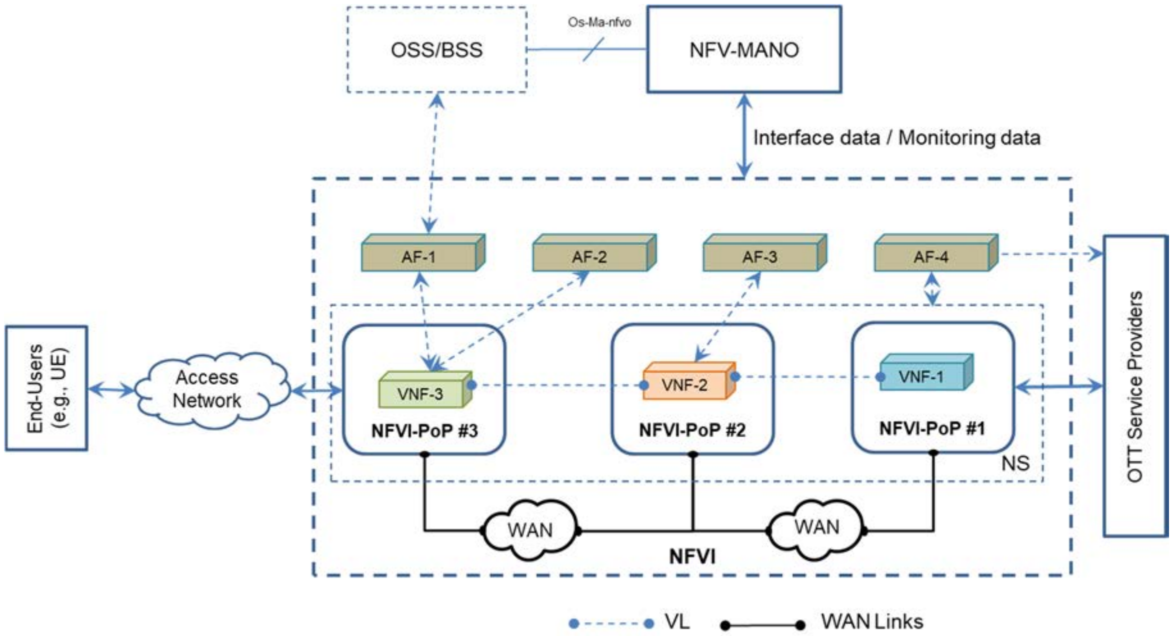

Figure 4.4-1: Overview of the use case 3 scenario

The use case scenario is illustrated in Figure 4.4-1 where AF instances are overlaid on a multi-site NS composed of 3 VNFs, namely VNF-1, VNF-2 and VNF-3 hosed in 3 NFVI-PoPs respectively. The 3 NFVI-PoPs are interconnected over WAN links. The AF instances 1 to 3 are coupled with a single VNF function whereas AF 4 is coupled with the whole NS. AF 4 could provide a low-latency service to an end-user over the NS while receiving application data from the NS or from an OTT service provider. The Figure shows the interaction of the AFs with the underlying NS and with the NFV-MANO system via the OSS/BSS over the Os-Ma-NFVO reference point, as specified in ETSI GR NFV-IFA 012 [i.1]. As an option, the AFs can interact with the NFV-MANO system over this reference point by providing application level latency information. This information could then trigger the NFV-MANO system via the OSS/BSS to take appropriate actions at the NS level to maintain the services' latency bounds.

5 Analysis

5.1 Introduction

The use cases presented in clause 4 describe scenarios in which a low-latency service is provided by a NS and the established connections get degraded by different events in different locations of the NS. To analyse the use cases this clause first explains what characterizes a low-latency service and what are the aspired reactions to keep up or re-establish the service guarantees laid down in the SLA of the NS providing the service.

The characteristics of a low-latency service add some end-to-end latency bounds to the NS that may be guaranteed for providing the quality of service the service provider wants to deliver to its customers. This is especially useful when a service provider is delivering real-time video or audio services. As a first step the NFV-MANO system may have to know the latency that the different NS components introduce when transmitting data. These components include network links as well as virtual functions that are used to provide the service. With this information, the NFV-MANO system is able to establish a NS that can provide the end-to-end latency bounds requested by the SLA describing the NS.

To ensure that the end-to-end latency bounds are obeyed during the whole lifetime of the service, the NFV-MANO system may have to regularly measure the latency of the NS components and monitor changes of the measurements to be able to recognize that is violating or even better is soon violating the end-to-end latency bounds defined by the NS QoS parameters. If such a violation occurs the NFV-MANO system should strive to keep the violation at a minimum and react on it to avoid that the latency bound are violated to such an extent that the service is disturbed severely. As a result, this means that NFV-MANO system should react fast on such violations to keep up or re-establish the guaranteed end-to-end latency bounds.

5.2 Considerations for low-latency service measurements and recovery

After sketching the overall picture of the important characteristics of a low-latency NS service with the NFV-MANO system in the previous clause, this clause will give a closer analysis of the different aspects that the characteristic pointed out.

At the first place, latency measurements are the key feature that is required to support low-latency services within the NFV-MANO system. The knowledge of the latency of network links and active network components like switches, VNFs, etc. would facilitate the establishment and maintenance of an end-to-end low-latency bound within a NS supporting low-latency service(s). Some latency bounds may have to be actively measured (network links) while others may be derived/included from/in other measurements. As an example, the network link latency can contain the latency introduced by the network switch, or the latency introduced by a VNF could be derived from the network link latency where the VNF is located between the endpoint of that link. Nevertheless, the more measurements could be taken at different locations the easier it would be to detect latency violation early and to locate it within the NS supporting low latency service. It is thus important for the NFV-MANO system to be aware of the monitoring points of a NS instance(s) supporting low latency service(s).

As a consequence, the measurement points should be carefully specified. Too few measurements or too few locations could slow down the detection time of a violation event and make the reestablishment more complicated. Too many measurement or unsuitable locations can slow down NFV-MANO system processing abilities and impact the management of services as such or even other NS's. It is noted that ETSI GS NFV-IFA 011 [i.5], ETSI GS NFV-IFA 014 [i.6] and ETSI GS NFV-IFA 031 [i.7] specify an attribute collection period, which describes the periodicity at which to collect the performance information. However, it does not specify, neither reflects, the monitoring frequency at which samples should be collected during the specified collection period. This is important to specify because two resources with same collection period may have different requirements on the granularity of the monitoring data, which will depend on the frequency at which samples are collected. Moreover, the sampling frequency for the same resource can be different for different collection period. Such a provisioning will enable to manage the monitoring load that the NFV-MANO system has to process. It should be noted that ETSI GS NFV-IFA 027 [i.8] and ETSI GS NFV-TST 008 [i.9] define the parameter *Tick Internal* for this purpose.

The location of the measurements can be distributed all over the components of the NS instance(s). To ensure the timely detection and reaction on latency violations, the measurements can be evaluated e.g. for performance prediction and analysis, by the local component that does the measurement and analysis or provided from elsewhere. What is important for a NFV-MANO system is that it should be able to utilize such information related to performance prediction and/or failures of low latency services supported by NS(s).

In addition to local actions/procedures, defining suitable thresholds that inform the individual components and/or the central NFV-MANO system early about potential latency violations that can occur in the near future should be taken into account. This would help in triggering actions even before a latency violation occurs. The source of such triggers might not be limited to NFV-MANO system itself but might be issued by external instances like an OSS/BSS system. It is therefore important for the NFV-MANO system to be able to react to external triggers in a timely manner.

A feature that could further improve the ability to react fast on the triggers is the prioritization of execution of actions of the NFV-MANO system. As an example instantiation of a new NS and/or LCM action on an active NS could wait in favour of restoring a low-latency service.

5.3 Delay points for latency measurements

This clause will highlight and analyse the possible delay incurring points in an infrastructure being managed by a NFV-MANO system. Figure 5.3-1 highlights the main delay points in a multi-site NFV infrastructure that was used as a reference for the use cases.

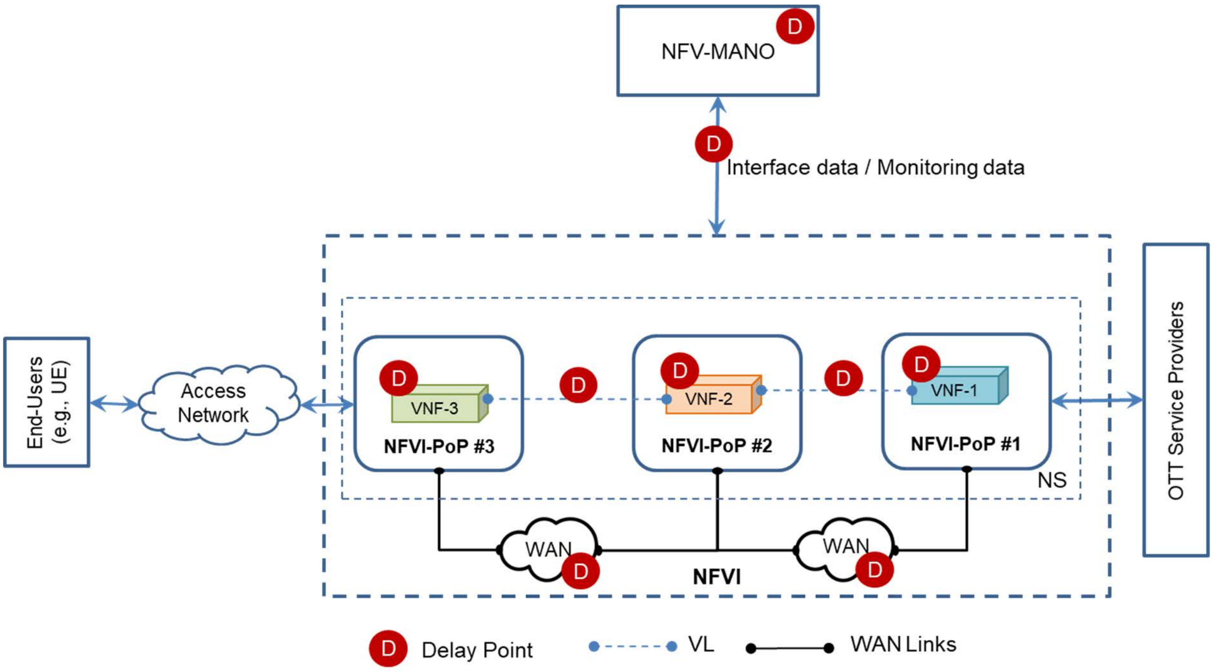

Figure 5.3-1: Possible Delay incurring points in the NFV-MANO managed NFV Infrastructure

Figure 5.3-1 gives a high-level overview of the delay points that, either alone or in combination with latencies experienced at other delay points, will have impact on the latency bounds of low latency service. As shown, there are different points within an NFV deployment with the potential of impacting delay bounds. These delay points exist in the NFV-MANO system itself, its interface(s) to the external elements, and also within the WAN infrastructure over which multiple NFVI-PoPs are connected.

Figure 5.3-2 gives a more detailed insight into the various delay points highlighted in Figure 5.3-1.

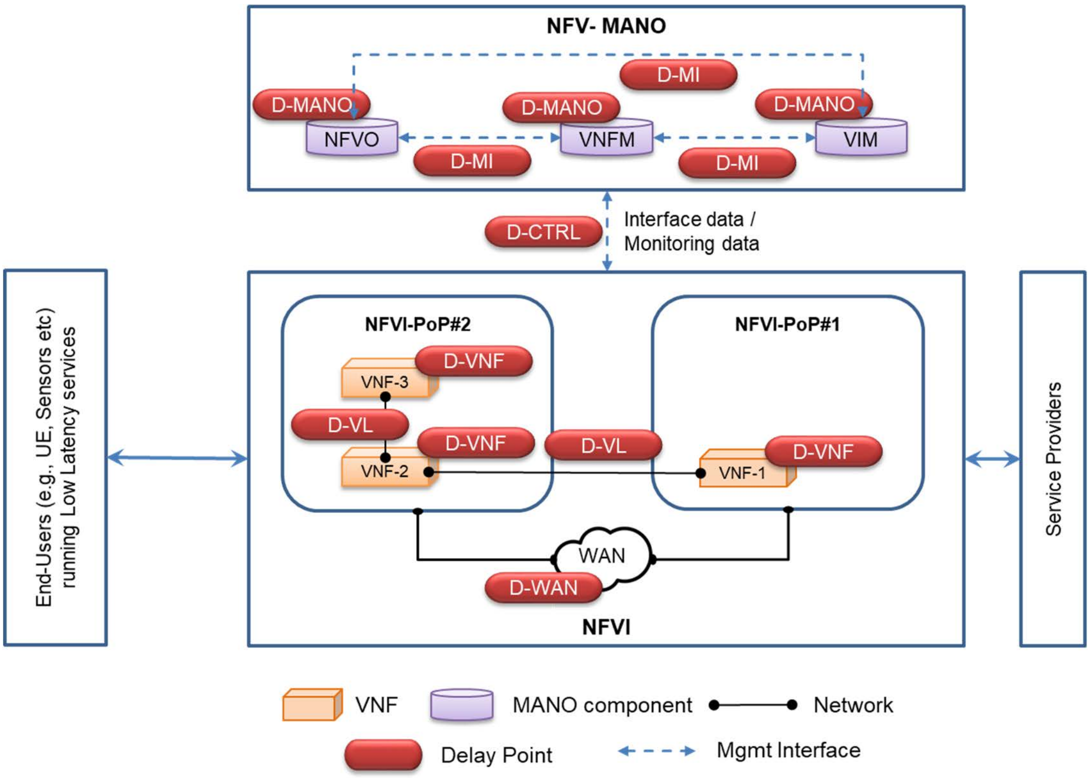

Figure 5.3-2: Possible Delay incurring points in the NFV-MANO managed NFV Infrastructure

With reference to Figure 5.3-2, the following delay points have been identified:

1) **NFV-MANO Interface Delay (D-MI):** This type of delay relates to the regular protocol delay experienced while transmitting protocol messages between the NFV-MANO functional blocks over the NFV-MANO interfaces. Increased protocol load over NFV-MANO interfaces will cause protocol delays, thereby delaying executing lifecycle management actions for service orchestration and resource orchestration.

2) **NFV-MANO Function Delay (D-MANO):** This type of delay relates to the processing delay of NFV-MANO messages by the respective NFV-MANO functional blocks, such as NFVO, VNFM, VIM and WIM. This introduces some unavoidable delay by every function executed by the functional block. In case of high load, this can affect the timely execution of lifecycle management actions for service orchestration and resource orchestration.

3) **NFV-MANO Control Delay (D-CTRL):** This is similar to the D-MI, however, D-CTRL denotes the delay experienced while transmitting messages between the NFV-MANO system and the managed objects. The managed objects are NFVI resources, the VNFs and the Network Service instances managed respectively by the VIM/WIM, VNFM and the NFVO.

4) **VNF Delay (D-VNF):** This type of delay characterizes the delay incurred within the VNF due to the processing workloads peculiar to the type of VNF itself. This delay is highly dependent on the VNF and its operational and functional capabilities. D-VNF cannot be managed by the NFV-MANO system itself, but the NFV-MANO system take the possible delay points into account when managing low latency services.

5) **Virtual Link Delay (D-VL):** This delay characterizes the delays incurring on virtual links interconnecting the VNFs. These delays do not account for the D-VNF but reflects the latencies due to issues in the underlying network resources.

6) **WAN Delay (D-WAN):** This is the delay occurring inside the WAN infrastructure and is pronounced in the multi-site scenarios where the NS instances are deployed over multiple NFVI-PoPS.

The delays within NS influence the operation of a low-latency service in one way or the other. These are delays introduced in the D-VNF, D-VL and D-WAN delay points. They characterize the delay the low-latency traffic experiences when processed by the NS. It is desirable that the NFV-MANO system knows these delays to be able to take them into account when setting up a low-latency NS.

The delays introduced in the D-MI, D-MANO and D-CNTRL delay points affect only the control traffic in the NFV-MANO system. These delays influence the time that the NFV-MANO system may require to react while dealing with service degradation or violation within a NS supporting low-latency service. It is desirable to keep these delays as low as possible or be able to prioritize the control traffic while restoring a degraded low-latency service.

5.4 NFV-MANO Deployment Considerations

5.4.1 Overview

This clause analyses the impact of the NFV-MANO system deployment on low latency network services. The analysis will be based on the NFV-MANO architectural options described in ETSI GR NFV-IFA 028 [i.2], with respect to the use cases in clause 4, and identify the influence of the architectural options on the delay points specified in clause 5.3 that may adversely impact the low-latency guarantees.

5.4.2 Summary of ETSI GR NFV-IFA 028

ETSI GR NFV-IFA 028 [i.2] reports on the potential architectural options to support the offering of NFV-MANO services across multiple administrative domains. In this regard, it has described a set of use cases based on which the interactions between NFV-MANO functional blocks belonging to different domains has been described.

According to ETSI GR NFV-IFA 028 [i.2], NFV-MANO services are offered and consumed by different organizations that can be different network operators or different departments within the same network operator domain. The functional requirements for the management of network services in a multiple administrative domains are specified in ETSI GS NFV-IFA 010 [i.4].

In the analysis, ETSI GR NFV-IFA 028 [i.2], differentiates between NFVIaaS Provider (NVFIaaS-P) and NFVIaaS Consumer (NVFIaaS-C), where the latter consumes the NFVI resources, such as compute, storage and networking, offered by the former. The relationship between NFVIaaS-P and NFVIaaS-C is depicted in Figure 5.4.2-1 (ETSI GR NFV-IFA 028 [i.2]).

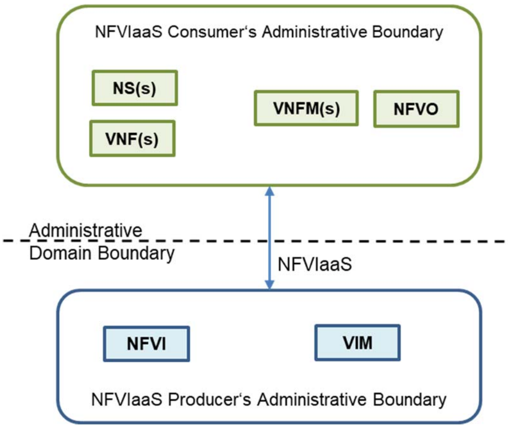

Figure 5.4.2-1: Relationship between NFVIaaS-P and NFVIaaS-C (ETSI GR NFV-IFA 028 [i.2])

As shown in Figure 5.4.2-1, the NFVIaaS-P has one or more VIM instances that are used to manage the NFVI resources within the NFVIaaS-P`s administrative domain. On the other hand the NFVIaaS-C administrative domain has the VNFM(s) and NFVO in order to manage the VNF(s) and NS(s) instance(s). The NFV-MANO components in the respective domains intercommunicate across the boundary between the two administrative domains.

Based on Figure 5.4.2-1, ETSI GR NFV-IFA 028 [i.2] specifies four architectural options based on the split in the responsibilities of the NFV-MANO management functions:

1) The NFVIaaS-C has access to Multiple Logical Points of Contacts (MLPOC) in the NFVIaaS-P's administrative domain, where the NFVIaaS-C has visibility to the NFVIaaS-P's VIMs and the VNF-related resource management is undertaken in *direct mode*. This is depicted in Figure 5.4.2-2.

   Note that NFVO interacts with the VIMs in a 1:N manner, whereas the VNFMs and VIMs interact in M:N manner.

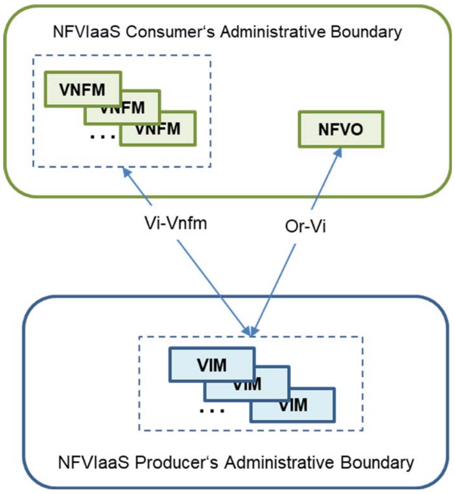

Figure 5.4.2-2: NFVIaas-C Management Functions interacting with NFVIaaS-P's VIM(s) via MLPOC in direct mode

2) The NFVIaaS-C has access to Multiple Logical Points of Contacts (MLPOC) in the NFVIaaS-P's administrative domain, where the NFVIaaS-C has visibility to the NFVIaaS-P's VIMs and the VNF-related resource management is undertaken in *indirect mode*. This is depicted in Figure 5.4.2-3.

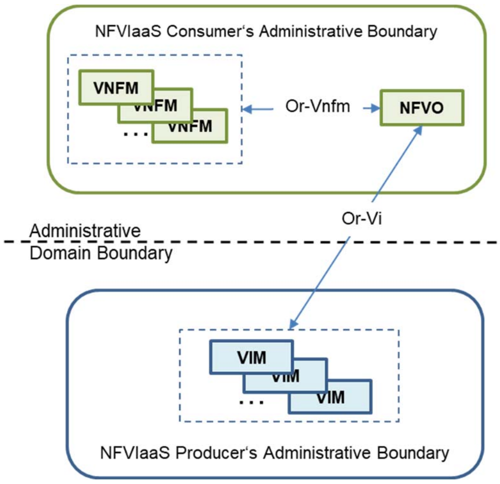

Figure 5.4.2-3: NFVIaas-C Management Functions interacting with NFVIaaS-P's VIM(s) via MLPOC in indirect mode

3) The NFVIaaS-C has access to Single Logical Points of Contacts (SLPOC) in the NFVIaaS-P's administrative domain, where the VIM(s) inside the NFVIaaS-P's domain are hidden from the NFVIaaS-C. The NFVIaaS-C interacts with the VIMs in *direct mode* for VNF-related resource management via unified interfaces exposed by SLPOC. This is depicted in Figure 5.4.2-4.

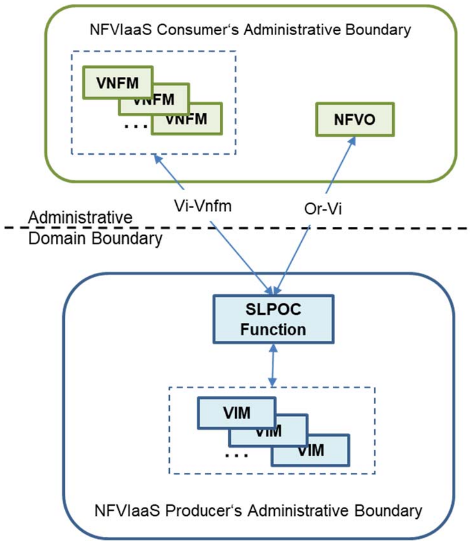

Figure 5.4.2-4: NFVIaas-C Management Functions interacting with NFVIaaS-P's VIM(s) via SLPOC in direct mode

4) The NFVIaaS-C has access to Single Logical Points of Contacts (SLPOC) in the NFVIaaS-P's administrative domain, where the VIM(s) inside the NFVIaaS-P's domain are hidden from the NFVIaaS-C. The NFVIaaS-C interacts with the VIMs in *indirect mode* for VNF-related resource management via unified interfaces exposed by SLPOC. This is depicted in Figure 5.4.2-5.

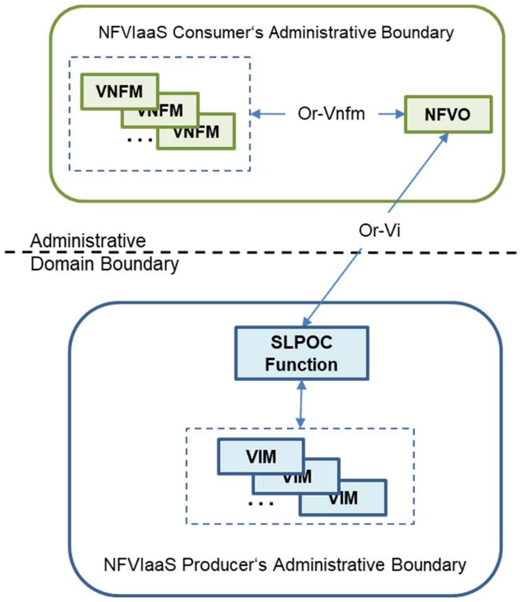

Figure 5.4.2-5: NFVIaas-C Management Functions interacting with NFVIaaS-P's VIM(s) via SLPOC in indirect mode

There are different possibilities on how the NFV-MANO functions are deployed with respect to each other, depending on how the respective administrative domains of both NFVIaaS-P and NFVIaaS-C are realized. Each deployment option will have an impact on the low-latency services, which are analysed in the proceeding clauses.

5.4.3 Intra-site NFV-MANO deployment considerations for low-latency services

In this option, the resources for both the NFVIaaS-P and NFVIaaS-C are co-located within the same NFVI-PoP site. In such a scenario, the NFVIaaS-P has better control over the design and configuration of internal resources, such as compute, network, storage, and thus control over the latency between the NFV-MANO components and the managed objects i.e. VNF and NS instance(s). For example, the NFVIaaS-P may allocate NFVI resources for the deployment of NS instance(s) in such a way so as to ensure minimum latency between the VNFCs. Moreover, the NFVIaaS-P can ensure that the proximity of the NFVIaaS-C's NFV-MANO components, such as the NFVO and VNFM(s) with respect to the VIM(s) are such so as to keep the *D-VL*, *D-MI* and *D-CTRL* to the minimum.

With respect to architectural option-1 (see Figure 5.4.2-2), D-MANO can be a possible prominent contributor of delays in case of high processing load. This can be the situation, where the NFVIaaS-C has to manage a large portfolio of VNF/NS instances resulting in higher monitoring/protocol load on the respective NFV-MANO components, including the VIMs, thereby accounting for higher D-MANO. Conversely, the VIMs provided by the NFVIaaS-P may also incur high D-MANO in case each VIM has to manage resources for multiple NFVIaaS-C services.

The above is also true for architectural option-2 (see Figure 5.4.2-3) in addition to the delay due to indirect mode of interaction between the NFVIaaS-C and the VIMs. Therefore, in addition to D-MANO such indirect mode of interaction will contribute towards D-MI as well.

The introduction of SLPOC in architectural option-3 (see Figure 5.4.2-4) and option-4 (see Figure 5.4.2-5) may also pose as a delay incurring entity. This is because the SLPOC, being a central entity through which NFVIaaS-C interacts with the VIM(s) can cause delays due to high NFV-MANO protocol load and thus the processing load. The SLPOC entity also results in one more additional processing entity within the NFV-MANO system. These delays will thus contribute towards D-MI and D-CTRL. The indirect mode of interaction in option-3 will further add to D-MI. Most crucial to the service integrity of low-latency services is that the SLPOC entity serves as a Single Point of Failure (SPoF) and in the event of any failure will result in the discontinuity and/or extended delays of NFV-MANO operations.

Thus for low latency NFV-MANO services, it is preferred for the NFVO and the VNFMs of the NFVIaaS-Cs to interact with the NFVIaaS-P's VIMs via MLPOC in direct mode.

5.4.4 Inter-site NFV-MANO deployment considerations for low-latency services

The delay incurring considerations highlighted for intra-site NFV-MANO deployment holds true for inter-site NFV-MANO deployments, where the NFVIaaS-C and NFVIaaS-P resources are located across multiple NFVI-PoP sites and interconnected via WAN infrastructure. The interfaces and their respective operations for the management of multi-site connectivity services over WAN has been specified in ETSI GS NFV-IFA 032 [i.3]. In terms of impact of low-latency services, multi-site deployments are expected to have the major influence on latency considerations, and thus the reason that the use cases in clause 4 have been described in the context of multi-site.

There can be several permutations of how the NSs and the NFV-MANO functions are deployed in a multi-site multi-operator environment depending on the ownership of the NS instances and the NFVI-PoPs. There are four main deployment options described and analysed below. In all these options it is assumed that the NFVIaaS-C and NFVIaaS-P are in different administrative domains and that the NS instances are deployed across multiple NFVI-PoPs characterizing the NFVIaaS-P administrative domains.

Option-1: The NFVIaaS-P's domain is characterized by multiple NFVI-PoPs whereas the NFVIaaS-C's domain is characterized by a different NFVI-PoP, where the latter is owned by the same or different NFVIaaS-P. This is shown in Figure 5.4.4-1, where the NFV-MANO functions of NFVO and VNFM(s) owned by the NFVIaaS-C is located in NFVI-PoP #1, which interacts with the NS instance(s) deployed across NFVI-PoP #2 and NFVI-PoP #3 via the VIM(s) located in them. The three NFVI-PoPs are interconnected over the WAN infrastructure.

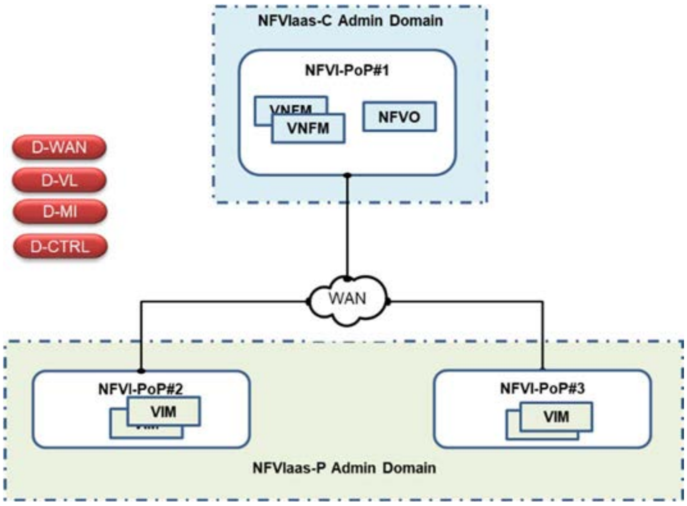

Figure 5.4.4-1: NFV-MANO deployment in multi-site environment with the NFVIaaS-C's NFV-MANO functions located in the same NFVI-PoP

In addition to intra-site latencies analysed in clause 5.4.3, the delays within the WAN infrastructure (D-WAN) becomes a prominent factor with a direct impact on D-MI and D-CTRL. The D-MI will remain high regardless of whether the NFVIaaS-C interacts with the VIMs via MLPOC or SLPOC as all NFV-MANO function interactions will go over the WAN infrastructure.

Option-2: This option, depicted in Figure 5.4.4-2, is similar to Option-1 with the exception that the VNFM(s) owned by the NFVIaaS-C is co-located with the NFVIaaS-P's VIMs within NFVI-PoP #2 and NFVI-PoP #3, while the NFVO is in NFVI-PoP #1.

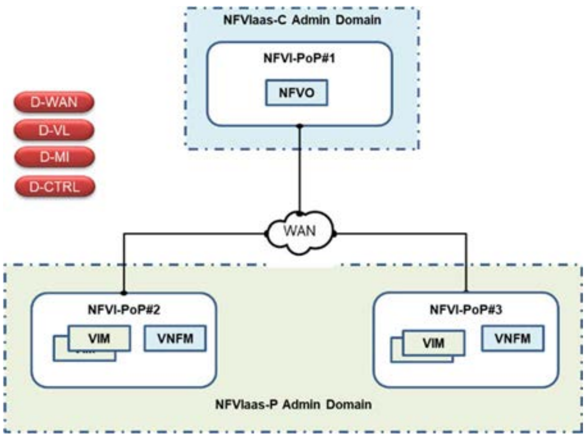

Figure 5.4.4-2: NFV-MANO deployment in multi-site environment with the NFVIaaS-C's NFV-MANO functions distributed in different NFVI-PoPs

Similar to Option-1, the D-WAN will influence both D-MI and D-CTRL thereby incurring delays on the management of low-latency services. However, to reduce D-MI it is recommended for the NFVIaaS-C to adopt direct mode of interaction as described in clause 5.4.2.

Option-3: In this deployment option, depicted in Figure 5.4.4-3, the NFVIaaS-C is managing the NS that is deployed across multiple NFVI-PoPs belonging to different NFVIaaS-Ps. Whether the NFVIaaS-C's management functions (i.e. NFVO and VNFM(s)) are deployed in same or different NFVI-PoPs, the WAN infrastructure delay (D-WAN) will have influence on all other delay points D-VL, D-MI and D-CTRL. In such highly federated scenario, it is important for the different NFVIaaS-Ps to negotiate trust agreements prior to deployment of low-latency NS instance(s). Once the trust relationships have been established, then the magnitude of delays will be the same as in Option-1 or Option-2 depending on how the NFVIaaS-C's management functions are deployed.

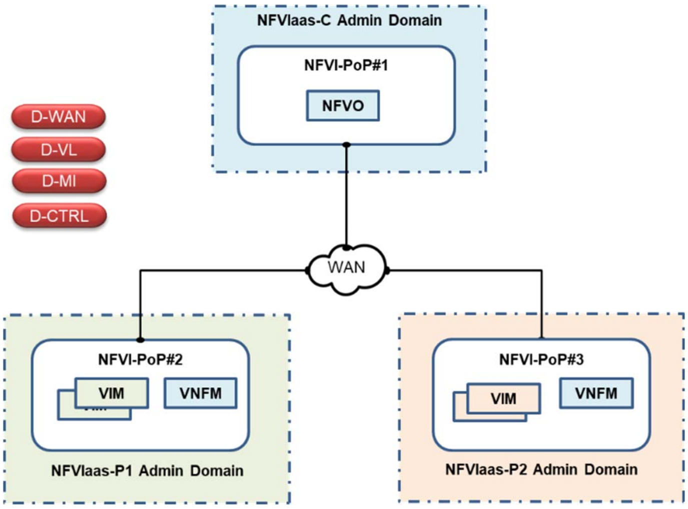

Figure 5.4.4-3: NFV-MANO deployment in multi-site environment with the NFVIaaS-C managing NS deployed across different NFVIaaS-P's administrative domains

Option-4: This deployment option, depicted in Figure 5.4.4-4, extends the scenario depicted in Figure 5.4.4-3 where the composite NS, described in ETSI GR NFV-IFA 028 [i.2], is composed by nested NS instances where each nested NS instance is owned and managed by a different NFVIaaS-C administrative domains deployed in different NFVI-PoPs. With reference to Figure 5.4.4-4, the NFVIaaS-C-1 and NFVIaaS-C-2 administrative domains are located in NFVI-PoP#1 and NFVI-PoP#4 respectively. For example, with reference to the use cases described in clause 4, VNF-1 can be one instance of a nested NS that is managed by while VNF-2 and VNF-3 compose the other instance of a nested NS instance.

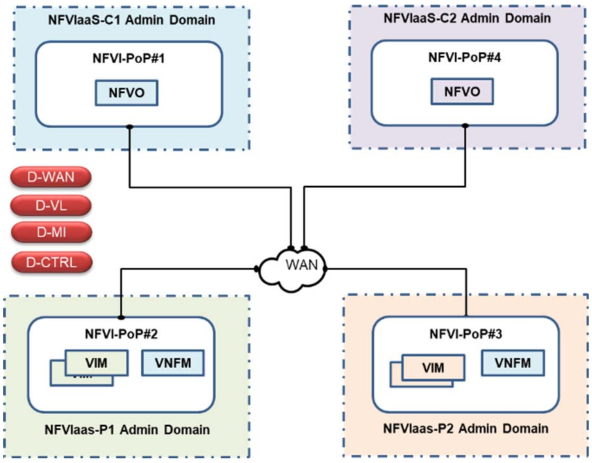

Figure 5.4.4-4: NFV-MANO deployment in multi-site environment with different NFVIaaS-Cs managing a composite NS deployed across different NFVIaaS-P's administrative domains

Similar to the previous deployment options, in this scenario D-WAN will influence D-VL, D-MI and D-CTRL. However, due to the added interactions and coordination required between the different NFVIaaS-C's administrative domains, especially over the Or-Or reference point, the D-MI is expected to be more pronounced than on other previous options.

6 Recommendations and Conclusions

6.1 General Recommendations

Table 6.1-1 provides a set of general recommendations that have been derived from the use cases.

Table 6.1-1: General Recommendations

S.No.	Recommendations
1	It is recommended to ensure that the NFV-MANO system is aware of the monitoring points on a NS instance supporting low-latency service (e.g. see ETSI GS NFV-IFA 027 [i.8]/ETSI GS NFV-TST 008 [i.9]).
2	It is recommended to ensure that NFV-MANO system monitoring the NS instances supporting low-latency services can recognize performance related failure events while they occur.
3	It is recommended to ensure that the NFV-MANO system is able to utilize the information related to prediction of performance of low-latency services supported by NS instances under NFV-MANO control.
4	It is recommended to ensure that the NFV-MANO system is able to utilize the information related to failures of low-latency services supported by NS instances under NFV-MANO control.
5	It is recommended to ensure that the NFV-MANO system is able to utilize the information related to remedy actions on NS instances under NFV-MANO control.
6	It is recommended to consider that NFV-MANO remedy actions may include prioritization.

6.2 Conclusion

The result of the study shows that supporting low latency services with NFV-MANO does not require major changes to the NFV-MANO system. The recommendations show that the most important factor to support low latency services is the monitoring of potential delay points within NS(s). Therefore, the above stated recommendations should be analysed in view of the impact on IFA specifications. For example, information elements related to monitoring, such as *MonitoringParameter* information element specified in ETSI GS NFV-IFA 011 [i.5] and ETSI GS NFV-IFA 014 [i.6] may be extended with additional attributes providing more control over configuring monitoring parameters as discussed in clause 5.2. This can be aligned with the definition of the parameter *TickInterval* defined in ETSI GS NFV-IFA 027 [i.8]. Furthermore, the respective specifications may be extended with specifying monitoring points on VNFC and VL instance(s) constituting a NS instance supporting low-latency service with impact on delays. Moreover, the *NsQoS* information element in ETSI GS NFV-IFA 014 [i.6] may be extended with additional attributes specifying thresholds on relevant QoS attributes for NS VL supporting low latency services. Similarly the *QoS* information element in ETSI GS NFV-IFA 011 [i.5] may be extended with additional attributes specifying thresholds on QoS attributes relevant to the VL that is part of a NS supporting low-latency services. Inclusion of an additional attribute may also be considered in ETSI GS NFV-IFA 014 [i.6] and ETSI GS NFV-IFA 013 [i.12] that may capture the end-to-end latency threshold for a NS.

History

Document history		
V1.1.1	August 2020	Publication