future internet

MDPI

Review

Intelligent and Autonomous Management in Cloud-Native Future Networks—A Survey on Related Standards from an Architectural Perspective

Qiang Duan

Information Sciences & Technology Department, Pennsylvania State University, Abington, PA 19001, USA; qduan@psu.edu

Abstract: Cloud-native network design, which leverages network virtualization and softwarization together with the service-oriented architectural principle, is transforming communication networks to a versatile platform for converged network-cloud/edge service provisioning. Intelligent and autonomous management is one of the most challenging issues in cloud-native future networks, and a wide range of machine learning (ML)-based technologies have been proposed for addressing different aspects of the management challenge. It becomes critical that the various management technologies are applied on the foundation of a consistent architectural framework with a holistic vision. This calls for standardization of new management architecture that supports seamless the integration of diverse ML-based technologies in cloud-native future networks. The goal of this paper is to provide a big picture of the recent developments of architectural frameworks for intelligent and autonomous management for future networks. The paper surveys the latest progress in the standardization of network management architectures including works by 3GPP, ETSI, and ITU-T and analyzes how cloud-native network design may facilitate the architecture development for addressing management challenges. Open issues related to intelligent and autonomous management in cloud-native future networks are also discussed in this paper to identify some possible directions for future research and development.

Keywords: network and service management; intelligent and autonomous management; cloud-native network design; machine learning

Citation: Duan, Q. Intelligent and Autonomous Management in Cloud-Native Future Networks—A Survey on Related Standards from an Architectural Perspective. *Future Internet* **2021**, *13*, 42. https://doi.org/10.3390/fi13020042

Academic Editors: Nane Kratzke and Choong Seon Hong
Received: 28 December 2020
Accepted: 1 February 2021
Published: 5 February 2021

1. Introduction

The rapid development of Internet technologies in the past two decades has enabled a broad spectrum of network applications with highly diverse service requirements and a wide variety of technologies for building heterogeneous network infrastructures. However, the ossification of the IP-based network architecture limits the Internet's flexibility and agility to face the challenges introduced by the diversity in both network applications and infrastructures [1].

Network virtualization has been proposed as a key attribute of future networks for overcoming the limitation of the IP-based network architecture. The key idea of network virtualization is to decouple the network functions for service-provisioning from the network/compute capabilities for data transportation, processing, and storage, which allows alternative network architectures and protocols to coexist upon shared underlying infrastructures [2]. The Network Function Virtualization (NFV) paradigm follows the network virtualization principle to realize network functions as software instances that can be deployed upon commodity servers and storage devices, thus enabling virtual networks (network slices) customized for multiple tenants [3].

A Software-Defined Network (SDN) is another innovative networking technology that may address the ossification of IP-based networks. The key idea of the SDN lies in decoupling between the data plane and control plane to enable a logically centralized

network controller with a global view of the entire network domain [4]. The SDN introduces a network operating system that provides an abstraction of data plane functionalities and resources upon which control applications may program network operations.

NFV and the SDN have rapidly become active research areas in networking that attract extensive interest from both academia and industry. Significant progress has been made in technology developments for enabling these two new paradigms in various networking scenarios, including data center networks, wide-area networks, and wireless mobile networks [5,6]. NFV and the SDN together form the foundation of future networks including the 5G and beyond networks. On the other hand, the current NFV and SDN implementations inherit some monolithic design patterns from the traditional network architecture that may constraint network and service flexibility and agility [7].

As the next step in the evolution of networking technologies toward fully exploiting the advantages of network virtualization and softwarization, cloud-native network design applies the service-oriented architectural principle together with virtualization and softwarization in networking, which enables network systems to be realized based on cloud technologies and network services to be provisioned following the cloud service model [8]. The cloud-native design principle has been embraced by all major network Standards Development Organizations (SDOs) in their latest work, for example ETSI NFV Release-4 and the 3GPP 5G core network architecture. Therefore, cloud-native design is expected to be a key attribute of future networks [9].

Cloud-native network design facilitates a significant transformation of communication networks from an infrastructure for data transportation to a versatile service platform for various industry verticals and customer segments, which plays a crucial role in supporting both cloud computing and the emerging edge computing paradigm. The same set of key technologies—virtualization and service-oriented architecture—now are widely applied in both cloud/edge computing and networking, thus allowing these two fields to converge together [10]. Therefore, network-cloud/edge convergence with a holistic vision of end-to-end service provisioning across the networking and computing domains is expected to be a key attribute of the future cloud-native networks.

Flexible and effective management is one of the most critical challenges to cloud-native future networks. Given the large scale, heterogeneity, and complexity of the converged communication-compute-storage systems in future networks, management solutions must be highly automated and intelligent. The integration of dynamic intelligent network telemetry and closed-loop control mechanisms leveraging machine learning (ML) and data analytics techniques offers promising new management technologies. A wide spectrum of technologies, often developed based on various ML methods, has been proposed to enable intelligent and autonomous management for various aspects of future networks [11]. However, the heterogeneous nature of ML techniques and the unique characteristics of future networking technologies impose a variety of requirements for integration between these areas. The current disparate management mechanisms for ML functionalities and network functions may degrade the effectiveness of intelligent network operations.

Therefore, it becomes vital that a wide range of intelligent management technologies is applied based on the foundation of a consistent architectural framework with a common holistic vision. This calls for standardization of new management architectures that support seamless integration of various ML/data analytics techniques for addressing the management challenges in a holistic and interoperable way. Currently, various relevant network SDOs, including 3GPP, ETSI, and ITU-T, are actively working on this important subject and making encouraging progress.

The goal of this paper is to provide a big picture that reflects the latest developments of architectural frameworks for intelligent and autonomous management in cloud-native future networks. The paper first gives an overview of the technical trend toward cloud-native network design and network-cloud/edge convergence in future networks. Then, the paper surveys the latest representative progress in the standardization of the management architecture for future networks, including works by 3GPP, ETSI, and ITU-T, and analyzes

how cloud-native network design may facilitate the architecture development for addressing the management challenges in future networks. Open issues related to intelligent and autonomous management from the architectural perspective are also discussed in the paper to identify some directions for future research and development.

The remainder of the paper is organized as follows. Section 2 introduces the incentive for cloud-native design and discusses its impacts on future networking. Then, the state-of-the-art developments of the management architecture, including 3GPP Network and Management Data Analytics Functions/Services (NWDAF/MDAS), the ETSI Experiential Networked Intelligence (ENI) architecture, the ITU-T Architecture for ML in Future Networks, and the ETSI Zero Touch Management (ZSM) Framework, are reviewed and analyzed in Sections 3–6, respectively. Open issues and possible directions for future research are discussed in Section 7. Section 8 draws the conclusions.

2. Cloud-Native Network Architecture in Future Networks

2.1. Network Function Virtualization and Software-Defined Networking

ETSI has been a leading force for realizing network virtualization since it defined the NFV architecture in 2012. The ETSI NFV Industry Specification Group (ISG) community has evolved through several phases from defining the initial architectural framework to developing detailed standard specifications. NFV enables network functions to be realized as software instances and hosted by virtual machines or containers running on commodity network/compute equipment such as standard servers and switches. The ETSI NFV architecture comprises a set of Virtual Network Functions (VNFs) deployed upon a virtualized infrastructure. The Management and Orchestration (MANO) system of the architecture contains a Virtual Infrastructure Manager (VIM) for unified management of network-compute resources in the virtualized infrastructure, a VNF Manager (VNFM) for managing VNFs, and an NFV Orchestrator (NFVO) for life-cycle management of network services [12].

Recent developments of the NFV specifications in Releases 3 and 4 have focused on enriching the NFV architecture for global deployment and operations. The set of new features in NFV mainly includes support for the latest network technologies such as edge computing and network slicing, new operational aspects such as multiple administrative domains and policy frameworks, advances in acceleration technologies and lightweight (e.g., container-based) virtualization platforms, and enhancement of NFV automation capabilities for improving life-cycle management and orchestration of VNFs and network services.

Multiple network SDOs, including ONF, ITU-T, and IETF, have developed their own versions of SDN specifications, but they all share the same basic architecture—an architecture that comprises three planes (the data plane, control plane, and application plane) and two standard interfaces (the southbound interface between the data and control planes and the northbound interface between the control and application planes). The key principles behind both SDN and NFV are decoupling with abstraction, but focus respectively on the layer dimension (for NFV) and plane dimension (for SDN) [13]. Therefore, NFV and the SDN are expected to be integrated in future networks in order to fully exploit the advantages of both paradigms. For example, multiple SDN controller instances may be realized upon a network hypervisor in order to control different virtual networks (network slices) implemented by NFV [14].

2.2. Cloud-Native Network Design

The current specifications for realizing network virtualization and softwarization, including the NFV and SDN architectures, mainly change how network functions are realized and deployed (as software instances hosted on VMs and/or containers), but do not change much how network functions are designed. The state-of-the-art NFV implementations often replace monolithic hardware-based network functions with their monolithic software VNF counterparts. Such a monolithic VNF design may introduce a

large number of common functionalities repeated across different VNFs, which causes some negative consequences that lead to sub-optimal resource usage and hinder network agility. In addition, the NFV and SDN architectures both comprise a set of predefined function blocks that are interconnected via point-to-point interfaces (reference points), which require standardization of a new set of reference points whenever a new function block is added into the architecture. Such monolithic and tight-coupling architectural features still introduce ossification, which limits the flexibility and agility of networking systems for service provisioning [7].

A promising strategy for addressing the ossification issue associated with the current NFV and SDN architecture is to enable finer granularity for network functions in the architecture and a common interface for loose-coupling interaction among network functions. The Service-Oriented Architecture (SOA) with its latest development as the Micro-Service Architecture (MSA) offers an effective approach to achieving this objective.

In general, the SOA principle advocates decomposing a large system into a collection of smaller units called services, which essentially are self-contained and platform-independent system modules that can be described, published, accessed, composed, and programmed through a standard interface and messaging protocol [15]. The MSA can be viewed as the second iteration of the SOA that aims to strip away unnecessary levels of complexity in service design in order to focus on the programming of simple services that effectively implement a single functionality [16]. Service instances in the SOA typically run on VMs, and the lightweight container-based virtualization offers an efficient platform for hosting microservices. The SOA and virtualization have been two key pillars of the foundation for cloud technologies [17], and the MSA is an enabler of the latest developments in cloud computing, including the emerging edge computing paradigm.

The service-oriented architectural principle has also been adopted in network design, which enables a Network-as-a-Service (NaaS) paradigm that abstracts various network functions and resources as services [18]. It is worth noting that the term "service" has a different meaning in the networking field than the "service" concept in the SOA/MSA context. In the networking context, a service typically refers to the data transport capability offered to a customer by an ordered set of network functions; while the service concept in the SOA/MSA emphasizes encapsulation of a system module into a self-contained component. On the other hand, the evolution of network design has been influenced by the SOA principle in the past decade [19], and the SOA service concept has been gradually adopted in the recent development of the network architecture with the NaaS paradigm [20].

In the ETSI NFV specifications, while a network service refers to an ordered set of (virtual) network functions specified by a service description (VNF forwarding graph), the SOA principle has been embraced by the NFV architecture at multiple levels. For example, NFV supports NFV Infrastructure-as-a-Service (NFVIaaS), Virtual Network Function-as-a-Service (VNFaaS), and Network Slice-as-a-Service (NSaaS), which all adopt the SOA service concept. In the 3GPP specifications for the 5G service-based architecture, usage of the term "service" becomes more aligned with the SOA service concept than its conventional definition in the networking field. Currently, the SOA-based NaaS paradigm encapsulates monolithic network functions into individual service components with coarse granularity. Adoption of the emerging MSA in network design allows the decomposition of monolithic network functions to fine-grained modules that can be flexibly orchestrated for achieving different functionalities for service provisioning.

The application of the virtualization and service-oriented principles in network design enables network systems to be realized based on cloud technologies and network services to be provisioned following the cloud service model. This emerging trend in networking technologies' development is often referred to as cloud-native network design, which is expected to be widely adopted in future networks including the design of 5G/6G networks. An important aspect of the recent development in the NFV specification is to support

cloud-native network design including the adoption of container-based virtualization technologies for deploying VNF components as microservices [21].

2.3. Service-Based Architecture for the 5G Network

The Service-Based Architecture (SBA) developed by 3GPP for the 5G core network is a representative case of cloud-native design in network architecture. The 5G SBA is based on network virtualization and softwarization: Network Functions (NFs) on the data plane and control plane are separated, and the NFs defined in the architecture may be realized as software instances deployed on a virtual infrastructure. On top of virtualization and softwarization, the 5G SBA adopts the service-oriented principle through the notion of NF services exposed by control plane NFs and the Service-Based Interface (SBI) for interactions among NFS services.

As depicted in Figure 1, the 5G SBA decouples the data plane and the control plane. The data plane comprises User Equipment (UE), the Radio Access Network (RAN), the User Plane Function (UPF), and the Data Network (DN). The interfaces between the data plane functions currently are still defined as reference points (N1–N6 in Figure 1). The 5G control plane comprises a set of NFs including the Access and Mobility Management Function (AMF), the Authentication Server Function (AUSF), the Session Management Function (SMF), the Policy Control Function (PCF), the Network Slice Selection Function (NSSF), Unified Data Management (UDM), the Unified Data Repository (UDR), the Network Repository Function (NRF), the Network Exposure Function (NEF), and the Application Function (AF) [22].

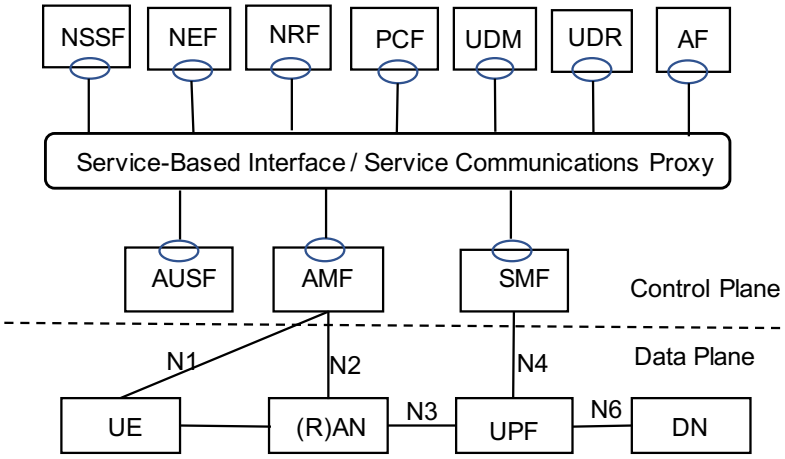

Figure 1. The Service-Based Architecture (SBA) for the 5G network.

Each control plane NF may expose one or multiple NF services as a service provider and also access the services exposed by other NFs as a service consumer. All the NF services are accessed through the SBI. With the SBI, the 5G SBA shifts from a point-to-point reference point-based interface toward web-based communications among NF services. The SBI may be realized by means of REST API calls using HTTP/2 on top of TLS/TCP as the transport protocol. The emerging transport protocol Quick UDP Internet Connections (QUIC) is paving the way toward HTTP/3 for supporting SBI communications [23].

The 5G SBA allows each service consumer to discover a suitable producer of a service instance. Service discovery in the SBA is supported by the NRF, which keeps a repository of all available NF instances and the services they expose. Each NF instance is required to register its profile at the NRF, and the NF profile contains relevant data about the NF including the services it provides and the binding information for accessing the services. The NEF may expose NF services provided inside the network to authorized external consumers such as AFs. The 3GPP introduced the service framework support function, referred to as Service Communication Proxy (SCP) in Release 16, which aims to extract the common NF

processes related to service registration, discovery, selection, and communication binding into a unified platform [24,25].

The 5G SBA supports two modes of interaction between the NF service producer and consumer. In the request-response model, an NF service consumer sends a request message to the target service provided by a producer and receives the corresponding response message back from the producer. In the subscribe-notify model, an NF service consumer subscribes to an NF service for a (set of) certain event(s) and receives a notification message from the service producer whenever a subscribed event occurs.

Although the current 3GPP specification limits the SBA to the control plane of the 5G core network, the cloud-native design principle is expected to be applied in a broader scope in the future 6G network. A natural evolution beyond the current 5G SBA would be an integration of the SBA framework for the core network into an end-to-end architecture that comprises multiple planes, including the data, control, and application planes as defined in the SDN architecture, and across multiple domains including the access, transport, and core networks.

2.4. Edge Computing

The emerging edge computing paradigm essentially deploys decentralized cloud computing capabilities in the network infrastructures typically at the network edge [26]. The Multi-access Edge Computing (MEC) architecture developed by ETSI is a representative architectural framework for edge computing. The MEC architecture comprises two levels: the lower level consists of a set of individual MEC hosts, and the higher level forms an MEC system comprising the MEC hosts and the network connections among the hosts. Each MEC host contains an MEC platform built on top of the virtualized network-compute infrastructure, and various MEC applications are deployed on the platform. The host-level management is responsible for managing the infrastructure, platform, and MEC applications at each host. The core function of system-level management is an MEC orchestrator that maintains a global view of the entire MEC system for coordinating host management for end-to-end service provisioning [27].

The true impact of the edge computing paradigm relies on its interaction with networking. The foundation of the MEC architecture is a virtualized network infrastructure upon which the MEC platform and applications are deployed. Most network operators want to consolidate VNFs and MEC applications on top of a shared infrastructure to the maximum possible degree for enhancing resource utilization and improving service performance. Toward this direction, ETSI MEC ISG has developed a reference architecture for deploying MEC in an NFV environment, which specifies how MEC entities can be integrated in the NFV architecture by fully leveraging NFV MANO capabilities [28]. In this MEC-in-NFV reference architecture, both the MEC platform and MEC applications can be treated by NFV MANO in the same way as VNFs being managed, which allows NFV MANO to be leveraged for unified management and orchestration of both VNFs and MEC applications.

2.5. Convergence of Networking and Cloud/Edge Computing

The cloud-native network design adopts the virtualization and service-oriented architectural principles in networking essentially in the same way as the principles being applied in cloud/edge computing. Therefore, cloud-native network design enables network systems to be realized based on cloud technologies and network services to be provisioning in the cloud service model, which facilitates convergence between networking and cloud/edge computing. Such convergence calls for a holistic vision across the networking and computing domains that supports integrated management of networking-computing functionalities and unified provisioning of network and cloud/edge services [10].

The NFV architecture provides a unified virtualization layer for network-compute infrastructures and common management of virtual network/compute functions, thus forming the basis for network-cloud/edge convergence as reflected in the MEC-in-NFV reference architecture. Through its embrace of cloud-native design in the SBA, the 5G

network introduces the very same characteristics about network function interactions, system operations, and service provisioning shared with cloud and edge computing [29]. Therefore, network-cloud/edge convergence with integrated network-compute infrastructures for end-to-end provisioning of composite network-cloud/edge services is expected to be a key feature of future networks.

The holistic architecture design, resource management, system operations, and service provisioning enabled by network-cloud/edge convergence may significantly improve resource utilization and service performance, lower capital/operational costs, and introduce opportunities for technical and business innovations. On the other hand, all these benefits can be realized only if the converged network-cloud/edge systems in future networks are properly managed. Standardization of the management architecture for cloud-native future networks plays a significant role in order to fully leverage various cognition techniques such as AI/ML to manage heterogeneous systems and integrate diverse management domains for end-to-end service provisioning. The following sections of this paper will review the representative standardization work on management architecture that addresses the challenges of intelligent and autonomous management in future networks.

3. Network and Management Data Analytics Functions in the 5G Network Architecture

The 3GPP defines the Network Data Analytics Function (NWDAF) in the 5G SBA as a core network function for supporting intelligent and autonomous network operations and service management [22]. NWDAF in the 5G architecture is a general system module through which various data-driven AI/ML-based analytics technologies may be integrated in 5G networks. A general framework for NWDAF-based 5G network automation developed by 3GPP [30] is depicted in Figure 2. As shown in this figure, the NWDAF collects data from various modules in the 5G system, including NFs, AFs, and data repositories (UDR), for performing the analytics. The obtained analytics information is then delivered by the NWDAF to the NFs/AFs that requested the analytics, which then leverage such information for making decisions about network operations and management actions. The NWDAF also works closely with the network Operation, Administration, and Management (OAM) system within this framework for collecting data from the OAM and exposing analytics information to the OAM.

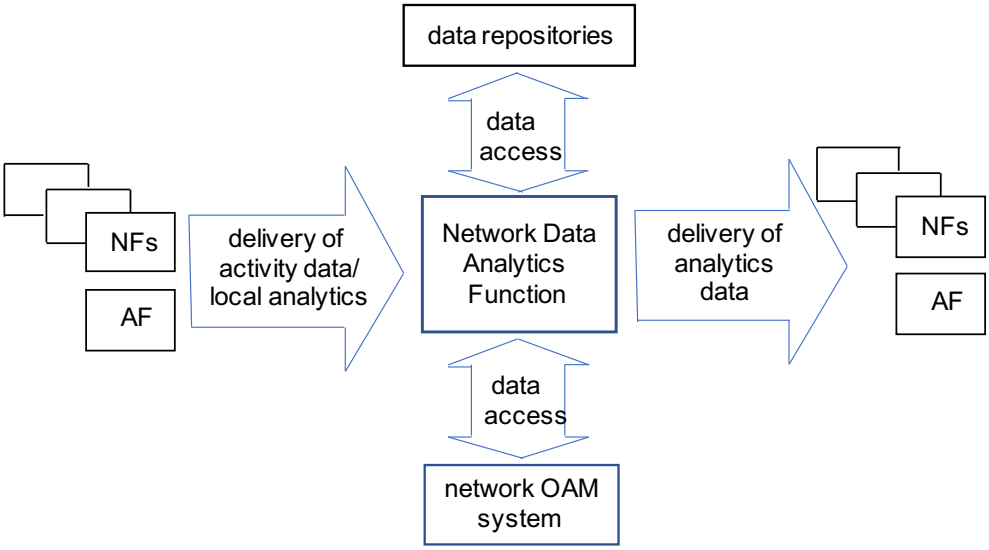

Figure 2. A Network Data Analytics Function (NWDAF)-based framework for 5G network automatic management.

The NWDAF-centric network automation framework is embedded in the 5G SBA. The NWDAF provides a set of analytics services that can be discovered, selected, and accessed by NFs/AFs and consumes the services provided by other system modules (NFs,

UDR, AFs, and OAM) for data collection. The NWDAF communicates with the rest of the framework via the SBI/SCP as defined in the SBA. Cloud-native design of the 5G SBA in principle allows any analytics function to be realized at NWDAF as long as its interactions with other NFs follow the SBI standard and allows NWDAF to leverage various AI/ML technologies for realizing its analytics functions without being constrained by other system modules in the framework [31].

The analytics information generated by NWDAF includes two main categories: statistical information and prediction information. Examples of statistical analytics information provided by the NWDAF include traffic load and resource utilization of network slice instances, network/service performance measurement, and user mobility pattern analysis. Examples of prediction information that can be obtained at the NWDAF include the temporal and spatial traffic distribution predicted for a future time period and user device location prediction for a future time instant.

The 5G SBA supports flexible deployment of the NWDAF as a single or multiple VNF instances in the same network domain. Multiple NWDAF instances can be deployed as either a central NF, or a collection of distributed NFs, or a combination of both. When multiple NWDAF instances exist, not all of them need to be able to provide the same type of analytics. Some instances can be specialized in providing certain types of analytics. An analytics ID information element is used to identify the type of analytics that an NWDAF instance can generate.

The 5G SBA enables consumers of analytics functions to discover and select an NWDAF instance that is able to provide a specific type of analytics via the NRF. After each NWDAF instance is instantiated, it should register itself at the NRF and provide the list of analytics IDs that it supports as part of the NWDAF profile that will be stored at the NRF. Other NFs may query the NRF for NWDAF discovery and include the analytics ID(s) for the required types of analytics in the query message. Then, the NRF will search the stored NWDAF profiles to discover and select an appropriate NWDAF instance for each query.

After discovering an NWDAF instance, the customer NF may employ the SBI request/response messaging protocol to obtain one time analytical information per request. The NF may also subscribe a certain type of analytic service offered by the NWDAF instance, then it will regularly receive notification messages from the NWDAF instance that deliver analytical information generated from the service. In addition to NFs, AFs and OAM functions may also be NWDAF customers and discover NWDAF instances. For an AF that is not registered in the same trust domain as the NWDAF, the NEF will be used as a gateway for all the interactions between the AF and the corresponding NWDAF instances.

In order to perform the required analytics functions, the NWDAF may need to collect data from various sources, including NFs (e.g., AMF, SMF, PCF, and UDR), AFs, and OAM functions, as a basis of the computation of network analytics. Typically, data collection is from NFs when the target data are related to individual User Equipment (UE) and network sessions. Data related to global network states, for example performance measurements of network slices and services/applications, are often obtained from OAM. Data collection may be reactive triggered by a request from a customer NF/AF/OAM function for a certain type of analytics or proactive for retrieving data needed for long-run analytics. NWDAF data collection is realized via the SBI using either the request/response mode for one-time data retrieval or the subscription/notification mode for regular data retrieval for a time period.

In order to obtain the proper data needed for the requested analytics, the NWDAF is configured with the corresponding NF/AF types and/or OAM measurement types for each analytic ID it supports; therefore, it is able to send the request or subscription message to the right data sources for each type of analytics function. The 5G SBA allows the NWDAF to discover the appropriate sources for data collection via the NRF using the data availability information provided to the NRF by NFs/AFs during their registration processes.

Data collection may cause additional communication overheads between the NWDAF and other system modules that may limit system scalability and degrade network performance. Therefore, it is important to achieve a balance between the completeness/preciseness of the collected data and the overheads introduced by data collection. The NWDAF should be able to efficiently obtain the required data with an appropriate level of granularity. Toward this objective, the following aspects of data collection are specified in the data retrieval request and subscription messages: granularity (individual event, a list of events, key performance indicators), temporality (periodical, on demand by the NWDAF, triggered by specific criteria, with validity period), and individuality (aggregated data vs. single scope data).

The NWDAF is defined as a control plane function in the 5G SBA architecture. In addition, 3GPP also developed a Service-Based Management Architecture (SBMA) and framework for 5G network management and orchestration following the service-oriented principle [32]. The fundamental building block of the SBMA is the Management Services (MnSs). An MnS is a set of capabilities for network management that are exposed and consumed via a standard service interface. A Management Function (MnF) is a logical entity playing the roles of MnS consumer and/or MnS producer. A set of MnSs can be freely grouped together in an MnF, and an MnF can be deployed as a stand-alone entity or embedded in an NF [33,34].

The data analytics capabilities in the SBMA are encapsulated as an MnS referred to as Management Data Analytics Service (MDAS). MDAS plays a similar role in the 5G management plane as the NWDAF does in the control plane and works with the NWDAF in a coordinated way. According to the current 3GPP standards, MDAS takes the responsibility of the management loop from the network wide view, while the NWDAF focuses on the control loop directly related to individual network functions. However, these two closely related aspects of data analytics capabilities in 5G networks might be integrated into one system module in practical implementations for achieving higher efficiency and better performance.

4. Experiential Networked Intelligence Architecture

The NWDAF/MDAS-centric network automation framework in the 5G SBA focuses on the data collection and analytics information dissemination aspects of intelligent management. The other important aspect is the process of transforming collected data to information and knowledge then to intelligence for managing network operations in an "observe-orient-decide-act" loop. Various ML techniques may be employed at the different stages in such a loop; therefore, it is important to standardize an architecture for integrating these techniques seamlessly for achieving context-aware, ML-driven, and policy-based autonomous management. This is the main objective of the Experiential Networked Intelligence (ENI) architecture developed by ETSI [34].

The ENI architecture is designed for meeting a wide range of requirements including requirements regarding network and service planning, deployment, optimization, and provisioning; requirements related to data collection/modeling/analysis, policy management, and interaction with other systems; and non-functional requirements regarding system performance (e.g., latency, accuracy, efficiency), reusability, extensibility, etc. [35]. The ENI architecture may be applied to virtually all management aspects of future networks with numerous use cases identified by the ETSI ENI ISG in the areas of infrastructure management, network operation, service orchestration and management, and network assurance [36].

As shown in Figure 3, the ENI architecture comprises three main modules—the input processing module, the analysis module, and the output generation module. The input processing module contains the data ingestion function and the normalization function. The analysis module consists of functions for knowledge management, context-awareness, cognition management, situation-awareness, model-driven engineering, and policy man-

agement. The output generation module comprises the denormalization function and the output generation functions [34].

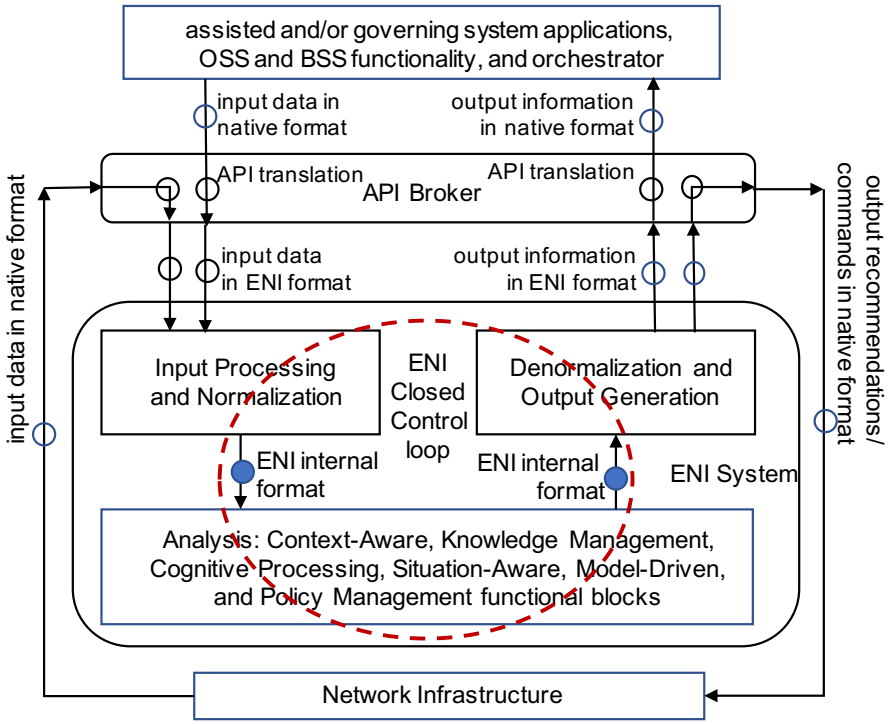

Figure 3. ETSI Experiential Networked Intelligent (ENI) architecture.

The data ingestion function is responsible for collecting data from multiple sources and processing the collected data to make them ready for being further processed and analyzed by other ENI functions. The normalization function translates the data received from the ingestion function into a form that other ENI functions can understand and use. These two functions for input processing may be combined into a single functional block for achieving more efficient data pre-processing in practical implementations of the ENI architecture.

Functions in the analysis module form the core of an ENI system, which can be further split to two categories: (i) functions for knowledge management and processing, including knowledge management, context-awareness, and cognition management functions; and (ii) functions for situation-aware model-driven policy generation, including situation-awareness, model-driven engineering, and policy management functions.

The knowledge management function provides a standard representation of the information about the ENI system and the systems under its management. Such a standard information representation forms a basis for leveraging various ML and data analytics techniques in the ENI architecture. The context-awareness function describes the environment in which the managed system exists or has existed, which is used to continuously update the context in which management decisions are made. The cognition management function is responsible for two key aspects of analysis: first understanding the normalized ingested data/information, as well as the context in which the data/information are produced, then evaluating the meaning of the data/information to determine if any operation/management action should be taken.

The situation-awareness function allows the ENI system to be aware of the current situation of the managed system, including how the ENI analytics results are impacting the system operations. The model-driven engineering function uses a set of models that collectively abstract all important concepts for managing the system(s) governed by ENI. The policy management function generates policies for guiding decision making about

network management based on the data/information provided by the set of knowledge management and processing functions.

In the output generation module, the denormalization function transforms the data received from other ENI functions, including policies, recommendations, and/or commands, into a form that may be communicated with the managed system(s). Then, the denormalized data are passed to the output generation function, which is responsible for handling data delivery to the managed system(s).

The convergence of networking and cloud/edge computing enabled by cloud-native network design implies the coexistence of heterogeneous network, compute, and storage systems with different management interfaces and APIs in future networks. If the ENI system has to understand all the different APIs in order to communicate with these heterogeneous managed systems, it will significantly increase system complexity and management overheads, thus degrading network and service performance. In order to address this challenge, an API broker mechanism is included in the ENI architecture. The API broker provides a common communication channel between the ENI functions and the external managed systems, which allows the ENI architecture to just define a single API. The API broker is responsible for the translation between this single API and the data communications with various managed systems and the transformation between the normalized data form used by all ENI functions and the specific data forms that can be understood by different external systems. Therefore, ENI supports autonomous the management of converged network-compute systems and composite network-cloud/edge services through a standard architecture in which various AI/ML-based technologies can be seamlessly integrated and fully leveraged.

5. Unified Architecture for Machine Learning in Future Networks

In order to facilitate the integration of ML capabilities into the future network architecture, the ITU-T Focus Group on ML for future networks including 5G (FG-ML5G) has developed a unified architectural framework for both ML functionalities and network functions. The following aspects of the architectural requirements have been considered in the design of this ML architecture. The architecture should be able to support various ML techniques/models and correlate the data in different formats obtained from heterogeneous resources. The architecture should be agnostic to the implementations of the underlying networks and support various future networking technologies. The architecture should also support the flexible deployment of ML functionalities in the underlying network including the distributed placement and composition of ML functionalities. Another important aspect of the requirement for the architecture is to enable flexible management of ML functionalities and the integration of the management and orchestration for both ML functions and network functions [37].

The high-level structure of the ITU-T architecture for integrating ML in future networks is depicted in Figure 4. This architecture comprises three subsystems—the ML pipeline subsystem, the management subsystem, and the ML sandbox subsystem [38].

An ML pipeline is a set of logical nodes, each with specific functionalities, that can be combined to form an ML application in a network. An ML pipeline comprises the following types of nodes. The Source (SRC) node is the source of data that can be used as the input to the ML pipeline. Examples of source nodes include user devices and some NFs defined in the 5G SBA such as SMF and AMF. The Collector (C) node collects data from one or more SRC nodes. The Pre-Processor (PP) node is responsible for cleaning, aggregating, and preparing the data to be in a suitable format for the ML model. The Model (M) node is an ML model used in the pipeline, for example a classification or prediction function. The Policy (P) node enables the application of policies to the output of the model node. The Distributor (D) node is responsible for identifying the sink node(s) and distributing the output of the M node to the corresponding sink node(s). The Sink (SINK) node is the target of the ML output on which it takes actions.

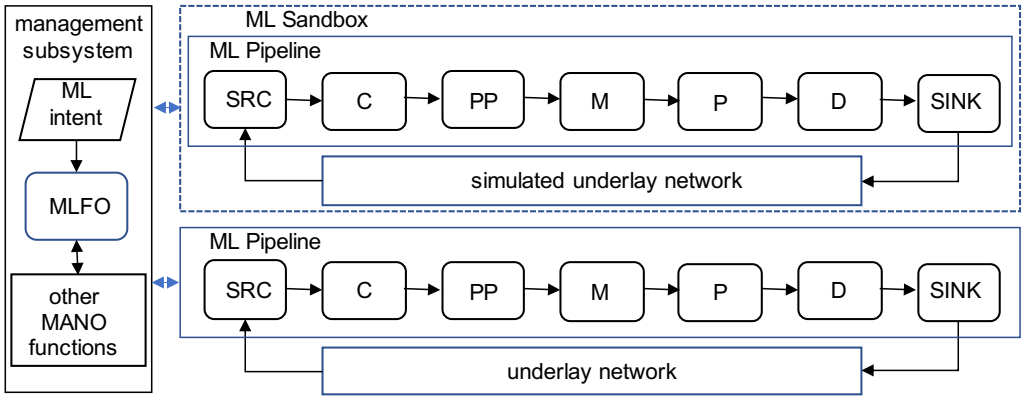

Figure 4. ITU-T unified architecture for ML in future networks.

The workflow of information process in an ML application can be represented by the chaining of nodes in an ML pipeline. The data from various source nodes in the underlying networks are collected (by the C node) and pre-processed (by the PP node) before the data are analyzed by the ML model (at the M node). The output of the ML model is then used to apply policies (by a P node) that will be implemented (by a SINK node). An ML application can be realized by instantiating the ML pipeline nodes as VNFs and associating them with the technology-specific underlying network functions, based on the corresponding requirements of the ML application and the capabilities of the underlying network functions.

The management subsystem extends the management and orchestration capabilities for network functions and services to ML pipeline nodes, thus enabling uniformity to the management of ML functionalities and network/compute functions in future networks. This subsystem includes NFV MANO functions such as VIM, NFVM, and VNFO and defines the ML Function Orchestrator (MLFO) as a core component. The MLFO is responsible for managing and orchestrating ML pipeline nodes based on the ML intent and/or dynamic network conditions. The MLFO selects the ML model for each ML pipeline based on factors such as the model performance, manages placement of ML pipeline nodes based on the corresponding network capabilities and ML application constraints, and provides chaining functions to connect ML nodes together to form an ML pipeline.

As shown in Figure 4, ML intent also plays an important role in the management subsystem as the input to the MLFO. ML intent is a declarative description that can be used by the network operator to specify an ML application and the corresponding ML pipeline. The technology-specific implementation of the specified ML pipeline is then determined and realized by the MLFO in coordination with other management and orchestration functions in the management subsystem.

An ML sandbox is an isolated domain that allows the hosting of separate ML pipelines to train, test, and evaluate them before deploying the pipelines in a live network. For training and testing, the ML sandbox can use data generated by a simulated underlay network and/or data collected from a live network. An ML sandbox may leverage ML techniques such as supervised learning to train and test ML models. The ML pipeline in the sandbox is also managed and monitored by the MLFO. Through having a sandbox subsystem, the ML architecture allows ML pipelines to adapt to the dynamic networking environment in future networks.

The ITU-T ML architecture defines data handling reference points between the ML pipeline and the underlay network (also between the ML pipeline in the sandbox and the simulated underlay network). Using these reference points, any impact to the underlay networks is localized to the source of the data and the target of configurations (as a result of ML pipeline execution). The ML architecture also defines standard interfaces between each pair of subsystems, between the ML pipeline and the underlay network, between the ML sandbox and the simulated underlay networks. The SBI defined in the 5G SBA may be

used for realizing the interfaces between subsystems, as well as the interfaces between ML pipeline nodes.

The ITU-T ML architecture embraces the cloud-native network design principle and is essentially aligned with the 5G SBA. In this ML architecture, decomposition of an ML-based management application to a pipeline of logical nodes allows finer grained modularity of the management system, which may enhance the flexibility and agility of network and service management. Each pipeline node in the ML architecture may be realized as an NF service as defined in 5G SBA, which can be deployed independently as a VNF instance interacting with other nodes via the service-based interface. The ML architecture may be realized based on the NWDAF/MDAS functions in 5G networks; for example, an ML pipeline may have AMF as the SRC node, PCF as the SINK node, and the other nodes hosted on the NWDAF. The cloud-native design of the ML architecture allows the MFLO to compose ML pipeline nodes for constructing various ML applications that meet the diverse management requirements in future networks. The unified orchestration of ML pipeline nodes and network/compute functions in the ML architecture may greatly facilitate the seamless integration of ML techniques in future networks for supporting network-cloud/edge convergence.

6. Architecture for Cross-Domain End-to-End Network and Service Management

The cloud-native future network with network-cloud/edge convergence makes the cross-domain network and service management particularly important and challenging. A key attribute of the network virtualization is the heterogeneous technology domains in the shared infrastructure; for example, the NFV virtualized infrastructure layer comprises the compute, storage, and network domains. Diverse management capabilities are required in the different technology domains for data transport, process, and storage. In addition to technology domains, end-to-end service provisioning in future networks often traverses different administration domains, e.g., domains managed by mobile network providers, edge service providers, transport network operators, cloud service providers, etc. Heterogeneous management mechanisms and polices are employed in different administration domains. Therefore, future networks call for unified resource management across heterogeneous technology domains and federated service management spanning over multiple administration domains.

Although the 3GPP NWDAF, ETSI ENI, and ITU-T ML architectures have defined capabilities that are able to handle the management diversity in technology domains, these standards all focus on the management scenarios within a single administration domain. An overarching framework is required to face the challenge of end-to-end cross-domain management. Toward this direction, ETSI is developing the ZSM architecture with the goal to define a holistic management framework that enables the integration of the management capabilities specified in various standards for supporting end-to-end management across technology and administration domains [39,40].

The ZSM architecture follows the service-oriented architectural principle in order to build a service-based framework for inter-domain network and service management. The set of specific principles for ZSM design include modularity, extensibility, scalability, simplicity, service composability, management concern separation, and functional abstraction. The ZSM architecture defines management domains as an abstraction of technology/administration domains to support the separation of management concerns. Each management domain exposes a set of management services that can be accessed via standard interfaces (service end-points). The self-contained management services are the basic ZSM system modules that can be deployed and scaled independently to accommodate the management load. Loose-coupling between management services allows new services and service capabilities to be added easily, thus making the system highly extensible. Through a consistent set of composition patterns and interfaces defined in the ZSM architecture, management services may be used collectively to construct more complex management functions [41].

As depicted in Figure 5, the ZSM reference architecture is composed of multiple Management Domains (MDs); each is responsible for managing a domain infrastructure and one End-to-End Service Management Domain (E2EMD) that orchestrates the management services provided by individual MDs for cross-domain end-to-end management. Each management domain, including the E2EMD, exposes a set of management services provided by the management functions in the domain [42].

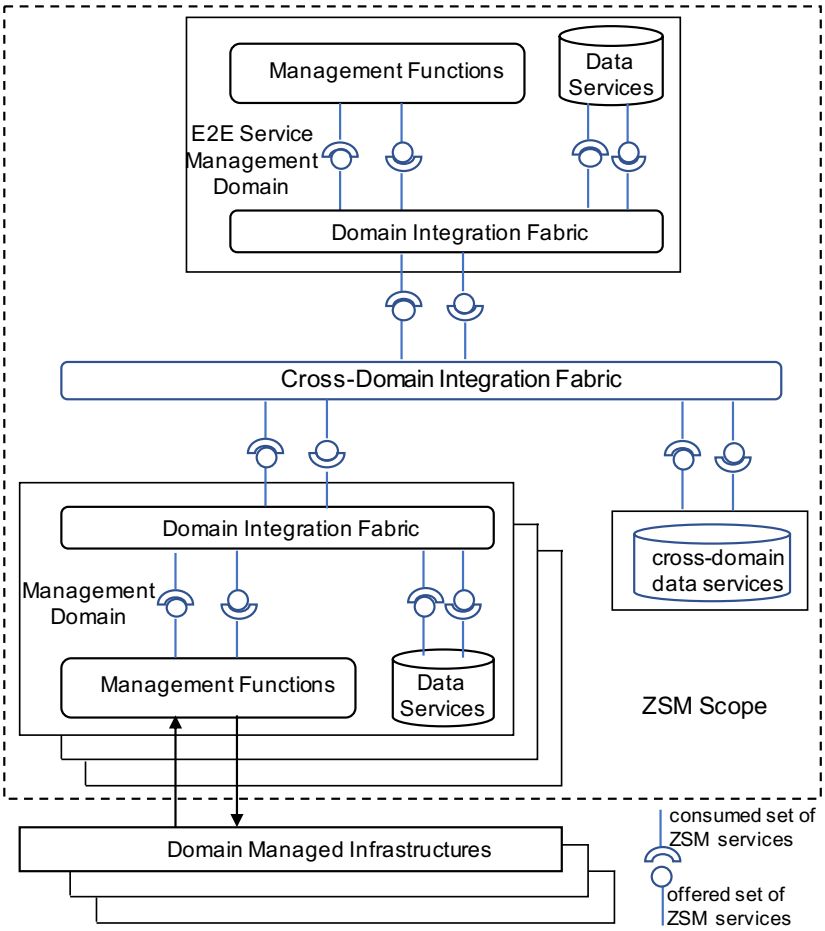

Figure 5. ETSI Zero Touch Network and Service Management (ZSM) architecture.

The management services realized by each MD can be categorized into the following groups. Domain data collection services monitor the managed infrastructure and collect data as requested by other management functions in the MD. Domain analytics services perform various data analytics on the collected data and/or data stored in the domain to provide domain-specific insights and generate domain-specific predictions. Domain intelligence services are responsible for driving intelligent closed-loop automation to enable autonomous management for the domain infrastructure. Domain orchestration services enable automatic life-cycle management of the virtual (and physical) functions in the infrastructure managed by the MD. Domain control services are consumed by other functions in the domain intelligence and domain orchestration service groups to control the states (including configuration, lifecycle, etc.) of the managed domain infrastructure.

The E2EMD is responsible for end-to-end management through orchestration of the management services provided by individual MDs. The E2EMD together with the set of MDs in the ZSM architecture form a two-level hierarchical structure where each individual MD directly manages the infrastructure resources within a single network domain, while the E2EMD composes the MD management services for managing customer-facing end-to-end network services. The end-to-end management services provided by the E2EMD support the following functions: E2E service orchestration that coordinates

lifecycle management of services across management domains, E2E service intelligence that drives intelligent close-loop automation of end-to-end network and service management, E2E service analytics that derives end-to-end service insights for managing end-to-end service performance, and E2E data collection that collects end-to-end service-related data.

As a service-based inter-domain management framework, the ZSM architecture must provide a flexible platform for supporting communications and interoperation of the management functions with respect to the offered and consumed management services, not only within each MD, but also across different MDs. The integration fabric is the mechanism defined in the ZSM architecture for achieving this objective. The integration fabric provides functions for supporting registration/deregistration, discovery, selection, and invocation of management services and providing both synchronous and asynchronous communications among management services. The ZSM integration fabric comprises two levels—a domain integration fabric in each individual MD (including the E2EMD) and a cross-domain integration fabric for supporting service orchestration across different MDs.

For achieving flexible data collection and maintenance in multi-domain management environments, the ZSM architecture provides cross-domain data services as well as data services within each MD. The key functions of ZSM data services include maintaining data persistence and supporting data sharing within and between management domains. ZSM Data services also enable abstraction of data process actions from management functions, which may eliminate the need to handle such actions on a per-function basis, thus allowing stateless management functions. The common cross-domain data services allow the separation of data storage and data processing, thus facilitating cross-domain data exposure and access.

It is worth mentioning that ETSI ZSM is not the only standard for cross-domain network and service management. Other network SDOs also developed architectures for service management and orchestration across technology and/or administration domains. For example, the Abstraction and Control of Transport Network (ACTN) specification [43] developed by IETF and the Lifecycle Service Orchestration (LSO) standard [44] developed by MEF both offer a framework for cross-domain management within which analytics techniques may be applied for supporting intelligent management. Compared to other standard architectures for cross-domain management such as ACTN and LSO, ZSM is more representative of the cloud-native design principle and is developed particularly for integrating the various ML techniques employed in individual domains for intelligent and autonomous management. Therefore, ETSI ZSM is chosen as the focus of this section for reflecting the state-of-the-art standardization of cross-domain management architecture for cloud-native future networks

7. Open Issues and Opportunities for Future Research

The management architecture for future networks and services has attracted extensive research interest from both academia and industry. In addition to the standardization work reviewed in previous sections, the open source community is also actively involved in developing platforms for network management and service orchestration. Open Source MANO (OSM) (https://osm.etsi.org/ (accessed on 1 February 2021))is an ETSI-hosted project to develop an open-source MANO software stack. ONAP (https://www.onap.org/ (accessed on 1 February 2021)) is a comprehensive open source platform for orchestration, management, and automation of composite network-cloud/edge services. The architecture for network service management and orchestration has been investigated in various research projects, for example the NGPaaS project (http://ngpaas.eu (accessed on 1 February 2021)) and the 5GTANGO project (https://www.5gtango.eu/ (accessed on 1 February 2021)), both sponsored by the 5G PPP research initiative. Numerous results of academic research on applying ML/analytics techniques for network management have also been published in the literature, for example [45–47].

Although encouraging progress has been made toward a holistic architectural framework that integrates various ML/data analytics techniques for intelligent and autonomous

management in future networks, there are still some open issues that need to be further studied, thus offering opportunities for future research and development. The discussions in this section focus on open issues and research directions particularly relevant to management for supporting network-cloud/edge convergence in cloud-native future networks from an architectural perspective.

7.1. Heterogeneity in the Managed Systems for Network-Cloud/Edge Convergence

The converged network-cloud/edge systems in future networks comprise various network-compute-storage infrastructures that are highly diverse in management capabilities; for example, NFV MANO functions for network slices, host-/system-level management for MEC infrastructures, and VM/container orchestrators for cloud data centers. In addition, various infrastructure systems may employ different models for resource/service abstraction; for example, the YANG model is often used in SDN networks, while the TOSCA model is typically adopted for cloud data centers. Therefore, how to design a unified management architecture for the heterogeneous infrastructure systems in future networks is a challenging problem.

Cloud-native network design based on service-oriented network virtualization certainly offers a promising direction for future research for addressing the heterogeneity issue. How to fully exploit the cloud-native principle to achieve optimal management architecture design still needs more thorough investigation. A possible research topic for the ENI architecture is service-based design for the API broker, which allows encapsulation of the diverse APIs of the managed systems in well-defined service interfaces that can be communicated via standard messaging protocols. For the ITU-T ML architecture, leveraging the latest progress in (micro)service composition and container orchestration for realizing the ML pipeline and MLFO offers interesting topics for future research.

7.2. Heterogeneity and Scalability of Management Functions

A large number of highly diverse management services are expected to be provided by various management functions for meeting the wide spectrum of service requirements in future networks. Therefore, scalable and flexible lifecycle management for the wide variety of management functions/services in the future network management architecture becomes an important topic for further study. Some important aspects of this topic include registration, discovery, selection, composition, deployment, invocation, scaling, and migration of management services.

The integration fabric in the ZSM architecture plays a crucial role in addressing most of the aspects listed above, but the current ZSM standard still lacks the specific mechanism needed for facilitating integration fabric implementation. A possible direction for future research to address this challenge is to exploit some technologies developed in (micro)service-oriented systems, for example the service communication proxy in the 5G SBA and the Kubernetes orchestrator in data centers. However, some special requirements for management services must be considered. Such requirements include declarative specification of intention for management service composition, capability and attribute exposure of management functions, and cooperation between the management service orchestrator (e.g., MLFO) and other MANO functions within the same network architecture.

7.3. Management Architecture for Cross-Domain Service Provisioning

Cross-domain management for end-to-end service provisioning in future networks with network-cloud/edge convergence is a particularly challenging problem. The design of a scalable structure for end-to-end service management across multiple autonomous domains is a key research topic from an architectural perspective. One possible development direction in this area is to follow the hierarchical structure defined in the current ZSM framework. Such a hierarchical structure requires an appropriate level of abstraction in terms of data/information and management capabilities between the two levels (individual MDs at the lower level and the E2EMD at the higher level) in order to achieve a scalable

management architecture, which has not been fully specified by the ZSM standard, thus offering a topic for future work.

Another research direction for addressing the cross-domain challenge is to explore alternative management architectures. Converged network-cloud/edge service provisioning across different service providers (e.g., network operators, edge/cloud service providers, and Internet service providers) makes a single end-to-end management domain unrealistic. Therefore, alternative structures for cross-domain end-to-end management should be explored. For example, peer-to-peer structures with federation between individual management domains and hybrid structures that combine federative and hierarchical management all deserve thorough evaluations in terms of their effectiveness and scalability for supporting intelligent management in future networks.

7.4. Data Collection in Future Networks

Data-driven AI/ML-based cognition techniques form the foundation of intelligent management in future networks, which requires instrumentation in order to provide the necessary telemetry data to fuel the ML engines. Future networks will involve high data rates and large scale, which imply that unprecedented volumes of network telemetry data will be generated, collected, and processed. In addition, the high-performance and often mission-critical applications supported by future networks often require telemetry data to be highly precise and comprehensive. Such requirements may limit the ability to apply sampling as a technique on a large scale, compounding the scalability challenge as the telemetry data volume explodes.

In order to address these challenges, it becomes imperative for the management architecture in future networks to support "smart" data collection that is able to obtain necessary data for analytics without generating vast volumes of raw data. This involves the development of functions that automatically adjust the resolution of data measurement based on the context and current conditions and allow data pre-processing in the managed network systems. Another possible research direction for tackling this challenging problem is to explore data analytics techniques that are able to support intelligent management without consuming a massive amount of raw data.

7.5. Data Sharing across Management Domains

Data sharing across multiple management domains is another open issue for future research. The data model and the data repository are two important aspects of this problem that deserve thorough investigation. Standard data models play a crucial role in data sharing among different management functions/services within and between management domains. The first question that needs to be answered is if there is a single preferred data model that can be used by all management functions or if multiple data models are needed. Then, for the former case, the common data model needs to be standardized and for the latter case, how different data models can be reconciled must be studied.

A cross-domain data repository is an important component in the management architecture to allow data sharing across management domains. A centralized data storage suffers scalability issues in a large network; therefore, the structures of distributed storage collaborating for exposing a common data service interface need to be investigated. In addition, the appropriate level of data abstraction for cross-domain data sharing with a trade-off between service performance and system scalability also deserves a more thorough study.

7.6. ML Collaboration across Management Domains

A single ML engine lacks the scalability for global management of a large network consisting of multiple management domains. Furthermore, a global ML process for the entire multi-domain network might not be realistic due to the administration autonomy of different network operators/service providers. Both the hierarchical structure as specified by the ZSM framework and alternative structures (e.g., peer-to-peer federative structure) assume that each domain has its own choice of ML/analytics models and techniques. The

combination of domain-specific knowledge and the aggregation of different analytics models across management domains are important issues that are still open for future research.

Recent progress in new machine learning technologies such as federated learning [48] offers some promising approaches to addressing these issues. The application of federated learning in networking has started attracting researchers' attention [49]. On the other hand, federated learning in large-scale networks faces some challenges such as expensive communications, system diversity, statistical heterogeneity, and privacy concerns [50]. How to design a management architecture that is able to fully leverage ML capabilities for cross-domain management while meeting the heterogeneity and scalability requirements becomes an important problem for future research.

8. Conclusions

Network virtualization together with the service-oriented architectural principle enable the trend of cloud-native network design that is transforming communication networks from a data transport infrastructure to a versatile service platform that supports the convergence of networking and cloud/edge computing. Intelligent and autonomous management of network operations and service provisioning play a critical role in cloud-native future networks. The wide variety of AI/ML-based technologies developed for addressing different management aspects and the multi-tenant multi-domain nature of future networking call for standard management architectures that may integrate various management technologies within a common holistic framework. This paper surveyed the current status of the standardization of management architectures for enabling intelligent and autonomous management in cloud-native future networks. The notion of cloud-native network design and its impacts on future networking were first introduced. Then, the paper reviewed the latest developments of representative management architectures, including the 3GPP Network and Management Data Analytics Function (NWDAF/MDAS), ETSI Experiential Networked Intelligence (ENI), the ITU-T Unified ML Architecture in future networks, and the ETSI Zero Touch Network and Service Management (ZSM), and analyzed how cloud-native network design facilitates the management architecture developments. The paper also discussed open issues related to intelligent and autonomous management from an architectural perspective and identified some possible directions for future research and development.

Funding: This research received no external funding.

Data Availability Statement: Not Applicable, the study does not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Turner, J.S.; Taylor, D.E. Diversifying the internet. In Proceedings of the GLOBECOM'05—IEEE Global Telecommunications Conference, St. Louis, MO, USA, 28 November–2 December 2005; Volume 2, pp. 6–12.
2. Feamster, N.; Gao, L.; Rexford, J. How to lease the Internet in your spare time. *ACM SIGCOMM Comput. Commun. Rev.* **2007**, *37*, 61–64. [CrossRef]
3. ETSI. *NFV ISG Introductory White Paper on Network Function Virtualization (NFV)*; ETSI: Sophia Antipolis, France, 2012.
4. Hu, F.; Hao, Q.; Bao, K. A survey on Software-Defined Network and OpenFlow: From concept to implementation. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 2181–2206. [CrossRef]
5. Kreutz, D.; Ramos, F.M.; Verissimo, P.E.; Rothenberg, C.E.; Azodolmolky, S.; Uhlig, S. Software-defined networking: A comprehensive survey. *Proc. IEEE* **2014**, *103*, 14–76. [CrossRef]
6. Yi, B.; Wang, X.; Li, K.; Huang, M. A comprehensive survey of network function virtualization. *Comput. Netw.* **2018**, *133*, 212–262. [CrossRef]
7. Chowdhury, S.R.; Salahuddin, M.A.; Limam, N.; Boutaba, R. Re-Architecting NFV Ecosystem with Microservices: State of the Art and Research Challenges. *IEEE Netw.* **2019**, *33*, 168–176. [CrossRef]
8. Duan, Q.; Wang, S. Network Cloudification Enabling Network-Cloud/Fog Service Unification: State of the Art and Challenges. In Proceedings of the 2019 IEEE World Congress on Services (SERVICES), Milan, Italy, 8–13 July 2019; Volume 2642, pp. 153–159.
9. Sharma, S.; Miller, R.; Francini, A. A cloud-native approach to 5G network slicing. *IEEE Commun. Mag.* **2017**, *55*, 120–127. [CrossRef]

10. Duan, Q.; Wang, S.; Ansari, N. Convergence of Networking and Cloud/Edge Computing: Status, Challenges, and Opportunities. *IEEE Netw.* **2020**, *34*, 148–155. [CrossRef]
11. Boutaba, R.; Salahuddin, M.A.; Limam, N.; Ayoubi, S.; Shahriar, N.; Estrada-Solano, F.; Caicedo, O.M. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *J. Internet Serv. Appl.* **2018**, *9*, 16. [CrossRef]
12. ETSI. *GS NFV 002 Network Function Virtualization Architectural Framework*, version 1.2.1; ETSI: Sophia Antipolis, France, 2014.
13. Duan, Q.; Ansari, N.; Toy, M. Software-defined network virtualization: An architectural framework for integrating SDN and NFV for service provisioning in future networks. *IEEE Netw.* **2016**, *30*, 10–16. [CrossRef]
14. Blenk, A.; Basta, A.; Reisslein, M.; Kellerer, W. Survey on network virtualization hypervisors for software defined networking. *IEEE Commun. Surv. Tutor.* **2015**, *18*, 655–685. [CrossRef]
15. Erl, T. *Service-Oriented Architecture (SOA)—Concepts, Technology, and Design*; Pearson Education Incorporated: London, UK, 2005.
16. Dragoni, N.; Giallorenzo, S.; Lafuente, A.L.; Mazzara, M.; Montesi, F.; Mustafin, R.; Safina, L. Microservices: Yesterday, Today, and Tomorrow. In *Present and Ulterior Software Engineering*; Springer: Berlin, Germany, 2017; pp. 195–216.
17. Letaifa, A.B.; Haji, A.; Jebalia, M.; Tabbane, S. State of the Art and Research Challenges of new services architecture technologies: Virtualization, SOA and Cloud Computing. *Int. J. Grid Distrib. Comput.* **2010**, *3*, 69–88.
18. Duan, Q.; Wang, S. *Network as a Service for Next Generation Internet*; Institution of Engineering & Technology: London, UK, 2017.
19. Magedanz, T.; Blum, N.; Dutkowski, S. Evolution of SOA concepts in telecommunications. *Computer* **2007**, *40*, 46–50. [CrossRef]
20. Duan, Q.; Yan, Y.; Vasilakos, A.V. A survey on service-oriented network virtualization toward convergence of networking and cloud computing. *IEEE Trans. Netw. Serv. Manag.* **2012**, *9*, 373–392. [CrossRef]
21. ETSI. *NFV Release 3 Definition*, version 0.14.0; ETSI: Sophia Antipolis, France, 2019.
22. 3GPP. *TS 23.501 System Architecture for the 5G System*, version 15.12.0; 3GPP: Sophia Antipolis, France, 2020.
23. Taleb, T.; Aguiar, R.L.; Grida Ben Yahia, I.; Chatras, B.; Christensen, G.; Chunduri, U.; Clemm, A.; Costa, X.; Dong, L.; Elmirghani, J.; et al. *White Paper on 6G Networking*; University GENT: Gent, Belgium, 2020.
24. 3GPP. *TR 23.742 Study on Enhancements to the Service Based Architecture, Release 16*; 3GPP: Sophia Antipolis, France, 2018.
25. 3GPP. *TS 23.501 System Architecture for the 5G System*, version 16.7.0; 3GPP: Sophia Antipolis, France, 2020.
26. Taleb, T.; Samdanis, K.; Mada, B.; Flinck, H.; Dutta, S.; Sabella, D. On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Architecture & Orchestration. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1657–1681.
27. ETSI. *GS MEC 003 Multi-Access Edge Computing (MEC) Framework and Reference Architecture*, version 2.1.1; ETSI: Sophia Antipolis, France, 2019.
28. ETSI. *GR MEC 017 Mobile Edge Computing (MEC) Deployment of Mobile Edge Computing in an NFV Environment*; ETSI: Sophia Antipolis, France, 2018.
29. ETSI. *White Paper No. 28: MEC in 5G Networks*; ETSI: Sophia Antipolis, France, 2018.
30. 3GPP. *TR 23.791 Study of Enablers for Network Automation for 5G*, version 16.2.0; 3GPP: Sophia Antipolis, France, 2019.
31. 3GPP. *TS 23.288 Architecture Enhancements for 5G System to Support Network Data Analytics Services*, version 16.5.0; 3GPP: Sophia Antipolis, France, 2020.
32. 3GPP. *TS 28.530 Management and Orchestration Concepts, Use Cases, and Requirements*, version 16.3.0; 3GPP: Sophia Antipolis, France, 2020.
33. 3GPP. *TS 28.533 Management and Orchestration Architecture Framework*, version 16.5.1; 3GPP: Sophia Antipolis, France, 2020.
34. ETSI. *GS ENI 005 Experiential Networked Intelligence (ENI) System Architecture*, version 1.1.1; ETSI: Sophia Antipolis, France, 2019.
35. ETSI. *GS ENI 002 Experiential Networked Intelligence (ENI) Requirements*, version 2.1.1; ETSI: Sophia Antipolis, France, 2019.
36. ETSI. *GS ENI 001 Experiential Networked Intelligence (ENI) Use Cases*, version 2.1.1; ETSI: Sophia Antipolis, France, 2019.
37. ITU-T. *FG-ML5G Technical Specification: Unified Architecture for Machine Learning in 5G and Future Network*; ITU-T: Paris, France, 2019.
38. ITU-T. *Recommendation Y.3172: Architectural Framework for Machine Learning in Future Networks including IMT-2020*; ITU-T: Paris, France, 2019.
39. ETSI. *GS ZSM 001 Zero Touch Network and Service Management (ZSM) Requirements Based on Documented Scenarios*, version 1.1.1; ETSI: Sophia Antipolis, France, 2019.
40. ETSI. *GS ZSM 004 Zero Touch Network and Service Management (ZSM) Landscape*, version 1.1.1; ETSI: Sophia Antipolis, France, 2020.
41. Benzaid, C.; Taleb, T. AI-driven Zero Touch Network and Service Management in 5G and Beyond: Challenges and Research Directions. *IEEE Netw.* **2020**, *34*, 186–194. [CrossRef]
42. ETSI. *GS ZSM 002 Zero Touch Network and Service Management (ZSM) Reference Architecture*, version 1.1.1; ETSI: Sophia Antipolis, France, 2019.
43. IETF. *RFC 8453 Framework for Abstraction and Control of TE Networks (ACTN)*; IETF: Fremont, CA, USA, 2018.
44. MEF. *Lifecycle Service Orchestration (LSO) Reference Architecture and Framework*; MEF: Los Angeles, CA, USA, 2016.
45. Pateromichelakis, E.; Moggio, F.; Mannweiler, C.; Arnold, P.; Shariat, M.; Einhaus, M.; Wei, Q.; Bulakci, Ö.; De Domenico, A. End-to-end data analytics framework for 5G architecture. *IEEE Access* **2019**, *7*, 40295–40312. [CrossRef]
46. Zafeiropoulos, A.; Fotopoulou, E.; Peuster, M.; Schneider, S.; Gouvas, P.; Behnke, D.; Müller, M.; Bök, P.B.; Trakadas, P.; Karkazis, P.; et al. Benchmarking and Profiling 5G Verticals' Applications: An Industrial IoT Use Case. In Proceedings of the 2020 6th IEEE Conference on Network Softwarization (NetSoft), Ghent, Belgium, 29 June–3 July 2020; pp. 310–318.

47. Trakadas, P.; Karkazis, P.; Leligou, H.C.; Zahariadis, T.; Vicens, F.; Zurita, A.; Alemany, P.; Soenen, T.; Parada, C.; Bonnet, J.; et al. Comparison of management and orchestration solutions for the 5G era. *J. Sens. Actuator Netw.* **2020**, *9*, 4. [CrossRef]
48. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Artificial Intelligence and Statistics, PMLR, Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
49. Niknam, S.; Dhillon, H.S.; Reed, J.H. Federated Learning for Wireless Communications: Motivation, Opportunities, and Challenges. *IEEE Commun. Mag.* **2020**, *58*, 46–51. [CrossRef]
50. Li, T.; Sahu, A.K.; Talwalkar, A.; Smith, V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. [CrossRef]