IEEE Open Journal of the Communications Society

Received 12 October, 2024; revised 20 November, 2024; accepted 19 December, 2024; Date of publication 23 December, 2024; date of current version 23 December, 2024.

Digital Object Identifier 10.1109/OJCOMS-02925-2024

Data usage control for privacy-enhanced network analytics in private 5G networks

HAMMAD ZAFAR [1], UMBERTO FATTORE [1,2] *(Member, IEEE)*, FLAVIO CIRILLO [1] *(Member, IEEE)* AND CARLOS J. BERNARDOS [2]

[1]Data Ecosystem and Standards (DES) team, NEC Laboratories Europe GmbH, Heidelberg 68115, Germany
[2]Telematics Engineering Department, Universidad Carlos III, Madrid 28911, Spain

CORRESPONDING AUTHOR: Umberto Fattore (e-mail: umberto.fattore@neclab.eu).

ABSTRACT With the rise of 5G networks and the ongoing evolution toward 6G, the proliferation of private networks has accelerated. While standalone deployments are possible, more efficient private network deployments often involve sharing parts of the private network with public operator's infrastructure, through approaches such as public network integration, hybrid private networks, and network slicing. However, these configurations introduce privacy concerns, particularly regarding the privacy and ownership of private network data that may need to be collected by the public network operator for analytics and joint optimization across private and public networks. This paper explores data usage control mechanisms to safeguard private network data when performing management data analytics. Specifically, we propose a framework for privacy-enhanced data analytics (PEDA) consisting of components that can complement the standard 5G analytics framework. Additionally, we provide a blueprint for privacy-enhanced orchestration of analytics services across public and private 5G networks utilizing NFV-MANO as the orchestration framework. To this end, we demonstrate orchestration of analytics services across distributed infrastructures belonging to private and public 5G networks, according to the data usage policies set by data owners.

INDEX TERMS 3GPP, private networks, non-public 5G networks, control plane, Data Usage Control, Management Data Analytics, mobile core, NFV, Standards, MDAF, NWDAF

I. INTRODUCTION

THE maturity in the Fifth Generation (5G) telecommunication technology and specifications is leading to various private 5G network deployments in the industry.

Private 5G, also termed as non-public 5G, offers benefits such as deterministic performance, service customization, data sovereignty, and enhanced security to different industry verticals like manufacturing, healthcare, airports, etc [1]. Manufacturing use cases (e.g., production line flexibility, machine-to-machine communication, connected workers) can be realized using private 5G network deployments, either via 'on-premises' networks or via public mobile networks. Similarly, airports can have their own private 5G network deployments to offer better services with dedicated resources to the hundreds and thousands of travelers using the facility [2]. Moreover, enhanced quality of service (QoS) offered by private 5G networks can be an enabler for advanced use cases such as AI-assisted computer vision for security and pandemic control in large airports [1], [2].

There are different deployment scenarios for private 5G networks identified by 5G Alliance for Connected Industries and Automation (5G-ACIA): (a) a standalone non-public network (NPN) which is isolated from the public network and where all network functions are hosted on the premises of the private organization such as an enterprise or a manufacturing plant; (b) an NPN with shared radio access network (RAN), where the RAN domain is shared with a public network but other core network functions remain inside the private organization; (c) an NPN with shared RAN and control plane (CP) functions where user plane traffic remains in the isolated premises but control functions are performed by the public network in addition to RAN sharing; and (d) an NPN hosted completely by the public network, a scenario where network slicing and network functions virtualization (NFV) can be utilized to ensure logical private network slices within the operator's public 5G network [3].

The private industries and enterprises can choose among the different above-mentioned deployment scenarios (or any

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

combination of them) for their private 5G networks, based on their specific requirements and evaluation of capital and operational expenditures (CAPEX and OPEX). In most cases, it may be more expensive to host a completely standalone 5G network on premises due to increased CAPEX and OPEX. Therefore, NPN owners may opt for the sharing scenarios owing to their cost benefits. However, it is still of paramount importance that the characteristics of private 5G networks, e.g., high QoS, isolation, security, privacy and accountability, are ensured in hybrid or public network deployments.

One of the major advancements in 5G and beyond 5G (B5G) networks over their 4G counterparts is the increasing use of AI/ML based analytics and automation. AI/ML based analytics are used for optimizing resource allocation, guaranteeing service level agreements (SLAs), proactively avoiding failures and performance degradation, performing root cause analysis, predicting traffic patterns and other purposes. The 3rd Generation Partnership Project (3GPP) specifies the data analytics framework for mobile networks in detail in [4]. Different services and functions, such as network data analytics function (NWDAF) and management data analytics function (MDAF) are defined in 3GPP specifications. These functions process and analyze various kinds of data related to 5G networks. Examples of such data include, but are not limited to, performance measurements, trace data, user equipment (UE) location data, quality of experience (QoE) and service experience data, and data associated to different 5G network functions (NFs) [4]. Enhancements proposed in 3GPP's MDAF framework cover more advanced use cases like slice coverage optimization, cross-slice resource optimization, and network slice traffic projection [5]. These use cases require different kinds of current and historical data from the 5G network slices serving private organizations and enterprises.

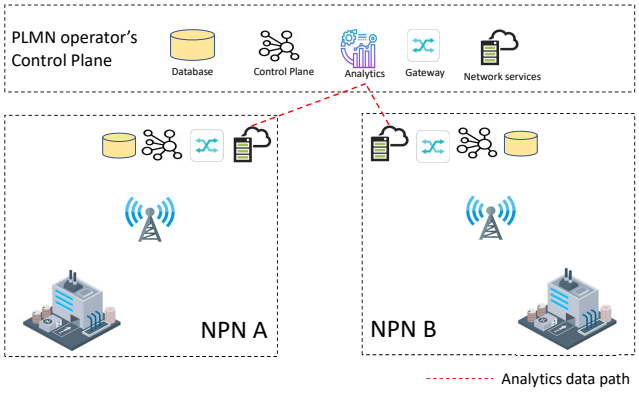

FIGURE 1. Two private networks NPN-A and NPN-B sharing the PLMN operator's Control Plane for management data analytics.

Sharing and processing of data belonging to private 5G networks - that are either sharing control plane with the public landline mobile network (PLMN) or being hosted completely by the PLMN in the form of a network slice - outside their respective trust zones can have serious implications. A PLMN may be serving two private 5G network deployments as network slices for two enterprises and may need to perform analytics based on operational data from the two slices to optimize resource allocation in the respective network slices. The example scenario shown in Figure 1 illustrates two manufacturing plants of market competitors, who have deployed their own in-house private 5G networks, NPN-A and NPN-B, with the control plane being offered by a common PLMN operator. This means that different kinds of analytics - provided by NWDAF and/or MDAF - for the two private networks are performed in the shared operator's network. These analytics may require different kind of data (trace, location, performance, etc.) to be transported out of the respective private premises of private network owners and be processed in the operator's domain. In such a scenario - where potential competitors are sharing the same network for 5G analytics - data owners would likely wish to keep their performance metrics and the information related to the end entities private. Furthermore, potential leak of fault, performance, and UE related data (number of registered users, location, etc.) to competitors or even the public network operator itself can be potentially damaging to the company's interests [6].

According to privacy and security attributes defined in [3], NPN owners can control the flow and processing of their private data. In addition to deciding what information goes where, owners can have several policies related to data privacy associated with their operational data. The operator serving the private 5G networks should enforce these policies by providing necessary techniques and frameworks to safeguard the privacy of operational data associated to private 5G networks while performing different kinds of analytics. Thus, there is a need for a privacy layer on top of the analytics framework provided by 5G that can perform data usage control, modeling and enforcement of policies related to data belonging to private 5G network owners before and during the life-cycle of analytics tasks and services performed in the public operator's domain. Such privacy use case is not limited to 5G analytics tasks and services and can extend to other use cases, for instance to operate distributed AI models over different networks while preserving the models' privacy (e.g., for Federated Learning). However, this aspect is out of scope of this paper.

In this paper, we consider the problem of privacy of data belonging to private 5G networks and its use in AI/ML based 5G data analytics. To this end, we propose a framework to ensure that data usage policies of data owners (i.e., NPNs) are enforced while performing analytics services in 5G networks. The main contributions of this paper are summarized below:

1) A comprehensive framework to ensure data usage control based on policies specified by data owners, enabling modifications to analytics service descriptions

as per the data usage policies and orchestrating these services in a distributed fashion.
2) A blueprint and workflow for privacy-enhanced data analytics (PEDA), featuring distributed orchestration of analytics services across public and private 5G networks using NFV-MANO as the orchestration system.
3) A PoC demonstration of the above-mentioned workflow and solution on a multi-cluster environment and on a single-cluster multi-node environment deployed in different zones of a public cloud, to showcase the efficacy of our framework in enforcing 'fenced data' usage control policies specified by private network owners.
4) An analytical evaluation of the configuration complexity and scalability of our solution showcasing how our PEDA solution performs better as compared to similar approaches applied to the use case considered in this paper.

The rest of the paper is structured as follows. Sections II and III provide relevant background information: the first on private networks in 5G/6G and their possible deployments, while the latter on data usage control (DUC) approaches applied to private networks, as well as examples of existing data usage control solutions. Section IV introduces reference use cases for private networks in which data usage control becomes relevant and the kind of policies that private networks owners may want to enforce on their network data. Section V presents the system architecture of the solution we propose, while Section VI provides a blueprint of how a privacy-enhanced orchestration of analytics services would work through our solution across public and private 5G networks, plus a concrete workflow example based on the previously introduced reference use cases. Section VII provides experimental evaluation and analytical comparison of our solution against relevant existing approaches. Finally, Section VIII concludes the paper and provides insights into our planned future activities.

II. PRIVATE NETWORKS IN 5G/6G

Starting from Release 16 [7], 3GPP specified the concept of a 5G private network using the term non-public network (NPN) and defining it as a network intended for non-public use [8]. The terms 'private networks' and 'non-public networks' (or the abbreviated form NPN) are used interchangeably in the rest of this paper, even when referring to 3GPP framework or use cases.

Reasons for the use of a private network could be various, ranging from security (i.e., keeping data within the boundaries of the organization) to improving performances due to the improved radio coverage (e.g., the organization has in general more control over the radio infrastructure parameters) or reduced latency in accessing services in the core network (i.e., edge-computing deployments) [9]. For instance, a private network could be a perfect deployment for a smart factory [10] [11], employed to connect sensors, robots, and control systems without exposing the network to public internet. The capabilities of 5G would allow the creation of such private network for a smart factory [12], resulting in enhanced security, as well as reliability and control over data flow, which is critical for sensitive operations such as real-time monitoring, automation, and system control.

A. Possible deployments for private networks

Private 5G networks can be deployed in various ways to cater to the specific needs of industrial applications, each with distinct characteristics, benefits, and challenges. The main difference between those deployments lies in the quota of network infrastructure and resources that are totally dedicated to the private network (i.e., owned and managed by the private network owner), ranging from the case of a totally private and isolated network to the other extreme of a private slice provided by a public network operator. Table 1 summarizes four different deployment options as described in 5G-ACIA [13]. While the first one is the standalone deployment, the other three differ in the degree of interaction and infrastructure sharing with the public network [11], [13].

The standalone non-public network (NPN) [13] consists of a fully private infrastructure where the enterprise owns and controls all elements, including the core network (CN) and radio access network (RAN). This setup is particularly suitable for industries that require full control over their network operations. This model allows for a full isolation of the infrastructure and therefore high security, as well as low-latency due to the on-premises data processing. However, the complexity and high costs associated with deploying and maintaining such a network can be significant, as the private owner needs to take care of all the management aspects.

The public network integrated (PNI) model [13] refers to integrating a private network with a public mobile operator's infrastructure. The enterprise maintains its own RAN but uses parts of the public operator's core network. This hybrid approach allows industries to benefit from some private network features while reducing costs and complexity. The major benefits include lower setup costs, easier management, and access to the capabilities of the public network. One of the main concerns is data secrecy and privacy as some components of the private network are shared with the public network. In that deployment scenario, it can be difficult to control access to the data that belongs to the private network. Furthermore, the statistics and information about the network itself are partially managed by the public network operator, restricting enterprise owners to exert full control over their network data.

Finally, hybrid private networks and network slicing represent two additional flexible models that combine private and public network elements [13]. In the first, some parts of the network are managed by the enterprise, while others are handled by the public operator. In the latter, the private

network is made up of a network slice in the operator's public network. The network slice is customized to meet the specific requirements of the enterprise, offering a quasi-private experience without the full responsibility of maintaining a standalone infrastructure. In both models, secrecy and privacy concerns stand regarding data belonging to the private networks, as the private enterprises by default do not have full control over their data.

While the considerations regarding data usage control in this paper apply similarly to all three non-standalone deployments listed in Table 1, we focus on the the PNI case when discussing the various examples and use cases for sake of simplicity.

TABLE 1. Deployment scenarios for private 5G networks as defined by 5G-ACIA [13]

Deployment Scenario	Description	Pros	Cons
Standalone Private Network	Fully private dedicated infrastructure.	Security; isolation; low-latency.	Setup costs; Complexity;
Public Network Integrated (PNI) Private Network	Private network integrated in a PLMN.	Setup costs; simpler management; public net access.	Less control over access; security.
Hybrid Private Networks	Some (not all) parts managed privately.	Flexibility; cost-effective scaling.	Complex management; security and data access.
Campus Networks with Slicing	Dedicated network slice running on public operator network.	Cost-effective, simple.	Less control over security and performance.

III. DATA USAGE CONTROL FOR PRIVATE NETWORKS

The previous section summarizes the advantages of the private network and the different possible deployments. Public network integrated (PNI) deployments enable utilization of different resources and functions of the public network infrastructure, with obvious benefits in term of costs and flexibility. However, the drawback comes in terms of security and privacy, with a totally standalone private network deployment having intuitively higher security and privacy due to its isolation [14]. On the contrary, shared sections of a private network that are managed by the public network need particular attention in terms of privacy and data usage control.

The 5G alliance [3] calls for the need for privacy and security in NPNs, with the first being able to decide "which information goes where" [3], i.e., that NPN owners are able to control the flow and processing of their data. The 5G alliance [3] also mentions the use of isolation as a mean to ensure data privacy, ranging from a physical isolation of deployments scenario in which the control plane of the NPN is totally independent, to a more logical isolation in which the control plane information flows are only logically separated (and protected) from each other.

A. Data usage in 5G Analytics

In 5G networks, a huge amount of data is produced and made available during network operations. Some information refers to users of the network (e.g., mobility patterns), while others to radio access components (e.g., radio channel parameters and configuration), and some more to telemetry data, such as real-time data from network equipment, sensors, and management systems (including alarms and operational states) [15]. Moreover, historical data may be collected to train AI models that are used for analytics in the network [16]. Such data plays a critical role in optimizing the overall performance of a 5G network. For example, mobility patterns and user location data can be used to predict user movement and optimize handover procedures, providing edge-to-edge service continuity to users [17]. Similarly, telemetry data from network equipment is valuable for fault detection and predictive maintenance [18], whereas energy consumption data can be used to improve energy efficiency, e.g., transitioning base stations in low-traffic areas to energy-saving modes without compromising connectivity [19].

Data collected by the 5G system can in some cases be exposed to external users of the network [20] or is typically consumed by two components, namely the network data analytics function (NWDAF) [15] and the management data analytics function (MDAF) [4]. The first provides analytics based on data collected from specific 5G network functions, while the latter is designed to provide data-driven insights and analytics for the overall 5G network [21]. Despite the similarities among the two components, NWDAF focuses on real-time, core network related analytics to optimize operational performance and user experience, while MDAF focuses on management plane analytics, addressing long-term planning of the network infrastructure mainly through telemetry and historical data analysis. Later in this paper, Section IV and particularly Table 3 describe some analytics use cases from 3GPP that are performed by NWDAF and MDAF using different types of data in a 5G network.

With the continuous evolution of mobile core networks and their increasing deployment in the cloud as virtual network functions, the management plane components and analytics services are expected to be deployed in the cloud infrastructure as well. For instance, the popular Telco Cloud technology, NFV, enables not only the orchestration of virtualized network functions but also any other kind of applications that can be hosted on the Telco Cloud (NFV)

infrastructure, including analytics applications. The associated standards organization ETSI ISG also foresees the use of Kubernetes as the default orchestration framework for managing containerized VNFs. Given the convergence of telecommunication network functions and applications on the same cloud infrastructure - a trend already started from 4G, continuing in 5G and foreseen also in the 6G networks - MDAF and NWDAF related applications are expected to be orchestrated in the same way as network functions and other control plane applications. Furthermore, the scopes of MDAF and NWDAF analytics are expected to become more and more interconnected with each other going forward for more complicated, end-to-end analytics use cases. For example, an overlap in the types of data required for both NWDAF and MDAF analytics use cases can already be seen in more complex analytics use cases spanning both management and data planes of the 5G mobile core network [22]. In this paper, we design our solution to be generic and agnostic of data analytics services and tasks that may be performed by either the MDAF or the NWDAF.

B. Data analytics across private and public networks

For the considerations made in the previous sections, standalone private networks are totally isolated 5G networks that therefore have their own NWDAF and MDAF components collecting network data and providing a set of analytics to improve the private network. In the case of standalone networks, there is total isolation of data and network information from a data secrecy and privacy perspective. However, data privacy and secrecy become significant challenges when the private network is not standalone, such as in the case of an integrated Public Network Integrated (PNI) described in Section II.

The example scenario described in Section I and depicted in Figure 1 can be prone to exploitation. NPN-A and NPN-B have their own radio and core infrastructure, but share a common control plane in the operator's network domain. This means that NWDAF and MDAF components are shared among the two NPNs. The operator's control plane collects and processes data coming from the two private networks to perform data analytics. This makes sense from an operator's point of view, since the operator's network needs to have some information of the private network in order to perform a joint optimization of the overall network and to respect SLA agreements with owners of both private networks. However, this poses significant problems from a data privacy point of view, as enterprises owning their private networks may not be, for competitiveness and privacy motivations, leaning towards sharing information about their networks. The collected data from the private networks can include subscription information about the users of the networks, data about the type of sessions that are active in the private network, as well as information about network devices that are operative in the private networks. All of which is information that is often classified as sensitive in an enterprise domain.

Therefore, a trade-off needs to be made between the need of privacy and security for the private network data, and the need to share some information to allow for a joint optimization of overall network performance. One robust way to do so is through the use of a data usage control approach, in which data from a private network are associated with access policies before being shared (if they can be shared) with the public infrastructure. The public owner will therefore have a limited or restricted access to the private network data, preserving the enterprise privacy, while still being able to use the data to perform some overall optimization. This approach is described in more detail in the next sub-sections, which also summarize different kind of data usage policies that can be specified by the data owners.

C. Data Usage Control for 5G Data Analytics

In data information systems, the data privacy and security is usually enforced on its provisioning, namely access control, with various types of policies and granularity. However, once the data is shared there is no further control on how the data is used. Data usage control (DUC) [23] is an evolution of access control. For instance, digital rights management (DRM) is a class of usage control systems aimed at protecting digital information shared in an open environment [24]. These systems pertain typically to specific scenarios and not to the data processing realm. Nowadays data exchange for data processing among different parties is regulated through agreements which can lead to legal action in case of misuse. However, detection of misuse and enforcement of regulations can be slow and cumbersome. Furthermore, data consumer might misuse the data unwittingly and, thus, creating harm without intention. Because of these challenges, data exchange across stakeholders is often hindered by high costs (on the clearance control and on the legal enforcement of the agreements) and slow processes.

The International Data Spaces Association (IDSA)[1] specifies a system aiming at technically enforcing data exchange in a trusted and secure manner through usage control [25]. A *data space* is a distributed network of data endpoints held by different stakeholders that enable the exchange of data. Whereas IDSA hosts a significant activity from many stakeholders on the trusted data sharing aspects, it has still not considered any specific application to 5G networks.

There can be multiple approaches on the data usage control enforcement depending on the usage scenarios. The network monitoring data of a private network is a stream of data record observed at different time steps. Therefore the data messages are passed to an analytics task incrementally and periodically. Controlling how a consumer uses every exchanged message, especially if the exchange frequency is high, can be cumbersome and resource-intensive. Furthermore, the control of the data usage can happen before

[1] https://internationaldataspaces.org/

(proactive) or after (reactive) the data usage takes place. Controlling after the data has been processed could still permit illegitimate processing although reaction to misuse might be enforced (such as dropping the invalid processed data [26]).

Considering the perspective of the data consumer, establishing a data processing flow that respects all the policies without unintentionally breaking some of them is not a trivial task. IntentKeeper [27] is a system based on a data processing orchestration approach that includes the policies into the orchestration decision. On the one side, the data provider specifies the policy at high level. On the other side, a data consumer designs its data analytics service describing input, output and used analytics operators in a topology. The system automatically and dynamically orchestrates the data processing pipeline and establishes the data flows in a manner that avoids any policy break and enabling the correct data analytics. The data processing and data exchange are assumed to happen within certified processing nodes that cannot be tampered or accessed. Once the flow is started, there is no additional check to be executed because the data usage control enforcement is proactively enforced by design, thus avoiding unnecessary overhead. Table 2 shows a list of policies that can be enforced with the IntentKeeper solution [27].

The IntentKeeper is an intent-based solution in which the data owner provides general policies for its data and then the usage control tool takes care of checking enforcement of these polices. Other existing solutions operate on a per-service (or per-flow) level of granularity, meaning that a policy needs generally to be defined per each new service or flow defined to utilize data of the data owner, resulting in a more complex and less automated configuration process. This is the case for instance of LUCON [28], a policy framework designed to govern data usage, which defines policies for each new flow of data. Another existing solution working with a similar approach to LUCON is the Dataspace Connector (DSC) [29] developed to implement data sovereignty rules when sharing telemetry data in an optical network.

The privacy-enhanced data analytics (PEDA) solution proposed in this paper adapts the same approach as the IntentKeeper and integrates it with the 5G network's MDAF and NWDAF elements to minimize the number of configurations while enforcing data usage polices summarized in Table 2.

TABLE 2. Data usage control policies in the data space of private networks and public network

Policy	Definition	Example application for Private Network
Secrecy	Classified data may not be forwarded to nodes that do not have the respective clearance	Network monitoring information is shared only if there are APIs to enforce data usage control
Time to live (TTL)	The persistence of data is limited to a given period of time	The private network owner might decide if the monitoring data may be used only for short-term network management or also for long-term management based on profiling.
Anonymize by aggregation	Personal data may only be used in an aggregated form by external parties. A sufficient number of distinct records must be aggregated to prevent deanonymization of individual records.	The private network owner might allow the public network to use its data without disclosing information that would allow profiling of the private network
Separation of duties	Two data sets from competitive entities must never be aggregated or processed by the same service	A private network owner might decide not to allow market analysis encompassing its data
Fenced data	Original data may be only processed by nodes within certain premises boundaries.	A private network owner might allow the processing of the data only within its own server (i.e. processing task is sent to the data owner)

IV. ANALYTICS USE CASE

This section describes a scenario involving different 5G analytics services that can be performed using the private data belonging to owners of private 5G networks, NPN-A and NPN-B, as shown in Figure 1. We consider some example data usage policies set forth by both data owners on their respective data. Analytics services involving both NWDAF and MDAF are studied in this scenario to illustrate a comprehensive privacy-preserving analytics framework. The orchestration blueprint described in section VI describes how to orchestrate the analytics services considered in our example scenario while ensuring data privacy and enforcement of data usage policies set by the data owners.

3GPP TS 28.288 [15] specifies architectural enhancements to support network data analytics including the interactions between NWDAF and other entities, the type of data to be collected for different kinds of analytics services and the procedures for performing these analytics. Few examples of analytics services that can be performed by the NWDAF are slice load level related network data analytics, NF load analytics, network performance analytics, UE related analytics etc. Type of data collected for these different kinds of analytics services and its sources are also specified for each analytics service in [15]. Similarly, 3GPP TS 28.104 [4] specifies capabilities and analytics services that an MDAF can perform in a 5G network. Examples of such analytics services are coverage related analytics, service experience analysis, energy saving related analytics etc. Data required

for these MDAF analytics services and their respective analytics outputs are also specified in [4].

Figure 1 shows two separate private 5G deployments, NPN-A and NPN-B. Owners of both private 5G network have their own set of policies associated with their data that should be enforced by the PLMN's control plane analytics services (i.e., NWDAF, MDAF) while performing data analytics. Table 3 summarizes the data usage policies set by data owners, types of data and desired analytics services to be performed on that data.

Suppose an application in the private 5G network NPN-A, wishes to use the PLMN's NWDAF services for NF load analytics of one or multiple NFs deployed in the private network. The analytics consumer – an application in NPN-A in this case - can request the 'NF load analytics' service from the NWDAF for concerned NF(s) and use the analytics output for its own purposes. For the analytics operations, the NF-related data is collected by the NWDAF for further processing to be done in the PLMN domain. However, the data owner – NPN-A – has set up data usage policies to anonymize the data as well as put a time-to-live (TTL) constraint on the data collected by the PLMN's control plane for respective analytics. Therefore, some pre-processing will be required as part of data usage control on the data collected from the NFs as well as monitoring the TTL constraints for the data to be kept by NWDAF for any future processing.

Another example of data usage policy can be a 'fenced data' policy, meaning the data cannot be process or taken out of the trust domain of data owner. In our use case, we assume that the data belonging to the private 5G network NPN-B, as shown in Figure 1 has the 'fenced data' policy. This means additional considerations should be taken if any analytics were to be performed on NPN-B data. We consider a scenario where either the owner of the private 5G network NPN-B or the PLMN operator on behalf of NPN-B needs to perform MDAF analytics related to energy saving in the network. To perform the energy saving analytics service, MDAF requires performance and minimization of drive test (MDT) related measurements as input. As per the 'fenced data' usage policy of NPN-B, measurements related to NF performance and data related to its private RAN should not be processed outside of the trust domain of NPN-B. In that case, MDAF related analytics services may need to be orchestrated inside NPN-B domain or in specific infrastructure zones that are considered 'trusted' by the data owner, i.e., owner of NPN-B. Existing network management and orchestration frameworks, such as Network Functions Virtualization – Management and Orchestrations Framework (NFV-MANO) [30] can be leveraged for such flexible and dynamic orchestration and distributed deployment of analytics services that can fulfill data owners' privacy requirements while performing 5G analytics services.

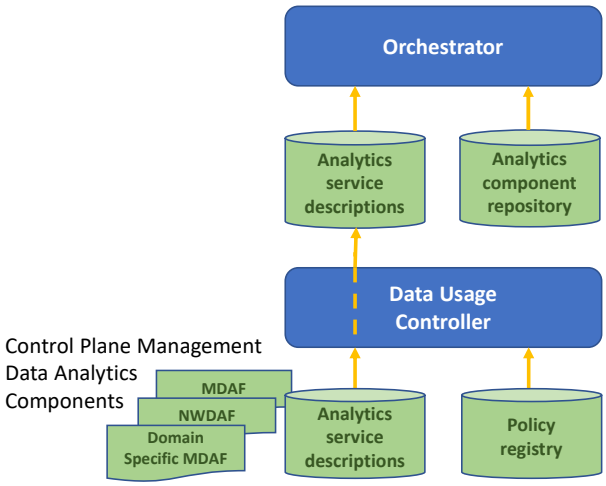

FIGURE 2. Privacy-enhanced management data analytics framework with Data Usage Control

V. SYSTEM ARCHITECTURE

To ensure that the privacy of data belonging to private 5G network owners is not compromised, additional entities and functions are considered in conjunction with the 3GPP specified data analytics framework. Together, these functions can ensure proactive data usage control and distributed policy enforcement as desired by the participating private 5G networks being served by a common public network. A high-level system architecture of the privacy framework introduced in this paper and its interplay with the data analytics components inside the operator network are illustrated in Figure 2. The system components of this privacy-enhanced data analytics (PEDA) framework are described below.

A. Policy Registry

The policy registry keeps track of all the data usage policies specified by data owners, i.e., NPN or private 5G network owners in this case. A high-level policy model, introduced in [27] can be used to describe characteristics of data itself, e.g., data owner, type of data, data consumer, purpose etc., as well as data usage constraints associated with said data, such as control actions that need to be taken for processing of the data. Examples of such control actions can be data anonymity, encryption, TTL constraints for the data after processing, fencing constraints etc. Fencing requires that the data does not leave the owners premises. In the case of private 5G networks, PLMN operators have to ensure private network data does not leave the premises of the network owner and analytics related tasks are instantiated inside the data owner's premises, if needed.

B. Analytics service descriptions

Analytics service descriptions refer to the different analytics process within the scope of 3GPP functions, NWDAF and

TABLE 3. Data analytics services and data usage policies of private 5G networks

Data owner	Analytics services and their scope	Types and sources of data	Data usage policies
NPN-A	Scope: NWDAF Analytics service: NF load analytics Analytics Consumer: another NF or an OAM service	• NF related data (e.g., load, status) from Network repository function (NRF) • NF resource usage and configuration data from operations and maintenance (OAM) services • Other kinds and sources of data as specified in [15]	• Anonymize • TTL
NPN-B	Scope: MDAF Analytics service: Energy saving analysis Analytics service consumer: private 5G owner, and/or PLMN operator	• Performance measurements (e.g., power consumption of network functions, UE throughput, traffic load variation, data volume of UPF, virtual resource usage of the NF, etc.) • MDT reports (e.g., RSRP and RSRQ measurements of UEs, UE location information, etc.) • Network analytics output from NWDAF (e.g., UE communication analysis, NF load analysis, etc.) • Other kinds and sources of data specified in [4]	• Fenced data

MDAF, that can be performed on data collected from private 5G networks. The framework defined in this paper assumes that a service is formed of one or multiple atomic processing tasks. A service might compose multiple tasks in a pipeline or in a more complex topology. Each task can be packaged and run in containers or VMs. In that case, the analytics applications are packaged as software images (either container or VM image), which are contained in the analytics component repository. This kind of modular approach can help the network operators in implementing data usage policies specified by the NPN owners on a granular level and offer flexibility in terms of deploying instances of analytics service according to the constraints. The description of analytics service also identifies the topology of the tasks and the input data for each task of the topology. The description might also contain additional metadata and directives on the service execution.

C. Data Usage controller

Data usage controller plays an important role in enforcing data usage policies described in the policy registry. It can implement pre- and post-processing actions on data sources according to the policies. For example, an NPN owner may have a policy to 'anonymize' their trace data collected from UEs inside their private 5G network. If the data or a subset of it belonging to that owner is required in the processing of an analytics task performed by the management data analytics layer of operator's public network, the data usage controller triggers necessary steps to ensure anonymity. For this purpose, the data usage controller alters the original analytics service description (as defined by the data consumer) to include such pre-processing and post-processing functions.

D. Analytics Components Repository

This repository stores analytics components that are available in the form of images. Orchestration layer can fetch these images for instantiating and deployment based on the requested (altered) description of the analytics service.

E. Orchestrator

The orchestrator is responsible for deploying the end-to-end data analytics service according to the analytics service descriptions. The descriptions interpreted by the orchestrator are already policy compliant due to the actions taken by the data usage controller. The composition of data analytics service depends to the kind of analytics requested from the operator's management data analytics functions, MDAF and/or NWDAF. Analytics components available in the repository are used as building blocks for orchestrating the end-to-end analytics service.

VI. PRIVACY-ENHANCED ORCHESTRATION OF ANALYTICS SERVICES

This section describes a blueprint for privacy-enhanced orchestration of analytics services across public and private 5G networks using NFV-MANO as the orchestration system. Respecting the data usage policies of data owners, i.e., enterprises using private 5G networks, described in section IV, the NFV-MANO system orchestrates analytics services in distributed cloud sites, belonging to private and public 5G networks.

The NFV-MANO framework specified by Industrial Specifications Group (ISG) of European Telecommunication

Standards Institute (ETSI) on Network Functions Virtualization (NFV) provides a comprehensive approach to orchestrating and managing the lifecycle of Network Services (NSs) and Virtual Network Functions (VNFs) in the form of virtualized or containerized workloads on the underlying cloud infrastructure, i.e., Network Functions Virtualization Infrastructure (NFVI). An NS can comprise of one or more VNFs, which are further comprised of VNF components (VNFCs). In the context of NFV, VNFCs form parts of the virtual network function and are deployed as either Virtual Machines (VMs) or Operating System (OS) containers, e.g., a Docker® container, in the NFVI.

To facilitate automated orchestration of NSs and VNFs, NFV-MANO uses descriptors. Both NS and VNF descriptors contain the requisite information required by the NFV-MANO components to orchestrate and instantiate NSs and VNFs. The NFV-MANO performs these tasks with the help of different management entities, e.g., NFV orchestrator (NFVO), VNF manager (VNFM), virtualization infrastructure manager (VIM), container infrastructure service manager (CISM) and CIS cluster manager (CCM) [31]. For example, to orchestrate a network service, NFVO can use the NS descriptor (NSD) which can further contain or reference VNF descriptors (VNFDs) to fetch required information necessary to instantiate the corresponding VNFCs. Information such as software images, required compute, network and storage resources, VM or OS container descriptions, deployment constraints, internal and external topology of VNFs etc. is part of the VNFD. The VNFDs are part of the VNF Package and are used by corresponding entities, e.g., NFVO and VNFM to instantiate VNF instances in the NFVI. The CISM function of NFV-MANO offers functionality to orchestrate workloads as managed container infrastructure objects (MCIOs) on container infrastructure service (CIS) clusters, e.g., Kubernetes clusters. Kubernetes resources such as Pods, DaemonSet, Deployment, Job, ReplicaSet are examples of compute related MCIOs in the context of NFV (full list is available in [32]).

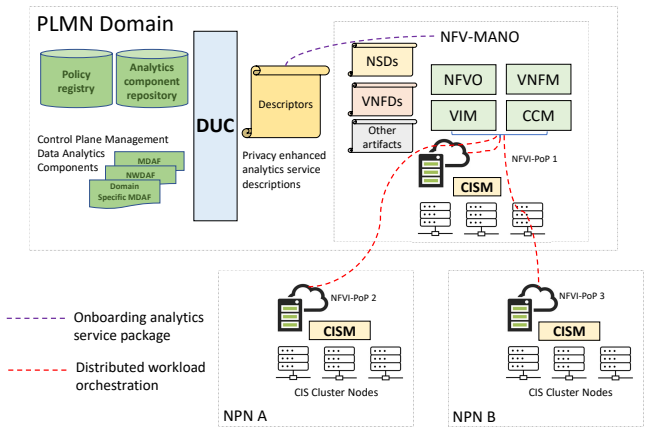

FIGURE 3. NFV-MANO based orchestration and management of analytics functions and services over distributed NFVI-PoPs

The descriptor-based orchestration of network services and functions performed by NFV-MANO can be utilized to orchestrate data analytics services provided by NWDAF and MDAF in the 3GPP framework. The analytics services can be composed as network services to be deployed on distributed NFV infrastructure (NFVI). In our use case, we consider two types of NFVI Points of Presence (NFVI-PoPs), on-premise NFVI-PoPs and off-premise NFVI-PoPs, belonging to infrastructure domains of private 5G networks and the PLMN respectively. The NFV-MANO system being used as the 'orchestrator' as described in section V of this paper, is used to orchestrate and schedule analytics services among different infrastructure domains while ensuring data usage control requirements for each data owner.

In this paper, we consider the case where analytics services are designed to be deployed as OS containers, e.g., Docker containers, in Kubernetes Pods over distributed Kubernetes clusters. Furthermore, we assume that the NFV-MANO is being used as the high-level orchestrator managing network functions and services deployed over different sites, i.e., NFVI-PoPs, belonging to public and private networks. To realize the analytics use case described in section IV, we consider the network topology as shown in Figure 3. Figure 3 illustrates a centralized NFV-MANO system comprising of NS/VNF orchestration and lifecycle management capabilities as well as infrastructure management capabilities managing distributed sites, i.e., NFVI-PoPs. The cloud deployment sites of the PLMN, NPN-A and NPN-B are denoted as NFVI-PoP 1, NFVI-PoP 2 and NFVI-PoP 3 respectively. Each NFVI-PoP comprises of one or more Kubernetes-based CIS clusters with one or more nodes in each CIS cluster. A CISM-like functionality including, e.g., Kubernetes control plane, API server and relevant Kubernetes resources, is managing the OS-container based VNFCs making up VNFs and NSs deployed in the NFVI.

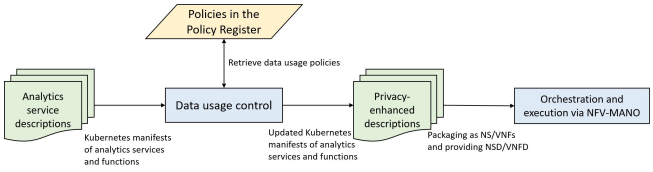

FIGURE 4. Generic process pipeline for orchestration of analytics services using NFV-MANO system as orchestrator

In addition to managing network functions deployed in the public and private 5G networks, the NFV-MANO system can also be utilized to orchestrate analytics related applications and services in the same cloud infrastructure. To meet privacy related requirements for the analytics services, the relevant components from our system architecture interact with the NFV-MANO system to enable 'customized' orchestration of analytics services as special network functions and/or network services (VNFs/NSs) in the NFVI. In that regard, the NFV-MANO system works as the 'orchestrator'

component in the system architecture described in section 3 of this paper. The relevant analytics service descriptions are fetched by the Data Usage Controller (DUC), which then alters these descriptions based on policies in the Policy Register. The DUC can make use of pre-existing components available in the analytics components repository, e.g., data pre-processing related functions like anonymizer, etc.

As described in Section IV of this paper, the NF load analytics service for the NPN-A needs to fulfill two privacy-related requirements, i.e., anonymity and TTL constraints. The analytics service related to NF load analytics need to be altered to offer some pre-processing functions to anonymize the data. Additionally, the orchestrator needs to fulfill the TTL requirements by destroying the relevant instances of analytics services and purging associated data, and re-instantiating them after a specific time interval has passed. In this case, the NF load analytics service of NWDAF is altered by the DUC to fulfill the data anonymity policy of NPN-A. We assume that this NWDAF service is packaged in the form of one or more containers, and the Kubernetes manifests for this service including container images are made available to the DUC for privacy-enabling alterations. During the service description alteration stage, the DUC attaches the pre-processing anonymizer functions to the analytics service descriptions, e.g., in the form of 'init containers' that perform anonymity on NF related data collected from NPN-A network. Furthermore, the DUC alters the service manifests to schedule (re)execution of analytics functions as per the TTL requirement using relevant Kubernetes resources, e.g., CronJobs. The altered descriptions in the form of Kubernetes manifests, e.g., Helm Charts, are provided to the orchestrator (NFV-MANO) to execute the analytics services. The updated Kubernetes manifests for the analytics services also contain deployment constraints for pods, e.g., affinity/anti-affinity constraints.

The Kubernetes control plane via the Kubernetes API supports specifying Pod scheduling constraints within a Kubernetes cluster [33]. However, in the case of multiple Kubernetes clusters, cluster-level isolation can be provided by the Kubernetes cluster API [34] by provisioning separate Kubernetes clusters per tenant for deploying tenants' workloads [35]. Within NFV, this multi-cluster setup and management is covered by the CIS cluster management (CCM) function. The NFVO can support cluster-level selection during instantiation of containerized VNFs by selecting the right existing CIS cluster or creating a new CIS cluster via the CCM function for each tenant to isolate workload deployments for multiple tenants [36]. Furthermore, NFV-MANO system supports specifying affinity/anti-affinity constraints per NFVI-PoP [37]. We extend this same concept and NFV-MANO feature to fulfill the 'fenced data' privacy requirement of NPN-B for performing data analytics. According to the analytics use case considered in section IV, critical data related to NPN-B network functions must not be taken outside of the trust zone of NPN-B. In this case, the DUC updates the manifests of relevant analytics service descriptions with the right affinity/anti-affinity constraints. In NFV-MANO, these constraints can be set for each MCIO (e.g., Pod) on multiple levels e.g., NFVI-PoP level, NFVI resource zone level, CIS cluster level, CIS node level etc [37]. Using the constraints specified in the VNFD, the NFV-MANO system orchestrates the 'Energy saving analysis' analytics service in the Kubernetes cluster that is deployed in NPN-B's NFVI-PoP 2.

For both NPN-A and NPN-B, the process pipeline for orchestrating the analytics services as per data privacy policies for each private 5G network can be generalized in the flow shown in 4.

A. Orchestration workflow for privacy-enhanced analytics services

This section discusses the workflow for privacy enhanced orchestration of 5G analytics services, including the interactions that take place between NFV-MANO components and those of 3GPP data analytics framework. The workflow is not to be considered as a standardized way of interaction between the two frameworks. It's only purpose is give an example of how the proposed data usage control approach can be implemented for orchestrating data analytics using NFV-MANO. For the scope of this workflow, we refer to the use cases introduced in Section IV and in particular the second example in Table 3. In this example, the PLMN operator needs to perform MDAF analytics on data coming from the NPN-B private network, data on which the NPN-B private owner has a 'fenced data' usage policy, i.e., the data owner mandates its data to be processed only within its administrative and/or trust domain. Figure 5 illustrates the step-by-step workflow for the interactions between the relevant components in such a scenario. The steps that take place in this flow are described below:

1) The request for analytics arrives at the MDAF or NWDAF within the PLMN domain, requesting desired analytics for the NPN-B network. The way such request can arrive according to 3GPP specifications (e.g., Subscription-based or Request-based [38]) is out of the scope of this paper.
2) The MDAF/NWDAF determines which data are needed from the NPN-B network. The requested analytics service is packaged and described in the form of NSD, that further contains VNFDs for each analytics application deployed as a VNFC (e.g., a VM or a Pod). Through the PLMN control plane, orchestration of the Network Service (NS) to perform the analytics is requested from the NFV-MANO. NSD and VNFD(s) are used for the NS on-boarding and instantiation as per NFV-MANO procedures, the exact details of those are outside the scope of this paper.
3) The Data Usage Controller (as described in Section V and depicted in Figure 2) examines the descriptors to check for the compliance of deployment constraints

This article has been accepted for publication in IEEE Open Journal of the Communications Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2024.3522379

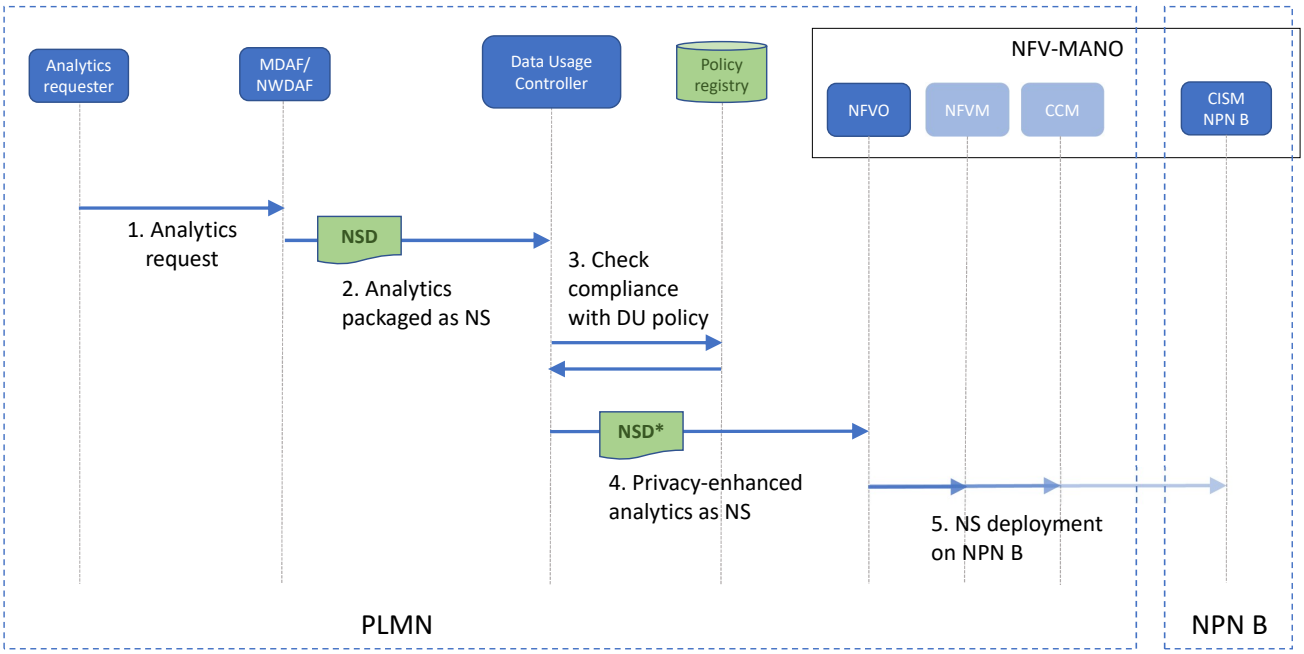

FIGURE 5. **Workflow example of a privacy-enhanced analytics network service deployed in NPN-B**

with NPN-B data usage policies. At this step, the Data Usage Controller interacts with the Policy Registry to retrieve data usage policies.

4) The Data Usage Controller updates the descriptors to align with the data usage polices. In our specific example, the Data Usage Controller adds necessary deployment constraints in the descriptors for the components of the analytics service (packaged as NS) to deployed in the NPN-B's NFVI-PoP. The adjusted descriptor NSD* is forwarded to the NFVO.
5) The NFVO takes the updated descriptor NSD* and starts orchestrating the NS deployment. The NFVO analyzes the NSD, and identifies that the NPN-B's NFVI-PoP has the necessary resources to deploy the requested service. The NFVO in turn interacts with other NFV-MANO functions, e.g., the VIM, the CCM and other relevant NFV-MANO components. The detailed interactions among the NFV-MANO components are out of scope of this example.

VII. Experimental Evaluation and Comparative Analysis

This section provides evaluation of the PEDA framework as well as comparative analysis with some existing relevant solutions. We demonstrate the performance of privacy-enhanced orchestration for one of the scenarios highlighted in our analytics use case described in section IV. We focus on NPN-B's data usage policy of 'fenced data' that it mandates to be respected when PLMN performs any kind of data analytics, whether in the scope of NWDAF or MDAF, on data that belongs to NPN-B. In that regard, we create two experimental setups based on minikube and Google Cloud and demonstrate our orchestration approach. Furthermore, we evaluate the 'scalability aspects' of configurations and their automation as the number of private 5G networks and their data usage policies increase. To this end, we perform an analytical comparison of PEDA framework with two existing alternative solutions.

A. Demonstration of Proof of Concept

This section demonstrates a proof of concept, which is described in section VI in the form of blueprint and workflow for privacy-enhanced orchestration of data analytics services. Two experimental setups are laid out to replicate cloud infrastructure sites of PLMN, NPN-A and NPN-B on which data analytics services can be orchestrated according to data usage policies specified by data owners. Figure 6 provides a simplified view of privacy-enhanced orchestration blueprint depicted in Figure 3 tailored to our experimental setup. For simplicity, in Figure 6 data usage control (DUC) components, e.g., policy registry, service descriptions, DUC controller, have all been contained within a single DUC block.

Following the standard orchestration process of NFV-MANO for containerized VNFs - composing the NWDAF and MDAF analytics services in our case - the NFVO interacts with the CISM components in different CIS Clusters that are under its scope of management. As previously mentioned, a CIS cluster is essentially a Kubernetes cluster, whereas CISM refers to Kubernetes control plane (CP). In our experimental setup, we deploy separate Kubernetes clusters for PLMN, NPN-A and NPN-B, each indicating

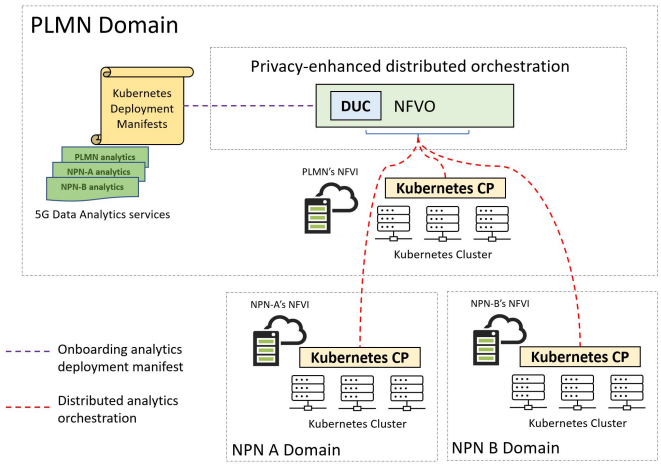

FIGURE 6. **Demonstration setup for privacy enhanced distributed orchestration**

the respective domains or trust zones of corresponding public and private networks. These clusters are set up using minikube[2] on a single host machine. To demonstrate privacy-enhanced orchestration, we emulate the role of NFVO orchestrating analytics services as network services on multiple Kubernetes clusters in a distributed fashion. We implement the functionality of DUC components that update the Kubernetes manifests of analytics services as per the data usage policies specified by each data owner in the policy registry. NFVO takes the updated, privacy-enhanced manifests from DUC and selects the right Kubernetes cluster for deploying analytics services by interacting with the Kubernetes CP, i.e., Kubernetes API server, of that cluster. This combined 'DUC + NFVO' functionality is implemented in the form of a Python application.

We focus on the data usage policy of 'fenced data' in our experimental evaluations to observe if the policy is being enforced by DUC-enabled, privacy-enhanced distributed orchestration. To this end, we orchestrate a number of different analytics services aimed to perform different kinds of NWDAF and MDAF analytics related to PLMN, NPN-A and NPN-B. These services are submitted to the combined 'DUC + NFVO' component in the form of Kubernetes deployment manifests, as shown in Figure 6. The component orchestrates necessary service components, i.e., Kubernetes pods, as per specified 'fenced data' policy. In case there is no 'fenced data' policy specified, the combined 'DUC + NFVO' component orchestrates services either within the PLMN domain or in the domain of corresponding private 5G network. Further details of the multi-cluster experimental setup are provided in Table 4.

Manifests for 15 distinct analytics services are created for the experiment, five services for each network domain. These services are assumed to be used for performing various kinds of MDAF and NWDAF related analytics as

[2]https://minikube.sigs.k8s.io/

TABLE 4. **Details and configurational parameters for multi-cluster experiment**

Components	Details and description
Kubernetes Platform	minikube version v1.34.0
Host Machine	OS: Ubuntu 20.04.6 LTS, Kernel: Linux 5.4.0-200-generic, Architecture: x86-64
minikube VM Driver	Docker
Container Runtime	docker://27.2.0
Kubernetes Clusters	3 (one per domain)
Cluster Details	Single node containing control plane, workload scheduling enabled
Node properties	CPUs = 2, Memory = 32100 MB
Kubernetes Version	v1.31.0
Fenced Data Policies	PLMN (fenced), NPN-A (non-fenced), NPN-B (fenced)
Analytics Services	15 Kubernetes Deployments (5 for each network)
Pods per deployment	2

per different analytics use cases in 5G and B5G networks. Three policies are specified in the policy registry, one for each network's domain. We specify the 'fenced data' policy for NPN-B as per our analytics use case described in section IV. Furthermore, we also specify the 'fenced data' policy for PLMN to avoid orchestration of PLMN's management domain analytics on NFVI-PoPs belonging to NPN-A and NPN-B.

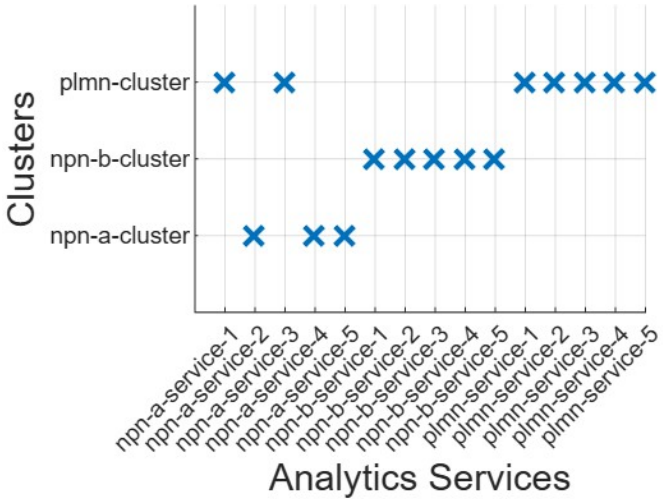

FIGURE 7. **Deployment of analytics services on across different minikube clusters as per DUC 'fenced data' policy**

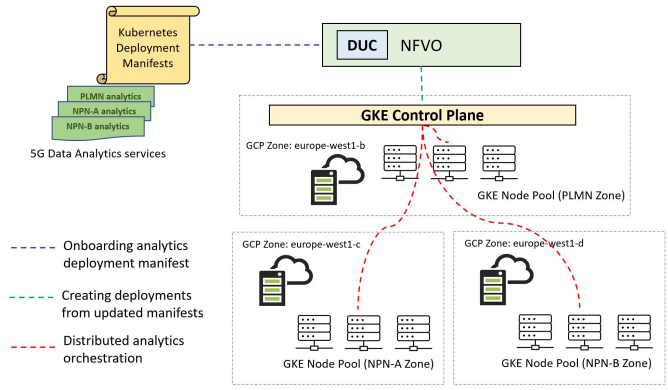

FIGURE 8. GKE setup for privacy enhanced orchestration of analytics services

Figure 7 demonstrates the efficacy of our DUC component when scheduling analytics services as per the specified 'fenced data' policies. Each plotted point on the graph represents all instances (i.e., pods) of the corresponding analytics service that are running on one of the cluster nodes. Analytics services targeted for NPN-B are only scheduled for execution on the NPN-B cluster, presumably residing in NPN-B's own infrastructure domain. Similarly, the fenced data policy of PLMN is respected by orchestrating its analytics services in the PLMN's domain. Since NPN-A has no 'fenced data' policy, its analytics services are distributed among its own cluster and that of the operator, i.e., PLMN. Our 'DUC + NFVO' orchestrator component is capable of orchestrating analytics services across multiple clusters as per the privacy constraints that are applicable to corresponding data. There are some existing frameworks for orchestration of workloads across multiple Kubernetes clusters, one such example being the mck8s framework [39]. However, mck8s distributes workloads across geographically distributed clusters with the objective of optimal resource allocation. Privacy preservation or data usage control aspects are not considered in the mck8s orchestration platform.

The analytics services used for this proof of concept (PoC) demonstration are not actual NWDAF or MDAF analytics services, rather 'dummy' analytics services made up of generic containerized applications in the form of Kubernetes deployments inline with cloud-native and micro-services architecture. Complex analytics services, composed of multiple micro-services, can be set up using service meshes. There are various tools for available for this, Istio[3] being one of them. However, the actual nature of analytics services, their implementation and complex compositions are all those aspects that lie outside the scope of this paper. Moreover, we chose to work on the granularity level of Kubernetes deployments for ensuring privacy so that the concept can be extended to complex analytics services that are made up of micro-services in the form of service meshes. Further evaluation of our framework on an even finer grained granularity is part of future work, where we can evaluate enforcement of other data usage policies like anonymity, TTL constraints, etc., on the actual data being used by some analytics services running in the cloud.

In a different experimental setup, we evaluate our PEDA framework by demonstrating privacy-enhanced distributed orchestration using public cloud. Google Cloud Platform (GCP) is used to create a single Google Kubernetes Engine (GKE) cluster distributed over different geographical locations, referred to as zones hereinafter. Each zone represents a different geographical location offered by the GCP, and node pools are created for each network domain, i.e., PLMN, NPN-A and NPN-B, in different zones. In this way, we try to replicate our reference scenario of different public and private 5G networks that we consider for the PEDA framework in this paper. Figure 8 illustrates this particular deployment setup, containing a single Kubernetes cluster comprising of Kubernetes CP (managed by Google) and worker nodes belonging to node pools in different zones, each zone representing a distinct network's domain. For the sake of simplicity, we have assigned one node for each network domain. Further details about the GKE setup are provided in Table 5.

The single cluster setup shown in Figure 8 also covers the scenario when the PLMN may host private 5G networks as network slices as described in Section II. These network slices can potentially be deployed, either partially or fully, in the public cloud by the PLMN for some use cases [40] [41]. From data privacy, secrecy and regulatory point of view, public cloud clusters can be configured appropriately to schedule workloads in certain geographical zones. Furthermore, specific node pools can be set up comprising of custom nodes with specialized hardware like trusted execution environments (TEEs) for offering privacy and confidentiality in the underlying hardware layer as well.

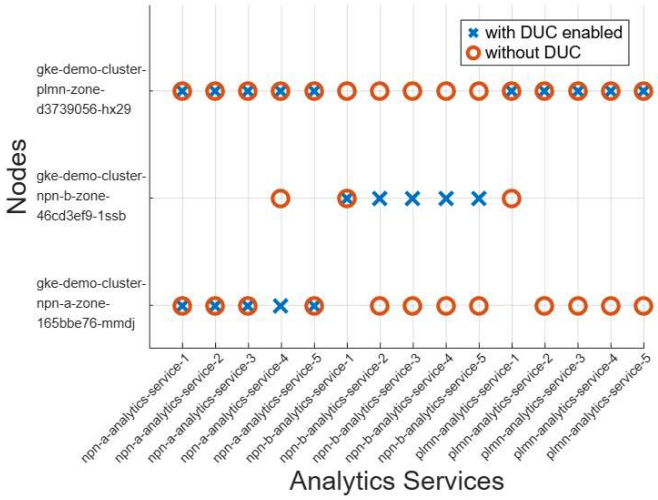

FIGURE 9. Deployment of analytics services on GKE nodes with and without DUC

[3]https://istio.io/latest/

TABLE 5. Details and configurational parameters for the GKE setup

Components	Details and description
Kubernetes Platform	Google Kubernetes Engine (GKE)
GKE Cluster Type	Zonal
Control plane zone	europe-west1-b
Default node zones	europe-west1-b, europe-west1-c, europe-west1-d
Nodes	3
Kubernetes Version	1.30.5-gke.1443001
Container Runtime	containerd://1.7.22
Operating System	OS: Ubuntu 22.04.5 LTS, Kernel: 5.15.0-1067-gke
Nodes per domain	PLMN (1 node in zone europe-west1-b), NPN-A (1 node in zone europe-west1-c), NPN-B (1 node in zone europe-west1-d)
Node properties	Machine type = e2-medium, Allocatable CPUs = 940 mCPU, Allocatable Memory= 2.92 GB
Fenced Data Policies	PLMN(fenced), NPN-A(non-fenced), NPN-B(fenced)
Analytics Services	15 Kubernetes Deployments (5 for each network)
Pods per deployment	2

The results from the GKE experiment are shown in Figure 9, showcasing the scheduling of different analytics services across the cluster nodes. There are two pods for each analytics service and each plotted point on the graph represents a pod of corresponding analytics service that is running on one of the cluster nodes. We chose the deployment granularity on the x-axis instead of pods to avoid the figure being too cluttered. This may result in multiple pods (maximum two in our case) being represented by a single point, especially in the case when DUC is enabled and pods of PLMN and NPN-B analytics services are running on the right nodes dedicated to those networks. It is evident from the figure that When DUC is enabled to enforce the 'fenced data' policies of NPN-B and PLMN, pods belonging to the analytics services concerning those network domains are only scheduled on the nodes that are dedicated to the corresponding domains. In the absence of DUC, default Kubernetes scheduling takes place across the cluster, resulting in orchestration of some NPN-B analytics services pods on nodes assigned to PLMN and NPN-A. Similarly, the 'fenced data' policy of PLMN is also not respected by the normal Kubernetes scheduler as pods of PLMN analytics services can be seen running on NPN-A and NPN-B nodes.

KubeFlower [42] is another framework that considers privacy of data when it is used for federated learning in single cluster environments with geo-distributed sites (nodes) across the IoT-edge-cloud continuum. According to our understanding, the kubFlower framework focuses strictly on federated learning applications, including in the telecommunication networks. In this paper, we aim to introduce a framework that is generic and can be applied on varying kinds of analytics applications in 5G and future 6G mobile networks. For our future works, we aim to further enhance the interplay between Kubernetes cluster components and the DUC components with the introduction of custom Kubernetes resources and specialized Kubernetes operators. This kind of Kubernetes-native approach would be necessary to implement fine-grained data usage policies like anonymity, TTL constrains, data isolation in the compute, networking and storage layers of the data plane, going beyond the scheduling aspects of control plane that we have considered so far in our PoC demonstration. Authors in [42] have also followed a similar Kubernetes-native approach for the kubeFlower framework.

B. Evaluation of configuration automation

We evaluate our proposed privacy-enhanced data analytics (PEDA) framework in terms of its configuration automation capabilities. Automation of configuration is a key KPI in mobile networks, as it significantly reduces operational expenditures (OPEX) for operators by minimizing the need for continuous network configuration adjustments. Specifically, we compare our solution with two existing solution, namely LUCON [28] and Dataspace Connector (DSC)[29]. The first is a flow-oriented approach, meaning that it specifies policies for individual data flows based on protocol, source, destination, and other parameters. Similarly, the second one defines a set of policies on data from a specific data owner and enforces them through an ad-hoc connector for every telemetry data exchange. Both solutions are described in Section III-C. In contrast, the proposed PEDA framework follows an intent-based approach, therefore reducing the complexity in terms of configuration due to choice of appropriate granularity.

Figure 10 depicts the reference scenario we consider for our comparison. Two private networks are considered (NPN-A and NPN-B) and one analytics service is assumed to be deployed in the PLMN domain. Assuming the private networks have a 'fenced data' policy as described in Table 2 (i.e., each private network limits the processing of its own data within its local domain), then the corresponding service needs to be allocated locally in each of the private networks' domains. With our solution, applying the 'fenced data' policies translates in modifying the descriptors related to the NS instantiation to be compliant with the policies from the Policy Registry, as described in Section VI-A.

This article has been accepted for publication in IEEE Open Journal of the Communications Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2024.3522379

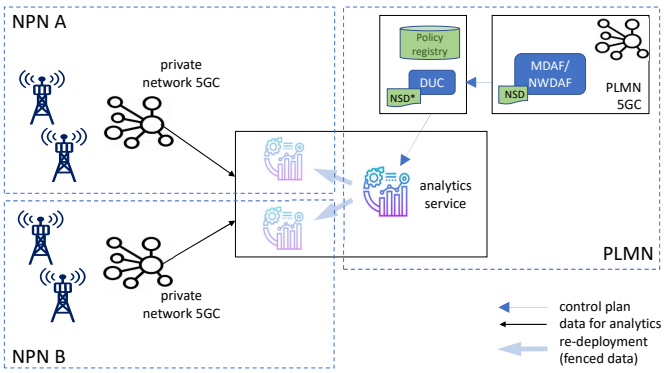

FIGURE 10. **Two NPNs reference setup for solution evaluation**

In such a scenario, we compare the number of configurations needed by the PEDA solution and alternate solutions, LUCON and DSC. Table 6 shows the results in terms of number of policies and function descriptors that need to be configured for each of the solutions while the number of private networks increases.

Both PEDA and DSC always require one policy for each private network. Therefore, in the general case (last column of the table) of n private networks, they will require n policies to be configured. Differently, LUCON is a flow-based approach that requires policies to be configured for each service instance. Therefore, if there is only one service as in Figure 10, the number of policies needed will be equivalent to PEDA and DSC cases. However, for s number of services, the number of policy configuration will be $s*n$, where again n is the number of private networks.

Similarly, when it comes to the number of function descriptors needed to deploy a service instance, PEDA requires one function descriptor for all the service instances to be deployed. In the case of s services it will always require s function configurations independently from the number of private networks n. Differently, both LUCON and DSC require one different function descriptor to be configured for each of the private networks domain in which the service need to be deployed due to the 'fenced data' policy. This results in $s * n$ configurations when considering s services and n private networks.

TABLE 6. **Operational configurations comparison between PEDA, LUCON [28], and DSC [29]**

private networks (n)		1	2	3	...	n
policies	PEDA	1	2	3	..	n
	LUCON	s	2*s	3*s	..	n*s
	DSC	1	2	3	..	n
functions	PEDA	s	s	s	..	s
	LUCON	s	2*s	3*s	..	n*s
	DSC	s	2*s	3*s	..	n*s

n is the number of private networks, s is the number of services.

The results described in Table 6 can also be observed in Figure 11. As the number of private networks increases, the LUCON (for both policies and function descriptors) and DSC (only for function descriptors) solutions need a higher number of configuration to be set, implying difficulty in implementing automation of these solution over larger scale. For example, in the case of 3 private networks ($n = 3$) and 5 analytics services ($s = 5$), LUCON will require five times the number of configurations as required by PEDA.

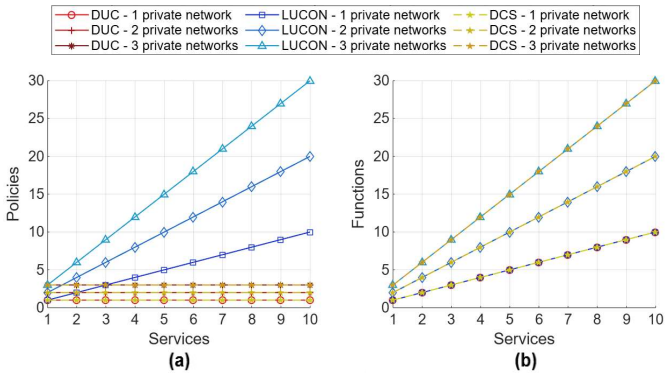

FIGURE 11. **Number of policy (a) and function descriptors (b) configuration needed with PEDA, LUCON [28] and DSC [29], for $n = 1$, $n = 2$ and $n = 3$ private networks**

VIII. CONCLUSION

In this paper, we highlight the data privacy and secrecy issues that may arise for 5G data analytics, especially when the private 5G networks are either hosted completely by the public network (e.g., in the form of network slices) or are sharing the operator's management and control plane and its associated analytics services. In that regard, we consider the data usage control approach enabling data owners to specify data usage policies that must be fulfilled by the operator while performing data analytics. Furthermore, we provide a detailed analytics use case describing different analytics services, kinds of data required for those analytics and the data usage policies that can be set by the data owners.

Additionally, we propose a novel framework for privacy-enhanced data analytics (PEDA) consisting of components that can complement the standard 5G data analytics framework specified by 3GPP. The PEDA framework makes use of existing orchestration systems like NFV-MANO, specified by ETSI ISG NFV, to perform data usage control for the data exchange between private and public 5G networks. To this end, a detailed blueprint for distributed orchestration of analytics services in the cloud infrastructure of both public and private 5G networks is introduced in this paper. Privacy-enhanced orchestration of analytics services enables enforcement of data usage policies set by private 5G network owners for the analyses that are performed on their data by the public network's control plane. Detailed workflow of this orchestration blueprint covering step-by-step interactions between all the components is also provided.

We have also demonstrated a PoC of privacy-enhanced orchestration approach proposed in this paper. Private and public 5G networks are replicated in the form of distributed cloud infrastructure clusters, on which the orchestrator deploys analytics services in a distributed manner while taking necessary data usage control measures for data privacy. We evaluated the efficacy of privacy-enhanced orchestration of analytics services while considering the 'fenced data' policies of data owners. To this end, experiments have been performed on multi-cluster and single cluster setups. In our evaluation, we focused on Kubernetes deployment granularity to ensure privacy, enabling extension to complex analytics services built as micro-service meshes. Our future work includes exploring finer-grained enforcement of data usage policies, such as anonymity and TTL constraints, on data used by cloud-based analytics services. Furthermore, we plan to enhance the integration of Kubernetes and DUC components using custom resources and operators. This Kubernetes-native approach will enable fine-grained data usage policies, such as anonymity, TTL constraints, and data isolation across compute, networking, and storage layers, extending beyond the control plane focus of our current PoC.

REFERENCES

[1] M. Wen, Q. Li, K. J. Kim, D. López-Pérez, O. A. Dobre, H. V. Poor, P. Popovski, and T. A. Tsiftsis, "Private 5g networks: Concepts, architectures, and research landscape," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 1, pp. 7–25, 2021.

[2] S. Ye, P. Pan, B. Bai, X. Zheng, Y. Huang, and G. Jia, "Research and application of 5g-based smart airport terminal security access," in *International Conference on Computer Application and Information Security (ICCAIS 2023)*, vol. 13090, pp. 1133–1137, SPIE, 2024.

[3] 5G Alliance for Connected Industries and Automation (5G ACIA), "Exposure of 5G Capabilities for Connected Industries and Automation Applications." online: https://perma.cc/3YCR-GD77. White Paper, 2021.

[4] 3GPP TSG SA5, "3GPP TS 28.104; Technical Specification Group Services and System Aspects; Management and orchestration; Management Data Analytics (MDA) (Release 18) V18.3.0 ." May. 2024.

[5] 3GPP TSG SA5, "3GPP TR 28.809; Technical Specification Group Services and System Aspects; Management and Orchestration; Study on Enhancement of Management Data Analytics (MDA); (Release 17); V17.0.0 ." Mar. 2021.

[6] S. Eswaran and P. Honnavalli, "Private 5g networks: a survey on enabling technologies, deployment models, use cases and research directions," *Telecommunication Systems*, vol. 82, no. 1, pp. 3–26, 2023.

[7] 3GPP TSG SA, "3GPP TR 21.916; Technical Specification Group Services and System Aspects; Release 16 Description; Summary of Rel-16 Work Items (Release 16) V16.2.0 ." Jun. 2022.

[8] 3GPP TSG SA5, "3GPP TS 23.501; Technical Specification Group Services and System Aspects; Management and orchestration; MSystem architecture for the 5G System (5GS) (Release 19) V19.0.0 ." Jun. 2024.

[9] 3GPP TS SA1, "3GPP TR 28.809; Technical Specification Group Services and System Aspects; Management and Orchestration; Study on Enhancement of Management Data Analytics (MDA); (Release 17); V19.7.0 ." Jun. 2024.

[10] Ericsson, "Unlocking the Smart Factory: Why 5G Private Networks are Essential for Autonomous Things." online: https://www.ericsson.com/en/blog/2024/7/unlocking-the-smart-factory-why-5g-private-networks, 2024. Accessed: Nov 2024.

[11] J. Harmatos and M. Maliosz, "Architecture integration of 5g networks and time-sensitive networking with edge computing for smart manufacturing," *Electronics*, vol. 10, no. 24, p. 3085, 2021.

[12] J. Sachs and K. Landernäs, "Review of 5g capabilities for smart manufacturing," in *2021 17th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–6, IEEE, 2021.

[13] 5G Alliance for Connected Industries and Automation (5G ACIA), "5G Non-Public Networks for Industrial Scenarios." online: https://5g-acia.org/whitepapers/5g-non-public-networks-for-industrial-scenarios/, 2021. White Paper, Accessed: August 23, 2024.

[14] F. Maggi, M. Balduzzi, R. Vosseler, M. Rösler, W. Quadrini, G. Tavola, M. Pogliani, D. Quarta, and S. Zanero, "Smart factory security: A case study on a modular smart manufacturing system," *Procedia Computer Science*, vol. 180, pp. 666–675, 2021.

[15] 3GPP TSG SA5, "3GPP TS 28.288; Technical Specification Group Services and System Aspects; Architecture enhancements for 5G System (5GS) to support network data analytics services (Release 18) v18.6.0." Jun. 2024.

[16] A. Narayanan, M. S. Korium, D. C. Melgarejo, H. M. Hussain, A. S. De Sena, P. E. G. Silva, D. Gutierrez-Rojas, M. Ullah, A. E. Nezhad, M. Rasti, *et al.*, "Collective intelligence using 5g: Concepts, applications, and challenges in sociotechnical environments," *IEEE Access*, vol. 10, pp. 70394–70417, 2022.

[17] G. Yilma, U. Fattore, M. Liebsch, N. Slamnik, A. Heider-Aviet, and J. Marquez-Barja, "5g automec-boosting edge-to-edge service continuity for cam in a sliced network," 2021.

[18] A. El Sayed, M. Ruiz, H. Harb, and L. Velasco, "Deep learning-based adaptive compression and anomaly detection for smart b5g use cases operation," *Sensors*, vol. 23, no. 2, p. 1043, 2023.

[19] Y. Zhu and S. Wang, "Joint traffic prediction and base station sleeping for energy saving in cellular networks," in *ICC 2021-IEEE International Conference on Communications*, pp. 1–6, IEEE, 2021.

[20] 3GPP TSG CT3, "3GPP TS 29.520; Technical Specification Group Core Network and Terminals; 5G System; Network Data Analytics Services; Stage 3 V19.0.0." Sept. 2024.

[21] 3GPP TSG SA5, "3GPP TS 28.531; Technical Specification Group Services and System Aspects; Management and orchestration; Provisioning; (Release 18) V18.5.0 ." Mar. 2024.

[22] Radcom, "Paving the way to autonomous 5g networks." online: https://radcom.com/paving-the-way-to-autonomous-5g-networks/, 2024. Accessed: November 2024.

[23] A. Lazouski, F. Martinelli, and P. Mori, "Usage control in computer security: A survey," *Computer Science Review*, vol. 4, no. 2, pp. 81–99, 2010.

[24] W. Ku and C.-H. Chi, "Survey on the technological aspects of digital rights management," in *International Conference on Information Security*, pp. 391–403, Springer, 2004.

[25] A. Eitel, C. Jung, R. Brandstädter, A. Hosseinzadeh, S. Bader, C. Kühnle, P. Birnstill, G. Brost, M. Gall, F. Bruckner, *et al.*, "Usage control in the international data spaces," *Aufl. IDS Association, Berlin*, 2021.

[26] C. Jung and J. Dörr, "Data usage control," in *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (B. Otto, M. ten Hompel, and S. Wrobel, eds.), ch. 8, Springer, 2022.

[27] F. Cirillo, B. Cheng, R. Porcellana, M. Russo, G. Solmaz, H. Sakamoto, and S. P. Romano, "Intentkeeper: Intent-oriented data usage control for federated data analytics," in *2020 IEEE 45th Conference on Local Computer Networks (LCN)*, pp. 204–215, IEEE, 2020.

[28] J. Schütte and G. S. Brost, "Lucon: Data flow control for message-based iot systems," in *2018 17th IEEE international conference on trust, security and privacy in computing and communications/12th IEEE international conference on big data science and engineering (TrustCom/BigDataSE)*, pp. 289–299, IEEE, 2018.

[29] B. Shariati, H. Qarawlus, S. Biehs, J.-J. Pedreno-Manresa, P. Safari, M. Balanici, A. Bouchedoub, H. Haße, A. Autenrieth, J. K. Fischer, *et al.*, "Telemetry framework with data sovereignty features," in *Optical Fiber Communication Conference*, pp. M3G–2, Optica Publishing Group, 2023.

[30] ETSI, "ETSI GR NFV-MAN 001; Network Functions Virtualisation (NFV); Management and Orchestration; Report on Management and Orchestration Framework." Sophia Antipolis, France, Dec. 2021.

[31] ETSI, "ETSI GS NFV 006; Network Functions Virtualisation (NFV) Release 4; Management and Orchestration; Architectural Framework Specification." Sophia Antipolis, France, May 2024.

[32] ETSI, "ETSI GS NFV-SOL 018; Network Functions Virtualisation (NFV) Release 5; Protocols and Data Models; Profiling specification

of protocol and data model solutions for OS Container management and orchestration." Sophia Antipolis, France, July 2024.
[33] Cloud Native Computing Foundation (CNCF), "Kubernetes API Reference Documentation; Resource Pod." online: https://perma.cc/9YWV-KRG3, 2024. Accessed: November 2024.
[34] Cloud Native Computing Foundation (CNCF), "Kubernetes Cluster API." online: https://perma.cc/E4VK-F2QH, 2024. Accessed: November 2024.
[35] Cloud Native Computing Foundation (CNCF), "Kubernetes Documentation; Multi-tenancy." online: https://perma.cc/AZU3-C3VM, 2024. Accessed: November 2024.
[36] ETSI, "ETSI GR NFV-EVE 018; Network Functions Virtualisation (NFV) Release 5; Evolution and Ecosystem; Report on Multi-tenancy in NFV." Sophia Antipolis, France, May 2024.
[37] ETSI, "ETSI GS NFV-IFA 011; Network Functions Virtualisation (NFV) Release 5; Management and Orchestration; VNF Descriptor and Packaging Specification ." Sophia Antipolis, France, June 2024.
[38] 3GPP TSG CT4, "3GPP TS 29.500; Technical Specification Group Core Network and Terminals; 5G System; Technical Realization of Service Based Architecture; Stage 3; V19.0.0 ." Sept. 2024.
[39] M. A. Tamiru, G. Pierre, J. Tordsson, and E. Elmroth, "mck8s: An orchestration platform for geo-distributed multi-cluster environments," in *2021 International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–10, 2021.
[40] M. Dalgitsis, N. Cadenelli, M. A. Serrano, N. Bartzoudis, L. Alonso, and A. Antonopoulos, "Cloud-native orchestration framework for network slice federation across administrative domains in 5g/6g mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 7, pp. 9306–9319, 2024.
[41] S. Arora, A. Ksentini, and C. Bonnet, "Cloud native lightweight slice orchestration (cliso) framework," *Computer Communications*, vol. 213, pp. 1–12, 2024.
[42] J. M. Parra-Ullauri, H. Madhukumar, A.-C. Nicolaescu, X. Zhang, A. Bravalheri, R. Hussain, X. Vasilakos, R. Nejabati, and D. Simeonidou, "kubeflower: A privacy-preserving framework for kubernetes-based federated learning in cloud–edge environments," *Future Generation Computer Systems*, vol. 157, pp. 558–572, 2024.

HAMMAD ZAFAR is a Senior Standardization Engineer at NEC Laboratories Europe in Heidelberg, Germany. His current research is focused on using standard orchestration and management frameworks to achieve energy-efficiency, performance optimization, trust and privacy in communication networks. He is an active delegate at the ETSI ISG NFV standards organization and has been involved in the development of NFV standards since 2020, and is currently Feature Prime and Rapporteur for multiple topics and work items in ETSI ISG NFV. He received his Masters in Electronic Engineering in 2018 and Bachelors in Electrical Engineering in 2015.

UMBERTO FATTORE (Member, IEEE) received both his B.Sc. and M.Sc. degree in Computer Engineering from University of Naples "Federico II", Italy, in 2018. Since then, he works for NEC Laboratories Europe GmbH, first as Early-Stage Researcher (ESR) for the EU SPOTLIGHT project - focusing on 5G mobile core flow optimization, then as Researcher for the EU 5G-CARMEN project on cross-border seamless 5G connectivity. From 2021, he is Standardization Engineer in NEC, mainly involved in the recently established IOWN Global Forum and in the ETSI NFV ISG, but also following several other groups' activities in ETSI, 3GPP, and IETF.

FLAVIO CIRILLO (Member, IEEE) is a Senior Research Scientist of the Data Ecosystem and Standards research team at NEC Laboratories Europe, Heidelberg (DE). His research topics include AI/ML and data platforms for the Digital Twin. He has worked on many IoT and Digital Twin related European research projects in the domain of smart city, as well as in open source research projects part of FIWARE Foundation. He is currently involved in Digital Twin standardization activities in FIWARE, as responsible for the Processing and Visualization chapter of the FIWARE Technical Steering Committee, and in the Innovative Optical and Wireless Networks (IOWN) Global Forum, as leader of the Data Space for Digital Twin Applications task force and of the Network Digital Twin activities.

CARLOS J. BERNARDOS received a Telecommunication Engineering degree in 2003, and a PhD in Telematics in 2006, both from the University Carlos III of Madrid, where he works as Professor. His research interests include 6G, deterministic networking, integrated sensing and communications, wireless mobility management, network virtualization and experimental evaluation of mobile wireless networks. He has published over 100 scientific papers in international journals and conferences. He is an active contributor to IETF since 2005.