AI-Driven Closed-Loop Service Assurance with Service Exposures

Min Xie*, Joan S. Pujol-Roig†, Foivos Michelinakis‡, Thomas Dreibholz‡, Carmen Guerrero§, Adrian Gallego Sanchez§, Wint Yi Poe¶, Yue Wang†, Ahmed M. Elmokashfi‡
*Telenor Research, Telenor, Norway; †Samsung Electronics R&D Institute, UK; §Universidad Carlos III de Madrid; ‡Simula Metropolitan Centre for Digital Engineering, Norway; ¶Huawei Technologies Düsseldorf GmbH, Germany

***Abstract*—Artificial Intelligence (AI) is widely applied in mobile and wireless networks to enhance network operation and service management. Advanced AI mechanisms often require high level of network service exposure in order to access data from as many network elements as possible and execute the AI recommended outcomes into the networks. However, in practice, it is not always feasible to expose the network services to 3rd parties or customers with AI ambitions. Considering that service assurance (SA) is a major area to which AI is applied, this paper describes how a closed-loop SA architecture is associated with the service exposure model in the 5G networks with network slicing. Then we investigate the impact and implication of service exposure on SA. Finally, a set of experiment results are provided to demonstrate the trade-off relationship between the AI ambition and the exposure level in SA.**

I. Introduction

5G aims to provide high-quality network services to vertical customers with diverse requirements. Provisioning such services is expected to be customized and cost-efficient, which brings a big challenge for service assurance (SA). Particularly, in network slicing with network function virtualization (NFV) and software defined network (SDN), virtualization and softwarization result in more dynamic network environments and service abstraction. SA needs to adapt to the dynamic network conditions and service requirements in a scalable way. Moreover, as a network slice is composed of several types of network elements (NEs) like resources, network functions (NFs) and network services (NSs), the generated data is extremely versatile and complex data in type, content, and volume. Consequently, it is not realistic to assure services in a manual way and and SA automation becomes necessary in 5G networks. Artificial intelligence (AI) is one of the enablers for SA automation.

SA can be implemented by network operators or communication service providers (CSP) inside the network or service management domains. However, network built-in SA is usually designed to deal with generic and fundamental problems and is thus not sufficient to meet 5G requirements on quality of experience (QoE) assurance given the high diversity of problems raised by vertical customers from different industries. More advanced SA is expected, *e.g.*, by including special SA services from authorized customers or 3rd parties. As a matter of fact, to handle the complexity of data generated by NEs of network slices (*e.g.*, infrastructure data vs. NF/NS data, control data vs user data, traffic and topology data vs. performance data, events vs. system logs), more powerful analytics tools such as AI have been applied to complement the network built-in SA. Customer-specific AI can significantly enhance the performance of the network built-in SA equipped with relatively simple analytics tools.

In AI-driven automated SA, the performance is maximized if i) AI can access all data of all NEs; ii) the AI recommendations can be immediately executed to manage these NEs directly. In other words, the control and management of all participating NEs should be *exposed* to AI, including data monitoring, configuration and lifecycle management (LCM). However, as most advanced AI modules are deployed in customers or 3rd parties, *external* to the network management domains, the two requirements for maximization are difficult to meet.

The capability and performance of AI-driven SA are constrained by *service exposure*, defined as the ability of CSP to securely expose management capabilities of the networks and services to authorized customers or third parties, called external customers in this paper. Service exposure decides to what extent the management services are exposed. Specified in ETSI-TS [1], it catalyses service innovations and is considered by GSMA as a key value-added feature for network slicing value chain [2]. Various service exposure models have been discussed in the literature, *e.g.*, by GSMA [2] and NGMN [3].

As NEs of different types are supplied and managed by different management entities, *e.g.*, virtualized infrastructure manager (VIM) for the infrastructure or NFV orchestrator (NFVO) for the NS, service exposure is characterized by the type of NEs. One type of NEs corresponds to one *exposure level*. The exposure level decides the scope of data that can be accessed by external AI and the type of NEs that can be managed by the AI recommendations.

In highly dynamic networks with continuously changing conditions and customer requirements, AI-driven SA is required to achieve *low latency* (the time to detect, identify, and restore a SA issue) and *high reliability* (the accuracy of detecting, identifying, and restoring the SA issue) with *high cost-efficiency*. For the first two requirements, external AI prefers high levels of service exposure in order to gather more data, run more precise analysis, and produce more effective recommendations. However, high levels of service exposure are costly and impose a risk for CSPs and network operators in terms of security, privacy and network safety. It is dangerous to allow external customers engage in network operations with potential threat to network stability. As a result, service exposure raises a question on tradeoff among the SA performance, cost-efficiency and network security.

This paper aims to investigate the impact of service exposure on AI-driven SA, from the perspective of SA performance and the achieved automation of network and service management in 5G networks with network slicing. First, a SA architecture is presented and associated with a service exposure model. Then a deep-dive study is conducted to evaluate the implications of different expo-

sure level on the two key SA services Monitoring and Analytics, as well as the overall impact on SA. This study is based on a practical experiment platform in the 5G-VINNI project. Furthermore, a testbed is built to generate preliminary results and demonstrate the tradeoff impact.

The paper is organized as follows. Section II introduces the SA architecture with the exposure model. Section III provides detailed analysis of impacts on individual SA functions, followed by the experiment results in Section IV. Several key challenges are discussed in Section V. The paper concludes with Section VI on future works.

II. Service Assurance with Service Exposure

A. Service Assurance Architecture and Service Exposure

SA automation can be achieved in different ways, *e.g.*, closed-loop (CL) SA to enable automation in [4]. Consistent with the recursive rule of the network slicing composition proposed for network slicing in 5G-VINNI [5], the SA architecture in [4] is composed of five layers, responsible for assuring customer-facing service (CFS), end-to-end (E2E) service, NS, NF, and infrastructure, respectively. Considering that in practice, CFS and E2E service may co-locate and be managed by the same entity (*e.g.*, OSS with network slicing orchestration), we modify the architecture to Fig. 1 with four layers, where the top layer E2E-SA covers both E2ES and CFS. The layered structure is convenient to demonstrate the differences in the NEs whose management services are to be exposed. One network slice often spans multiple technology domains, such as access network (AN), core network (CN) and transport network (TN), each of which controlled by the corresponding domain controller that orchestrates the underlying networks operations at the application level.

Service exposure defines how each NE exposes its management and control capabilities, *e.g.*, provided by the corresponding orchestrator or controller. These management capabilities contain SA services like monitoring, analytics and intelligence. Fig. 1 illustrates how the SA architecture [4] is leveraged with the four exposure levels proposed in 5G-VINNI [6]:

1) Level 1 (E2E service level) exposes configuration and management capabilities of CFS/E2ES and applications, *e.g.*, via E2E service orchestration.
2) Level 2 (domain level) exposes configuration and management capabilities of network domains (AN, CN, or TN) from the application management aspect, *e.g.*, via domain controllers.
3) Level 3 (network level) exposes management of both NFs and NSs from the resource management aspect, *e.g.*, via NFVO or VNF manager (VNFM).
4) Level 4 (infrastructure level) exposes resource control and management of the (physical or virtual) infrastructure, *e.g.*, via VIM in clouds or WIM (WAN infrastructure manager).

The CSPs and network operators manage the exposure level of each external customer. At each exposure level, the corresponding manager manage how the management capabilities of individual NE are exposed. Obviously, the hierarchy of the exposure levels implies that the higher the exposure level, the deeper the customer can intervene with network management. Besides, customers with Level-i exposure are automatically granted Level-j ($1 \leq j < i$) exposure, *e.g.*, customers with Level-2 exposure have Level-1 exposure by default.

In practice, the basic exposure level is Level-1 where customers only manage the CFS or E2E service that they directly interact with. Ideally, many AI applications of the customers desire to manage network resources as a means to optimize their E2E service performance, which, however, requires a Level-4 exposure. The gap between reality and expectation with respect to the exposure level significantly impacts the performance of the AI-driven SA.

B. SA Services to be Exposed

The impact of service exposure can be analyzed from two aspects, *consuming services* and *managing services*. Based on service-based architecture, *e.g.*, ETSI ZSM [7], an orchestrator or controller provides a set of management services, which could be consumed (*e.g.*, data collection, analytics and intelligence) and/or managed (orchestration, control). Exposing services for consumption is relatively feasible than for management as the former does not directly intervene with network management.

Fig. 2 displays five main SA functions proposed in [4]. Among them, Data Fabric and SA Interpretation play a role of assisting SA and are not directly impacted by the exposure level. Monitoring and Analytics are mainly exposed for consumption whereas Policy Management might be exposed to customers for managing the policies on monitoring and analytics. The exposure level decides the scope and content of monitoring data and analytics results that can be accessed by customers.

For network built-in SA, a CL is formed between SA and orchestration to automate the management of the managed NEs. With service exposure, customers' AI receive the exposed monitoring and analytics services, run their own analytics, and produce recommendations that are fed back into either the SA (Policy Management) or orchestrator to complement the network built-in SA. For example, customers' AI recommend to reallocate resources at Level-4 to improve the QoE of the E2E service, or adjust the monitoring policy to collect more infrastructure data. Then a bigger CL is formed among network orchestrator, network built-in SA and customers' AI to facilitate customer-driven SA (Fig.2).

III. Impact of Service Exposure on SA

To understand how service exposure impacts the SA performance, we conduct an in-depth impact analysis of the exposed SA services, Monitoring and Analytics.

A. Impact of Exposing Monitoring Service

The Monitoring service measures and collects relevant information and key metrics of the managed NE that can be i) used by the internal Analytics service; or ii) consumed by external customers with AI function (the red block in Fig. 1). A direct impact on SA lies in the monitoring data exposed to the external AI. Simply speaking, a higher exposure level implies more comprehensive data exposed and potentially better AI performance.

Although advanced AI usually desire higher levels of Monitoring exposure, it is not always ideal for SA. One of the main reasons is the *monitoring cost*. As indicated in Fig. 1, Level-4 Monitoring exposure allows customers to access monitoring data of Levels 1-4, from all NEs. The cost in data storage and transmissions is extremely

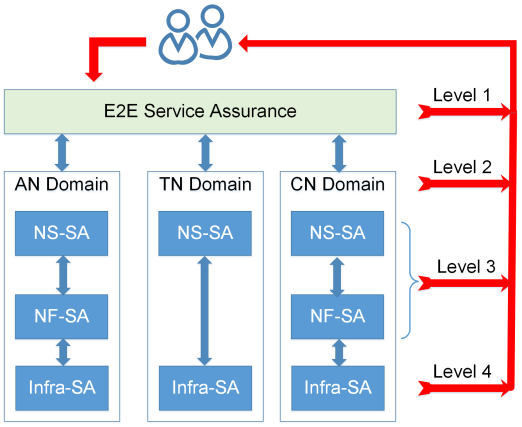

Fig. 1. SA Architecture of network slicing with exposure levels

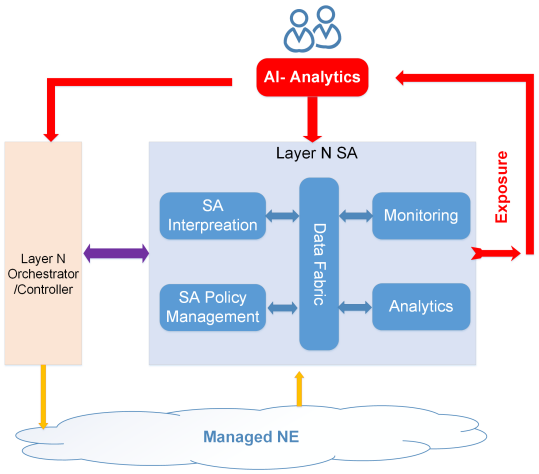

Fig. 2. SA functions with service exposure

high. Moreover, these data sets are produced by NEs that are tightly correlated, due to the associated relationship: network slice $\longrightarrow$ network domain $\longrightarrow$ NS $\longrightarrow$ NF $\longrightarrow$ infrastructure. Higher levels of monitoring exposure is not only costly but also unnecessary as the data sets of different levels are essentially correlated.

High levels of Monitoring exposure significantly complicates monitoring framework for both customers and CSPs. On one hand, customers need efficient tools to retrieve and store the large volume of exposed monitoring data of different types, formats, time scale, and contents from different NEs. The large quantity of data also increases the retrieval delay, which impacts the AI operation efficiency. On the other hand, in a multi-tenant environment where customers are granted with different levels of Monitoring exposure, the CSP needs to differentiate the monitoring data exposed to multiple customers simultaneously. For example, if customer A and B subscribe to two isolated network slices sharing one common infrastructure, then Level-4 exposure of infrastructure monitoring should divide the infrastructure data set into two isolated subsets for customer A and B, respectively. In network slicing, isolation assurance is part of SA. Services for customers A and B should be isolated in a way that i) data related to customer A is exposed to and only to customer A; ii) data related to customer B is kept private from customers A (differential privacy). Then very fine-grained and detailed monitoring is required, which is challenging for high level monitoring, *e.g.*, infrastructure monitoring.

To elaborate the impact of different levels of Monitoring exposure, we use the 5G-VINNI Spanish facility [8] as an example (Figure 3).

Level 1 allows access to the general performance measurements of the E2E service such as QoE and QoS metrics (*e.g.*, throughput and E2E delay). The monitoring data provides an outlook on the overall E2E performance but without details on how the performance is achieved. For example, if an SLA violation is detected but the root cause is at the infrastructure layer, the E2E performance data is not sufficient to pinpoint the root cause. 3rd party E2E monitoring tools (*e.g.*, virtual probes (vProbes)) can be deployed to collect and expose Level-1 data.

Level 2 exposes data on the performance of the applications in the network domains, *e.g.*, deployment management, life-cycle duration, service creation time, service path viewing, trouble ticketing. This level enables more complex analytics (*e.g.*, fault diagnosis) and orchestration/control activities. vProbes are also used at this level.

Level 3 grants access to NS- and NF-specific data, *e.g.*, on flow control, link control, NS topology or chain of VNFs. It enables configuration and customization of the managed NSs and NFs to change their behaviours. Level-3 monitoring can be achieved via built-in tools, such as Open Source MANO (OSM) [9] whose performance management stack contains the VNF Metrics collection tool. As an open source project, OSM exposes the VNF information through the *Juju (VNFM)* metrics system.

Level 4 exposes information regarding the (physical or virtual) infrastructure, whose monitoring is usually provided by VIM. For example, OpenStack telemetry tools Stein [10], *Ceilometer* and *Aodh* services provide collections of data on resource utilization and performance (*e.g.*, CPU and memory utilization, Cores Hyper-Threading, networking I/O speed) and alarms, respectively. This level provides the most detailed data and enables the most direct and efficient actions for SA.

The monitoring tools and cost of these four levels are summarized in Table I.

B. Impact of Exposing Analytics Service

The Analytics service processes and analyzes the data collected by the Monitoring service to solve various problems, including anomaly detection, fault localization, resource allocation, and network traffic forecasting. The outcome of Analytics can be i) fed into orchestration to take action; or ii) exposed to customers' AI for further analysis. Tools used in Analytics range from traditional Statistical Learning (SL) to more recent AI techniques.

Analytics can be classified into *reactive* and *proactive*. Reactive analytics solves reactive problems like detection and forecasting. As an informative approach, it does not directly impact orchestration and control decisions. Exposing Reactive Analytics mainly makes customers aware of how various services behave or raises alarms. On the other hand, Proactive Analytics aims to prevent incidents by recommending decisions on orchestration and control. Typical proactive analytics mechanisms include deep reinforcement learning (DRL) that are envisioned to be applied in real-time for online learning, *i.e.*, learning on

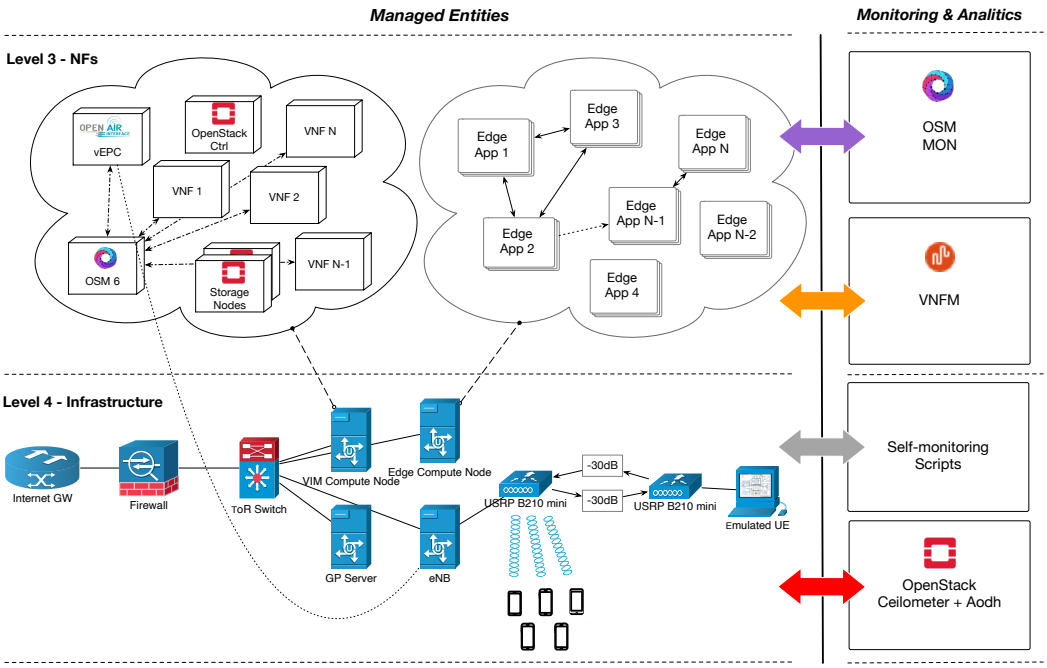

Fig. 3. 5G-VINNI Spain facility site M&A framework

the fly through a feedback loop. With Proactive Analytics, an ideal CL can be formed in built-in SA (Fig. 2).

The capability of Analytics depends on the quality and quantity of the gathered data and information via Monitoring. Therefore, the analysis of exposing Analytics services is linked to the Monitoring exposure.

Level 1 usually exposes limited Analytics capability as the Monitoring data is of low value. It is more suitable for SL-based Reactive Analytics, such as logistic regression for classification of E2E service performance. Then the customers' AI has to be proactive in order to achieve full SA automation. For example, Leve-1 Analytics exposure notifies the customer of a detected QoE degradation based on threshold crossing. Then the customer's AI acts proactively to recommend solutions from the customer's perspective, *e.g.*, adjusting the customer behaviour.

Level 2 can expose more complex Reactive Analytics results given the access to the network domain data. It still relies on customers' AI for proactive actions. Nevertheless, the outcomes of Level-2 Analytics are more detailed and accurate, allowing for more detailed reactions and thus requiring less proactive customers' AI. For example, by diagnosing a congestion in the RAN, Level-2 notifies customers to initiate congestion control from the applications' perspective, *e.g.*, using WebRTC. Or the customers' AI aggregates data of joint Level-1 and Level-2 Monitoring and runs more comprehensive analytics, such as joint forecast of network load, device, application and use behavior to pinpoint the congestion location, by using long-short term memory cells neural nets (LSTMs) [11].

Level 3 encourages Proactive Analytics solutions of managing NSs and NFs as more detailed data is available. Then the reliance on customers' proactive AI is lowered. For example, DRL can be used to proactively scale up or down VNFs to avoid congestion or over-provisioning [12], [13]. Its outcome is more efficient as it directly instructs the VNF LCM. Then it is sufficient for the customer to deploy simpler reactive analytics.

Level 4 exposes highly detailed decisions on infrastructure management via Proactive Analytics, *e.g.*, scale VM in or out to improve the VNF/NS performance, migrate VM to a healthy network, attach/detach a new network interface to the VM, change VM firewall rules. Then AI is not necessary for customers as all decisions are already made by the built-in SA. However, Level-4 Proactive Analytics is significantly complex as it needs to aggregate Monitoring data of all four levels prior to analyzing and making decisions. The high-dimensional dataset makes DRL and deep neural network (DNN) necessary tools to deal with the resulted large action spaces.

Note that although high level of Analytics exposure allows more advanced proactive analytics, it is not mandatory for each CSP to equip with such complex analytics in the built-in SA. For instance, OSM opens Level 3 LCM via Juju charms, but it does not have proactive analytics.

Level	Monitoring Tool	Monitoring Cost	Automation Level	Isolation Level
1	3rd-party vProbe	Low	Descriptive	High
2	3rd-party vProbes	Medium	Diagnostic	Medium
3	built-in EMS/ VNFM/NFVO	High	Prescriptive/ Predictive	Low
4	built-in VIM	Very High	Prescriptive/ Predictive	Very Low

TABLE I
IMPACT OF SERVICE EXPOSURE ON SA

C. Overall Impact Analysis

The analysis of exposing Monitoring and Analytics services leads to two conclusions. Given a higher exposure level: 1) the monitoring cost is higher and the monitoring framework is more complex; 2) the analytical results could be more valuable and insightful. As a consequence, the customers need less proactive and complex AI. Since

proactive analytics is more complex than reactive analytics, we can see that increasing the exposure level essentially shifts the analysis complexity from the customer side to the CSP side. Overall, *high complexity* is exposed to CSP that exposes high level of SA services to customers.

In terms of SA automation, that can be classified into *descriptive* (describe when a problem occurs), *diagnostic* (diagnose why the problem occurs), *prescriptive* (prescribe how the problem can be fixed), and *predictive* (when a similar problem will occur again), the higher level of SA certainly indicates higher level of automation, due to the increase in the analytics capability.

In terms of SA performance, low-level exposure implies *low accuracy* and *long latency*. Low-level Monitoring exposure provides data of low value and makes it hard for customers' AI to run accurate analysis. Low-level Analytics exposure has limited capability to influence orchestration and control. Any decisions or requirements from customers have to be forwarded to and approved by all orchestrator and controllers in a top-down way. For example, if customers with Level-1 exposure plan to scale a VM, the command is passed from E2E-SO to domain controller, NFVO/VNFM, and finally to VIM. Unsurprisingly, it takes a long time to react and respond. Moreover, some Level-1 commands are not fully accepted by Level-4 manager. Via OSM, customers are only allowed to run a small set of basic configurations towards OpenStack. On other hand, low-level exposure also implies *high security* and *high isolation* as less shared infrastructure and data are exposed to multiple customers.

Table I summarizes the major impacts of the four exposure levels. Clearly a tradeoff exists among cost/complexity, SA automation level, and security/isolation protection. In general, high level exposure may not be the best solution even though it is expected by customers' AI. Networks with scarce resource cannot afford highly complex proactive analytics with intensive resource consumption to process millions of parameters. The proper exposure level should be selected based on the customer demand, network capability, and SA requirements. Note that Policy Management could be exposed to customers and adapt the monitoring and analytics policies dynamically to gradually improve the tradeoff, *i.e.*, securely enhance automation with high cost-efficiency.

IV. Experiment Results

In order to investigate the tradeoff impact, an experiment testbed is built (Fig. 4 [14]). It experiments network slicing with an EPC based on OpenAirInterface (OAI). The management entities are OSM [9] for the EPC and OpenStack cluster as the VIM [10]. The EPC NS is composed of four components: Home Subscriber Server (HSS), Mobility Management Entity (MME), Control Plane of the Packet Data Network Gateway (SPGW-C) and User Plane of the Packet Data Network Gateway (SPGW-U). SA-Monitoring is realized via OSM, OpenStack, and `sysstat`, a VM performance metrics collection tool that continuously collects and stores data about CPU and memory usage, disk and network I/O, etc. The testbed is capable of exposing Monitoring at Level-4 via the OpenStack Controller (*e.g.*, the instantiation status of the virtual networks and VMs), Level 3 via OSM (*e.g.*, status of NS deployment, VNF instances, and day-1/day2 configuration with Juju, etc.) and `sysstat` (*e.g.*, statistics collected in the VMs, analysis of the logs of HSS, MME, SPGW-C, SPGW-U service). Table II lists some key metrics collected from the testbed.

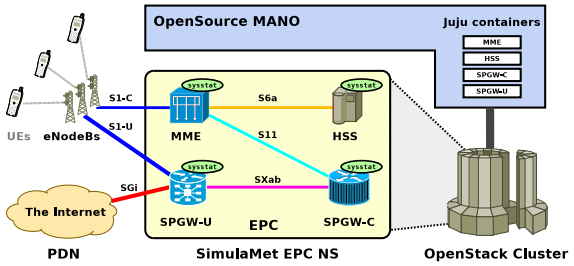

Fig. 4. The SimulaMet EPC Network Service

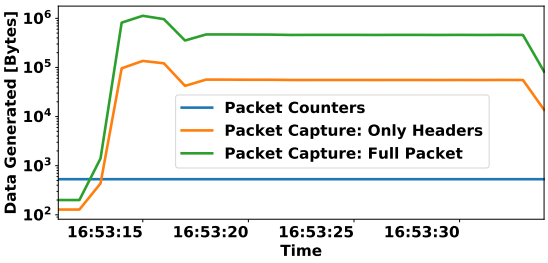

Fig. 5. Overhead of transmitting network KPIs during a high-bitrate UDP download as monitored at the SPGW-U.

Level	Monitoring Tool	KPIs
1	OSM	active users, database size app/buffer status
3	OSM	operational status configuration status
3	sysstat	TCP statistics Packet counters CPU, memory, buffer size
4	OpenStack	CPU/memory usage Network activity

TABLE II
Monitoring Data Collected from the Testbed

The experiment is designed to illustrate the impact of Monitoring exposure regarding monitoring cost. A high-bitrate UDP traffic is generated from a well-provisioned server, using NetPerfMeter [15]. It traverses the Internet and enters the testbed through the SPGW-U and terminates at a UE within the testbed. Monitoring is applied to the traffic that crosses the egress interface of the SPGW-U (*i.e.*, the internet-facing interface of the VNF that handles the user traffic). By saturating the link, a peak traffic scenario is simulated as with multiple users. The traffic is monitored in three levels of granularity: i) per-second aggregated statistics of network KPIs of the interface as reported by `sysstat`; ii) packet captures with only the first 64 bytes of the sniffed packets; and iii) packet captures with all the bytes of the sniffed packets.

Assume at the current time slot, the data captured during the previous time slot is transmitted at the three levels. The communication overhead is the amount of bytes sent during the experiment. Fig. 5 displays the communication overhead of transmitting the monitoring data to an analytics service. As expected, full packet capture is the

most expensive operation, followed by capturing only the headers whereas the cost of reporting packet counters is constant, independent of the traffic load, even if there is no traffic. Then monitoring cost could increase significantly as the monitoring granularity and monitored NEs increase.

V. Challenges

AI-driven CL SA is a challenging topic itself. With service exposure, even more challenges are brought up. First, in the multi-tenant environment, monitoring isolation and differentiation is mandatory for high-level exposure but is time- and resource-consuming. More importantly, this is a contradictory requirement against resource sharing, the essence of network slicing. A carefully designed solution is needed to resolve this conflict.

Since monitoring data of different NEs are correlated, low-level exposure of Monitoring may expose information implicitly hidden for high-level exposure. For example, the data on a NF may shed light on the server hosting the NF. Then Level-3 Monitoring implicitly exposes Level-4 information, which is a double-edged sword. Positively, it allows customers with lower exposure level to access data and gain extra knowledge about the higher levels of exposure. Then the customers can improve their analytics performance with low monitoring cost. Negatively, it impairs the abstraction feature designed for NFV and network slicing, and data security specified for each customer.

In terms of security, the exposure model poses a significant threat to the network. Granting vertical customers with high exposure levels risks the network operation with potential mishandling by customers. Furthermore, when multiple customers plan to modify the operations of the managed NEs simultaneously, their decisions may contradict with each other and thus destabilize the network. For example, many customers desire to apply their AI to optimize infrastructure resource allocation and maximize their own service performance. Therefore, a unified and harmonized security framework is demanded to provide state-of-the-art security solutions in networks with service exposure.

Trustworthy and the corresponding credential system is another key issue. With a large number of customers concurrently accessing the network with different exposure levels, a trusted digital identity system is required to establish a circle of trust and successfully authenticate these customers and their exposure levels. Such a system should take into account the business relationships in the 5G ecosystem and may work in a similar way as the certificate authorities (CA) of the traditional Internet.However, the difference in roles, scope, particularities of different actors should be considered. For customers with high exposure levels, the communications with CSPs have to be extremely secure, which demands more advanced cryptography techniques like fast symmetric cryptography algorithms, *e.g.*, AES, 3-DES, or RSA are envisaged to enable the exchange of public or shared symmetric keys once the credentials are confirmed.

VI. Conclusions

This paper investigates the impact and implication of four levels of SA service exposure from several perspective, including the cost and complexity, the achieved SA performance and automation level, the potential risk to network security and slice isolation, and the requirements on external customers' AI. In general, a tradeoff exists among these factors. Therefore, the exposure level should be decided after a joint evaluation of these factors.

Our analysis shows that Monitoring exposure and Analytics exposure lead to opposite outcomes, *i.e.*, the performance gain of Analytics is achieved at the cost of Monitoring in monitoring overhead and complexity. It is hard to harmonize and balance these two types of exposures. Considering that the customers' AI can complement to the Monitoring and Analytics services of the built-in SA, it may be more practical to further refine the service exposure model by selectively exposing management services at each level. Using a service-based exposure model, Monitoring and Analytics can be exposed individually. In other words, a customer can be granted with Level-2 Monitoring and Level-1 Analytics. In this way, the CSP can shift the pressure of high complexity of Analytics to customers, who are willing to pay for better services. This will be studied in the future work.

Acknowledgment

This work has been supported by the European Community through the 5G-VINNI project (grant no. 815279) within the H2020-ICT-17-2017 research and innovation program.

References

[1] ETSI, "ETSI TS 129 522 V15: Network Exposure Function Northbound APIs; Stage 3 ," 2018.
[2] GSMA, "White Paper: An Introduction to Network Slicing," 2017.
[3] NGMN Alliance, "Security Aspects of Network Capabilities Exposure in 5G, v1.0," Sep 2018.
[4] M. Xie *et al.*, "Towards Closed Loop 5G Service Assurance Architecture for Network Slices as a Service," in *European Conference on Networks and Communications (EuCNC)*, Jun 2019.
[5] 5GVINNI, "D1.2 Design of network slicing and supporting subsystems v1," https://www.5g-vinni.eu/.
[6] ——, "D3.1 Specification of services delivered by each of the 5G-VINNI facilities," https://www.5g-vinni.eu/.
[7] ETSI ZSM, "ZSM 002 Draft: Zero-touch Network and Service Management (ZSM); Reference Architecture," Jan 2019.
[8] 5GVINNI, "Spain Main Facility Site," https://www.5g-vinni.eu/spain-main-facility-site/, last accessed: 2020-01-31.
[9] OSM, "Open Source MANO Release SIX," https://osm.etsi.org/wikipub/index.php/OSM_Release_SIX, Last accessed: 2020-02-03.
[10] OpenStack, "OpenStack Stein," https://www.openstack.org/software/stein/, Last accessed: 2020-02-03.
[11] A. Dalgkitsis, M. Louta, and G. T. Karetsos, "Traffic forecasting in cellular networks using the lstm rnn," in *Proceedings of the 22nd Pan-Hellenic Conference on Informatics*, 2018, pp. 28–33.
[12] J. S. P. Roig, D. M. Gutierrez-Estevez, and D. Gündüz, "Management and orchestration of virtual network functions via deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, 2019.
[13] P. Tang, F. Li, W. Zhou, W. Hu, and L. Yang, "Efficient auto-scaling approach in the telco cloud using self-learning algorithm," in *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, pp. 1–6.
[14] T. Dreibholz, "Flexible 4G/5G Testbed Setup for Mobile Edge Computing using OpenAirInterface and Open Source MANO," in *Proceedings of the 2nd International Workshop on Recent Advances for Multi-Clouds and Mobile Edge Computing (M2EC) in conjunction with the 34th International Conference on Advanced Information Networking and Applications (AINA)*, Caserta, Campania/Italy, Apr. 2020.
[15] T. Dreibholz, M. Becke, H. Adhari, and E. P. Rathgeb, "Evaluation of A New Multipath Congestion Control Scheme using the NetPerfMeter Tool-Chain," in *Proceedings of the 19th IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Hvar, Dalmacija/Croatia, Sep. 2011, pp. 1–6, ISBN 978-953-290-027-9.