Zero-touch AIOps in multi-operator 5G networks

Katherine Barabash * Gino Carrozzo † Dean Lorenz * Kalman Meth * Shuaib M. Siddiqui ‡
* IBM, Haifa, Israel, † Nextworks, Pisa, Italy
‡ i2CAT, Barcelona, Spain
emails: {kathy,dean,meth}@il.ibm.com, {g.carrozzo}@nextworks.it, {shuaib.siddiqui}@i2cat.net

***Abstract*—Automation and intelligence are key functionalities for implementing smart network management decisions in 5G/6G softwarised networks. In this work, we present the 5GZORRO approach for applying Artificial Intelligence to network operations (AIOps). The goal of the 5GZORRO AIOps is to implement truly zero-touch automation of operation and maintenance tasks in the specific context of multi-operator networks. The approach goes beyond the current state of the art of single domain 5G networks and paves the way to an evolution of Network management from 5G towards 6G.**

In this paper, we shortly cover the motivation, explain the technology, and show how heavily it relies on secure, reliable, and intelligent data collection, aggregation and processing. We put forward our assessment of domain specific requirements and challenges and introduce 5GZORRO Operational Data Lake, one of the major innovations towards fully automated multi-operator networks required for smooth progression into 6G.

***Index Terms*—5G, Network Management Network Slicing, AIOps**

INTRODUCTION

Since 2017, when Gartner has coined the term AIOps to stand for Artificial Intelligence for IT Operations and have given it an official definition [1], AIOps has seen a surge in adoption in various large scale operational environments such as Cloud environments, distributed storage systems and services [2–4]. Multiple AIOps products and services have been created for various use-cases but according to Gartner Market Report gartner-AIOps-market, all AIOps platforms incorporate three main aspects: 1) Data ingestion and handling (Observe); 2) Machine Learning (ML) analytics (Engage); and 3) Remediation (Act) and are inherently data-driven.

As part of cognitive slice management research in SLICENET [5], we have started experiments with introducing data-driven analytics into slice orchestration and management as part of FCAPS (Fault, Configuration, Accounting, Performance, Security) operations loop. This work brought us to understanding that AIOps is a promising approach to satisfy operational needs of 5G and future 6G networks as it allows to implement a data-centric solutions for implementing smart network management decisions. This work found appreciation in the 5G community [6] and was also resented in standardization groups such as ETSI ENI, ETSI ZSM, and ITU.

Figure 1 presents the major architectural properties of the AIOps conceived in SLICENET, which replaced the common linear pipeline-like representation of the solution. Usually, the architecture depicts a series of data processing steps beginning with data capture and moving to data ingestion (with aggregation, filtering, encoding, etc.), then to modeling and analytics,

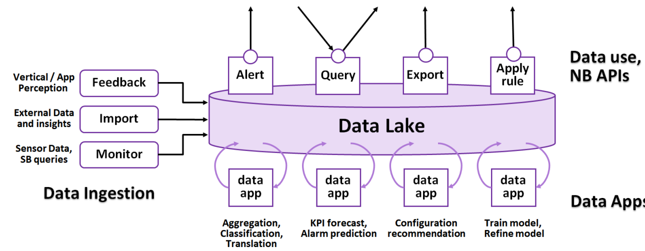

Fig. 1: Operational Datalake for Cognitive Slice Management

and then to insight that can affect the operational system through actuators. Our approach differs in several important ways and and can be described as data-centric. First, we introduced several types of feedback loops required to infuse data into multiple types operational flows. For example, we conceived a feedback loop from analytical components back to the data capture components so that the former can affect data collection dynamically, e.g. to reduce data collection rate when and where the system is quite and healthy and dig deeper into areas and at times of turbulence. Another important feedback loop ensured adding data from the system users or vertical applications, e.g. QoE metrics received in out-of-bound manner from customer analytics[7]. To implement all the feedback loops in efficient and timely manner, we had to decouple data processing flows one from another as much as possible so they can be performed simultaneously and asynchronously whenever possible. As seen in Figure 1, this architecture not only allows to stream data from multiple sources, e.g. system's own sensors, external sensors, and vertical feedback sources. This architecture follows the data lake concept and allows creation and dynamic incorporation of data processing agents, e.g. aggregation, filtering, formatting, and data analytics agents, e.g. contextualizing, model creation, model training, etc. These agents are compute elements that act independently on already existing datasets and, as a result create new datasets and/or invoke actions outside the data lake, e.g. raise alerts to operations and invoke orchestration workflows. Having introduced the Operational Data Lake architecture in SLICENET, we've created and demonstrated an initial prototype based on serverless technologies. More on SLICENET architecture and its AIOps focus can be found in [8].

5GZORRO OPERATIONAL DATA LAKE

Leveraging on the background of SLICENT in adopting AIOps paradigm in 5G networks, we started designing solutions in 5GZORRO project for infusing intelligence into the

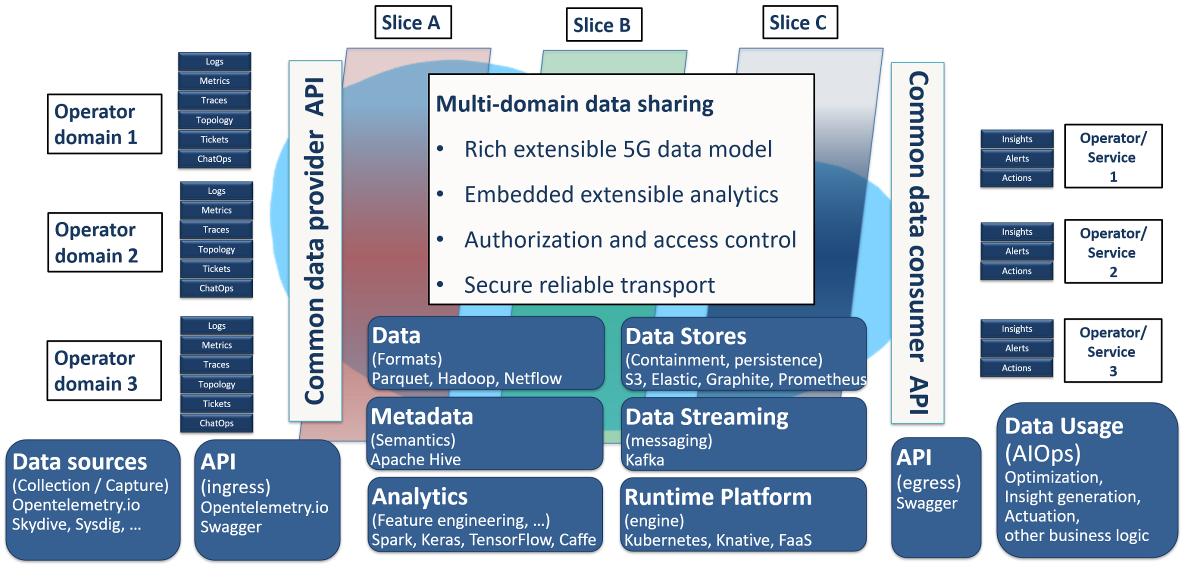

Fig. 2: 5GZorro Operational Data Lake Concept

slice management process and extend this to multi-operator scenarios. Believing that Operational Data Lake is the correct concept to center the AIOps architecture in this domain, we continue to evolve it and to integrate it into more scenarios and use cases. Figure 2 represents the complexity of the environment 5GZORRO Operational Data Lake must function in. In addition, the figure presents architectural highlights such as: 1) Uniform data ingestion and data consumption APIs; 2) Asynchronous even-driven data processing engine that can be extended with custom data-dependent computations and flows, e.g. online and offline model training, model validation, SLA breach detection, amendment, and prevention, etc.; 3) Technology building blocks the solution will be created from, listing the target technologies and their concrete implementation we evaluate as candidates for 5GZORRO integration.

Here is a brief summary of the most important aspects we address to make 5GZORRO AIOps architecture suitable for multi-operator global network and services beyond 5G:

First, the multi-operator aspect of 5GZORRO use case raises issues around *muti-party data collection and sharing*. We envision that beyond 5G operational data will come from multitude of devices and services deployed in different geographical, operational, and ownership domains and that not always these players will be willing to share the data openly. On the other hand, providers that offer services composed from lower level services, in order to satisfy SLAs contracted with their customers, will need data from underlying devices and services to be incorporate into their automation pipelines. There can be several approaches to resolve this tension. We choose to create zero trust environment where each player's data can be stored safely and privately, fully protected, so it can be retrieved and operated on only by parties and for reasons the data owner has agreed to through 5GZORRO multi-party smart contracts.

Second, to facilitate analytics across multiple operators and technology domains, data received from different sources must be enriched, contextualized, and sometimes translated into a unified format. For this, 5GZORRO introduces and implements an *extensible domain specific data model for 5G*, starting with our target use cases and following industry standards, e.g. Generic Network Slice Template created by GSMA[9]. 5GZORRO Operational Data Lake includes metadata service component to keep the metadata for cross-referencing and correlation. A shown in the figure, we evaluate Apache Hive[10] for this role.

Third, multivariate time series measurements, such as network flows, compute, memory, spectrum, and other resource usage, coming from multiple geographies and devices, will inherently loose some of its time context due to lack of clock synchronization between devices, due to communication/processing delays, or due to variances in timestamp encoding etc. On the other hand, many learning algorithms rely on having data prepared in form of fixed-sized time windows. In such cases, there is a need align and correlate these asynchronous time-series streams coming from multiple subsystems of multiple providers' systems. For this, Operational Data Lake ingestion pipeline must normalize the timestamps of all the data sources and aggregate the measurements into elementary time windows, e.g. per minute. These aggregated data blocks must be annotated with their time window identity as part of their metadata. In addition, we propose to create an SQL-like API that allows analytic components to selectively fetch time-series data from multiple streams based on time window indicator as a parameter.

All these data preparation steps make sure operational data is easily transformable into domain specific Machine Learning features and is governed under strict control of zero-trust multi-party sharing. 5GZORRO Operational Data Lake also contains

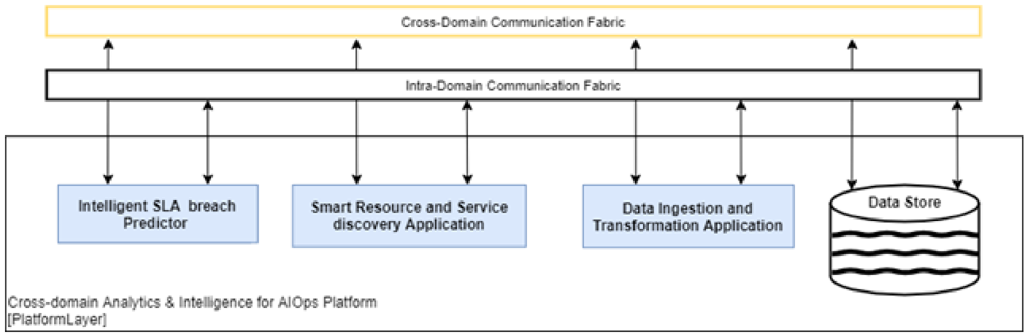

Fig. 3: 5GZORRO Cross-domain Analytics & Intelligence for AIOps platform

an event-driven workflow engine whereby data scientists can create, train and deploy ML models directly into the operational domain and, if needed connect them back to orchestration tools for actuation and to data capture tools for influencing the data collection cadence and/or richness. One example of such analytic scenario is SLA breach detection presented in a recent 5GZORRO publication [11]. Additional use cases are described in the next section and will be realized and and demonstrated as project progresses.

5GZORRO Cross-domain Analytics Intelligence for AIOps

The 5GZORRO Operational Data Lake includes - as any data lake service - several coordinating services, which can be used to create analytical pipelines. Form a standpoint of any specific analytical and/or operational functionality, the cycle of operations is:

1) Specify what data is required to realize this specific functionality. This can be made through Data Lake APIs and/or by configuring data capturing and monitoring tools in the managed environment. Also specify whether the data will be required in batching or in streaming mode or both.
2) (out-of-band) Data required by this specific functionality flows to Data Lake.
3) Fetch data from the Data Lake.
4) Perform some analysis on the data to obtain insights.
5) Perform some action based on the result of the analysis. Actions can be of different types such as raising an alert, invoking orchestration service, writing new data-sets into the Data Lake, etc.

5GZORRO architecture [12] specified a number of functionalities that we have modelled with the same generic idea. These are presented in Figure 3 and include:

- Intelligent SLA monitoring breach predictor [11],
- Intelligent Network Slice and Service optimizer,
- Intelligent 3rd party resource planner,
- Data Ingestion Transformation.

For example, for the Intelligent SLA monitoring breach predictor, we can have the following specific pipeline stages:

1) Provide monitoring data.
2) Aggregate the monitoring data.
3) Using the monitoring data, evaluate SLA satisfaction.
4) Perform some analytics to predict violation of SLA.
5) Raise an event for actions to be performed upon prediction of SLA violation.

Similar pipelines need to be defined for all the other analytics workflows.

The implementation and deployment of the stages of the pipeline is coordinated. The output from one stage typically serves as the input for another stage. The information transferred between stages may be the actual data upon which to operate, or it may simply be a notification that the previous processing stage has completed with an indication of where to find the data in the data store.

A general framework adopted in 5GZORRO to allow easy connectivity between services defined in a Data Lake is Argo [13], generally used to implement pipeline architectures for Kubernetes. It follows this approach:

1) Register the service; specify the data that is to be used by the service (some kind of pointer to the data or other type of description of the data).
2) Use messaging for communication between services.
3) Each service has a channel from which it receives input and has a channel to which it sends output. The input/output might be the actual data, or it may include a pointer to the location of the relevant data to be processed.
4) The output produced by a single component could be consumed by multiple consumers. For example, the output of a data aggregator service may be forwarded both to the data store as well as to a service that checks for SLA compliance.
5) Finally, suppose the analytics (or action) wants to trigger some functionality on the client. We need to define a way for these to interact. This could be encapsulated in a message delivered via an output channel.

The various parameters to set up the microservice are all contained in a (e.g. yaml) configuration file. This configuration file is provided together with the compiled microservice module. Upon registration, variable fields of the configuration file are provided to customise the particular pipeline.

I. CONCLUSIONS

In 5GZORRO, we are building the Operational Data Lake for AIOPs by integrating open-source components, as presented in Figure 2, similarly to how a generic open source data lake, Open Data Hub [14], is put together. In fact, we deploy objects storage, messaging, monitoring, analytics, metadata, and other services to create a 5G specific data lake, while Open Data Hub offers a complete generic solution out of the box. The choice to create our own solution instead of using ready-to-go one is dictated by the need to specialise the data lake for 5G use in many different places to satisfy domain specific needs as described in this work. In addition, working with components and not with hardened integrated solution makes it easier to decompose the work and to integrate with other 5GZORRO components and gives us freedom to make different component selection choices as we go, to eventually find the best candidate technology for every need. For example, for the AI/ML part we evaluate Seldon [15], the open source framework for deploying AI/ML models on Kubernetes and OpenShift. The model can be created and trained using many tools such as Apache Spark, scikit-learn and TensorFlow. Seldon also provides metrics for Prometheus scraping. Metrics can be custom model metrics or Seldon core system metrics.

Some particular APIs and services are being introduced to support zero-touch, which are specific to 5GZORRO.

The solution will be demonstrated in prototypes planned for Q3-2021 which will demonstrate intelligent network slicing creations and SLA breach predictions across multiple 5G operators.

ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871533 (5GZORRO).

REFERENCES

[1] A. Lerner, "Gartner aiops platforms," https://blogs.gartner.com/andrew-lerner/2017/08/09/aiops-platforms/, 2017, 2017-08-09.

[2] A. Levin, S. Garion, E. K. Kolodner, D. H. Lorenz, K. Barabash, M. Kugler, and N. McShane, "Aiops for a cloud object storage service," in *2019 IEEE International Congress on Big Data (BigDataCongress)*, 2019, pp. 165–169.

[3] H. Wang and H. Zhang, "Aiops prediction for hard drive failures based on stacking ensemble model," in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 2020, pp. 0417–0423.

[4] Y. Dang, Q. Lin, and P. Huang, "Aiops: Real-world challenges and research innovations," in *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, 2019, pp. 4–5.

[5] "End-to-end cognitive network slicing and slice management framework in virtualised multi-domain, multi-tenant 5g networks," https://slicenet.eu/.

[6] "Machine learning based end-to-end network slice management and orchestration," https://www.innoradar.eu/innovation/38201, 2020.

[7] Kenneth Nagin et al., "Estimating client QoE from measured network QoS," in *Proceedings of the 12th ACM International Conference on Systems and Storage*, 2019.

[8] Qi Wang et al., "SliceNet: End-to-end cognitive network slicing and slice management framework in virtualised multi-domain, multi-tenant 5G networks," *2018 IEEE international symposium on broadband multimedia systems and broadcasting (BMSB)*, 2018.

[9] GSMANG.116, "Generic Network Slice Template Version 3.0," 2020.

[10] "ApacheHive TM." [Online]. Available: https://hive.apache.org//

[11] V. Theodorou et al., "Blockchain-based Zero Touch Service Assurance in Cross-domain Network Slicing," 2021.

[12] 5GZORRO, "Design of the 5GZORRO Platform for Security & Trust," 2021.

[13] "Argo workflows and pipelines." https://argoproj.github.io/projects/argo, 2021.

[14] "Open data hub," https://opendatahub.io/, 2021.

[15] "Seldon - Machine learning deployment for enterprise." [Online]. Available: https://www.seldon.io///