Enhancing Network Data Analytics Functions: Integrating AIaaS with ML Model Provisioning

Ali Nadar, Jérôme Härri
EURECOM, 450 route des Chappes, 06904 Sophia-Antipolis, France
ali.nadar, haerri@eurecom.fr

***Abstract*—The Network Data Analytics Function (NWDAF) is a new 5G Core System (5GS) application function providing network analytics via data collection and exposure APIs and predefined data analytics or AI/ML models. 3GPP rel.16 only supports locally trained and inferable AI/ML models, which might be a limiting factor in decentralized, multi-vendors/tenants 5G architectures. 3GPP rel.17 proposes APIs for AI/ML sharing between NWDAF to address these limitations. In this paper, we present a 3GPP rel.17 architecture extending OpenAirInterface (OAI) NWDAF towards AI-as-a-Service and supporting AI/ML model sharing. We integrate an AI/ML ontology for AI/ML and introduce APIs between the NWDAF and an AIaaS platform. We finally demonstrate its feasibility via AI/ML model sharing abnormal traffic behavior use case.**

***Index Terms*—5G, NWDAF, AI-as-a-Service, Machine learning, abnormal behavior, OAI.**

I. Introduction

The rapid evolution of telecommunications technologies, notably the advent of 5G networks, has ushered in a new era of connectivity characterized by unprecedented data volumes, diverse services, and dynamic network environments. Central to the efficient operation and optimization of these networks is the Network Data Analytics Function (NWDAF), a critical component responsible for extracting actionable insights from network data to drive intelligent decision-making and resource management.

In parallel, the rise of Artificial Intelligence as a Service (AIaaS) platforms has democratized access to advanced machine learning and AI capabilities, offering organizations scalable and cost-effective solutions for deploying and managing AI models. AIaaS platforms leverage cloud infrastructure and sophisticated algorithms to provide on-demand access to AI services, enabling organizations to harness the power of AI without the need for extensive infrastructure or expertise.

The integration between NWDAF ML model provision and AIaaS platforms represents a compelling opportunity to enhance the capabilities of NWDAF and unlock new avenues for network optimization and management. By leveraging the scalability, flexibility, and advanced AI capabilities offered by AIaaS platforms, NWDAF can augment its analytical capabilities, accelerate model development and deployment, and adapt more effectively to evolving network dynamics.

However, the integration between NWDAF ML model provision and AIaaS platforms also presents unique challenges and considerations, including data privacy, security, interoperability, and performance optimization. Addressing these challenges will be crucial to realizing the full potential of this integration and ensuring its successful implementation in real-world network environments.

In this paper, we explore the integration between NWDAF ML model provision and AIaaS platforms, investigating its potential benefits, challenges, and implications for network analytics and management in 5G networks. We present a comprehensive analysis of the synergies between NWDAF and AIaaS, demonstrate the feasibility and effectiveness of their integration by an operational prototype exchanging AI/ML models for traffic anomaly detection, and provide insights into best practices and recommendations for deploying and operate integrated NWDAF-AIaaS[1] solutions.

Our contributions are threefold: first we describe the architectural integration of an AIaaS platform and a microservices-based NWDAF; second we introduce our methodology extending a NWDAF architecture, adapting it to domain of knowledge of an AIaaS platform as well as the message exchanging between the two architectures; finally, we demonstrate the feasibility of AI/ML model exchanges for abnormal traffic detection with a real trained ML model. The platform is available as open-source under an Apache 2.0 license.

The rest of this paper is organized as follows: Section II introduces NWDAF, while Section III outlines the architectural design between AIaaS ad NWDAF. In Section IV we detail the integration methodology, while in Section V we presents a proof-of-concept for AI/ML model exchange to detect abnormal traffic behavior. Section VI concludes the paper.

II. Literature Survey

NWDAF has been introduced by 3GPP in TS 23.288 [1] Rel-15 and further extended in Rel-16 and Rel-17 to addresses operator needs for supporting analytics based on data collected from 5G network, application or management functions. The analytics information supported by NWDAF are either statistical information of the past events or predictive information. TS 23.288 [1] Rel-16 provided analytics APIs for slice load level, UE mobility or abnormal detection, and APIs for AI/ML models. The NWDAF architecture and functionalities have been further extended in TS 23.288 [1] Rel-17 to (i) support multiple instances of NWDAF, (ii) logical decomposition into two logical functions such as Model Training logical function (MTLF) and Analytics logical function (AnLF), and (iii) trained ML model sharing between multiple NWDAF instances. 3GPP however does not intent to standardize the format or process of sharing AI/ML across vendors. This paper precisely aims at providing a domain agnostic solution to that aspect.

In NWDAF research topic, *Manias, et al.* [2] proposed a 3GPP rel.15 NWDAF using clustering for similarity analytics between NF-NF interactions. A 3GPP rel.16 NWDAF architecture supporting APIs for 3GPP analytics and AI/ML is

[1] https://gitlab.eurecom.fr/oai/cn5g/oai-cn5g-nwdaf/tree/ML-provision

introduced by *Mekrache et al.*. An operational prototype based on OpenAirInterface (OAI) [2] open 5G Core System (5GS) is proposed, which uses an LSTM Auto-encoder algorithm as AI/ML model to detect abnormal traffic events. On AIaaS side, *Nadar et al.* presented in [3] an AI-as-a-Service architecture providing APIs and components to subscribe to and exchange AI/ML models, as well as orchestrate decentralized AI/ML training. This paper therefore proposes to integrate this AIaaS framework to a 3GPP rel.17 compliant NWDAF supporting AI/ML sharing between NWDAF entities.

III. Architectural Integration

The integration of AIaaS platform with NWDAF involves seamless interaction between various components. AIaaS platforms provide scalable infrastructure and expertise for ML model development and training, while NWDAF offers domain-specific knowledge and access to network telemetry data. This collaboration enables the development, deployment, and orchestration of ML models within NWDAF, facilitating real-time analytics and decision-making in 5G networks. Fig.1 shows an architecture overview for the three main services provided by 3GPP for network data analytics. In this work we focus on the implementation of ML model provision service which play a crucial role to find the most relevant AI models based on the use case context of NWDAF consumer request.

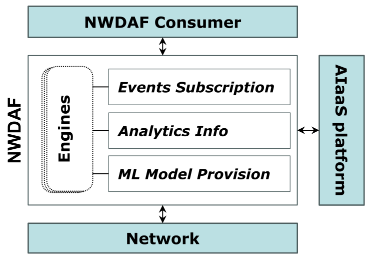


Figure 1: AIaaS-NWDAF architecture overview

IV. Methodology

The integration of AIaaS and NWDAF represents a critical step towards advancing 5G network analytics capabilities. This section introduces the methodology employed to seamlessly integrate AIaaS platforms with NWDAF, facilitating the development, deployment, and orchestration of machine learning models for enhanced network management and optimization. Our work builds upon and complements previous studies by proposing a novel framework for integrating AIaaS platforms with NWDAF ML model provision.

A. Extending NWDAF architecture

In [4] *Mekrache et al.* proposed a microservices-based architecture for NWDAF, at exposure layer as depicted on Fig.2 they implemented only the analytics info and events subscription services as standardized by 3GPP. For instance, when a client subscribes to an event requiring complex analytical tasks, such as detecting abnormal traffic behavior generated by a UE, at analytics layer upon event received, the abnormal behavior engine search inside the local repository for the requisite ML model. This model is then utilized to infer and thus detect abnormal behavior based on the UE input data.

[2]https://openairinterface.org/

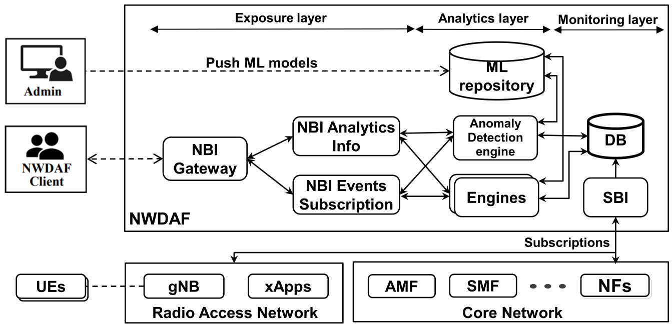


Figure 2: microservices-based architecture for NWDAF

Employing an approach reliant on NWDAF local storage of ML models presents several challenges, including issues related to dynamic adaptability, scalability, freshness of ML models, resource efficiency and interchangeability. To effectively tackle these challenges, we propose an extension for the existing NWDAF architecture including the design and implementation of new NWDAF ML model provision microservice which is fully 3GPP-standardized, as well as integrating an "*AIaaS agent*" microservice within NWDAF components to provide an HTTP RESTful interface fulfilling the contextual interaction with the AIaaS platform. Fig.3 illustrates our new proposed microservices-based architecture for NWDAF, we removed the Name-based ML repository in order to replace it by exposing a "*ML model provisioning*" service that handles the process of finding the best fitted ML model based on the contextual description (ML metadata).

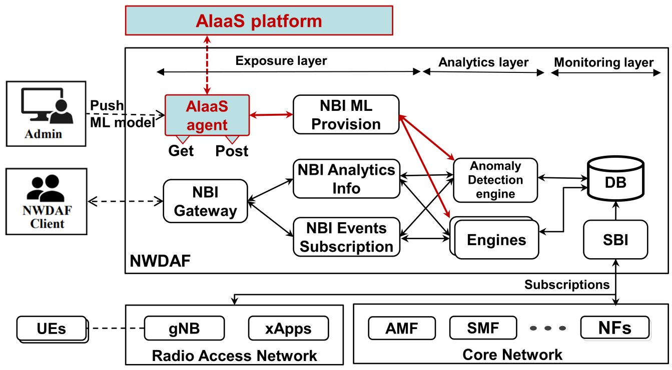


Figure 3: AIaaS-powered microservices-based architecture for NWDAF

B. AIaaS architecture: Adapting knowledge domain Structures for NWDAF

In this section, we describe an existing architecture for AIaaS platform that could expose the context-based ML models required by NWDAF, then we define a new specific ontology to fit the knowledge domain of NWDAF. To this end, we adopt a previous work in [3], where *Nadar et al.* proposed an architecture and implementation for AIaaS platform with a particular focus on vehicular networking domain. As shown in Fig. 4, the two main components of AIaaS platform are as following:

- AIaaS server; corresponds to the backend processing of the AIaaS framework, it consists of a message passing component (MQTT), a data access management engine and a Knowledge Management (KM) engine. For data storage, it has a RDF knowledge graph database to store the metadata of AI models as well as a file database to store byte-codes AI models.

- AIaaS agent; can communicate with many AI clients, it manages user request in threefold; first it exposes AIaaS-API to receive request content. Second, using the semantic functions, it interprets the graph-based query to validate, analyse and align the query with the structure of the associated ontology. Finally, it communicates with AIaaS server microservices using publish/subscribe messaging pattern. The described AIaaS-API is an HTTP RESTful service, all requests must be in JSON/RDF-format that describes the context of client request.

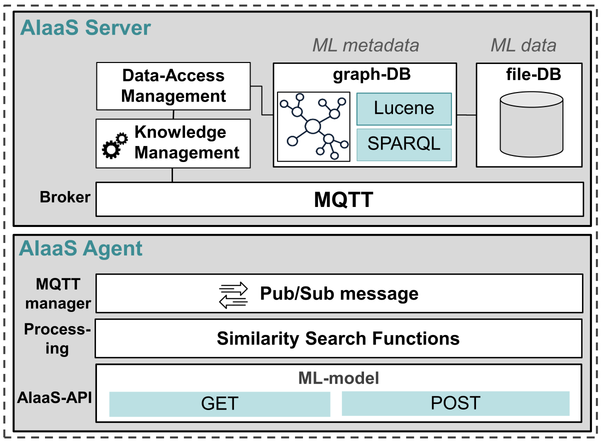


Figure 4: AIaaS architecture

On the other hand, ontology speaking, to adapt an existing AIaaS platform to fit the requirement of NWDAF and keeping the proximity metric strategy to find the most relevant ML model, several changes on AIaaS knowledge domain are required. To this end, we follow the 3GPP technical specifications and resources related to NWDAF standards and protocols as introduced in the contents of *ML Model Provisioning*[3] in 3GPP TS 23.288 (Release 17) in order to define a 3GPP-Compliant Ontology for NWDAF ML Model Provisioning service.

To ensure broad applicability, we have devised an ontology that integrates two key components: Firstly, an adapted ontology derived from the work of *Braga et al.* [5] which focuses on Machine Learning concepts. Secondly, an ontology that captures the contextual aspects introduced by the NWDAF ML Model Provisioning service. In this ontology, each entity corresponds to a parameter outlined in 3GPP specifications. From these specifications, we have selectively chosen attributes that are conducive to proximity metrics techniques within the AIaaS platform. As illustrated in Fig. 5 the ontology encompasses both *mandatory* parameters, such as Event and EventFilter, as well as supplementary *optional* parameters. These optional parameters are aimed at enriching the process of identifying the most relevant ML model for the context of NWDAF consumers. We describe below the content the we take into consideration in our proposed 3GPP-compliant ontology as the following:

[3]http://www.tech-invite.com/3m23/toc/tinv-3gpp-23-288_za.html

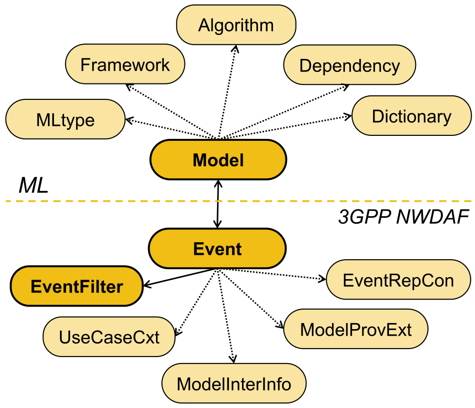
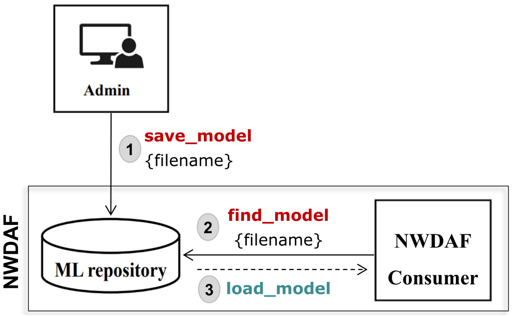


Figure 5: NWDAF-ML-based ontology for AIaaS

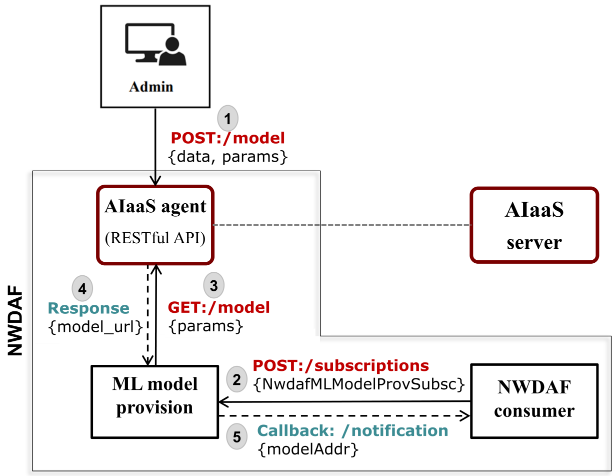
- Event; [required] describes the NWDAF Events, such as *"ABNORMAL_BEHAVIOUR"*.
- EventFilter; [required] represents the event filters used to identify the requested analytics, such as {*"excepIds": "UNEXPECTED_LARGE_RATE_FLOW"*} .
- UseCaseCxt; [optional] indicates the context of usage of the analytics to select the most relevant ML model. The value and format of this parameter are not standardized.
- EventRepCon; [optional] indicates the ML event reporting condition.
- ModelInterInfo; [optional] (Model Interoperability Information) this is vendor-specific information that conveys, e.g., requested model file format, model execution environment, etc..
- ModelProvExt; [optional] (Model Provision Parameters Extension) indicates the extended ML model parameters that a service consumer optionally sets when subscribing to an ML model to be provisioned.

C. NWDAF Messaging with AIaaS

AIaaS platform revolutionizes the process of finding ML models by prioritizing relevance over exact matches. Unlike traditional methods that strictly adhere to exact model matches, our AIaaS platform leverages an advanced proximity metrics techniques to identify and recommend the most suitable ML models based on the contextual relevance to the NWDAF client context. Through proximity metrics and contextual understanding, AIaaS platforms empower NWDAF consumers to discover ML models that align closely with their specific needs and application scenarios. Technically speaking, emphasising on the role of NWDAF ML model provision microservice, upon receiving a model provisioning subscription it translates the subscription content (e.g.including event ID, filter, input features, output threshold, ...) to a JSON format context parameters in order to use it as RESTful GET method to find the most relevant ML model. Fig. 6b shows how NWDAF consumer can interact with AIaaS platform via ML model provision microservice, message 2 and 5 are following 3GPP specifications as introduced in *Nnwdaf_MLModelProvision* API service, while message 1, 3 and 4 are following AIaaS specifications to POST/GET messages. In the next section, we describe an experiment demonstrating how ML provision microservice interacts with AIaaS platform to get the most relevant ML model for abnormal traffic detection.



(a) Repository-based approach (3GPP rel.16 default).



(b) Novel AIaaS-based approach.

Figure 6: Approaches to store/find NWDAF ML models.

V. Experimentation: Identifying Abnormal Behavior ML model via AIaaS

Without loss of generalities, to illustrate the interoperability and data exchange between NWDAF microservices and an AIaaS platform, we conducted a simplified experiment. We employed the identical ML model as utilized by *Mekrache et al.* in [4] for Long-Short-Term-Memory(LSTM) Auto-encoder model trained with real network data extracted from the Milano dataset [6]. Initially, we pushed the LSTM model to the AIaaS platform using the RESTful POST method encapsulated in bytes format along with its contextual parameters *"params"* described in JSON format as following:

```
{
  "event": "ABNORMAL_BEHAVIOR",
  "event-filter" : {
     "excepIds": "UNEXPECTED_LARGE_RATE_FLOW"},
  "use-case-cxt":{
     "features":["weekday", "hour", "minute",
                 "internet_data"]}
}
```

As shown, we specified the NWDAF event, event filter and model features in the optional use-case-context attribute. We added the usecase context to underscore the significance of aligning the context of NWDAF consumer with the expected ML model, thereby adding more granularity to facilitate the proximity search process in AIaaS platform. The message flow diagram illustrating the interaction between the AIaaS platform and the NWDAF microservices is depicted in Fig. 7.

In summary, the diagram illustrates the message flow and their respective contents, commencing with the client's

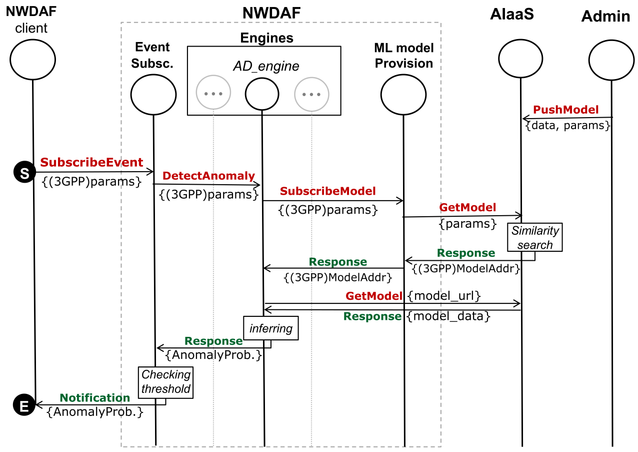


Figure 7: NWDAF-AIaaS message flow.

NWDAF Anomaly Detection request. Subsequently, the request traverses various NWDAF microservices, facilitating communication with the AIaaS platform to identify the optimal ML model. Following model selection, the system infers the anomaly detection probability and periodically notifies the client based on a predefined threshold.

VI. Conclusion

The integration of AIaaS to provide ML models to the NWDAF ML model provision service offers significant advantages for network data analytics. By leveraging AIaaS platforms, NWDAF can access a diverse repository of ML models tailored to various network monitoring and data analytics tasks. This integration enhances the agility and flexibility of NWDAF by enabling the dynamic provisioning of ML models based on evolving network conditions and requirements. Overall, the collaboration between AIaaS and NWDAF ML model provision service represents a promising approach to advancing network intelligence and enhancing the resilience and security of modern communication networks.

VII. Acknowledgements

This work has benefited from a government grant managed by the "*Agence Nationale de la Recherche*" under the France 2030 program with NF-FITNESS, grant "ANR-22-PEFT-0007".

References

[1] 3GPP, "3rd generation partnership project; technical specification group services and system aspects; architecture enhancements for 5g system (5gs) to support network data analytics services (release 18)," 3GPP, Tech. Rep. TS 23.288 V18.4.0 (2023-12), 2023.

[2] A. S. Dimitrios Michael Manias, Ali Chouman, "An nwdaf approach to 5g core network signaling traffic: Analysis and characterization," in *GLOBECOM 2022, IEEE Global Communications Conference, 4-8 December 2022, Rio de Janeiro, Brazil*, 2022.

[3] A. Nadar and J. Härri, "An ai-as-a-service platform for an artificial intelligence of things (aiot)," in *VTC 2024, IEEE Vehicular Technology Conference, 24-27 June 2024, Singapore*, 2024.

[4] A. Mekrache, K. Boutiba, and A. Ksentini, "Combining network data analytics function and machine learning for abnormal traffic detection in beyond 5g," in *GLOBECOM 2023, IEEE Global Communications Conference, 4-8 December 2023, Kuala Lumpur, Malaysia*, 2023.

[5] "Machine Learning Ontology (mlonto)," https://osf.io/chu5q/, accessed: 2022-11-17.

[6] G. B. et al, "A multi-source dataset of urban life in the city of milan and the province of trentino," *Science Data 2*, p. 150055, 2015.