

Um *framework* NWDAF para algoritmos de análise de dados de rede 5G e além *

Júnia Maísa Oliveira¹, Jônatan Almeida², Daniel Fernandes Macedo¹,
José Marcos Nogueira¹

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brasil

²Departamento de Engenharia Elétrica
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brasil

{juniamaisa, damacedo, jmarcos}@dcc.ufmg.br, jonatandavi2009@gmail.com

Abstract. *The data analytics function for 5G networks, as defined by the 3GPP, is responsible for analyzing network data and enabling the execution of actions on the network as a result of the analysis. 5G network data analysis systems of the literature has algorithms for a single analysis context and does not allow the inclusion of new analysis algorithms. As an alternative, this work presents a system for 5G networks data analysis function, which allows the inclusion and editing of data analysis algorithms, in a open-source software based architecture. As a case study, artificial intelligence techniques in the context of network anomalies detection are incorporated into the system.*

Resumo. *A função de análise de dados de redes 5G, definida pela 3GPP, é responsável por analisar os dados de rede e propiciar a execução de ações na rede como resultado da análise. Os sistemas de análise de dados de rede 5G da literatura utilizam algoritmos para um único contexto de análise e não permitem a inclusão de novos algoritmos de análise. Como alternativa, este trabalho apresenta um framework para a função de análise de dados de redes 5G que permite a inclusão e edição de algoritmos de análise de dados, em uma arquitetura composta por software open-source. Como um estudo de caso, técnicas de inteligência artificial no contexto de detecção de anomalias de redes são incorporadas ao framework.*

1. Introdução

A necessidade de análise de dados de redes e de automatização de ações fez com que uma nova função fosse considerada pelo *3rd Generation Partnership Project* (3GPP) para a operação de redes 5G e além, denominada função de análise de dados de rede *NetWork Data Analytics Function* (NWDAF). A NWDAF é uma função de rede que visa coletar dados de outras funções de rede e fazer análises com o auxílio de técnicas de inteligência artificial, visando gerar novas informações e propiciar ações que levem ao aumento de eficiência operacional da rede

*Os autores agradecem às seguintes instituições brasileiras pelo apoio: MPMG - Ministério Público de Minas Gerais, CAPES, CNPq, FAPEMIG e FAPESP/MCTI (processos 2023/13518-0, 2020/05182-3 e 2018/23097-3).

[3rd Generation Partnership Project (3GPP) TS 29.520]. Um processo de análise de dados de rede definido por padrões, implantado com uma arquitetura escalável e otimizada, simplifica o gerenciamento de redes 5G/híbridas e é fundamental para garantir a melhor experiência ao usuário. Para isso, é muito importante escolher o sistema e os algoritmos de processamento certos para lidar com diferentes tipos de conjuntos de dados [Ma and Chi 2022]. Porém, as soluções existentes para a NWDAF são em geral proprietárias e fechadas, com abrangência limitada a certos serviços e casos de uso [Donertasli and Medithe 2022]. Isso limita a evolução das análises de dados para gerenciamento e ações de melhorias automáticas nas redes 5G e além.

Para a função NWDAF, alguns trabalhos propõem arquiteturas para execução em núcleos 5G concentradas em contextos de análise tais como qualidade da experiência do usuário, tarifação de serviço, sinalização de rede, gerenciamento autônomo da rede, segurança, fatiamento de rede ou eficiência energética. Ademais, poucos trabalhos apresentam uma arquitetura que utiliza componentes *open-source*.

Este trabalho apresenta um *framework* de análise de dados que possibilita a inclusão e edição de algoritmos para análise de dados em diferentes contextos, com arquitetura composta por software *open-source* e adequada às especificações técnicas (TS) 3GPP. Em engenharia e programação de software, um *framework* é uma coleção de componentes de software reutilizáveis que tornam mais eficiente o desenvolvimento de novas aplicações. Um vídeo de demonstração parcial do *framework* está disponível no YouTube¹.

Este trabalho está organizado da seguinte maneira: A Seção 2 apresenta trabalhos da literatura. A Seção 3 apresenta o *framework*, sua arquitetura e componentes. A Seção 4 apresenta técnicas de exploração de dados necessárias para inclusão de novos algoritmos no *framework*. A Seção 5 apresenta um contexto de análise para compreensão da inclusão de algoritmos no *framework*. Por fim, a Seção 6 conclui o trabalho.

2. Trabalhos relacionados

A literatura apresenta propostas de *frameworks* que viabilizam a análise de dados de rede 5G. Os autores em [Aumayr et al. 2022] apresentam uma arquitetura de *Analytics* flexível e escalável, baseada em microsserviços. Testes foram realizados em um núcleo 5G NSA, 5GENESIS. Para gerenciamento do ciclo de vida dos algoritmos de aprendizado de máquina, foram utilizadas Acumos e Kubeflow. Para detecção de anomalias, os programas Z-Score e MAD.

Os autores em [Perez et al. 2020] apresentam uma arquitetura de monitoramento escalável, confiável, de baixa latência, distribuída, reconfigurável e de agregação de múltiplas fontes de dados. A coleta de dados é realizada com Apache Kafka e o Data shipper. O monitoramento é feito com Elastic Stack (ELK). Testes com duas VMs, Docker, Apache Kafka, Elastic Stack e Sangrenel mostraram que variações nos parâmetros de projeto afetam o consumo e desempenho da CPU. Os autores em [Samaras et al. 2022] apresentam a demonstração de uma Plataforma de Monitoramento e Predição de QoS (QMP) para detecção e mitigação proativa de quedas de QoS na operação de fatia de rede. Os algoritmos de IA utilizados foram *Long short-term memory* (LSTM), *Recurrent Neural Network* (RNN) e *Support Vector Regression* (SVR). As tecnologias utilizadas para

¹<https://youtu.be/BsO4gVLfY9c>

execução dos algoritmos e exibição de resultados foram Kubernetes, Kafka, Prometheus e Grafana. Os autores em [Abbas et al. 2021] aplicam a análise de dados no gerenciamento do ciclo de vida de uma fatia de rede para previsão de carga e dimensionamento de recurso de rede. A arquitetura é composta pelos softwares: Open Source MANO (OSM), Flexran controller, Grafana, Prometheus e ElasticMon. O ambiente 5G utilizado foi OpenAirInterface (OAI).

Diferentemente das abordagens existentes, este trabalho apresenta um *framework* de análise de dados de rede 5G e além, implementado com componentes *open source* e tecnologia de contêineres, o que possibilita a inclusão de algoritmos para novos contextos de análises. Seu diferencial é a robustez do repositório de dados e o processamento de dados que possibilitam análises de grandes volumes de dados, de fácil instalação em qualquer rede de núcleo com arquitetura baseada em serviços (SBA).

3. O *framework*

O *framework* coleta e armazena dados em formato .pcap, analisa os dados, salva e expõe as informações obtidas na análise. A análise depende de algoritmos de IA que são treinados a partir de dados que pertencem a um contexto específico de análise. O desenvolvedor de contextos deve explorar esses dados e conhecer quais informações precisam ser extraídas e, dentre as técnicas existentes, qual ele deve aplicar para processar os dados. Ele deve utilizar uma interface de Experiência de desenvolvedor para editar códigos de processamento e análise de dados. O usuário interessado nas informações de análise consegue visualizá-las em um *dashboard*. O *framework* é adequado para redes 5G e futuras, devido à possibilidade de inclusão de novos algoritmos e à sua arquitetura baseada na tecnologia de contêineres. Isso o possibilita ser incorporado a qualquer outro sistema que esteja em Docker, por exemplo a um núcleo 5G. A arquitetura do *framework* é apresentada a seguir.

Para garantir a integridade e eficiência da infraestrutura da rede 5G foi aplicado o conceito de microsserviços utilizando a tecnologia de contêineres, em específico o Docker². Foram escolhidos os softwares *open-source* HDFS³, Flink⁴, Apache Zeppelin⁵, Grafana⁶ e PostgreSQL⁷. Os componentes integrados formam o *framework* que possibilita a inclusão de algoritmos para análise de dados de rede. A Figura 1 apresenta a arquitetura do *framework* com as funções lógicas do NWDAF definidas pelo 3GPP, exceto a Experiência de desenvolvimento. A descrição de cada componente da arquitetura é apresentada a seguir. **1. Interface de rede:** é a interface que recebe ou envia os pacotes do tráfego de rede do computador para a Internet. O *framework*, que pertence ao núcleo de rede 5G, conecta-se a esta interface. **2. Coletor de tráfego de rede:** responsável pela coleta de dados de tráfego da rede. Os dados no formato .pcap são coletados na **Interface de rede** e armazenados no **Repositório de dados**. **3. Repositório de dados:** armazena dados coletados pelo **Coletor de tráfego de rede**. Armazena e disponibiliza dados a qualquer momento e em qualquer quantidade. Os tipos de dados são: (i) para treino de

²<https://www.docker.com>

³<https://hadoop.apache.org>

⁴<https://flink.apache.org>

⁵<https://zeppelin.apache.org>

⁶<https://grafana.com/docs/>

⁷<https://www.postgresql.org>

modelos de IA, (ii) para processamento, e (iii) para análise. **4. Processamento e análise de dados:** responsável por analisar dados de rede armazenados no **Repositório de dados**. Ele executa algoritmos e armazena modelos de IA treinados no **Repositório de algoritmos e modelos de IA**. **5. Repositório de resultados de análises:** armazena os resultados das análises de dados feitas no Processamento e análise de dados. Disponibiliza dados para uma API de exposição de informações. **6. Exposição de informações:** apresenta informações oriundas do **Repositório de resultados de análise** em um *dashboard* para visualização pelo usuário. **7. Repositório de algoritmos e modelos de IA:** armazena algoritmos e modelos de análise prontos para execução, os quais estão no formato .json na pasta de instalação do *framework*. **8. Experiência de desenvolvimento:** interface para inclusão e edição de algoritmos no *framework* pelo desenvolvedor do contexto de análise.

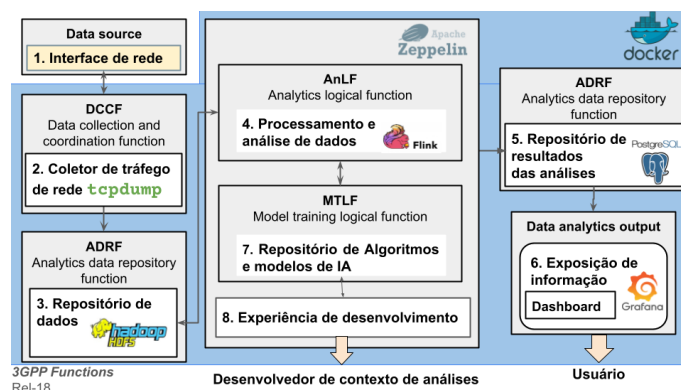


Figura 1. Arquitetura do *framework* de análise de dados de rede 5G relacionado com os componentes NWDAF apresentados pelo 3GPP e softwares *open-source*.

4. Técnicas de exploração de dados de tráfego de rede no *framework*

As técnicas de exploração de dados de tráfego consistem na apresentação das etapas para inclusão de algoritmos para novos contextos de análise no *framework*. Primeiramente, um contexto de análise deve ser definido pelo desenvolvedor de contextos. Depois, o desenvolvedor deve executar as etapas a seguir para desenvolver os algoritmos de análise. As técnicas estão a seguir.

1. Seleção de dados: escolha de um conjunto de dados relevantes para o contexto de análise. Os dados são utilizados para treinar o modelo de análise do *framework*. **2. Preparação dos dados:** definição de etapas de transformação do conjunto de dados para que sejam interpretados pelos algoritmos de aprendizado de máquina, além de conversão do arquivo .pcap para arquivo .CSV. **3. Processamento de dados:** especificação de técnicas de processamento de dados, como seleção de atributos ou componentes. **4. Seleção de algoritmos de aprendizado de máquina:** seleção de algoritmos apropriados aos dados e ao contexto de análise. **5. Treinamento e avaliação do modelo:** treinamento e avaliação do aprendizado de máquina usando o conjunto escolhido de dados para verificar a sua eficiência.

As técnicas são similares às técnicas de implementação de algoritmos de aprendizado de máquina. Isso facilita a utilização pelo usuário do *framework* que poderá adicionar novos algoritmos para diferentes contextos de análise no núcleo 5G e redes do futuro. A seguir, apresentamos o contexto de análise desenvolvido no *framework* para comprovar sua funcionalidade.

5. Um contexto de análise: detecção de anomalia de rede

A escolha do contexto de análise é a etapa que antecede o desenvolvimento dos algoritmos no *framework* proposto. No caso, escolhemos a detecção de anomalias na segurança cibernética 5G como o contexto de análise. A escolha deste contexto devida à sua importância, ao crescimento atual dos ataques cibernéticos e à necessidade da proteção do núcleo da rede 5G.

- **Seleção e preparação de dados:** O conjunto de dados escolhido, denominado 5GAD-2022, consiste em dados de rede 5G interceptados por um invasor que inclui dados de tráfego normal e malicioso capturados pelo software Wireshark em formato .pcap e disponibilizado no GitHub [Coldwell et al. 2022]. A preparação de dados é realizada da mesma forma que os autores dos dados fizeram em seu trabalho.
- **Processamento dos dados:** A Análise de Componentes Principais (PCA) foi utilizada para redução de dimensão, uma vez que é uma técnica estatística amplamente utilizada para análise exploratória de dados. A redução de dimensionalidade é capaz de simplificar modelos, reduzir o tempo de treino e reduzir o *overfitting*. Neste trabalho, utilizamos PCA de três componentes, ou seja, três novas variáveis que foram construídas como combinações lineares (ou misturas) das variáveis iniciais. Como não utilizamos métodos para seleção de recursos (*features*), o PCA transforma todos os recursos do conjunto de dados em componentes.
- **Seleção de algoritmos de aprendizado de máquina:** Selecionamos os algoritmos de IA que tiveram melhor desempenho em nossos trabalhos anteriores [Tavares et al. 2022] [Oliveira et al. 2023], para o contexto de análise escolhido em nosso *framework*, sendo eles o **Bagging Regressor (BR)** e o **Elliptic Envelope (EE)**.
- **Treinamento e avaliação dos algoritmos:** Para o *framework*, são utilizadas as métricas de avaliação em dois momentos: o primeiro é para a escolha do algoritmo de aprendizado de máquina não supervisionado; o segundo é para avaliação do modelo criado.

Os resultados da avaliação dos algoritmos são armazenados e, a partir daí, o modelo para detecção de anomalias de rede estará pronto para ser utilizado pelo *framework*. É importante ressaltar que esses algoritmos foram adicionados para detecção de anomalias de rede e, portanto, devem ser atualizados de acordo com o tipo de análise desejada.

6. Conclusão

Este trabalho apresentou a proposta de um *framework* para inclusão de algoritmos de análise de dados para redes 5G. O *framework* desempenha a função de de uma NWDAF - *Network data analytics function* especificada pelo 3GPP. O *framework* foi implementado utilizando softwares *open-source* e contêineres na integração. Com o um contexto de análise inicial, foram selecionados e integrados ao *framework* algoritmos de aprendizado máquina para detecção de anomalias de rede. O *framework* foi inicialmente concebido para a detecção de anomalias. Contudo, novos algoritmos podem ser adicionados ao *framework* para outros contextos de análise de dados. A inclusão é possível devido à interface que possibilita a “Experiência de desenvolvedor” com uma interface conhecida, o Apache Zeppelin, para que os seus usuários com conhecimentos em linguagem

de programação Python e contêiners possam desenvolver outros algoritmos. A Exposição de informação no Grafana é por meio de um *dashboard* que no futuro fará ações na rede para uma gestão automática da rede. O *framework* pode ser facilmente incorporado ao núcleo de 5G SBA como uma nova função de rede devido à sua arquitetura em Docker. A continuidade imediata do trabalho é a sua avaliação funcional e eficácia na detecção de ataques, bem como na execução/sugestão de ações na rede 5G decorrentes das análises.

Referências

- 3rd Generation Partnership Project (3GPP) TS 29.520. "5g system; network data analytics services, version 18.3.0 release 18". https://www.3gpp.org/ftp/Specs/archive/29_series/29.520/29520-i30.zip. Acesso em outubro de 2023.
- Abbas, K., Khan, T. A., Afaq, M., Diaz Rivera, J. J., and Song, W.-C. (2021). Network data analytics function for ibn-based network slice lifecycle management. In *2021 22nd Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pages 148–153.
- Aumayr, E., Caso, G., Bosneag, A.-M., Zayas, A. D., Özgü Alay, Garcia, B., Kousias, K., Brünstrom, A., Gomez, P. M., and Koumaras, H. (2022). Service-based analytics for 5g open experimentation platforms. *Computer Networks*, 205:108740.
- Coldwell, C., Conger, D., Goodell, E., Jacobson, B., Petersen, B., Spencer, D., Anderson, M., and Sgambati, M. (2022). Machine learning 5g attack detection in programmable logic. In *2022 IEEE Globecom Workshops (GC Wkshps)*, pages 1365–1370.
- Donertasli, H. and Medithe, M. (2022). Nwdaf udi (use-case development interface) for end-to-end ai enabled 5g and beyond networks. In *2022 International Conference on Artificial Intelligence of Things (ICAIoT)*, pages 1–6.
- Ma, C. and Chi, Y. (2022). Knn normalized optimization and platform tuning based on hadoop. *IEEE Access*, 10:81406–81433.
- Oliveira, J. M., Almeida, J., Macedo, D., and Nogueira, J. M. (2023). Comparative analysis of unsupervised machine learning algorithms for anomaly detection in network data. In *2023 IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1–6.
- Perez, R., Garcia-Reinoso, J., Zabala, A., Serrano, P., and Banchs, A. (2020). A monitoring framework for multi-site 5g platforms. In *2020 European Conference on Networks and Communications (EuCNC)*, pages 52–56.
- Samaras, G., Theodorou, V., Laskaratos, D., Psaromanolakis, N., Mertiri, M., and Valantasis, A. (2022). Qmp: A cloud-native mlops automation platform for zero-touch service assurance in 5g systems. In *2022 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*, pages 86–89.
- Tavares, I., Oliveira, J. M., Teixeira, A. F., de Arruda Pereira, M., Kakitani, M. T., and Nogueira, J. M. (2022). Machine learning algorithms applied to telemetry data of scd-2 brazilian satellite. In *Proceedings of the 2022 Latin America Networking Conference, LANC '22*, page 50–58, New York, NY, USA. Association for Computing Machinery.