Exploring time series xAI techniques for 5G/6G Networks

Fábio Silva[1][0000−0001−9872−7117], José Ribeiro[1][0000−0002−2457−1567], Cesar Analide[2][0000−0002−7796−644X], and Ricardo Santos[1][0000−0002−2139−5414]

[1] CIICESI, ESTG, Insituto Politécnico do Porto
{jfr,rjs,fas}@estg.ipp.pt
[2] Department of Informatics, ALGORITMI Center, University of Minho, Braga, Portugal
analide@di.uminho.pt

Abstract. This study explores the applicability of explainable artificial intelligence (xAI) techniques in the analysis of deep learning models for anomaly detection in 5G/6G networks. With the increasing complexity of networks and network traffic, the mission to guarantee the security access points and devices against attacks and intrusions is also larger. Models used for these tasks operate like black boxes, making it difficult to understand and interpret their decisions at a human level. To address this challenge, we devised a case study with a real world dataset and a performant deep learning anomaly detection algorithm and implemented strategies to generate humam understandable explanation through xAI algortihms. xAI can provide insights into the factors that lead to the detection of anomalies, allowing for greater transparency and reliability in the process. This work is part of the context of intelligent networks and is aligned with initiatives such as the Privateer project, contributing to the evolution of security in 5G/6G infrastructures. The integration of deep learning and xAI facilitates interaction between human operators and automated systems, promoting greater control over decision-making in modern networks.

Keywords: 5G/6G networks· anomaly Detection · xAI

1 Introduction

Artificial Intelligence (AI) has addressed challenges in 5G and 6G networks, particularly in traffic monitoring and anomaly detection. Deep learning models have shown effectiveness in identifying patterns that may suggest faults, congestion or attacks [7]. However, these models are opaque, making it difficult to understand their decisions [4]. The interpretability of AI models, known as eXplainable Artificial Intelligence (xAI), has been an area of great interest in the scientific and industrial community. The goal of xAI is to increase the transparency of deep learning models, making their predictions understandable to human operators

and network experts [6]. In 5G/6G network security and optimization applications, explainability plays a role in providing justification for anomaly detections and recommendations generated by algorithms. Techniques such as SHAP (SHapley Additive Explanations) [10] and LIME (Local Interpretable Model-agnostic Explanations) [14] have been widely used to provide insights into the contribution of each variable to the final result of the models. There is however, a time dependency that most models for network traffic anomaly detection implement as a timeseries problem which also require personalization of common xAI algorithms.

In this study, we explore the application of interpretability methods to the analysis of traffic anomalies in 5G and future 6G networks. In particular, we investigate how xAI techniques can be used to explain decisions made by deep learning models that detect anomalous patterns in network traffic. The problem addressed involves the analysis of traffic filtered at access points, where it is possible to distinguish between normal behavior and potential intrusion attempts or cyber attacks.

2 State of the Art

Advances in artificial intelligence techniques have led to the development of increasingly complex models, applied in various fields, including intelligent communication networks. In the context of 5G networks and future 6G infrastructures, deep learning models have been used to analyze and monitor traffic, enabling the detection of patterns and possible anomalies. However, the opaque nature of these models represents a significant challenge for their adoption in these scenarios, where transparency of decisions is useful [14], [12].

Explainable artificial intelligence (xAI) has emerged as a key field for increasing the interpretability of models, providing the means to understand the reasons behind the predictions made by algorithms. Techniques such as SHAP (SHapley Additive Explanations), a method based on Shapley values from game theory that allows the influence of each variable on the model's final prediction to be quantified, and LIME (Local Interpretable Model-agnostic Explanations) have been widely used to reveal the influence of different variables on the decisions of anomaly detection models. In addition, strategies based on autoencoders and attention mechanisms have been explored to improve model transparency in modern network scenarios.

In this context, the applicability of xAI in 5G/6G networks enables greater control over intelligent infrastructures, allowing effective interaction between human operators and automated systems, promoting security and reliability in decision-making processes.

2.1 Anomaly dectection in Network Traffic

Detecting anomalies in 5G/6G networks is a current challenge, given the increasing complexity and dynamism of data traffic. Different approaches have been

explored for this purpose, including methods based on deep learning, statistical techniques and hybrid models. Deep neural networks, such as autoencoders and attention-based models, have proven effective in identifying anomalous patterns by learning latent representations of traffic data. Models such as LSTM(Long Short-Term Memory), a type of recurrent neural network (RNN) designed to capture long-term temporal dependencies in time series, [5], [13] and GRU (Gated Recurrent Unit), a simplified variant of LSTM that also captures temporal patterns in sequential data with less computational complexity, have been widely used to capture temporal dependencies in mobile network scenarios, enabling early detection of attacks such as DDoS and malicious intrusions [8], [3].

In addition to deep learning-based techniques, statistical approaches such as threshold rules and clustering algorithms (e.g. DBSCAN and k-means) are often used to identify atypical behavior. However, the opaque nature of deep learning models poses challenges in terms of interpretability.Explainability methods such as SHAP and LIME have been applied to understand the decisions of these algorithms, allowing network operators to identify the most influential factors in detecting anomalies. xAI enhances transparency and monitoring reliability in 6G networks.

2.2 Explicability in Machine Learning Models

The explainability of machine learning models has become useful for ensuring transparency and reliability, especially in sensitive applications such as telecommunications and network security [1]. Explainable artificial intelligence (xAI) seeks to make models more comprehensible to humans, allowing the interpretation of decisions made by "black box" models such as deep neural networks. In the context of traffic analysis in 5G/6G networks, where models need to detect anomalies in continuous data flows, explainability should try to help understand situations.

The interpretation of models applied to time series is still an open challenge. Most xAI techniques were originally developed for conventional supervised models, which requires adaptations when applied to continuous data streams [12]. However, methods such as SHAP and LIME have been used to evaluate the influence of input variables on the predictions of anomaly detection models in networks [10], [14]. Other approaches include gradient-based methods such as Integrated Gradients and Layer-wise Relevance Propagation (LRP), which analyze the model's sensitivity to variations in inputs, and visualization techniques such as saliency maps and feature visualization, widely used in convolutional networks but adaptable to temporal networks [15], [17]. Although not applied here, these approaches can be integrated into the modular architecture designed to support various xAI techniques.

Applying these techniques to 5G/6G time dependent network traffic anomaly detection models provides a better understanding of the decisions made, providing justification for detecting attacks or anomalous patterns. The use of interpretability techniques in 5G/6G networks is useful due to the complexity of these infrastructures [...] making the decisions of these models more understandable

to network operators and experts. 6G implementation proposals, such as in [11], and [9] have explored the integration of xAI techniques into intelligent telecommunications systems, with the aim of improving the transparency of anomaly detection algorithms in high-connectivity environments.

3 Explainable anomaly Detection in 5G/6G Networks

Growing connectivity and device diversity make securing mobile networks a relevant challenge. The complexity of traffic patterns and the need for real-time monitoring require advanced solutions that combine machine learning and explainable artificial intelligence (xAI) techniques to make the models more understandable and reliable for operators and experts. The architecture proposed, figure 1 consists of flexible microservices, containerized, responsible for every step for the production of xAI insights through different techniques and strategies.

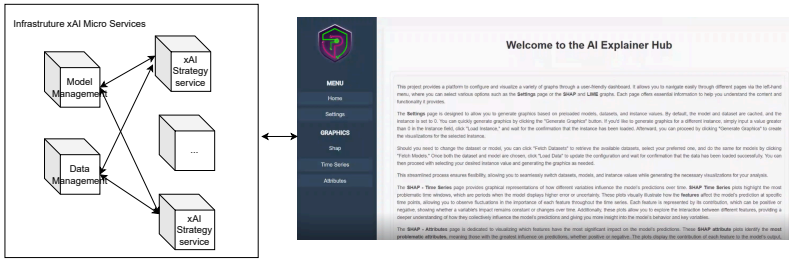

Fig. 1: Architecture

The architecture developed consists of a set of microservices, organized to ensure flexibility and scalability in the explainable analysis of anomalies. The microservices are divided by function: (1) a database management microservice, responsible for managing the data collected from the 5G infrastructure; (2) a model management microservice, responsible for loading the trained models and making them available for analysis; and (3) microservices dedicated to explainability techniques. This structure allows for modular integration, facilitating the future introduction of new detection algorithms or xAI techniques.

3.1 Dataset and anomaly Detection Models

The test environment consisted of two cells with a total of nine UEs connected to the same core network. The 5G network infrastructure was implemented with the Amarisoft Callbox Mini and Amarisoft Classic solutions, the latter responsible for hosting the core of the 5G network. The set of UEs included Huawei P40 smartphones, Raspberry Pi 4 microcomputers equipped with 5G modules, industrial 5G routers, WiFi-6 mobile hotspots and a CPE Box. During data recording periods, the devices generated network traffic in a programmed way, simulating

real communication patterns producing the NCSRD-DS-5GDDoS dataset [16]. This dataset contains detailed radio and core metrics from 5G networks, including sporadic Distributed Denial of Service (DDoS) attacks initiated by malicious users (UEs).

The anomaly detection model used was trained with an autoencoder architecture based on the LSTM deep learning algorithm, using a pre-processed version of the NCSRD-DS-5GDDoS [2] dataset. Training was carried out with benign test data from the dataset, organized in a time series format by device, selecting the 8 most relevant variables in time windows of 120 instants. Each input vector was thus built with 960 values (120×8), representing the temporal context of each device. For interpretability purposes, during the application of the SHAP technique, the vector is again segmented into its original components, making it possible to analyze the influence of each variable over time and identify the instants with the greatest impact on the model's decision. The model identified anomalies with 99% accuracy.

3.2 xAI strategies for 5G/6G time series environements

This study uses a pre-trained model to apply interpretability techniques. The approach adopted used SHAP exploring the nature of the anomaly detection algorithm and the properties of the timeseries dataset to devise strategies that can help explain the influence of input features accross data windows to the final detection. The model analyzed takes as input a time window of 120 instants, with 8 attributes per instant, resulting in an input vector of dimension 960. Each instance analyzed by XAI corresponds to this structure, showing how temporal variations of features influence the model's anomaly detection.

Two main strategies were implemented to explain the predictions. Figure 2 shows two different approaches to explaining the model's decisions. The image on the left represents the "Loss by Attribute" strategy, which evaluates the average impact of each attribute over the time window. The image on the right represents the "Loss per Instance" strategy, which assesses the influence of a specific variable over multiple instances of the time series. Both representations aim to visually illustrate the contribution of temporal data to the model's decision.

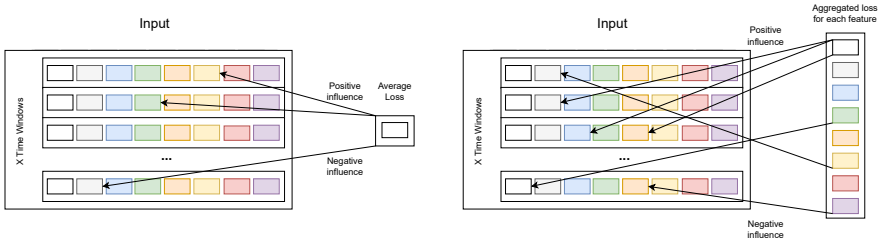

Fig. 2: xAI timeseries strategies

3.3 Analysis of Influence by Attribute (Loss by Attribute)

In this section, we present the first xAI-based explanation approach, in which we analyze the global contribution of all attributes in anomaly detection. To do this, we use techniques such as SHAP to calculate the influence of each variable over the time sequence and assess its relevance in predicting the model. This approach makes it possible to see which attributes, in general, have the most influence on the model's decision [10]. As a result, it was possible to see that certain attributes, such as cell_x_ul_retx and cell_x_ul_tx, had a significantly greater impact on the classification of anomalies, while other attributes had a lesser influence. In addition, it was observed that the importance of certain variables can vary over time, reflecting changes in network traffic patterns.

These 5G radio-layer metrics help interpret traffic and identify anomalies. The cell_x_ul_retx variable refers to the number of retransmissions on the uplink, which can indicate failures or congestion in data transmission. Cell_x_ul_tx represents the total volume of data sent on the uplink, while cell_x_dl_tx refers to the volume of data transmitted on the downlink, both direct indicators of network activity and performance. The variables dl_total_bytes_non_incr and ul_total_bytes_non_incr indicate the volume of data that has not increased over a period of time, and may be associated with periods of inactive traffic or irregular patterns that suggest anomalous behavior. Explaining these attributes clarifies their influence and justifies the model's output.

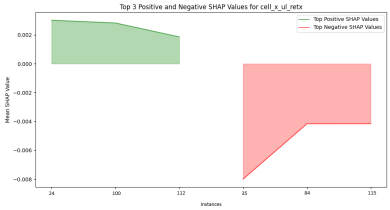

Fig. 3: Top 3 SHAP values for cell_x_ul_retx (Area chart)

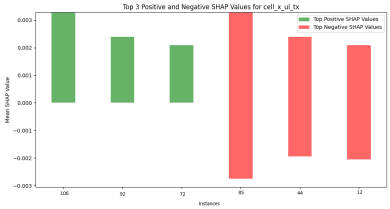

Fig. 4: Top 3 SHAP values for cell_x_ul_tx (Bar chart)

Figures 3 and 4 show the three main attributes with the greatest influence on the model's prediction. These graphs illustrate the positive and negative contribution of the attributes selected on the basis of the SHAP values. Figure 3 shows the cell_x_ul_retx variable in the time windows (24, 100, 112, 35, 84, 115). The instances with a positive impact on the model's prediction are 24, 100 and 112, while instances 35, 84 and 115 have a negative impact. The area graph visually highlights these differences, where green areas represent positive contributions and red areas represent negative contributions. The slope of the areas shows the magnitude of the influence of each instance.

Figure 4 shows the cell_x_ul_tx variable in the time windows (106, 92, 72, 85, 44, 12). Instances 106, 92 and 72 have a positive impact on the model's

forecast, while instances 85, 44 and 12 contribute negatively. The height of the bars indicates the magnitude of the influence of each instance, with green bars representing positive SHAP values and red bars representing negative values.

This analysis shows that the variables associated with uplink traffic (ul_tx and ul_retx) significantly influence the model's predictions and are determining factors in detecting network anomalies. This information can be used to improve the interpretation of the model's decisions and refine future analyses.

3.4 Analysis of Influence by Instance (General Loss)

In this section, we present the second explanation approach applied in this study, which focuses on the influence of a specific variable over the different instances analyzed. Instead of considering all attributes simultaneously, this strategy aims to understand how a single attribute impacts the model's decision in different time windows [14].

This analysis makes it possible to check the variability of an attribute's importance at different points in the time series. For example, the variable cell_x_dl_tx was found to have a high impact in certain time windows associated with traffic peaks, indicating a strong correlation with anomalous events.

In this way, the influence-by-attribute approach offers a detailed perspective on the behavior of individual variables, helping to identify specific patterns that may be related to the occurrence of anomalies.

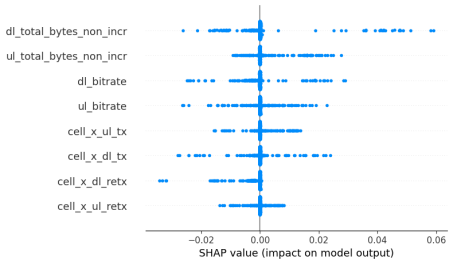

Fig. 5: Scatterplot of SHAP values for different attributes

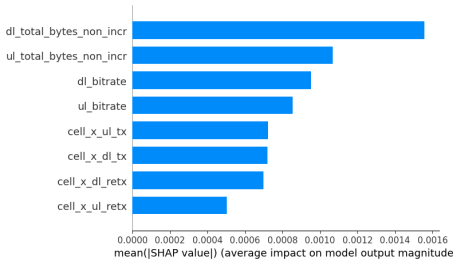

Fig. 6: Average impact of attributes on the model based on SHAP values

Figure 5 and figure 6 illustrate the importance of the attributes for predicting the model in instance 0, considering the SHAP values. In figure 5, each point represents a SHAP value associated with a specific attribute. The distribution of points along the horizontal axis indicates the relative impact of each variable on the model's decision. It can be seen that attributes such as dl_total_bytes_non_incr and ul_total_bytes_non_incr show greater variability, suggesting that they have more influence in predicting anomalies. In figure 6 shows the average absolute value of the SHAP coefficients for each variable, highlighting the attributes with the greatest overall influence. It can be seen that the attributes dl_total_bytes_non_incr and ul_total_bytes_non_incr

have the highest impacts, while variables such as cell_x_ul_retx have a lower influence. The analysis highlights how attribute impact varies over time, aiding interpretation and anomaly detection.

3.5 Critical Review

In the first strategy, the analysis of a specific variable throughout the instances revealed interesting patterns. For example, when selecting the variable dl_bitrate, it was observed that its influence varied according to the instance analyzed. In certain instances, this variable had a very high impact on predicting the anomaly, while in others its effect was less significant. This behavior suggests that the relevance of the variables may be sensitive to the context of the instance analyzed, reinforcing the need for interpretive techniques to understand the dynamics of the model.

In the second strategy, it was possible to observe that certain variables had a consistent impact throughout the time sequence. For example, the variable dl_bitrate proved to be one of the most influential in predicting anomalies, indicating that significant variations in the downlink transmission rate may be strongly associated with unusual events in the network. Other variables, such as ul_bitrate and cell_x_dl_retx, also had a significant influence, standing out at specific moments in the time sequence as determining factors for the model's classification.

These interpretability techniques improve transparency, allowing experts to understand and validate automated decisions in 5G/6G networks. The combination of these two strategies allowed for an in-depth analysis of the model's behavior, providing valuable insights into the factors that most influence anomaly detection in 5G/6G networks. The results obtained demonstrate the importance of explainability in machine learning models, especially in scenarios such as network security, where the correct interpretation of predictions can help in decision-making and risk mitigation.

The evaluation of the explanations considered the SHAP values, assuming that values close to zero indicate irrelevance and values close to one reflect high influence, thus allowing an estimate of the fidelity of the explanations generated. Although the detection model used only returns a binary classification (anomaly or normal), the introduction of explainability techniques such as SHAP has made it possible to overcome this limitation, providing a detailed view of the variables that influenced each decision. This capability represents an added value compared to traditional approaches without interpretability, which do not make it possible to understand the factors behind the classifications. Although not compared directly, the architecture enhances model understanding and supports real-world use where transparency matters.

4 Conclusion and Future Work

The implementation of xAI strategies enabled an in-depth analysis of the interpretability of the LSTM model applied to anomaly detection in 5G/6G networks.

Two main approaches were applied throughout this study: loss per attribute, which assesses the relative importance of each variable over the time series, and general loss, which measures the relevance of attributes in identifying anomalous patterns.

The first approach identified key variables by analyzing their contribution over the 960-entry window. It was observed that certain attributes had a considerable impact at specific moments in the time series, directly influencing the model's decisions. This understanding is intended to improve the explainability of the model and direct efforts towards improving the quality of the data and pre-processing techniques. In the second approach, the importance of attributes in detecting anomalies was analyzed globally, making it possible to see which variables were most strongly associated with identifying anomalous events. This strategy highlighted the relationship between the influence of certain attributes and the occurrence of unusual patterns, providing valuable insights into the factors that contribute to the classification of an anomaly.

This study opens the way to several possibilities for future work which include analysing the importance of each data windows towards the final outcome. Another aspect for future research is adapting the model to different traffic and attack scenarios in 5G/6G networks. Validating xAI strategies in new contexts could help improve the robustness of the model and its applicability in real systems. This reinforces the importance of interpretable models in security contexts, supporting trust in automated systems and informed decision-making. Future work will evaluate latency, overhead and scalability for deployment in real-time distributed systems.

Acknowledgments. This work was performed in the scope of the Smart Networks and Services Joint Undertaking (SNS JU) under the EU Horizon Europe programme PRIVATEER under Grant Agreement No. 101096110, This work has also been supported by national funds through FCT – Fundação para a Ciência e Tecnologia through project UIDB/04728/2020.

References

1. Adadi, A., Berrada, M.: Peeking inside the black-box: A survey on explainable artificial intelligence (xai). IEEE Access **6**, 52138–52160 (2018). https://doi.org/10.1109/ACCESS.2018.2870052
2. Argyriou, L., Karamatskou, A.: Privateer deliverable d3.1: Decentralised robust security analytics enablers - rel.a. (Jun 2024). https://doi.org/10.5281/zenodo.12530825, https://doi.org/10.5281/zenodo.12530825
3. Cho, K., Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation (06 2014). https://doi.org/10.3115/v1/D14-1179
4. Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv: Machine Learning (2017)

5. Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R., Schmidhuber, J.: Lstm: A search space odyssey. IEEE Transactions on Neural Networks and Learning Systems **28**, 2222–2232 (10 2017). https://doi.org/10.1109/TNNLS.2016.2582924
6. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM Computing Surveys **51** (9 2018). https://doi.org/10.1145/3236009
7. Heaton, J.: Ian goodfellow, yoshua bengio, and aaron courville: Deep learning. Genetic Programming and Evolvable Machines 2017 19:1 **19**, 305–307 (10 2017). https://doi.org/10.1007/S10710-017-9314-Z
8. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**, 1735–1780 (11 1997). https://doi.org/10.1162/NECO.1997.9.8.1735
9. Liyanage, M., Porambage, P., Zeydan, E., Senavirathne, T., Siriwardhana, Y., Yadav, A.K., Siniarski, B.: Advancing security for 6g smart networks and services pp. 1169–1174 (2024). https://doi.org/10.1109/EuCNC/6GSummit60053.2024.10597131
10. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 4768–4777. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (2017)
11. Masouros, D., Soudris, D., Gardikis, G., Katsarou, V., Christopoulou, M., Xilouris, G., Ramón, H., Pastor, A., Scaglione, F., Petrollini, C., Pinto, A., Vilela, J.P., Karamatskou, A., Papadakis, N., Angelogianni, A., Giannetsos, T., Villalba, L.J.G., Alonso-López, J.A., Strand, M., Grov, G., Bikos, A.N., Ramantas, K., Santos, R., Silva, F., Tsampieris, N.: Towards privacy-first security enablers for 6g networks: The privateer approach pp. 379–391 (2023). https://doi.org/https://doi.org/10.1007/978-3-031-46077-7_25
12. Molnar, C.: Interpretable Machine Learning. Leanpub (2020)
13. Olah, C.: Understanding lstm networks (2015), https://research.google/pubs/understanding-lstm-networks/
14. Ribeiro, M.T., Singh, S., Guestrin, C.: "why should i trust you?" explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining **13-17-August-2016**, 1135–1144 (8 2016). https://doi.org/10.1145/2939672.2939778
15. Samek, W., Wiegand, T., Müller, K.R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models (2017), https://arxiv.org/abs/1708.08296
16. of Scientific Research "Demokritos", N.C., (Greece), S.H.: Ncsrd-ds-5gddos: 5g radio and core metrics containing sporadic ddos attacks. https://doi.org/10.5281/ZENODO.10671494, https://zenodo.org/records/10671494
17. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. p. 3319–3328. ICML'17, JMLR.org (2017)