Survey on intrusion detection systems in 5G

Sara Chennoufi[1], Gregory Blanc[1], Houda Jmila[1], and Christophe Kiennert[1]

[1]SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France
{firstname.surname}@telecom-sudparis.eu

***Abstract*—5G is the new generation of mobile networks intended to improve the performance of earlier generations while integrating a variety of use cases with varying requirements into a single network. With such a wide coverage and its pervasiveness into society, it is crucial to assess the cyber-risks inherent to its implementations. In particular, we question the suitability of network monitoring solutions such as intrusion detection systems with regards to 5G requirements.**

This survey presents an analysis of intrusion detection systems in contexts related to 5G networks and propose a taxonomy to determine a set of suitable features. Among these features, it motivates the need for collaboration in order to overcome some of the challenges imposed by 5G networks such as heterogeneity or low-latency. We study Federated Learning (FL) as a candidate to enable collaboration in intrusion detection for 5G networks and discuss future research directions.

***Index Terms*—5G, intrusion detection system, machine learning, network security, federated learning.**

I. Introduction

5G is the new mobile network generation designed to enable new services and connect almost everything and everyone while providing the highest possible performance to meet each use case needs, which are grouped into three classes [1]: Enhanced Mobile Broadband Connectivity (eMBB) which represents a continuity of 4G with improved performance, Massive Machine Type Communications (mMTC), which includes applications with high devices density that needs energy optimization, and Ultra-Reliable Low-Latency Communication (URLLC) which is used for mission-critical services that require a real-time connection and high reliability.

Beyond providing increased performance, 5G is expected to accommodate a wider range of use cases on top of a common physical network, thanks to a set of enablers [2]. These include slicing, which allows for different services with varying requirements to use the same network; software-defined networking (SDN) for flexible network management; multi-access edge computing (MEC) for improved latency and performance by bringing computing closer to the user; and massive MIMO for scalability using time-division duplexing (TDD) and spatial multiplexing.

As 5G is rolled out, it is important to evaluate potential cyber risks due to its complex technology stack and increased attack surface. Security incidents could have severe consequences, as 5G will be used for critical applications. Previous security measures may not be adequate for the specific requirements of 5G, such as heterogeneity, high connection density, and low latency. In particular, we are interested in determining whether existing IDS can satisfy the challenges of 5G networks, which are highly distributed, connecting heterogeneous devices across domains [3].

Therefore, current works are increasingly interested in the development of collaborative IDS [4]–[6] to (i) enhance performance by training on more data collected from multiple points, (ii) enable information sharing and detect new attacks through the utilization of threat intelligence, leading to the anticipation of attacks that happened elsewhere or that may be coordinated, (iii) reduce false positive rate by preventing overfitting on small datasets learned in isolation, and (iv) enable collaboration between different parts of the same system. However, mixing data from different sources raises privacy issues. Federated Learning (FL) is a such collaborative machine learning framework that supports privacy-preserving aggregation. Although successful, FL also comes with its deployment challenges.

In the remainder of this paper, we will study and classify a subset of our survey of IDS in 5G related contexts, and the potential of FL to enforce collaboration.

II. Survey

This section presents a taxonomy to classify and compare works from the literature on IDS in 5G contexts as shown in Table I. We collected related papers from well-known scientific libraries, to which we added works on IDS related to 5G technologies, characteristics, and challenges, as well as articles from the selected papers' references. For brevity, in this paper, we only provide a couple of examples for each column as follows.

A. IDS Objective

5G IDS is designed to secure a 5G use case with respect to its characteristics or to detect attacks related to an enabling technology. For example, for the IoT use case, in mMTC class, Fan et al. [5] proposed an IDS that solves three related problems: (i) power and resource constraints by using MEC platforms, (ii) heterogeneity by using transfer learning to get a personalized model for each IoT network and (iii) privacy issues by using Federated Learning, which will be detailed later. For enabling technologies, Kuadey et al. [9] proposed a framework to detect DDoS attacks in 5G network slicing systems and Alamri et al. [12] in the SDN controller.

B. IDS requirements

5G introduces new challenges that have been considered in designing an IDS, such as privacy which will be detailed in the

TABLE I
Classification of some representative state of the art papers based on the proposed taxonomy

Research paper	IDS Objective							IDS requirements						Deployment		Detection method				Evaluation			
	5G class				Technology			5G related challenges						Architecture		Machine Learning model				Dataset		Metric	
	mMTC	URLLC	eMBB	All/ Not specific	Slicing	SDN	MEC	Big Data	Privacy	Robustness	Personalization	Mobility	Adaptivness	Collaborative	Not collaborative	Shallow supervised	Shallow unsupervised	Deep supervised	Deep unsupervised	General	Specific	ML metric	System performances
[5]	x						x		x		x			x				x		x	x	x	
[7]		x							x					x				x		x		x	
[8]	x														x		x			x		x	
[9]				x	x									x				x		x		x	
[10]				x		x								x		x	x			x		x	x
[11]				x				x					x	x					x	x		x	x

next section, personalization [5], Big Data and heterogeneity [8], and robustness [13].

C. Architecture

The objectives and constraints lead to a target deployment architecture for monitoring selected threats. IDS can be collaborative or non-collaborative. Collaborative IDS can be centralized, decentralized, or distributed. Centralized IDS sends data to a central unit for analysis but lacks scalability (single point of failure). Decentralized IDS eliminates this issue by using a hierarchy of processing units. Distributed IDS uses a network of analysis units in a peer-to-peer manner.

D. Detection model

Selecting an appropriate detection algorithm requires consideration of specific use case requirements, and adapting or combining algorithms accordingly. The survey emphasizes ML models, which can learn from large datasets to detect new and complex attacks. ML techniques fall into categories of supervised or unsupervised and shallow or deep learning. Supervised learning is easier to train and evaluate, but requires labeled data. In contrast, unsupervised learning is better with big data, but is harder to train and evaluate. Deep learning is effective with complex data but requires more resources, while shallow learning is faster but less effective.

E. Evaluation

Evaluating IDS involves measuring commonly used metrics after submitting a data set. Typically, general-purpose network intrusion detection data sets are used (KDD, CTU, UNSW, etc.), but few papers utilize specific 5G use case data sets, such as [5] Most publications evaluate general ML metrics [7], such as accuracy, precision or recall. False Positive Rate (FPR) was less used even if it represents an important metric to evaluate IDS. The system performance is often evaluated in terms of execution time, the number of packets analyzed per second and memory consumption for example.

F. Results

This study found that many articles do not specify the 5G class or use case they study and use datasets that do not represent the 5G use case. The metrics used are mostly ML metrics, while system metrics that are crucial to evaluate 5G Key Performance Indicators and requirements are not considered. FPR is seldom considered, and some challenges of 5G IDS such as mobility have not been thoroughly studied. The commonly used ML models are deep supervised models because most used datasets are labelled and necessitate complex models. These models perform better with big data, which may not be achievable locally. This leads to a trend towards collaborative IDS aiming to mutualize enough data to enhance performance, reduce FPR, and detect new attacks.

III. Towards Federated Learning for collaborative 5G IDS

As we have seen in the previous section, different users need to collaborate to form better IDS in terms of performance and ability to detect unknown attacks. However, this raises a major issue in areas where privacy is critical, such as in many 5G services (healthcare, smart grids, etc.). To solve this, FL seems to be a suitable candidate: it builds a global model from a number of local client models that were trained on client devices containing private data. FL preserves data privacy and reduce communication costs compared to centralized collaboration where all user's data are sent to a central unit (rather than only the model updates that are sent in FL). Its architecture consists of a number of clients that have private data and a central server. FL is an iterative process where each client trains its local model on its own data, then sends that model to the server for aggregation. The server sends the global model (aggregation result) to the clients, which use it to update their local models. And the process repeats.

In the literature, only a few studies have used FL in 5G IDS. FL was used in 5G smart metering networks, where FL clients were Local-ID, and the server was a concentrator in the neighboring area network [14]. FL was also used in IoT networks, where data from each IoT network was sent to the MEC platform as FL clients, and the 5G security cloud platform was the FL server [5]. Sun et al. [7] proposed a hierarchical FL system where smart grids were low-level FL clients, base stations were upper layer clients that grouped lower-level clients, and the FL server was at the top.

Although FL usage has been motivated, among other advantages, by its privacy-preserving aggregation mechanism, some open issues remain, including: (i) privacy itself when malicious users try to get information on other users' private data using inference attacks; (ii) robustness where attackers try to poison the global model by modifying their local models or data; (iii) heterogeneity, in devices, systems, and data that leads to decreased model performance and increased FPR; the main difficulty is creating high-quality models with non-IID (independent and identically distributed) and inaccessible data; (iv) synchronization, between devices that have different training time and resources; (v) client selection, which is the process of selecting the best clients to participate in each FL round to maximize learning efficiency and decrease convergence time; (vi) communication overhead, which arises from frequent model updates over a large number of iterations; (vii) fairness, which means preventing any group from being discriminated against; (viii) hyper parameters adjustment and optimization, by determining the optimal parameters that align with the studied use case, desired performance, and 5G characteristics; (ix) differences in confidentiality and security requirements between participants.

IV. Open issues and future research directions

After the study of different solutions proposed in the state of the art in 5G IDS, we found that this field remains open due to the following open problems.

With respect to *data sets*, Most publications use established IDS datasets such as KDD, CTU, and UNSW, which were not collected from 5G networks and do not contain new 5G attacks. These datasets are not reflective of 5G enabling technologies, unique characteristics, and performance. Researchers need appropriate datasets to evaluate IDS effectiveness in a 5G network. The lack of such datasets poses a real challenge that can be addressed by either requesting real data from ISPs, which raises privacy concerns, or by simulations.

The need for cross-domain IDS in 5G is due to device, data, and model heterogeneity challenges and the need for collaboration between different domains that we have seen previously. However, some techniques like transfer FL are proposed to train the model collaboratively and then personalize it [5]. Other methods like Multi-task Learning, Meta-Learning, and Knowledge Distillation are proposed in [15]. Some of those techniques have not been used in a 5G scenario. Exploring them to personalize 5G IDS represents an interesting research direction. A hierarchical personalization by 5G slice and subslice can also be an open direction to explore.

Real-time IDS is necessary for most 5G applications, mainly when reliability and security are highly required (URLLC). Proposing an IDS that can detect attacks in real-time with respect to the required latency represents an open challenge.

In many 5G applications, such as medecine, it is necessary to protect data privacy. While FL can be used, it may still suffer from privacy leakage, so complementary techniques can be used, such as Homomorphic Encryption, Multi-Party Computation and Differential Privacy.

Other notable open issues related to the development of FL in 5G IDS concern its security, robustness and explainability.

V. Conclusion

In this paper, we summarized a few results from review of IDS in 5G networks. We established the need to implement specific IDS for 5G networks because of its inherent characteristics, use cases, and enabling technologies, which increase the attack surface and introduce new challenges in the IDS research field. Then, using a new taxonomy, we evaluated and classified existing 5G IDS works. We studied FL IDS in 5G and evaluated the challenges that remain in this field because FL represents an interesting research orientation due to the necessity for collaboration and privacy protection in 5G. Finally, based on the literature, we suggested potential study directions. We intend to create a hierarchical, personalized cross-domain FL IDS for the 5G network in the future.

References

[1] D. Jiang and G. Liu, "An overview of 5g requirements," *5G Mobile Communications*, pp. 3–26, 2017.

[2] B. Blanco, J. O. Fajardo, I. Giannoulakis, E. Kafetzakis, S. Peng, J. Pérez-Romero, I. Trajkovska, P. S. Khodashenas, L. Goratti, M. Paolino *et al.*, "Technology pillars in the architecture of future 5g mobile networks: Nfv, mec and sdn," *Computer Standards & Interfaces*, vol. 54, pp. 216–228, 2017.

[3] X. Li, S. Zhao, C. Chen, and Z. Zheng, "Heterogeneity-aware fair federated learning," *Information Sciences*, vol. 619, pp. 968–986, 2023.

[4] R. Sharma, C. A. Chan, and C. Leckie, "Evaluation of collaborative intrusion detection system architectures in mobile edge computing," *Mobile Edge Computing*, pp. 359–384, 2021.

[5] Y. Fan, Y. Li, M. Zhan, H. Cui, and Y. Zhang, "Iotdefender: A federated transfer learning intrusion detection framework for 5g iot," in *2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE)*. IEEE, 2020, pp. 88–95.

[6] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated learning of deep networks using model averaging," *arXiv preprint arXiv:1602.05629*, vol. 2, 2016.

[7] X. Sun, Z. Tang, M. Du, C. Deng, W. Lin, J. Chen, Q. Qi, and H. Zheng, "A hierarchical federated learning-based intrusion detection system for 5g smart grids," *Electronics*, vol. 11, no. 16, p. 2627, 2022.

[8] N. Hu, Z. Tian, H. Lu, X. Du, and M. Guizani, "A multiple-kernel clustering based intrusion detection scheme for 5g and iot networks," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 11, pp. 3129–3144, 2021.

[9] N. A. E. Kuadey, G. T. Maale, T. Kwantwi, G. Sun, and G. Liu, "Deepsecure: Detection of distributed denial of service attacks on 5g network slicing—deep learning approach," *IEEE Wireless Communications Letters*, vol. 11, no. 3, pp. 488–492, 2021.

[10] J. Li, Z. Zhao, and R. Li, "A machine learning based intrusion detection system for software defined 5g network," *arXiv preprint arXiv:1708.04571*, 2017.

[11] L. F. Maimó, Á. L. P. Gómez, F. J. G. Clemente, M. G. Pérez, and G. M. Pérez, "A self-adaptive deep learning-based system for anomaly detection in 5g networks," *Ieee Access*, vol. 6, pp. 7700–7712, 2018.

[12] H. A. Alamri, V. Thayananthan, and J. Yazdani, "Machine learning for securing sdn based 5g network," *Int. J. Comput. Appl*, vol. 174, no. 14, pp. 9–16, 2021.

[13] J. Suomalainen, A. Juhola, S. Shahabuddin, A. Mämmelä, and I. Ahmad, "Machine learning threatens 5g security," *IEEE Access*, vol. 8, pp. 190 822–190 842, 2020.

[14] P. H. Mirzaee, M. Shojafar, Z. Pooranian, P. Asefy, H. Cruickshank, and R. Tafazolli, "Fids: A federated intrusion detection system for 5g smart metering network," in *2021 17th International Conference on Mobility, Sensing and Networking (MSN)*. IEEE, 2021, pp. 215–222.

[15] V. Kulkarni, M. Kulkarni, and A. Pant, "Survey of personalization techniques for federated learning," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*. IEEE, 2020, pp. 794–797.