Digitalization and Management Innovation IV
A. Lau and A.J. Tallón-Ballesteros (Eds.)

doi:10.3233/FAIA250724

From Machine Learning to Federated Knowledge Systems – Current Challenges and Future Architectural Directions for Cybersecurity Applications

Felix Härer [1] and Noah Agostinis
University of Applied Sciences and Arts Northwestern Switzerland, 4002 Basel, Switzerland
ORCiD ID: https://orcid.org/0000-0002-2768-2342

Abstract. Machine learning (ML) is foundational to cybersecurity today, underpinning critical applications such as intrusion detection systems. Despite their widespread adoption, however, established ML approaches are beginning to show critical limitations in handling evolving cyberthreats. In this paper, we (1.) conduct a data-driven analysis of ML for intrusion detection to understand and demonstrate current limitations, (2.) discuss the literature and contextualize the findings, and (3.) identify and outline architectural directions for advancing local ML to federated knowledge systems for cybersecurity applications. The analysis results indicate significant challenges within established systems, notably in defining normal behavior, high false alarm rates, and detection rates. These shortcomings highlight intrinsic constraints of localized ML approaches, as exemplified in intrusion detection and beyond. Toward addressing these issues, the properties of a federated knowledge system architecture can provide distributed data inputs rather than siloed sources, data sharing, and federated learning to derive widely distributed, structured knowledge bases.

Keywords. Artificial Intelligence, Machine Learning, Federated Learning, Knowledge Sharing, Intrusion Detection

1. Introduction

Artificial Intelligence (AI) has become an integral component of technical systems, with Machine Learning (ML) serving as its foundational paradigm. By enabling systems to recognize patterns, make predictions, generate outputs, and improve over time, ML has the potential to transform numerous industries. In certain fields such as cybersecurity, these advancements are particularly critical, as cyberthreats are evolving and intrusions are occurring with increasing complexity at an accelerating pace. ML-based Intrusion Detection Systems (IDS) have become the primary line of defense against these threats;

[1]Corresponding Author: Felix Härer, University of Applied Sciences and Arts Northwestern Switzerland; E-mail: felix.haerer@fhnw.ch.

however, the effectiveness of ML remains a major challenge, particularly in handling constantly evolving attack patterns and novel zero-day threats.

More broadly, ML effectiveness is critical across domains, e.g., when relying on process automation within distributed settings such as in healthcare or critical infrastructure. In future environments that are likely to become increasingly dynamic and distributed, advancements in ML will be crucial, potentially leveraging distributed and dynamic learning approaches to enhance effectiveness.

Problem Statement: Combating globally occurring cyberthreats on an internet-wide scale is, today, limited by challenges of current-generation machine learning approaches that face accuracy challenges and operate locally.

In the current discussion on countering increasingly distributed threats, novel architecture concepts that are equally distributed could be considered. Motivated by the possibility of federated learning for threat intelligence, this paper highlights current challenges through a data-driven analysis in the domain of intrusion detection and outlines a conceptual architecture towards federated and knowledge-based systems. The paper seeks to contribute in the discussion on combating distributed cyberthreats by highlighting the potential of such an architecture.

In particular, we aim to: (1.) systematically evaluate and demonstrate the limitations of local ML approaches in the context of intrusion detection; (2.) contextualize our findings within the broader literature; and (3.) identify future directions for the transition from isolated ML systems to federated knowledge systems. Such advancements would enable distributed data inputs, local learning with global aggregation of learned models across distributed nodes, and the structured redistribution of knowledge within knowledge bases.

In the following, Section 2 introduces background on ML for intrusion detection along with related literature. Section 3 investigates current ML-based intrusion detection approaches by a data-driven analysis with measurements of performance metrics and contextualizes the findings. In Section 4, future directions are discussed in an architecture overview for federated knowledge systems. Section 5 reflects and concludes.

2. Background and Related Work

AI enhances Network Intrusion Detection Systems (NIDS) by automating tasks traditionally requiring human intelligence [1]. AI-based NIDS use Machine Learning (ML) methods, including Deep Learning (DL), and follow three key stages: data pre-processing, model training, and model testing [2]. During training, network traffic or malware samples serve as inputs to learn patterns and predict threats [2,1]. Feature selection is crucial, as relevant data elements impact detection performance [2]. DL-based NIDS improve detection by automatically learning complex features through additional neural network layers, reducing the need for manual feature engineering [1,2].

NIDS analyze network traffic to detect threats, unlike Host-Based IDS (HIDS), which monitor individual devices [1,3]. Detection methods include signature-based, anomaly-based, and hybrid approaches. Signature-based detection matches traffic against predefined attack patterns, while anomaly-based methods flag deviations from normal behavior, making them effective against novel threats but sensitive to changing network

conditions [1,4,5]. Hybrid methods combine both techniques for improved accuracy, and further specialized approaches continue to evolve in research [1,3,2].

AI-enhanced NIDS have been the subject of various prior studies. Notably, a survey [3] reviews recent research trends in NIDS and categorizes the surveyed works based on the three detection approaches discussed in Section 2. [2] is another notable survey that also provides a comprehensive overview over the topic and focuses on different ML-based or DL-based NIDS approaches. Regarding particular solutions towards distributed approaches, the paper [4] focuses on lightweight IDS solutions for IoT, where IDS is performed at the edge. The need to address current challenges in anomaly detection is underscored by works on its current limitations such as [5]. These challenges are partially addressed, in particular by the need for self-adaptive models, e.g. by [6], however, few works address solutions in the context of combating internet-scale cyberthreats. While federated approaches such as [7] present first solutions, they are not based on recent developments that integrate knowledge-based approaches such as ontologies with federated learning. The feasibility of such approaches has already been demonstrated in combination with Large Language Models (LLMs) [8,9], however, since this field is rapidly evolving, these approaches have not been applied together with federated learning and knowledge-based systems, yet. To point out this future architectural direction that can potentially combat evolving cyberthreats by global sharing and learning with accuracy beyond traditional ML approaches, this paper suggests considering a federated knowledge system architecture utilizing global knowledge sharing and learning based on ontologies and LLMs. In the following, currently established approaches are analyzed in order to understand the challenges and motivate the architecture concept.

3. Analysis of Machine Learning Approaches for Intrusion Detection

This section analyzes the strengths and weaknesses of intrusion detection approaches used in current AI-enhanced systems, involving signature-based and anomaly-based detection as well as state-of-the-art hybrid approaches. To understand and demonstrate current limitations, we trained and applied learned models using the NSL-KDD dataset.

3.1. Dataset.

The NSL-KDD dataset is a widely recognized and well-understood dataset. It is applied due to its well-known characteristics and limitations while containing data that is commonplace and, ideally, should be detected by IDS systems today. NSL-KDD is a refined version of the KDD'99 dataset, designed to benchmark threat detection systems by removing redundancies and balancing difficulty levels [3,10]. It includes 125,973 training records and 22,544 testing records across 43 columns, with 41 features and two labels for attack type and difficulty level. The dataset contains 52.04% normal traffic and 47.06% attack traffic, categorized into DoS (35.82%), Probe (9.43%), R2L (2.53%), and U2R (0.17%) [10]. Some attacks appear only in testing data, simulating zero-day threats.

3.1.1. Data preparation.

All models are trained for binary classification, distinguishing threats from normal behavior, with the added "threat" label and `attack_type` for evaluation. The same NSL-

KDD training and testing split is used, with 41 connection features as inputs and corresponding labels as targets. No further preprocessing is needed due to the dataset's quality, but categorical features are one-hot encoded, and min-max normalization scales all features between zero and one [1].

3.1.2. Training.

The signature-based model is trained with the Random Forest Classifier machine-learning algorithm. The standard hyper-parameters and the random state of 42 are used.

The Isolation Forest algorithm is used to train the anomaly-based model. For the training, only the non-malicious entries of the training set were used, with the contamination of 0.25 and random state of 42.

In the hybrid-based model both the Random Forest Classifier and Isolation Forest algorithm are used. The Random Forest model used the same configuration as the signature-based model. For the Isolation Forest model, the only change from the anomaly-based model is the contamination parameter, set to 0.5. During prediction, both models estimate the confidence of an attack being non-malicious or malicious. These values are then normalized between 1 (threat) and 0 (no threat). The Isolation Forest score contributes 92% to the final confidence score while the Random Forest score contributes only 8%. If the final confidence score is higher or equal to 0.5 the entry is flagged as a threat.

3.2. Performance measurement.

The key metric to evaluate the proposed approaches is their F-measure. Accuracy is not used as key metric to account for the imbalance of malicious and non-malicious entries [3]. The False-Alarm Rate (FAR) shows how many non-malicious entries were wrongly classified. The detection of lesser known attack types (R2L and U2R) and the simulated zero-day are also compared using their recall values.

3.2.1. Results of evaluation.

The evaluation shows that the anomaly-based and hybrid-based models have similar reliability, with F-measures of 88.04% and 87.31% respectively. The signature-based model performs worse with a F-measure of 75.23%. However, the model achieves the lowest FAR (2.78%), outperforming the anomaly-based (19.32%) and hybrid-based (11.26%) models.

The lesser-known attacks in the signature training set, U2R and R2L, have a low recall for the signature-based model at 9.00% and 3.78%, respectively. In contrast, the anomaly-based model detects these attack types much better, with recall values of 90.50% for U2R and 61.22% for R2L. The hybrid approach achieves 86.00% recall for U2R and 36.42% for R2L. The anomaly-based model also excels at detecting zero-day threats, as reflected in its high recall rate of 96.59%. The signature-based approach has a zero-day attack recall of 23.47%, while the hybrid approach falls between the two at 79.87%.

This indicates that the anomaly-based model is far better in detecting lesser known or zero-day threats with the cost of more false alarms than the signature-based model. The

anomaly-based model also achieves a better overall accuracy. The hybrid-based model balances the strengths of both models, offering strong threat detection with a moderate false-alarm rate. These findings are consistent with recent studies [1,3,2].

3.2.2. Implications for Intrusion Detection

ML-enhanced IDS appear to counter unknown threats by anomaly-based approaches, however, detection performance is not nearly comparable to the detection of known threats, which require signature-based approaches in addition. The introduction of hybrid IDS has been pursued for this reason and the analysis supports this conclusion as a necessary first step for baseline security. However, it is not sufficient due to the problems observed in rising False Positive (FP) rates and declining accuracy when involving anomaly-based methods. Recently, advancements in AI have impacted the threat landscape. In the background, known threats remain constantly active and produce the highest share of internet-side attack attempts[11], however, they tend to be constant threats that can be countered by proven hybrid approaches. Constant threats undergo comparatively little evolutionary changes, are typically based on well-known predecessors, and developed or merged in new variants over months or years.

In recent years, adaptive threats were increasingly observed [12] that change their behavior according to new environments and to evade detection. In this paradigm, threats are adjusted partially or fully automatically on the attacker side or in the field, e.g., by changing code through command and control remotely, or by AI-enhanced self-modifying code and code generation. Consequently, threats tend to develop at a faster rate and undergo significant changes over time. While ML detection approaches demonstrated to be effective within stable environments, learning new normal or malicious behavior presented severe challenges for detecting evolving threats. Previously unknown threats in zero-day attacks are especially problematic in this regard, since they are present in few local traffic flows only.

3.2.3. Limitations.

The dataset and analysis aim to provide principal strengths and weaknesses for the discussion. Detailed conclusions are limited since (1.) further attack types exist beyond the ones included in the dataset, and they evolve constantly, and (2.) further method and parameter optimizations could have been carried out towards optimizing NIDS for production, e.g., systematic evaluations for hyper parameter configurations, input features, and learning methods. The results reflect performance in a laboratory environment in order to demonstrate strengths and weaknesses of the system classes under comparable conditions.

4. Advancing toward a Federated Knowledge System Architecture

In the recent years, AI-enhanced approaches have been introduced across disciplines and are being integrated into business systems and technical systems. At the present, ML is frequently applied such as for intrusion detection [3]. From our analysis in this domain, significant limitations are apparent that are, however, not limited to this domain. They are common challenges in ML today, inherent to ML based on limited, local data sources

in an evolving environment. Effectiveness is especially impacted by challenges of False Positive (FP) rates, accuracy, and recognizing complex patterns under dynamic conditions. In particular, the coupled effect of rising FP rates and declining accuracy is highlighted by the analysis, e.g., when detecting lesser known attack variants, for few known samples of signatures or patterns, and for benign application traffic not seen before.

4.1. Architectural Approach to Federated Knowledge System

Advancing ML-based systems beyond architectures reliant on siloed data sources presents an opportunity to leverage shared data for improved learning on the basis of shared data and, eventually, evolve in the direction of further dependable and structured knowledge representations. In certain fields such as cybersecurity, these advancements are critical to join data sources on larger scales, introducing global networks that sense appearing threats, collect patterns, update models, distribute them, and apply them for detection. The criticality of ML effectiveness is the primary concern, evident in many other areas such as business processes of critical industrial systems, critical infrastructure, or healthcare. In addition, future systems are likely to operate in increasingly heterogeneous and distributed environments. In these settings, data and events are generated and processed in local systems, which can then serve as inputs for federated learning strategies. These strategies learn and integrate based on local datasets and events, enable knowledge extraction into higher-level knowledge representations, and re-distribute them back to local systems. Such an architecture has the potential to create federated knowledge systems, utilizing widely distributed knowledge bases through federated learning. The architecture of such a federated knowledge system architecture encompasses three components:

4.1.1. Data Collection and Event Handling in Distributed Nodes.

Distributed nodes are part of heterogeneous local systems, where they serve as the primary interface for data and events by fulfilling three functions: (1.) Gathering diverse data sources, including process events, sensor outputs, and network traffic logs. (2.) Local Event Handling: Processing inputs in real time to respond to local events and classify them appropriately. In cybersecurity contexts, this may involve detecting specific traffic patterns or system events, such as authentication activities. (3.) Application of local ML models: Utilizing local and efficient ML models to extract knowledge from collected data, e.g., by lightweight approaches such as [13,14,15]. Subsequent processing tasks may be executed or triggered on external systems.

4.1.2. Federated learning

enables the distributed nodes to function as a collective intelligence system rather than isolated collection points. Each node acts as a client that contributes to combined learning and knowledge sharing using aggregation and learning, based on state-of-the-art protocols such as [13,15]. In particular by (1.) computing updates for local models based on locally available data, (2.) performing local training iterations, (3.) apply model updates to the locally maintained model, and (4.) transmitting these updates for aggregation by receiving nodes. The continuous cycle of local training, aggregation, application, and transmission facilitates the propagation of updated models throughout the network.

4.1.3. Distributed Knowledge Bases.

For the redistribution of knowledge in structured representations, the updated model — post-aggregation and learning — is applied locally. By globally identified patterns, the model is now applied to the collected data and events for knowledge representation. This process entails: (1.) knowledge extraction and (2.) mapping of the extracted entities and relationships in existing graph or model representations, and (3.) knowledge distribution that transmits the updated representations to nodes for integration. Knowledge extraction and mappings are well supported through recent LLM approaches [8,9] or named entity recognition [16]. On the basis of standardized structures, integration will be supported, e.g., on the basis of standardized knowledge graphs and ontologies, resulting in a distributed and coherent knowledge system.

4.2. Implications

In combination, these components have the potential for wide-ranging and large-scale learning efforts, towards jointly learning from locally available data and events for creating dependable and structured knowledge bases that improve over time. The architecture concept is inherently decentralized and scalable, with knowledge bases being present at the nodes. Siloed data and event sources, also of larger data volumes, are supported by federated learning through local learning, aggregation, and the transmission of model updates. Thus, this architecture vision advocates for further research in the direction of federated knowledge systems towards an architecture and its realization, combining knowledge-based approaches with federated learning.

5. Conclusion

This paper presented current Machine Learning (ML) approaches in the domain of intrusion detection by a data-driven analysis involving training and model application, it contextualized the findings leading to identified challenges of effectiveness, and suggested an architecture vision for advancing to federated learning and knowledge-based systems. The analysis findings identified challenges of effectiveness, especially when applying ML with local, siloed data sources under dynamic conditions. Based on these findings, the need for distribution and dynamic environments in future systems results in the proposed architecture vision. The architecture supports decentralization and scalability by a local training and aggregation through federated learning and by re-distributes structured knowledge representations back to nodes. Further research is needed towards future systems that have the potential to combine federated learning and knowledge representations for widely-distributed future federated knowledge systems.

References

[1] Arun Kumar Silivery, Ram Mohan Rao Kovvur, Ramana Solleti, LK Suresh Kumar, and Bhukya Madhu. A model for multi-attack classification to improve intrusion detection performance using deep learning approaches. *Measurement: Sensors*, 30, 2023.

[2] Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, and Farhan Ahmad. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, 32(1), 2021.

[3] Oluwadamilare Harazeem Abdulganiyu, Taha Ait Tchakoucht, and Yakub Kayode Saheed. A systematic literature review for network intrusion detection system (ids). *International Journal of Information Security*, 22(5), 2023.

[4] Nadia Chaabouni, Mohamed Mosbah, Akka Zemmari, Cyrille Sauvignac, and Parvez Faruki. Network intrusion detection for iot security based on learning techniques. *IEEE Communications Surveys & Tutorials*, 21(3), 2019.

[5] Jiayi Liu, Donghua Yang, Kaiqi Zhang, Hong Gao, and Jianzhong Li. Anomaly and change point detection for time series with concept drift. *World Wide Web*, 26(5), 2023.

[6] Alok Kumar Shukla. Detection of anomaly intrusion utilizing self-adaptive grasshopper optimization algorithm. *Neural Computing and Applications*, 33(13), 2021.

[7] Beibei Li, Yuhao Wu, Jiarui Song, Rongxing Lu, Tao Li, and Liang Zhao. Deepfed: Federated deep learning for intrusion detection in industrial cyber–physical systems. *IEEE Transactions on Industrial Informatics*, 17(8):5615–5624, 2021.

[8] Bowen Zhang and Harold Soh. Extract, Define, Canonicalize: An LLM-based Framework for Knowledge Graph Construction. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, Florida, 2024.

[9] Benedikt Reitemeyer and Hans-Georg Fill. Leveraging LLMs in Semantic Mapping for Knowledge Graph-based Automated Enterprise Model Generation. In *Modellierung 2024 Satellite Events*. Gesellschaft für Informatik e.V., 2024.

[10] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani. A detailed analysis of the kdd cup 99 data set. In *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009.

[11] Cloudflare. Mitigated traffic sources. Technical report, March 14 2025. [Online]. https://radar.cloudflare.com/security-and-attacks. Accessed March 14 2025.

[12] Maryam Roshanaei, Mahir R. Khan, and Natalie N. Sylvester. Navigating AI Cybersecurity: Evolving Landscape and Challenges. *Journal of Intelligent Learning Systems and Applications*, 16(3):155–174, 2024. Number: 3 Publisher: Scientific Research Publishing.

[13] Hongyi Zhang, Jan Bosch, and Helena Holmström Olsson. EdgeFL: A Lightweight Decentralized Federated Learning Framework. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2024.

[14] Pian Qi, Diletta Chiaro, and Francesco Piccialli. Small models, big impact: A review on the power of lightweight Federated Learning. *Future Generation Computer Systems*, 162:107484, 2025.

[15] Jinhyun So, Chaoyang He, Chien-Sheng Yang, Songze Li, Qian Yu, Ramy E. Ali, Basak Guler, and Salman Avestimehr. LightSecAgg: a Lightweight and Versatile Design for Secure Aggregation in Federated Learning. *Proceedings of Machine Learning and Systems*, 4:694–720, 2022.

[16] Tareq Al-Moslmi, Marc Gallofré Ocaña, Andreas L. Opdahl, and Csaba Veres. Named Entity Extraction for Knowledge Graphs: A Literature Overview. *IEEE Access*, 8:32862–32881, 2020.