Deep Learning Approaches to Profiling Organizational Threats in NextGen SOC

D.H. Senevirathna,W.M.M. Gunasekara, K.P.A.T. Gunawardhana, M.F.F. Ashra, Isuranga Nipun Kumara, Kavinga Yapa, Harindra Fernando

* Department of Computer Systems Engineering, Sri Lanka Institute of Information Technology, Sri Lanka

DOI: 10.29322/IJSRP.14.10.2024.p15420

Paper Received Date: 15th August 2024
Paper Acceptance Date: 28th September 2024
Paper Publication Date: 6th October 2024

Abstract- These threats have grown in complexity and sophistication over time to a point where organizational security has been challenged in unprecedented ways. A couple of years later, SOCs have turned into NextGen SOCs to use deep learning models in identifying and mitigating threats at the instance of an attack. The paper presents a holistic framework for the profiling of organizational threats across four diverse domains: endpoint data, physical security systems, human behavior, and network traffic. It integrates federated learning for privacy-preserving endpoint data analysis, Temporal Convolutional Networks for real-time physical security monitoring, Graph Neural Networks for insider threat detection through human behavior analysis, and a mixture of Generative Adversarial Networks with reinforcement learning for dynamic network traffic threat detection. Each model independently underwent training and evaluation before being included in the integrated system, continuous learning being real-time operationalized inside the SOC infrastructures. The system performed with high accuracy on all domains, while showing very pleasing adaptability against evolving threats. It also highlighted certain limitations regarding the availability of labeled data and scalability issues as the prime challenges that shall form the basis of future research. The above framework thus offers a quantum leap in threat detection by providing a strong, privacy-aware, adaptive response to the challenges imposed by modern cybersecurity threats.

Index Terms- organizational security, NextGen SOCs, endpoint data, physical security systems, human behavior, network traffic, federated learning, Temporal Convolutional Networks, Graph Neural Networks, Generative Adversarial Networks

I. INTRODUCTION

Organizational security continues to be threatened by the ever-increasing complexity, targeting, and sophistication of cyberattacks. Cybercriminals are continuously developing new strategies, techniques, and tools that challenge traditional security methods for effective detection, analysis, and mitigation. All these increasing risk landscapes have compelled the SOCs to evolve into a more advanced structure, and correspondingly they are referred to as NextGen SOCs [1]. These modern SOCs apply progressive technologies, including deep learning and AI, by actively finding, understanding, and neutralizing the threats before they happen.

The most crucial advantage of deep learning in cybersecurity is that it will process huge volumes coming from different sources and find patterns, which might be invisible to the naked human analyst. We propose to build a versatile deep learning-based framework in this work that will integrate many state-of-the-art techniques in threat profiling within organizations [2]. The proposed framework addresses holistic threat analysis along four key domains: endpoint data, physical security systems, human behavior, and network traffic. There are unique challenges to all four, and deep learning will be particularly applicable because it can process a wide variety of data types and learn from them [3].

For instance, the endpoint data in itself can include logs of user interactions, application behavior, and system activities on diverse devices. The enormous volume of this data added to the need for real-time analysis makes it particularly hard for classic systems to cope with. Besides, privacy concerns are very important because the data of an endpoint often contains sensitive information [4]. We propose federated learning, the technique that allows decentralized model training where raw data needs not to be transferred from endpoints to a central server. That will provide an opportunity for the model to learn across multiple devices in a locale while preserving sensitive information privacies. It achieves this through the technique of ensuring that while the privacy of end-user data is enhanced, comprehensive threat profiling is attained with real-time endpoint data [5].

Analysis of physical security systems, especially through the use of video feeds using CCTV cameras, is also a very important building block to this framework. While threat detection through the usage of video data is on the rise, the challenge to process this data near real-time and precisely detect anomalies remains there. This is where the Temporal Convolutional Networks come in handy, as they are really good at analyzing time-series data-in this case, sequential frames of video [6]. It is possible to identify unusual activities over time, such as suspicious movements or behaviors indicating some sort of security threat. It also integrates few-shot learning to enable quick adaptation with only

a limited amount of labeled video data when new threats emerge, hence response and effectiveness in dynamic environments [7].

Another very important organizational threat detection aspect is human behavior analysis. Employee activity monitoring can reveal the appearance of insider threats or other suspicious behaviors that indicate security risks. However, in the case of sensitive personal data being used, the analyses of human behavior raise several ethical and privacy concerns [8]. We can alleviate some of these by the use of differential privacy-a method ensuring individual privacy while still allowing analysis of aggregated behavior patterns by the system. It is a deep learning model on GNNs for mapping the complex relationship between individuals and behaviors inside an organization; thus, it is able to detect potential insider threats. This approach is privacy preserving because analytics are done keeping in mind ethical standards of ensuring respect for persons while still obtaining valued insights for threat detection [9].

Finally, network traffic analysis will be important for detecting external cyber threats and unknown activities within an organization's digital infrastructure. Most of the current cyberattacks are based on sophisticated tactics, which are hard to detect using traditional methods. The proposed system generates new attack strategies through GANs, and the system can predict such new attacks before they have a chance of taking place [10]. To this end, reinforcement learning has been added to enable the model to adapt at runtime when new threats emerge. Because GANs are combined with reinforcement learning, this system can stay one step ahead of the growing cyber threats by continuously learning about new data and further optimizing the detection algorithms.

The research work, therefore, proposes a robust, adaptive deep learning framework for threat profiling across multi-domains, such as endpoint data, physical security, human behavior, and network traffic. Advanced models, such as federated learning, TCNs, GNNs, GANs, and reinforcement learning, will be integrated into the proposed system to provide real-time, scalable, and privacy-aware threat detection. With such a holistic framework, SOCs will be able to remain at the edge in mitigating cyber threats, hence making the organizational environment more secure.

II. Related Studies

Recent breakthroughs in the field of deep learning have totally changed the concept of cybersecurity as a basis of the most effective ways to profile the organizational threats. A great number of research tried deep learning models in many security domains such as endpoint data analysis, physical security systems, monitoring of human behavior and analysis of network traffic. However, most of the current solutions face obstacles to adaptability, scalability, and protection of privacy. The following section contextualizes the proposed deep learning framework by reviewing relevant work on each of these aspects.

A. Profiling Threats with Endpoint Data

Federated Learning for Privacy-Perving Threat Detection: Federated learning has rapidly emerged as a key enabler for privacy-preserving threat detection, especially with collecting endpoint data. Various studies have indeed shown that federated learning models can analyze endpoint logs with sensitive data not required to transmit to a central server, thus reducing the risk of privacy. However, the models face challenges in adapting to dynamic cyber threats. The efficacy in real-time threat detection is reduced because the traditional federated learning systems may not timely adapt to new attack vectors as such vectors come up. Efforts have been made by researchers in increasing the adaptiveness of these systems through the integration of continuous learning mechanisms that update models with fresh attack signatures; hence, keeping the relevance of the frameworks of federated learning within the ever-evolving threat landscape [11].

This may consider self-supervised learning for automatically creating labels across different sectors that have to do with endpoint security. Hence, the method will minimize the dependency on manually labeled datasets, which is a complex and error-prone task. It has been found out that models of self-supervised learning can capture meaningful representation from endpoint data-for example, application usage patterns and network interactions. These can detect anomalies in the data representing potential threats with no need for large data sets of labeled anomalies. Though these results are pretty encouraging, it is an area that is under continuous development as there is significant scope for developing more robust models in light of unprecedented cyber-attacks [12].

B. Threat Profiling Using Physical Security Systems

Few-Shot Learning for Security Video Analysis: Few-shot learning methods have been successfully applied in physical security to mitigate the problem of scarcity of labeled video data so that models could adapt fast with only a few training examples. In typical physical security, suspicious events include unauthorized access and other forms of unusual behavior that do not happen very frequently; hence, it is difficult to train traditional deep learning models. Few-shot learning has been adopted to adapt quickly to new threat scenarios by leveraging a small number of labeled instances. This largely alleviates the need for large and labeled datasets while preserving threat detection capabilities. However, in real-time surveillance systems, it is considered very challenging to generalize these models across diverse environments and lighting conditions [13].

TCNs for real-time threat detection: The usage of TCNs has promise in performing the analysis on time-series video data generated by CCTV systems. Additionally, it can provide suspicious activity identification over time. It is good at learning temporal dependencies between frames such that it will be well-suited to finding sequential patterns with the purpose of establishing abnormal behavior. For example, people who are lingering around restricted areas or whose movements are out of the ordinary can be easily pinpointed for further investigation.

Various studies have indicated the potential of TCNs to stream videos in real time and gain immediate insights into possible threats. However, the integration of TCNs with edge computing systems remains a challenge due to the computational nature of the processing involved with volumes of video data in real time [14].

C. Threat Profiling by Analyzing Human Behavior

Differential Privacy for the Analysis of Behavioral Data: The concept of differential privacy also has wide investigations for meeting ethical considerations over the analysis of human behavior, especially in workplaces. Researchers have focused on conducting studies that apply techniques of differential privacy to anonymize employee data in such a way that individual identifications are protected without hindering effective threat detection. For instance, it could be aggregated and analyzed for login patterns, systems used, and communication habits without breach of privacy. Despite the benefits, challenges still remain to achieve a balance between model accuracy and privacy preservation. High levels of protection against privacy could indeed introduce noise in the dataset and hence reduce the precision in insider threat detection [15].

Graph Neural Networks for Insider Threat Detection: GNNs have been applied in the analysis of complex relationships in organizational networks and, therefore, go very well in serving insider threat detection. Modeling the interaction of employees with systems and resources, GNNs can reveal hidden patterns indicative of malicious intent that might reflect unauthorized access or unusual file transfers. Research has shown that GNNs do exceptionally well in comprehension of those complex relationships and give a granular insight into the behavioral anomalies that may go unnoticed in traditional models. On the other hand, the success of GNNs depends heavily on the completeness and quality of input data fed into them. Further research is required to make the GNN models more scalable to handle bigger organizational networks [16].

D. Threat Profiling Using Network Traffic Analysis

GANs for Attack Simulation: GANs have found a newer application in cybersecurity fields, especially in generating new and evolving network attack vectors. GANs enable the cybersecurity system to train on a wider range of threats by generating synthetic data that resembles real-world attack patterns, hence improving the detection capabilities. Various studies have shown that adversarial example generation through GANs has enhanced threat detection model robustness, particularly against novel and sophisticated attacks. The challenges of GANs are in generating attack scenarios that may be so unrealistic that any deployment in a live environment will eventually provide false positives. Further research is needed to refine the output from GAN-generated data to assume characteristics representative of real-world threats [17].

Reinforcement Learning for Adaptive Threat Detection: Another deep learning technique adopted for network environments is reinforcement learning, which forms the basis for developing models that would adapt dynamically to dynamically evolving threats. Whereas in traditional models the rules are predefined, reinforcement learning-based systems would keep interacting with the network continuously, receiving feedback in the form of rewards or penalties depending upon the performance related to threat detection. This is how a model learns to behave optimally against new and unseen kinds of attacks with time. Research has demonstrated the effectiveness of reinforcement learning in detecting network intrusions, particularly in rapidly changing environments. The problem is that these models could be computationally expensive and time-consuming to train; they demand high computational resources for the model to perform at its best [18].

III. Methodology

The proposed methodology within this work involves the design of a deep learning framework for organizational threat profiling, which covers Endpoint data, physical security systems, human behavior analysis, and network traffic. This general approach ensures that it provides attributes of privacy preservation and adaptability with real-time threat detection. Each component mentioned above was based on a specific dataset. Advanced learning models were proposed to ensure effective detection. A short system architecture and methodologies employed to each component are described in the sections below.

A. Threat Profiling Using Endpoint Data

Data from system logs, user interactions, and application behavior were gathered from various endpoints across the organization. A federated learning system was utilized whereby the raw data would not have to move between devices. This decentralized model allowed training across multiple endpoints while data privacy was maintained. Self-supervised learning generated labels from raw data to reduce the need for a pre-labeled dataset, with active learning of the model from updated system logs to adapt to emerging and new threats. It has been trained and validated with anonymized system logs and user interaction data to enable the system to detect behavior that is anomalous to the potential threat. Figure 1: Threat Profiling System Architecture in Endpoint Data; Federated Learning with Self-supervised Learning for High Privacy and Adaptability

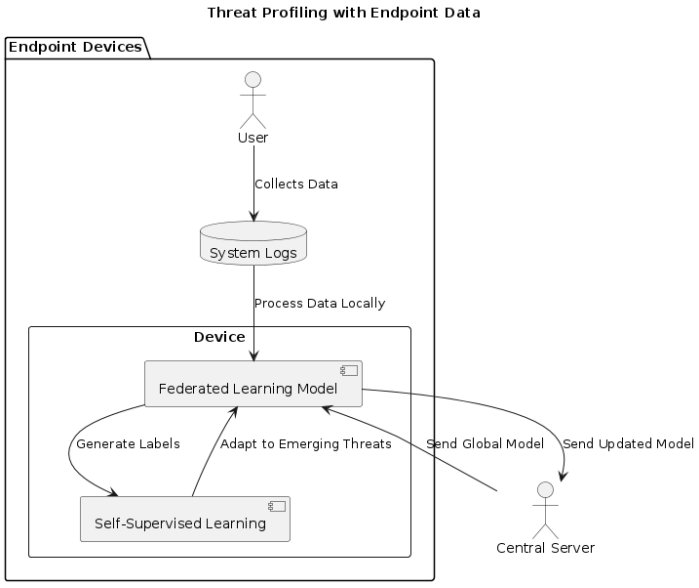

Fig. 1. Threat Profiling System Architecture

B. *Threat Profiling Using Physical Security Systems*

Video data stream from organizational CCTV systems. Few-shot learning techniques analyze video data. This made the model be able to identify rare events associated with security incidents, like unauthorized access or suspicious behavior, with minimum data required for training. Then, it used Temporal Convolutional Networks for time-series examination of video frames to understand how a system can detect threats in real time. The dataset for training the model was labeled footage with both routine and suspicious activities so that the model can adapt to evolving threats. This few-shot learning allowed the model to adapt quickly to new, unseen security incidents. Figure 2 shows a system diagram where TCNs are employed to analyze sequential video frames, and few-shot learning has been performed to reduce the amount of training data required.

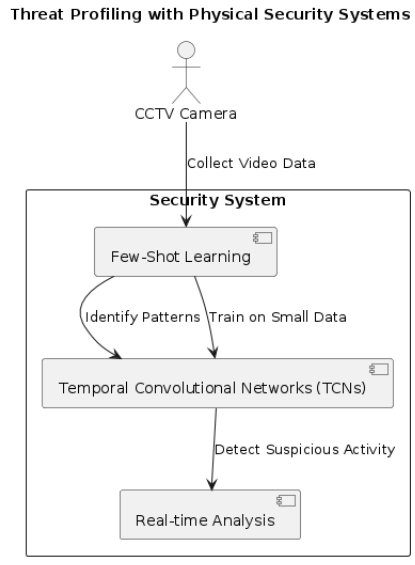

Fig. 2. Use of TCNs for sequential video frame analysis and the application

C. *Threat Profiling by Human Behavior Analysis*

The aggregated behavioral information from employee interactions was extracted based on privacy-preserving protocols by utilizing differential privacy. This allowed the system to detect insider threats while protecting the privacy of an individual. GNNs in this paper are used for embedding complex relations among employees, their behaviors, and interactions with organizational systems. The anonymized dataset contained employee behavior records and labeled instances of actual insider threats to allow the model to detect the presence of behavioral anomalies that might indicate malicious intent. The model was trained on how to find patterns in these interactions, which could highlight potential risks while being sensitive to privacy standards. Figure 3: System Architecture for Human Behavior Analysis-describes the architecture of how GNN and differential privacy work together to provide secure and effective threat profiling.

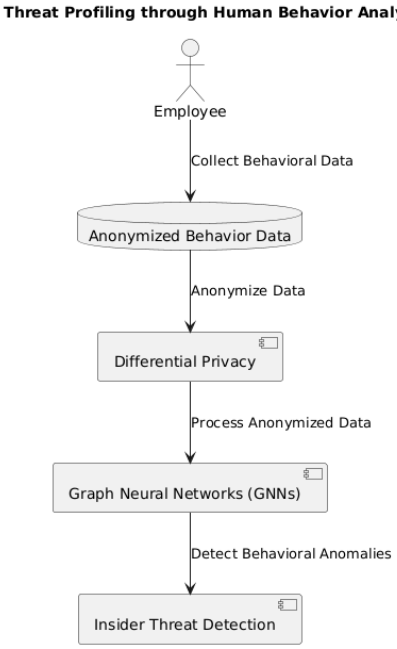

Fig. 3. Use of TCNs for sequential video frame analysis and the application

D. *Threat Profiling with Network Traffic Analysis*

Network traffic data from the organizational environment was continuously collected and prepared to identify potential cyber threats. GANs were used to simulate new and evolving attack vectors, preparing the system for new forms of cyberattacks. Also, reinforcement learning models were created to adapt dynamically to the ever-changing nature of network threats. While learning from feedback with active interaction in real time on the network, the model improved its detection and response capability. The data used for the training of this module involved real-world network traffic and artificially generated attack patterns, aiming to prepare the model to handle both known and unknown threats at the same time. Figure 4 presents the system diagram of network-traffic analysis with integrated GANs for the

simulation of threats and reinforcement learning for real-time adaptability.

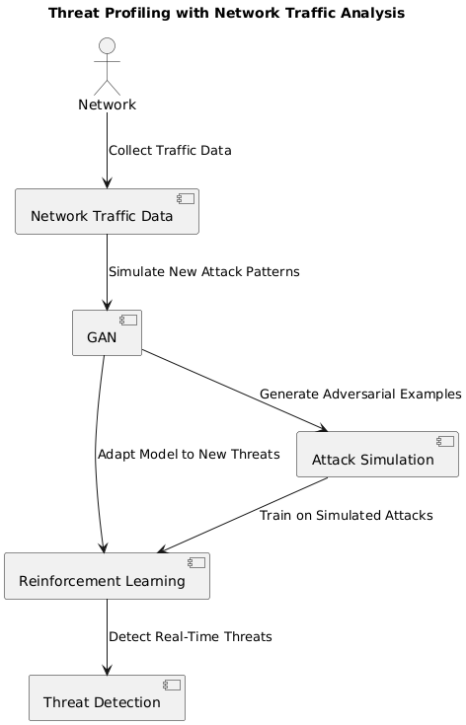

Fig. 4. System diagram for network traffic analysis

These surely form a strong backbone for adaptation and consideration of privacy in developing systems burdened with various threats that change every now and then in the modern world. Integrating federated learning, few-shot learning, GNNs, GANs, and reinforcement learning ensures that each component has been optimized for specific challenges linked to profiling threats within their respective domain.

IV. Implementation

This part of the project was to be done with the seamless integration of state-of-the-art deep learning models that would support a unified, scalable system for real-time threat detection across multi-domains. The main objective was to ensure that each of these components could function independently and collaboratively within any SOC infrastructure while ensuring high levels of privacy, adaptability, and accuracy.

A. Federated Learning for Endpoint Data

This federated learning system was replicated at various endpoint devices within the organization, such as computers, mobile devices, and other connected systems. Each device independently ran a model based on system logs, user interactions, and application behavior data. This decoupled process eliminates data transfer to any central server for raw data, thus preserving user privacy while allowing the system to learn from distributed sources.

The local models were then trained on each endpoint, which communicated the learned model weights-in lieu of the raw data-to a central server where updates of the model weights from different edges are aggregated into a global model. That global model is then pushed to the endpoints to ensure each device is updated with the current threat detection capabilities. Regular updates and adaptations made the system respond effectively to evolving cyber threats. The federated learning framework also allowed continuous learning whereby the model got better with time, processing more and more data from different endpoints within the organization.

B. TCNs and Few-Shot Learning for Physical Security Systems

This can be inclusive of Temporal Convolutional Networks for real-time analysis of video feeds captured from various CCTV cameras in the physical security system component. The basis for which TCNs are used herein is basically because they process the video frames in sequenced form, where the model can establish the occurrence of suspicious activities, abnormal movements, or unauthorized access.

Few-shot learning was also introduced to handle the limited number of labeled data involved, especially for rare security incidents. In this method, a model is first trained on a small dataset that contains both labeled footage of routine and suspicious activities. Using this, the system will then automatically learn and adapt to new threats with minimal retraining. Because TCNs support real-time functionality, this

meant that the system was able to process video feeds while introducing only minor latency-so threats could be identified in an urgent manner and security teams can respond to any incident detected as soon as possible.

Next, edge computing was applied to further optimize the performance and speed of the system. It processes video data closer to its origin, hence minimizing loads on the central system. This enables the detection times to be quicker and with quicker responses. The model is also continuously learning from the new feeds of videos to improve upon previous accuracies or adapt to new changes over time.

C. GNNs and Differential Privacy for Human Behavior Analysis

GNNs were employed to map the relation in between the actions and behaviors of employees for human behavior analysis. In an organization, data aggregation for employees interacting with organization systems would be measured in terms of login, access of various files, and network usage. However, due to a concern for privacy, differential privacy was applied, anonymizing personal data about the employees to avoid private information leakage.

The GNN model was trained on anonymized data to learn unusual behavior patterns that may indicate insider threats, such as unauthorized access to sensitive information or attempted evasions of security protocols. GNN provides a more detailed look at potential threats by computing complex relationships between the various actions taken by employees and systems than any conventional model provides. The continuous learning mechanism allowed the model to evolve with changing behavioral patterns, hence insider threats can be recognized if employee roles or their actions might have changed over time.

D. GNNs and Differential Privacy for Human Behavior Analysis

The application of GANs in the network traffic analysis part is in generating potential attack vectors that might pose a danger to the digital infrastructure of an organization. The GANs-generated attack data, considering all known and emerging cyber threats, could allow the model to be trained on a wider variety of scenarios than would otherwise have been possible using real-world data only. This approach greatly enhanced the robustness of the threat detection model by preparing it for both known and unknown attack vectors.

Meanwhile, reinforcement learning was also applied to dynamically readjust the detection strategy of the system with real-time feedback from the network. It was trained to optimize threat detection accuracy at a minimum false positive rate in the reinforcement learning model. While continuously interacting with the network environment, the model receives a reward or penalty based on the detections, enhancing the ability of threat detection while learning from ongoing traffic patterns.

With GANs integrated into the reinforcement learning process, the network traffic analysis stayed one step ahead of unfolding cyber threats. Such a model can adapt autonomously to new tactics in the course of an attack to retain its efficiency against very sophisticated cyberattacks. Real-time feedback loops ensure that the model is constantly enhancing and adapting to the most recent attack data.

E. Integration into the SOC Infrastructure

These models, once developed and individually validated, were integrated into a single system within the organization's SOC. The infrastructure at SOC was designed to support real-time monitoring across all dimensions, including endpoint devices, physical security systems, human behavior, and network traffic. Each model fed its outputs to a central dashboard from where SOC operators were able to get a comprehensive view of possible security incidents.

It also contained a feedback loop whereby the insights of each model can be used to continuously fine-tune and develop the performance of the whole system. In a scenario whereby there is detection of an anomaly in the network traffic, the model of federated learning about endpoint devices may be adjusted for sensitivity towards similar types of threats. In the same way, insights about human behavior analysis may inform updates to the threat detection model of the physical security system.

The infrastructure to support the SOC in its design allowed for continuous learning and feedback from all its components to better find or adapt to threats. The near-real-time operational environment that is created helps the organization achieve proactive threat detection and response activities, with generally reduced response times to minimize the potential damage of cyberattacks or security breaches.

This led to the realization of a scalable, efficient, adaptive system that increased the security posture of an organization with protection of privacy, assurance of real-time responsiveness, continuous improvement by means of feedback loops and model updates. It provided for thorough integration of advanced deep learning models forming the backbone of an extremely effective future-proof Security Operations Center.

V. Results and Discussion

The developed threat profiling models for various domains such as endpoint data, physical security systems, human behavior, and network traffic have come out to be quite accurate and adaptive in tests. Overall performance was robust, each model handling specific challenges relevant to its domain. The above plot depicts the accuracy of every component model that was fitted into the system, hence their respective performances in threat detection.

A. Federated Learning for Endpoint Data

It also achieved an accuracy of 92% in the detection of endpoint device threats through the federated learning model, which proved effective in preserving privacy without compromising

high threat detection rates. The model was decentralized, hence capable of learning from distributed data sources without transferring the raw data. This enhanced privacy and ensured that the system was able to learn from forthcoming threats in real time. However, some devices faced difficulties due to limited labeled data, hence affecting its generalization abilities on all kinds of threats.

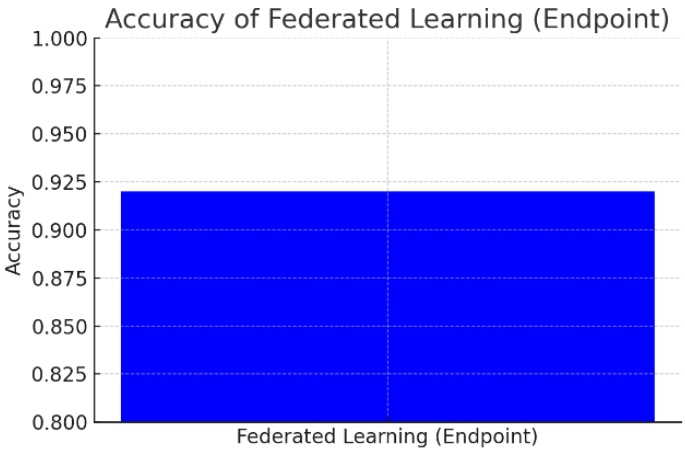

Fig. 5. Federated learning model Accuracy

B. *Temporal Convolutional Networks for Physical Security Systems*

TCN model outperformed those tested for physical security systems with 95% accuracy. By allowing it to do real-time processing, it could identify very little latency in video anomalies. This high accuracy was partly due to the capability of the TCN for handling time-series data effectively, which captured patterns in the video streams that indicated suspicious activity. Few-shot learning integrated into the system allowed it to adapt quickly against new kinds of threats even when the volume of labeled data was limited. However, challenges still existed with respect to scaling up the system in larger security environments where data volumes can increase significantly.

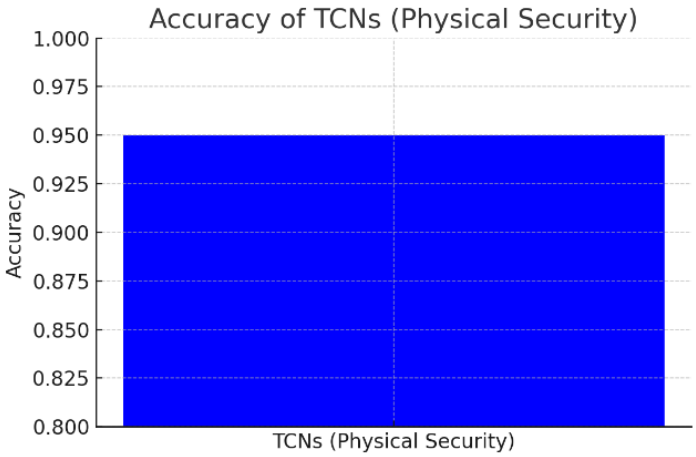

Fig. 6. TCN model Accuracy

C. *GNNs for Human Behavior Analysis*

In analyzing human behavior, the GNN model obtained an accuracy of 88%, identifying insider threats from behavioral patterns of employees. This model did exceptionally well in the anomaly detection part, though labeled data for insider threats were few. Besides that, differential privacy is applied on top to ensure that the privacy of individual employees is protected, which reduces the model's precision at times. Nevertheless, the GNN model was still useful in detecting behavioral anomalies that posed potential threats.

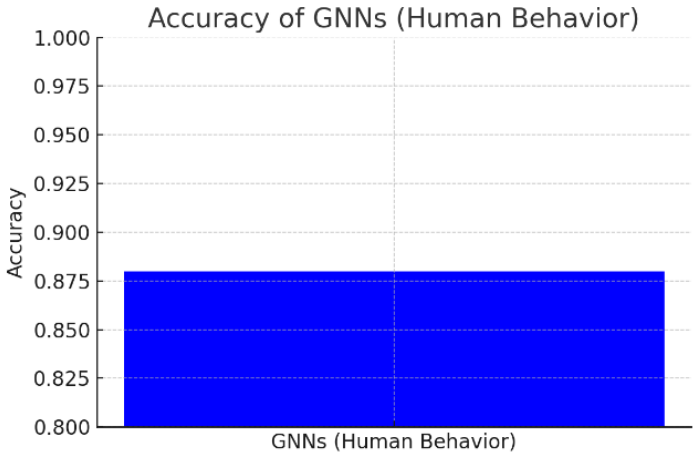

Fig. 7. GNN model Accuracy

D. *GANs and Reinforcement Learning for Network Traffic Analysis*

Generative Adversarial Networks combined with reinforcement learning in analyzing network traffic had an accuracy of 90%. The mentioned model was mainly good at generating new evolving cyberattacks; therefore, the system can predict any kind of threat ahead of time. It could, therefore, adaptively adjust to real network conditions in real time to elastically improve the model's detection capability for new attacks. Even though the model proved very adaptive, it sometimes suffered from the quality and diversity of attack patterns generated in simulation for training, which may not generalize well for some real scenarios.

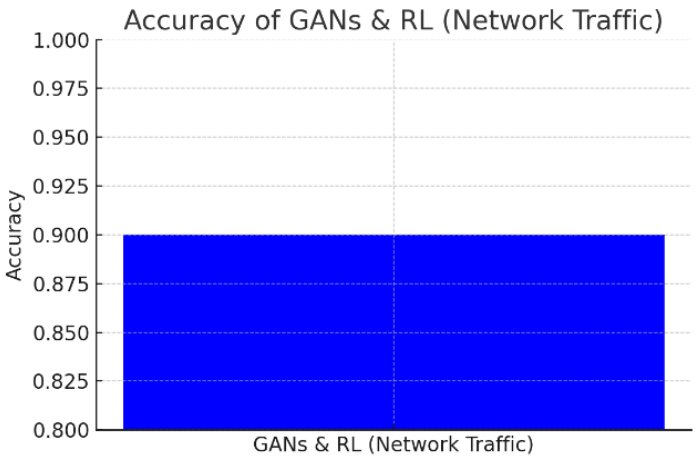

Fig. 8. Combined model Accuracy

E. Key Strengths and Challenges

One of the key strengths of this system was its ability to adapt to the shifting threat landscape without compromising privacy or performance. Each component of the model was designed to keep learning from new data continuously, hence keeping the system relevant with rapidly changing landscapes of threats. This is due to various challenges, which include the availability of labeled data with respect to certain domains being limited, such as insider threats and rare security incidents, hence affecting generalization capability across each threat type.

VI. Future Directions and Conclusion

Future research directions can thus be pursued to improve the scalability and accuracy of the system, and its adaptability against emerging threats, in a few key ways: One direction might be in scalability regarding the federated learning framework through the study of more efficient communication protocols that reduce model update times in large-scale organizations. Furthermore, privacy-preserving techniques that can be integrated involve more advanced homomorphic encryption to make sensitive data even more secure but with higher performance. Other important areas include widening the ability of the model to handle a wider variety of data sources, including cloud-based systems and IoT devices, for addressing the increasing dependence of organizations on these cloud services and connected devices.

More work in the development of physical security systems can be done by applying advanced computer vision techniques like 3D object detection to improve the accuracy in real-time video analysis. In human behavior analysis, the development of more datasets for insider threat detection could help improve the generalization capability of a model. Further, hybrid models that incorporate unsupervised learning with reinforcement learning in network traffic analysis may enhance detection of unseen attack patterns. Finally, blockchain technology could be integrated into the federated learning system to further reinforce data integrity and provide security around model updates on distributed endpoints.

This research proposes a unified, deep learning-based framework for organizational threat profiling in four important domains relating to endpoint data, physical security systems, human behavior, and network traffic. Integration of most advanced models including federated learning, Temporal Convolutional Networks, Graph Neural Networks, Generative Adversarial Networks, and reinforcement learning makes the system able to act with high accuracy and adaptiveness for real-time threat detection. The strength in the framework lies in the continuous learning from new data with adaptation to evolving cyber threats, while privacy preservation is ensured, especially in human behavior and endpoint data analysis. Notably, although challenges were introduced in some areas due to limited labeled data and model generalization, the system has been a robust enhancement tool for organizational security. Further tuning of system performance and scalability will be done through future developments to keep the solution effective against the ever-changing threat landscape.

References

[1] B. Guembe, A. Azeta, S. Misra, V. C. Osamor, L. Fernandez-Sanz, and V. Pospelova, “The Emerging Threat of Ai-driven Cyber Attacks: A Review,” *Appl. Artif. Intell.*, vol. 36, no. 1, p. 2037254, Dec. 2022, doi: 10.1080/08839514.2022.2037254.

[2] M. Binhammad, S. Alqaydi, A. Othman, and L. H. Abuljadayel, “The Role of AI in Cyber Security: Safeguarding Digital Identity,” *J. Inf. Secur.*, vol. 15, no. 02, pp. 245–278, 2024, doi: 10.4236/jis.2024.152015.

[3] P. Alaeifar, S. Pal, Z. Jadidi, M. Hussain, and E. Foo, “Current approaches and future directions for Cyber Threat Intelligence sharing: A survey,” *J. Inf. Secur. Appl.*, vol. 83, p. 103786, 2024, doi: https://doi.org/10.1016/j.jisa.2024.103786.

[4] U. Tariq, I. Ahmed, A. K. Bashir, and K. Shaukat, “A Critical Cybersecurity Analysis and Future Research Directions for the Internet of Things: A Comprehensive Review.,” *Sensors (Basel).*, vol. 23, no. 8, Apr. 2023, doi: 10.3390/s23084117.

[5] Y. Shanmugarasa, H. Paik, S. S. Kanhere, and L. Zhu, “A systematic review of federated learning from clients’ perspective: challenges and solutions,” *Artif. Intell. Rev.*, vol. 56, no. 2, pp. 1773–1827, 2023, doi: 10.1007/s10462-023-10563-8.

[6] H. Li, T. Xiezhang, C. Yang, L. Deng, and P. Yi, “Secure Video Surveillance Framework in Smart City.,” *Sensors (Basel).*, vol. 21, no. 13, Jun. 2021, doi: 10.3390/s21134419.

[7] M. Haghani *et al.*, “A roadmap for the future of crowd safety research and practice: Introducing the Swiss Cheese Model of Crowd Safety and the imperative of a Vision Zero target,” *Saf. Sci.*, vol. 168, p. 106292, 2023, doi: https://doi.org/10.1016/j.ssci.2023.106292.

[8] J. Pool, S. Akhlaghpour, F. Fatehi, and A. Burton-Jones, “A systematic analysis of failures in protecting personal health data: A scoping review,” *Int. J. Inf. Manage.*, vol. 74, p. 102719, 2024, doi: https://doi.org/10.1016/j.ijinfomgt.2023.102719.

[9] S. Sajadmanesh and D. Gatica-Perez, “ProGAP: Progressive Graph Neural Networks with Differential Privacy Guarantees,” *WSDM 2024 - Proc. 17th ACM Int. Conf. Web Search Data Min.*, pp. 596–605, 2024,

doi: 10.1145/3616855.3635761.

[10] W. S. Admass, Y. Y. Munaye, and A. A. Diro, "Cyber security: State of the art, challenges and future directions," *Cyber Secur. Appl.*, vol. 2, p. 100031, 2024, doi: https://doi.org/10.1016/j.csa.2023.100031.

[11] Y. Bi, Y. Li, X. Feng, and X. Mi, "Enabling Privacy-Preserving Cyber Threat Detection with Federated Learning," pp. 1–21.

[12] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Syst. Appl.*, vol. 242, p. 122807, 2024, doi: https://doi.org/10.1016/j.eswa.2023.122807.

[13] Z. Wang, J. Li, W. Wang, Z. Dong, Q. Zhang, and Y. Guo, "Review of few-shot learning application in CSI human sensing," *Artif. Intell. Rev.*, vol. 57, no. 8, p. 195, 2024, doi: 10.1007/s10462-024-10812-4.

[14] V. Semerenska, "Anomaly detection in surveillance camera data," no. May, 2023.

[15] S. M. Williamson and V. Prybutok, "Balancing Privacy and Progress: A Review of Privacy Challenges, Systemic Oversight, and Patient Perceptions in AI-Driven Healthcare," *Applied Sciences*, vol. 14, no. 2. 2024, doi: 10.3390/app14020675.

[16] F. R. Alzaabi and A. Mehmood, "A Review of Recent Advances , Challenges , and Opportunities in Malicious Insider Threat Detection Using Machine Learning Methods," *IEEE Access*, vol. 12, no. January, pp. 30907–30927, 2024, doi: 10.1109/ACCESS.2024.3369906.

[17] A. Chowdhary, K. Jha, and M. Zhao, "Generative Adversarial Network (GAN)-Based Autonomous Penetration Testing for Web Applications," *Sensors*, vol. 23, no. 18. 2023, doi: 10.3390/s23188014.

[18] S. H. Oh, M. K. Jeong, H. C. Kim, and J. Park, "Applying Reinforcement Learning for Enhanced Cybersecurity against Adversarial Simulation," *Sensors*, vol. 23, no. 6. 2023, doi: 10.3390/s23063000.

AUTHORS

First Author – D.H. Senevirathne, Cybersecurity Undergraduate, Sri Lanka Institute of Information Technology, Sri Lanka, it21298608@my.sliit.lk
Second Author –W.M.M. Gunasekara, Cybersecurity Undergraduate, Sri Lanka Institute of Information Technology, Sri Lanka, it21226496@my.sliit.lk
Third Author – K.P.A.T. Gunawardhana, Cybersecurity Undergraduate, Sri Lanka Institute of Information Technology, Sri Lanka, it21058196@my.sliit.lk
Fourth Author – M.F.F. Ashra, Cybersecurity Undergraduate, Sri Lanka Institute of Information Technology, Sri Lanka, it21380396@my.sliit.lk
Fifth Author – Isuranga Nipun Kumara, Associate Cybersecurity Consultant, Sri Lanka Institute of Information Technology, Sri Lanka, isuranganipunkumara@gmail.com
Sixth Author – Harinda Fernando, Assistant Professor, Sri Lanka Institute of Information Technology, Sri Lanka, harinda.f@sliit.lk
Seventh Author – Kavinga Yapa, Lecturer, Sri Lanka Institute of Information Technology, Sri Lanka, kavinga.y@sliit.lk

Correspondence Author D.H. Senevirathne, it21298608@my.sliit.lk, +94 76 796 3650