**Rajarambapu Institute of Technology, Rajaramnagar**



**Department of Information Technology**

1. **Y. 2023-24**

***Capstone Project Synopsis***

|  |  |
| --- | --- |
| **Domain Area** | Machine Learning, |
| **Title of the project** | Federated Machine Learning for Anomaly Detection and Intrusion Detection in Modbus-Enabled Industrial Networks |
| **Project Guide Name** | Prof. D.T. Mane |
| **Team Leader’s Name** | Sayali Aundhakar |
| **Group Number** | 15 |

**Members**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Roll No.** | **Name** | **Email** | **Phone** | **Sign** |
| 1 | 2260002 | Pranita Patil | 2260002@ritindia.edu | 9322220210 |  |
| 2 | 2260003 | Swati Patil | 2260003@ritindia.edu | 9763473033 |  |
| 3 | 226000 | Sayali Aundhkar | 2260004@ritindia.edu | 7385939810 |  |
| 4 | 2110040 | Tejal Yadav | 2110040@ritindia.edu | 8485842959 |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Sign of Project Guide** | **Sign of Project In-charge** | **Sign of HOD, IT** |

1. **Introduction**

The integration of Federated Learning (FL) and a Machine Learning-Based Intrusion Detection System (ML-IDS) within Modbus-enabled industrial networks marks a significant advancement in enhancing cybersecurity for industrial environments.

Federated Learning, a decentralized machine learning approach, stands out as a key technology in this context. It enables the training of machine learning models across multiple devices or nodes without centralizing sensitive data. In the realm of industrial networks, this approach is particularly crucial as it allows for the development of robust intrusion detection models while respecting the privacy and security of sensitive information.

The focus on Modbus-enabled industrial networks is strategic, considering the prevalence of Modbus communication protocol in industrial automation systems. This protocol is widely used for communication between various devices and sensors in industrial settings, making it a prime target for potential cyber threats. By integrating Federated Learning with a Machine Learning-Based Intrusion Detection System specifically tailored for Modbus environments, the aim is to create a more adaptive and responsive security framework.

The machine learning component of the system adds an extra layer of sophistication. Through continuous learning from network behaviors and anomalies, the Intrusion Detection System becomes adept at identifying and mitigating potential security breaches. This amalgamation of technologies not only enhances the overall cybersecurity posture of Modbus-enabled industrial networks but also aligns with the dynamic and evolving nature of cyber threats in the industrial landscape. In essence, the collaborative synergy between Federated Learning and Machine Learning-Based Intrusion Detection System for Modbus-enabled networks represents a forward-thinking approach in fortifying the security infrastructure of critical industrial processes.

**2**. **Objectives**

1. Conduct a comprehensive literature review on existing approaches and methodologies for decentralized anomaly detection in industrial networks, specifically focusing on Modbus attacks.

1. Build a machine learning-based Intrusion Detection System for detecting and

classifying Modbus attacks in real-time.

3. To prepare and publish research paper.

**3. Literature Survey**

[1] Aitor Belenguer, Jose A. Pascual, and Javier Navaridas. GöwFed A novel federated network intrusion detection system.

Introduces GöwFed as an innovative solution to the challenges faced by network intrusion detection systems (NIDS). The primary objective of the study is to address the evolving landscape of NIDS by proposing a novel approach that combines Federated Learning and Gower Dissimilarity matrices. The authors emphasize the potential of GöwFed in enhancing the detection of network threats within industrial-level networks. The key objectives include addressing privacy constraints, reducing communication overheads, and providing a privacy-friendly infrastructure. The methodology employed in the research involves two main components: GöwFed system design and evaluation, and a comparative analysis with a centralized approach. For the GöwFed system design and evaluation, the authors utilize the TON\_IOT dataset. In the first approach, a vanilla Multilayer Perceptron (MLP) model is employed, achieving exceptionally high accuracy, precision, and recall rates (up to 99%). This success justifies the viability of Gower distances for effective threat detection purposes. The second approach involves a Federated Learning model designed to operate in simulated distributed environments, achieving commendable accuracy metrics, albeit to a slightly lesser extent compared to the centralized approach. The research extends its applicability to real streaming data adaptation, implementing GöwFed as a Fully Distributed FL-IDS without Central Orchestration. This approach aims to minimize dependence on a centralized coordinator, offering a decentralized and scalable solution for intrusion detection. The paper highlights the significance of this approach in addressing the limitations of centralized models in real-world scenarios.

[2] Priyanka Verma, John G. Breslin, and Donna O'Shea in 2022. FLDID: Federated Learning Enabled Deep Intrusion Detection in Smart Manufacturing Industries.

Introduces the Federated Learning Enabled Deep Intrusion Detection (FLDID) framework as a solution to the challenges of detecting cyber threats in smart manufacturing industries. The paper begins with an overview of Smart Manufacturing Industries and Industry 4.0, highlighting the need for robust cybersecurity measures in such environments. The proposed FLDID framework addresses the limitations of detecting cyber threats in smart manufacturing, particularly emphasizing the issues of limited data and privacy preservation. The authors employ the "X-IIoTID" intrusion dataset to evaluate their framework, achieving an impressive accuracy of 99.79%. The paper also discusses privacy and security measures embedded in FLDID, emphasizing the utilization of federated learning and deep learning techniques. A comparison with baseline studies and traditional machine learning classifiers further validates the effectiveness of the FLDID framework. The conclusion highlights the success of FLDID in enhancing intrusion detection and outlines potential future research directions in this domain. Overall, the paper provides a comprehensive exploration of addressing cybersecurity challenges in smart manufacturing through the integration of federated learning and deep learning techniques.

[3] A. Shubha, Dr A. Kanagaraj, S. Sathiyapriya , N. Balakumar , P. Karthiga. Federated learning for intrusion detection system: concept, challenges and future direction

Published in 2023, delves into the exploration of Federated Learning (FL) for Intrusion Detection Systems (IDS) within the context of smart manufacturing industries. The primary objectives of the research include conceptualizing FL for IDS, identifying associated challenges, and providing potential solutions for future research endeavors. Serving as a baseline for prospective studies, the paper underscores the importance of developing robust algorithms and techniques to safeguard data privacy and mitigate communication overhead in FL-based IDS implementations. The methodology involves the utilization of Machine Learning (ML) and Deep Learning (DL) in conjunction with IDS to achieve elevated classification accuracy. The paper advocates for decentralized ML solutions, aligning with the increasing distribution of data across networks of interconnected devices in smart manufacturing industries. Despite the absence of a specified dataset in the paper, the authors emphasize the potential to enhance the accuracy of data privacy through the implementation of techniques such as differential privacy, homomorphic encryption, and secure multi-party computation. The identified challenges in FL implementations for IDS within smart manufacturing industries encompass communication efficiency, data privacy, and model heterogeneity. The paper underscores the necessity for further research to address these issues and advance the development of FL-based IDS systems. Overall, the research contributes to the foundation of knowledge in FL for IDS, offering insights into challenges and proposing avenues for future research aimed at improving the security and efficiency of intrusion detection in smart manufacturing environments.

[4] Ohammed Aledhari (Member, Ieee), Rehma Razzak Reza M. Parizi (Senior Member, Ieee), And Fahad Saeed (Senior Member, Ieee). Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications

It provides a comprehensive survey aimed at expanding the understanding of Federated Learning (FL) on a technical level. The key objectives include elucidating the current platforms, protocols, and architectures of FL, offering a deeper insight into its technical aspects, and presenting a comprehensive overview of FL in terms of definition, applicability, and usefulness. The methodology involves data collection, organization, and presentation to contribute valuable insights into FL architectures, use cases, and design practices. The authors utilized the BraTS 2018 dataset, containing MRI scans of nearly 300 individuals with brain tumors, and data from the Caltrans Performance Measurement System (PeMS) database, offering traffic flow information from over 30,000 detectors. The paper highlights that the accuracy of FL compared to traditional Machine Learning methods varies based on specific use cases, data, and implementation. The study identifies the need for further research, emphasizing challenges related to communication, system heterogeneity, and privacy within the FL framework. Additionally, it calls for increased focus on incorporating FL into various wireless networks, including IoT, while ensuring the reliability of participants involved in FL. In summary, the paper serves as a valuable resource, presenting a comprehensive survey of FL technologies, protocols, and applications, and outlining key areas for future research and development.

[5] Helio N. Cunha Neto, Jernej Hribar, Ivana Dusparic, Diogo Menezes Ferrazani Mattos, And Natalia C. Fernandes (2023). A Survey on Securing Federated Learning:

Analysis of Applications, Attacks, Challenges, and Trends.

Internet Of Things (IOT) is growing Technology right now, results in increasing vast amounts of data, it is significantly enhancing the performance of Machine Learning Models. However, it is challenging to collect vast amounts of data and train model due to privacy laws. Federated Learning is a collaborative machine leaning approach where data is trained on local devices. It is beneficial because data is not shared to any device. Unfortunately, several in-design vulnerabilities have been exposed that gives chance to attackers to infer private data. It is negatively impacts to Federated Learning model. So,it needs a solution to tackle those problems. This paper provides a comprehensive analysis of secure Federated Learning Proposals which protects user privacy and enhance the model performance. They performed systematic review and comparative analysis based upon predefined data of vulnerabilities addressed by extracting data from electronic devices and secure analysis of Federated Learning applications. To analyze the data ML/DL algorithms, IDS model and Federated models are used.

[6] Enrique Mármol Campos, Pablo Fernández Saura, Aurora González-Vidal, José L. Hernández-Ramos, Jorge Bernal Bernabé, Gianmarco Baldini, Antonio Skarmeta. Evaluating Federated Learning for intrusion detection in Internet of Things: Review and challenges.

The application of Machine Learning Technique to well-known Intrusion Detection Systems plays an vital role to enhance the cybersecurity by detection of cyber attacks.ML based IDS used centralized approach where IOT devices share data to central server which arises a privacy concern. To enhance privacy of data Federated approach is used in many sectors like HealthCare, Education. Although, using Federated Learning approach arises security challenge. So, this paper evaluates Federated Learning Approach based on multi-class classifier with consideration of data distributions for detection of attacks in IOT. And evaluate impact of different aggregation functions using IBMFL framework.

[7] Slavica V. Boštjančič Rakas, Mirjana D. Stojanović, and Jasna D. Marković-Petrović. "Review of Research Work on Network-Based SCADA

offers a comprehensive review of research work on network-based security controls (SC). The primary objectives include critically evaluating Intrusion Detection System (IDS) solutions and identifying research gaps within signature-based, anomaly-based detection methods, and machine-learning-based algorithms such as Genetic Algorithms (GA), clustering, outlier detection, as well as knowledge-based approaches like Fuzzy Logic (FA) and expert systems, and statistical-based methods including time series models and cumulative sum. Unfortunately, the paper does not specify the dataset used, but it attains a high level of accuracy. The study highlights the inadequacy of signature-based techniques, emphasizing the need for improvement. Performance evaluation and completeness of the proposed methods are discussed, and the applicability of these techniques to the Future Internet Environment (FIN) is considered. In summary, the paper provides a thorough analysis of IDS solutions, identifies gaps in existing methodologies, and emphasizes the need for advancements in signature-based techniques for enhanced security in network environments.

[8] Maria Nawaz Chohan1, Usman Haider2, Muhammad Yaseen Ayub3, Hina Shoukat3, Tarandeep KaurBhatia4 and Muhammad Furqan Ul Hassan3. Detection of Cyber Attacks using Machine Learning based Intrusion Detection System for IoT Based Smart Cities.

This research paper explores the evolving dynamics of the world with the integration of artificial intelligence (AI) into smart cities, powered by the Internet of Things (IoT). While smart cities offer convenience, security concerns hinder their development. The paper focuses on addressing these concerns through an Intrusion Detection System (IDS) that monitors network traffic and alerts in case of anomalies. The study presents a comparative analysis of Machine Learning-based IDS algorithms, including ADA Boost, Linear Support Vector Machine, Auto Encoder Classifier, Quadratic Support Vector Machine, and Multi-Layer Perceptron, using the UNSW-NB15 dataset. ADA Boost stands out with an impressive accuracy of 98.3%. The introduction highlights the significance of addressing vulnerabilities in smart cities, emphasizing the diverse applications such as e-government, smart homes, intelligent transportation, and more. The role of IDS in ensuring secure communication channels and combating cyber-attacks is crucial in the context of smart city infrastructure.

[9] D. Srinivas, Prabhakar Kumar Thakur, Pramod Kumar, K. Sindhu Sri Vani, Mr. S. Nirmal Sam. Network Intrusion Detection Using Machine Learning.

This research paper addresses the growing security threats in wireless communication and proposes an efficient Intrusion Detection System (IDS) using Principal Component Analysis (PCA) and the Random Forest classification algorithm. The approach aims to enhance IDS accuracy by organizing the dataset through dimensionality reduction with PCA and utilizing Random Forest for classification. The results indicate superior efficiency compared to other techniques such as SVM, Naïve Bayes, and Decision Tree, with a performance time of 3.24 minutes, an accuracy rate of 96.78%, and an error rate of 0.21%. The introduction emphasizes the significance of network security in the face of increasing data integration and network services, highlighting the role of Network Intrusion Detection Systems (NIDS) in detecting attacks by monitoring network activities. The paper underscores the necessity for accurate and swift IDS responses to distinguish between internal and external attacks, contributing to the overall security of computer networks.

[10] Nabila Farnaaz∗ and M. A. Jabbar. Random Forest Modeling for Network Intrusion Detection System.

This research paper addresses the increasing significance of intrusion detection in the context of growing technological usage. Focusing on the development of an effective and robust Intrusion Detection System (IDS), the study employs the Random Forest (RF) classifier as an ensemble method, highlighting its superiority over traditional classifiers in classifying attacks. Utilizing experiments on the NSL-KDD dataset, the proposed model demonstrates efficiency with a low false alarm rate and a high detection rate. The paper emphasizes the rising threat of intrusion activities in computer systems and categorizes IDS into host-based and network-based systems. It introduces the RF algorithm to detect various attack types and applies 10-fold cross-validation and feature selection to enhance classification accuracy. Symmetrical uncertainty is employed for feature selection, addressing issues of information gain. Comparative analysis with the J48 classifier showcases the superiority of the proposed RF model in terms of accuracy, detection rate, false alarm rate, and Matthews correlation coefficient. The paper concludes by suggesting future work involving the application of evolutionary computation for feature selection to further improve classifier accuracy.

[11] Haicheng Qu, Jitao Qin1(&), Wanjun Liu, and Hao Chen. Instruction Detection in SCADA/Modbus Network Based on Machine Learning

The research paper addresses the increasing sophistication of cyber security threats in industrial control systems and the limitations of intrusion detection based solely on network communication behavior. Machine learning, specifically One-Class Support Vector Machine (OCSVM), is applied to detect anomalies in the Supervisory Control and Data Acquisition (SCADA) network. The study evaluates various supervised learning methods and builds an anomaly detection model using OCSVM, analyzing its effectiveness through metrics such as recall rate, accuracy rate, false positive rate, and false negative rate. The paper emphasizes the unique challenges of industrial control network intrusion, highlighting the need for machine learning and data mining to address defects in industrial hardware and control systems. The results suggest that the OCSVM-based model achieves high accuracy, while the Decision Tree method performs commendably in detecting intrusion behavior. The paper concludes with insights into the limitations of unsupervised learning in intrusion detection and proposes the exploration of semi-supervised approaches for improved models.

[12] Wenbin Yu , Yiyin Wang and Lei Song. A Two Stage Intrusion Detection System for Industrial Control Networks Based on Ethernet/IP.

The research paper addresses the security challenges posed by the implementation of standard Ethernet in Industrial Control Systems (ICS) and proposes an improved Intrusion Detection System (IDS) tailored to specific industrial scenarios. The paper outlines three attack models (infiltration, creative forging, and false data injection) and introduces a two-stage IDS comprising a traffic prediction model based on ARIMA and an anomaly detection model using OCSVM. The traffic prediction model detects infiltration attacks by forecasting abnormal changes in traffic patterns, while the anomaly detection model identifies malicious control instructions in Ethernet/IP packets. The proposed method demonstrates outstanding performance in detecting various attacks compared to other IDSs, showcasing its effectiveness in ensuring the security of modern ICS facing advanced threats from the Internet. The paper emphasizes the need for IDS that reflects ICS behavior characteristics and copes with vulnerabilities for satisfactory overall accuracy and false alarm rates.

[13] Mukesh Kumar Yadav1, Mahaiyo. Ningshen2. Enhancement of Intrusion Detection System using Machine Learning

The paper addresses the growing challenges of internet-based attacks by proposing a machine learning-based Intrusion Detection System (IDS) capable of predicting four different attack types: Denial of Service (DOS), Probe, Remote to Local (R2L), and User to Root (U2R). To enhance IDS performance, the paper introduces an ensemble model using the chi-squared feature selection method on the NSL-KDD dataset. The proposed model, AdaBoost with Logistic Regression, outperforms other models in terms of accuracy, precision, recall, and F1-measure. The study emphasizes the effectiveness of machine learning algorithms in processing vast data, detecting harmful behavior, and promptly identifying various attacks. The proposed IDS demonstrates superior performance compared to existing state-of-the-art models, and the paper concludes with discussions on challenges and future scope.

[14] Dapeng Man ,Fanyi Zeng, Wu Yang, Miao Yu,Jiguang Lv, and Yijing Wang. Intelligent Intrusion Detection Based on Federated Learning for Edge-Assisted Internet of Things.

The research paper discusses the challenges in the security of edge-assisted IoT, considering the exponential growth of data generated at the network edge. Traditional centralized cloud computing faces limitations in terms of data transmission costs, network bandwidth, and data privacy, leading to the emergence of edge computing. Edge-assisted IoT involves migrating network functions and data processing to the edge of the network, providing lower latency, flexible access, and enhanced data privacy. However, security threats, including malware, hacking, system vulnerabilities, and unauthorized access, pose significant risks to the benefits of edge computing. The paper proposes an intelligent intrusion detection mechanism called FedACNN, based on Federated Learning (FL) aided Convolutional Neural Network (CNN), designed to address these security challenges in edge-assisted IoT. The proposed mechanism demonstrates high accuracy in intrusion detection while protecting data privacy, showcasing its potential contributions to security research in edge-assisted IoT. The research concludes with expectations for future contributions to the protection of edge-assisted IoT, including intrusion detection on encrypted traffic data.

[15] Solane Duquea\*, Dr.Mohd. Nizam bin Omarb .Using Data Mining Algorithms for Developing a Model for Intrusion Detection System (IDS)

A study utilized k-means clustering for an Intrusion Detection System (IDS) using the NSL-KD dataset, consisting of 25,192 entries with 22 data types. Results showed varying efficiency rates, false positive rates, and false negative rates for different cluster numbers. Optimal performance occurred when the number of clusters matched data types. The challenge lies in dynamically determining the cluster number. Recommendations include exploring other data mining techniques, combining k-means with a signature-based approach to reduce false negatives, and developing a system for automatic cluster number identification. Identifying the correct number of clusters is crucial for optimal results in this model.

[16] Mohanad Sarhan a,\*, Siamak Layeghy a, Nour Moustafa b, Marcus Gallagher a, Marius Portmann. Feature extraction for machine learning-based intrusion detection in IoT networks

The paper addresses the unreliability of current Network Intrusion Detection Systems (NIDSs) in IoT networks. Researchers often aim to improve classification results through various combinations of Feature Reduction (FR) and Machine Learning (ML) techniques on diverse NIDS datasets. Six ML models and three FE algorithms are evaluated using UNSW-NB15, ToN-IoT, and CSE-CIC-IDS2018 datasets. Optimal dimensions for feature extraction are identified for each dataset, with Linear Discriminant Analysis (LDA) degrading performance in two datasets. The results emphasize that no single FE method or ML model universally excels across all datasets, highlighting the need for a benchmark feature set to advance research in the field. The paper also calls for increased focus on the generalizability of proposed algorithms, advocating for generic feature sets applicable in various NIDS datasets and practical network settings.

[17] Imran Hidayat1, Muhammad Zulfiqar Ali2 and Arshad2,\* Machine Learning-Based Intrusion Detection System: An Experimental Comparison

This research focuses on enhancing intrusion detection systems (IDS) using machine learning (ML) techniques on the TON\_IoT dataset, which features new and recent attack types and features. A hybrid feature selection method combining Pearson correlation coefficient and a random forest model is proposed. ML algorithms, including decision tree, AdaBoost, and K-nearest neighbor (KNN), are trained and tested, with decision tree achieving nearly 99.6% accuracy. Deep learning (DL) techniques, such as multilayer perceptron (MLP) and long short-term memory (LSTM), are also applied, with MLP reaching an accuracy of almost 99.2% on the TON\_IoT dataset. The study concludes that ML techniques, particularly decision tree and MLP, provide optimal accuracy with minimal false-positive and false-negative rates for effective intrusion detection in networks.

[18] Tirthankar Ghosh , Sikha Bagui , Subhash Bagui , Martin Kadzis and Jackson Bare Anomaly Detection for Modbus over TCP in Control Systems Using Entropy and Classification-Based Analysis

The article proposes a statistical approach for detecting anomalies in industrial control systems traffic, emphasizing the limitations of traditional methods in such environments. It introduces an extended analysis using cluster-based entropy and a classification-based approach on Modbus over TCP/IP data from an HVAC system. The study focuses on Denial of Service (DOS), Man-in-the-Middle (MITM), and Reconnaissance attacks. While univariate and bivariate entropy analysis detects anomalies in DOS and MITM attacks, it falls short in identifying reconnaissance stages. The classification-based analysis with eight classifiers reveals that J48 and random forest perform well, with the best results achieved in classifying DOS attacks (97.87% accuracy) using an unbalanced data split (60–40). MITM attacks achieve a lower accuracy of 82.81% using an 80–20 non-attack vs. attack data split with a 75–25 training vs. testing split. The research provides insights into improving anomaly detection in industrial control systems, specifically addressing the challenges in the Cyber Kill Chain.

[19] Vasiliki Kelli, Vasileios Argyriou, Thomas Lagkas , George Fragulis , Elisavet Grigoriou and Panagiotis Sarigiannidis . IDS for Industrial Applications: A Federated Learning Approach with Active Personalization

This paper addresses the security challenges posed by the widespread integration of Internet of Things (IoT) devices, particularly in critical infrastructures. The focus is on developing an Intrusion Detection System (IDS) for safeguarding these infrastructures using a combination of federated learning and active learning. Federated learning is employed for private model training in a collaborative fashion, while active learning serves as a semi-supervised approach for global model adaptation to individual participant's network traffic. Experimental results demonstrate that the global models significantly enhance performance for each participant when locally personalized with just a few active learning queries. The study emphasizes the efficacy of combining federated and active learning to achieve better IDS model performance with minimal data samples, ensuring security, privacy, and adaptability to diverse local traffic patterns.

[20] Shaojie Yang ÿID ,Wei Zheng,Meiyun Xie, and Xueyang Zhang. Research of Federated Learning Application Methods and Social Responsibility

This research provides a comprehensive review of federated learning's application patterns across various fields, addressing challenges such as trust risks, regulatory disparities, and data-sharing requirements. The study outlines four dimensions of social responsibility for federated learning: compliance application, system security mechanisms, trust mechanisms, and ethical security. By examining current characteristics and regulatory demands, the paper discusses federated learning as a cross-platform privacy protection solution. It emphasizes the growing acceptance of federated learning as a practical encrypted computing model and explores its potential future developments. The study also identifies challenges, including regulatory compliance, cybersecurity, ethical concerns, and a crisis of confidence, proposing attributions and guidelines to address these risks.

[21] Léo Lavaur, Yann Busnel, Fabien Autrel, Marc-Oliver Pahl. The Evolution of Federated Learning-based Intrusion Detection and Mitigation: a Survey

This research addresses the challenges in securing Information Technology (IT) and Operational Technology (OT) infrastructures due to the complexity of interconnected heterogeneous networks. The paper emphasizes the importance of collaboration and information-sharing for effective cybersecurity. Intrusion Detection Systems (IDSs) are essential for monitoring and detecting attacks, but traditional processes are slow. The introduction of Federated Learning (FL) presents a promising solution by enabling local detection and mitigation with low latency while preserving privacy. The paper reviews existing FL-based intrusion detection approaches and proposes the term Federated Intrusion Detection System (FIDS) for FL-based IDS. It aims to structure and complete the understanding of FIDS through a structured literature review, qualitative and quantitative analyses, taxonomy development, and reference architecture establishment. The survey addresses questions about FIDS usage in different domains, differences in FIDS architectures, and the state of the art in FIDS. The contributions include a comprehensive review and a structured framework for understanding FIDS.

[22] MOHAMED AMINE FERRAG, OTHMANE FRIHA, LEANDROS MAGLARAS, (Senior Member, IEEE), HELGE JANICKE, (Member, IEEE), AND LEI SHU, (Senior Member, IEEE) ] Federated Deep Learning for Cyber Security in the Internet of Things: Concepts Applications, and Experimental Analysis.

This article conducts a comprehensive study on federated deep learning approaches for cybersecurity in IoT applications. The review covers various IoT domains, including Industrial IoT, Edge Computing, Internet of Drones, Healthcare Things, and Vehicles, highlighting federated learning's security and privacy systems. The discussion includes federated learning with blockchain and malware/intrusion detection systems for IoT. Vulnerabilities in federated learning-based security systems are examined. An experimental analysis using three deep learning approaches (RNN, CNN, DNN) is presented, comparing centralized and federated learning under real IoT traffic datasets. Results show that federated deep learning approaches outperform classic/non-federated versions, ensuring privacy and providing higher accuracy in attack detection. The study aims to provide valuable insights into federated deep learning with emerging technologies for IoT cybersecurity.

[23] PHAM KHANH QUAN , (Graduate Student Member, IEEE), MAJID KUNDROO , (Graduate Student Member, IEEE), AND TAEHONG KIM , (Senior Member, IEEE). Experimental Evaluation and Analysis of Federated Learning in Edge Computing Environments.

This study addresses the limited research on federated learning (FL) performance in edge computing environments. Utilizing a KubeEdge-based FL framework, the research conducts experimental evaluations, emphasizing aspects like system and statistical heterogeneity, communication bandwidth, and various factors affecting model convergence time. The results reveal significant impacts on FL model convergence in edge computing settings. The study contributes to the field by providing comprehensive insights and guidelines for optimizing FL performance in heterogeneous environments, offering developers valuable tools to build efficient systems in edge computing. The findings emphasize the need for further research to develop robust and scalable FL systems capable of handling heterogeneity challenges in edge computing, highlighting FL's potential importance in this context.

[24] Léo Lavaur, Benjamin Costé, Marc-Oliver Pahl, Yann Busnel, Fabien AutrelFederated Learning as enabler for Collaborative Security between not Fully-Trusting Distributed Parties  
This paper explores the integration of heterogeneous distributed data sources using federated learning, emphasizing trust-building without revealing sensitive information. Federated learning allows collaborators to share machine learning models while maintaining privacy. However, the challenge lies in determining justified trust levels for each contributor to optimize joint models effectively. The paper suggests addressing this issue through experiments on a platform with generated datasets, highlighting various use cases and proposing envisioned solutions for future work. The contributions aim to increase the security of organizations by leveraging federated learning to create a common trusted security model for inherently distributed systems, fostering collaboration without compromising critical internals.

[25] SAFDAR HUSSAIN JAVED, MAAZ BIN AHMAD, MUHAMMAD ASIF ,WASEEM AKRAM2, KHALID MAHMOOD, (Senior Member, IEEE), ASHOK KUMAR DAS, (Senior Member, IEEE), AND SACHIN SHETTY, (Senior Member, IEEE). APT Adversarial Defence Mechanism for Industrial IoT Enabled Cyber-Physical System.

This study addresses the challenge of detecting Advanced Persistent Threat (APT) attacks in Industrial Internet of Things (I-IoT)-enabled Cyber-Physical Systems (CPSs). Leveraging Graph Attention Network (GAN), a novel multidimensional algorithm, the proposed approach captures behavioral features in a manner not achieved by prior Deep Learning (DL) methods. Evaluation on DAPT2020 malware and Edge I-IoT datasets demonstrates high detection accuracy (96.97% and 95.97%) with a prediction time of around 20 seconds. Compared to conventional ML algorithms, the GAN-based approach significantly improves performance in the I-IoT-enabled CPS realm. The study emphasizes the superiority of DL in detecting complex APT malware in the hazardous domain of I-IoT-enabled CPSs, suggesting potential future exploration areas such as other neural network variants, attention mechanisms for interpretability, and incorporating edge features for enhanced problem-solving.

[26] Dapeng Man , Fanyi Zeng, Wu Yang, Miao Yu , Jiguang Lv, and Yijing Wang. Intelligent Intrusion Detection Based on Federated Learning for Edge-Assisted Internet of Things

The introduction highlights the rapid growth of the Internet of Things (IoT) and the challenges arising from the massive data generated at the network edge. Traditional centralized cloud computing faces limitations such as data transmission costs and privacy concerns. To address these issues, edge computing has emerged, bringing data processing closer to end-users. However, security challenges persist, with threats including malware, hacking, system vulnerabilities, and unauthorized access. The paper emphasizes the importance of intrusion detection in the edge-assisted IoT architecture to prevent potential damage, especially concerning personal safety. The subsequent sections provide an overview of intrusion detection for IoT, federated learning, and detail the proposed intrusion detection method, experimental setup, and results.

[27] Jos´e L. Hern´andez-Ramos, Georgios Karopoulos, Efstratios Chatzoglou, Vasileios Kouliaridis, Enrique Marmol, Aurora Gonzalez-Vidal and Georgios Kambourakis Intrusion Detection based on Federated Learning: a systematic review.

The development of Intrusion Detection Systems (IDSs) has evolved with the integration of machine learning (ML) techniques, especially in the context of big data and sophisticated cybersecurity threats. Traditional approaches involve centralized learning architectures, sharing data with data centers for analysis. However, the emergence of federated learning (FL) has garnered significant interest, allowing collaborative intrusion detection without the need to share raw data. This paper presents a comprehensive taxonomy for FL-enabled IDS approaches, based on a thorough literature survey from 2018 to 2022. It analyzes ML models, datasets, aggregation functions, and implementation libraries used in FL-enabled IDS approaches, offering insights into the current state, challenges, and future directions in this rapidly evolving field.

[28] Muhammad Asad 1,\* , Ahmed Moustafa 1,2 and Chao Yu . A Critical Evaluation of Privacy and Security Threats in Federated Learning

This paper addresses privacy concerns in Federated Learning (FL) due to the massive data exchange in the Internet of Things (IoT). While FL prevents direct data leakage from clients during model aggregation, the study investigates privacy and security threats to the FL execution process. Practical solutions to counter these attacks and protect individual privacy are proposed, supported by experimental results from a publicly available dataset. The work aims to shed light on potential threats in FL, providing evidence of their significance and suggesting future research directions for secure FL deployment.

[29] Viraaji Mothukuri , Graduate Student Member, IEEE, Prachi Khare , Reza M. Parizi , Senior Member, IEEE, Seyedamin Pouriyeh , Associate Member, IEEE, Ali Dehghantanha , Senior Member, IEEE, and Gautam Srivastava , Senior Member, IEEE. Federated-Learning-Based Anomaly Detection for IoT Security Attacks

This article introduces a federated learning-based anomaly detection approach for enhancing IoT network security. Addressing privacy concerns, the approach employs decentralized on-device data and utilizes federated training rounds on gated recurrent units (GRUs) models. Only the learned weights are shared with the central server, preserving local IoT device data privacy. The ensembler part aggregates updates from multiple sources to optimize the global machine learning model's accuracy. Experimental results demonstrate the approach's superior performance over classic/centralized machine learning versions in attack detection, emphasizing improved user data privacy in IoT networks. Future work aims to enhance the approach with an IoT device testbed and live data evaluation.

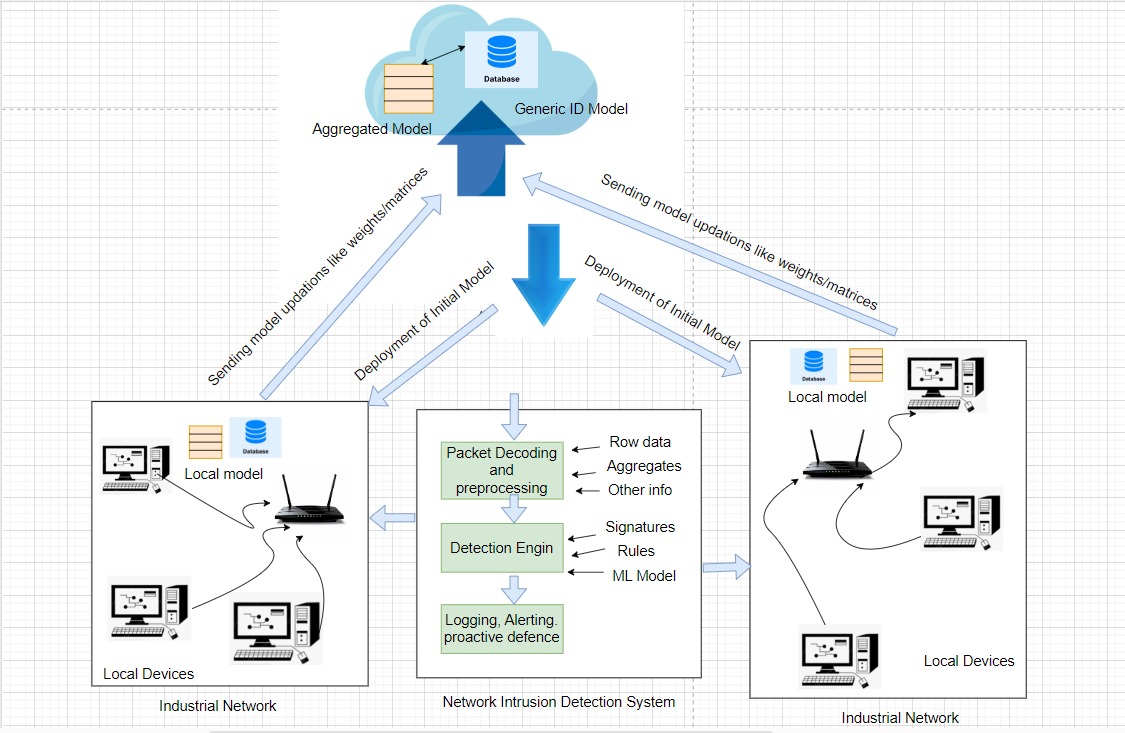
[30] Milad Nasr, Reza Shokri, Amir Houmansadr. Comprehensive Privacy Analysis of Deep Learning:Passive and Active White-box Inference Attacks against Centralized and Federated Learning

This study focuses on privacy analysis of deep learning models, particularly susceptibility to white-box inference attacks. The research evaluates various scenarios, including centralized and federated learning, passive and active inference attackers, and different adversary prior knowledge levels. The study presents novel white-box membership inference attacks designed to trace training data records, emphasizing that well-generalized models are significantly susceptible to such attacks. The research investigates privacy vulnerabilities in the stochastic gradient descent algorithm, revealing susceptibility even in stand-alone and federated settings. The findings underscore the need for addressing privacy concerns in deep learning models to mitigate potential inference attacks.

1. **Problem Statement**

Securing industrial networks, especially against Modbus protocol attacks, is crucial for protecting critical infrastructure like substations. However, the lack of a comprehensive understanding of decentralized anomaly detection and the absence of specialized Intrusion Detection Systems (IDS) for real-time Modbus attack classification pose significant challenges. This research aims to fill these gaps through a literature review and the implementation of a machine learning-based IDS, contributing essential insights for research and practical implementation to enhance the cybersecurity of Modbus-enabled industrial networks.

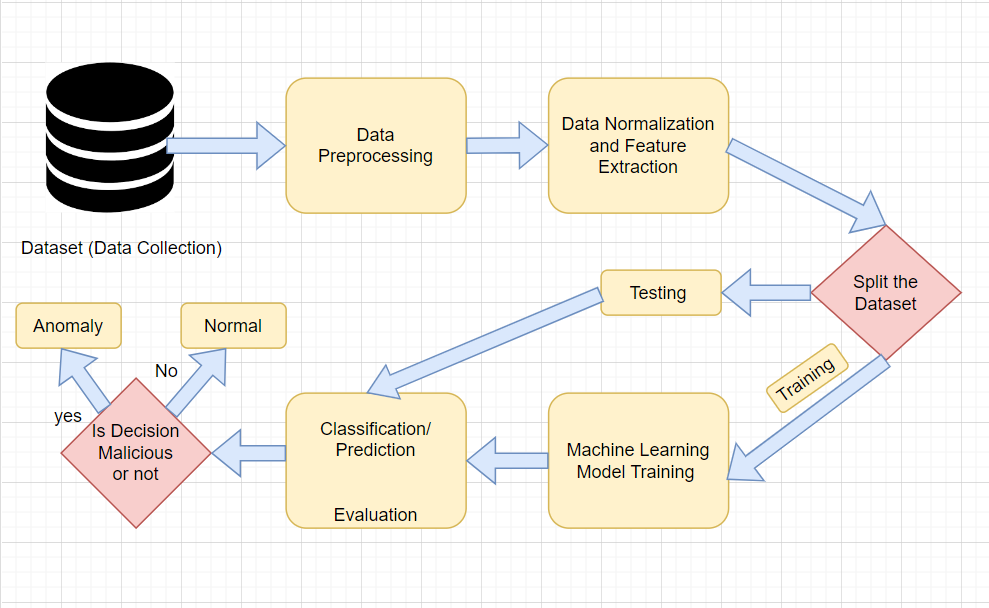
1. **Problem Description**



**Fig 1. System Architecture**

The increasing digitization of critical infrastructure, particularly in substation networks utilizing the Modbus protocol, has exposed vulnerabilities to cyber threats. Modbus, lacking native security features, becomes a potential entry point for unauthorized access, leading to risks of data manipulation, service disruptions, and even physical damage. Traditional intrusion detection systems struggle to keep pace with sophisticated attacks, and centralized solutions raise concerns about data privacy and network resilience. Addressing these challenges, our project aims to deploy Federated Machine Learning for Anomaly Detection and Intrusion Detection in Modbus-Enabled Substation Networks, ensuring robust cybersecurity measures while preserving data integrity and privacy.

1. **System Architecture**

****

**Fig 2. System Architecture**

1. **Modules**
2. **Data Preparation**

**Dataset Selection:**

* + 1. Utilize both the attack and benign datasets from the CIC Modbus Dataset.
    2. Ensure a balanced representation of attack and normal traffic in the dataset.

1. **Feature Selection:**

**Identify Relevant Features:**

* + 1. Select features crucial for intrusion detection, including source IP, destination IP, protocol, port numbers, and other Modbus-specific attributes.

1. **Data Preprocessing:**

**Handling Missing Data:**

* + 1. Check for and handle missing or null values in the dataset.

**Normalization/Scaling:**

1. Normalize or scale numerical features to bring them to a similar scale.
2. **Model Selection:**

**Choose Supervised Learning Algorithm:**

* + 1. Select a suitable supervised learning algorithm for classification (e.g., Random Forest, Support Vector Machine, Gradient Boosting).

1. **Training the IDS:**

**Splitting the Dataset:**

* + 1. Divide the dataset into training and testing sets to assess the model's generalization.

**Model Training:**

1. Train the selected classification algorithm on the training dataset using identified features.
2. **Hyperparameter Tuning:**

**Grid Search or Random Search:**

* + 1. Utilize grid search or random search to fine-tune hyperparameters of the selected algorithm.
    2. Optimize parameters for achieving better model performance.

1. **Model Evaluation:**

**Performance Metrics:**

* + - 1. Evaluate the IDS model on the testing dataset using metrics such as accuracy, precision, recall, and false positive rate.
      2. Assess the model's ability to distinguish between normal and attack traffic.

1. **Scenario-Specific Evaluation:**

**Different Attack Scenarios:**

1. Assess the model's performance for each specific attack scenario present in the dataset.
2. Identify any variations in classification accuracy across different types of Modbus attacks.
3. **Feasibility Analysis:**

**Real-world Deployment Considerations:**

* + - 1. Analyze the feasibility of deploying the trained IDS in a real-world substation network.
      2. Consider factors such as computational resources, latency, and adaptability to dynamic network conditions.

1. **Results and Findings:**

**Assessment of Model Accuracy:**

1. Summarize the results, highlighting the accuracy, precision, recall, and false positive rate achieved by the IDS.

**Hyperparameter Tuning Impact:**

1. Discuss the impact of hyperparameter tuning on the model's performance.

**Feasibility Insights:**

1. Provide insights into the practicality of deploying the trained IDS in substation networks.
2. **Conclusion and Recommendations:**

**Summary of Findings:**

1. Conclude the study by summarizing the key findings.

**Recommendations:**

1. Offer recommendations for further improvements or adjustments to enhance the IDS's effectiveness.
2. **System requirements with justification**

Table No. 1 Hardware Requirements

|  |  |  |
| --- | --- | --- |
| Sr.No. | Hardware Requirements | Specifications |
| 1 | Processing Unit (CPU) | Multi-core processor(i5 core) |
| 2 | Memory (RAM) | Minimum 8 GB RAM |
| 3 | Storage | SSD storage minimum 512GB SSD |
| 4 | Network Interface Cards (NIC) | High-speed NICs |
| 5. | GPU | NVIDIA or AMD GPU |

|  |  |  |
| --- | --- | --- |
| Sr.No. | Software Requirements | Specifications |
| 1 | Python Environment with Package management tools (pip, conda) | **Python 3.12.2** version or less |
| 2 | Federated Learning Framework | TensorFlow Federated (TFF) or PySyft |
| 3 | Machine Learning Libraries | TensorFlow or PyTorch |
| 4 | Modbus Library | pyModbus |
| 5 | Data Preprocessing Libraries | Pandas, NumPy |

Table No. 2 Software Requirements

1. **Project Outcome**
2. Identification of existing approaches, challenges, and advancements in decentralized anomaly detection for industrial networks, with a focus on Modbus attacks, culminating in recommendations for future research and practical implementation.
3. Successful development and implementation of a machine learning-based Intrusion Detection System (IDS) for real-time detection and classification of Modbus attacks, accompanied by an in-depth analysis of model performance, hyperparameter tuning impact, feasibility for real-world deployment, and insights for future research.
4. Research Paper publication in reputed journal/conference.
5. **References**

[1] Aitor Belenguer, Jose A. Pascual, and Javier Navaridas. GöwFed A novel federated network intrusion detection system.

[2] Priyanka Verma, John G. Breslin, and Donna O'Shea in 2022 .FLDID: Federated Learning Enabled Deep Intrusion Detection in Smart Manufacturing Industries.

[3] A. Shubha, Dr A. Kanagaraj, S. Sathiyapriya , N. Balakumar , P. Karthiga. Federated learning for intrusion detection system: concept, challenges and future direction

[4] Ohammed Aledhari (Member, Ieee), Rehma Razzak Reza M. Parizi (Senior Member, Ieee), And Fahad Saeed (Senior Member, Ieee) .Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications

[5] Helio N. Cunha Neto, Jernej Hribar, Ivana Dusparic, Diogo Menezes Ferrazani Mattos, And Natalia C. Fernandes(2023). A Survey on Securing Federated Learning:

Analysis of Applications, Attacks, Challenges, and Trends.

[6] Enrique Mármol Campos, Pablo Fernández Saura, Aurora González-Vidal,José L. Hernández-Ramos, Jorge Bernal Bernabé, Gianmarco Baldini, Antonio Skarmeta. Evaluating Federated Learning for intrusion detection in Internet of Things: Review and challenges.

[7] Slavica V. Boštjančič Rakas, Mirjana D. Stojanović, and Jasna D. Marković-Petrović. "Review of Research Work on Network-Based SCADA

[8] Maria Nawaz Chohan1, Usman Haider2, Muhammad Yaseen Ayub3, Hina Shoukat3, Tarandeep KaurBhatia4 and Muhammad Furqan Ul Hassan3. Detection of Cyber Attacks using Machine Learning based Intrusion Detection System for IoT Based Smart Cities.

[9] D. Srinivas, Prabhakar Kumar Thakur, Pramod Kumar, K. Sindhu Sri Vani, Mr. S. Nirmal Sam. Network Intrusion Detection Using Machine Learning.

[10] Nabila Farnaaz∗ and M. A. Jabbar. Random Forest Modeling for Network Intrusion Detection System

[11] Haicheng Qu, Jitao Qin1(&), Wanjun Liu, and Hao Chen. Instruction Detection in SCADA/Modbus Network Based on Machine Learning

[12] Wenbin Yu , Yiyin Wang and Lei Song. A Two Stage Intrusion Detection System for Industrial Control Networks Based on Ethernet/IP.

[13] Mukesh Kumar Yadav1, Mahaiyo. Ningshen2. Enhancement of Intrusion Detection System using Machine Learning

[14] Dapeng Man ,Fanyi Zeng, Wu Yang, Miao Yu,Jiguang Lv, and Yijing Wang. Intelligent Intrusion Detection Based on Federated Learning for Edge-Assisted Internet of Things.

[15] Solane Duquea\*, Dr.Mohd. Nizam bin Omarb .Using Data Mining Algorithms for Developing a Model for Intrusion Detection System (IDS)

[16] Mohanad Sarhan a,\*, Siamak Layeghy a, Nour Moustafa b, Marcus Gallagher a, Marius Portmann. Feature extraction for machine learning-based intrusion detection in IoT networks

[17] Imran Hidayat1, Muhammad Zulfiqar Ali2 and Arshad2,\* Machine Learning-Based Intrusion Detection System: An Experimental Comparison

[18] Tirthankar Ghosh , Sikha Bagui , Subhash Bagui , Martin Kadzis and Jackson Bare Anomaly Detection for Modbus over TCP in Control Systems Using Entropy and Classification-Based Analysis

[19] Vasiliki Kelli, Vasileios Argyriou, Thomas Lagkas , George Fragulis , Elisavet Grigoriou and Panagiotis Sarigiannidis . IDS for Industrial Applications: A Federated Learning Approach with Active Personalization

[20] Shaojie Yang ÿID ,Wei Zheng,Meiyun Xie, and Xueyang Zhang. Research of Federated Learning Application Methods and Social Responsibility.

[21] Léo Lavaur, Yann Busnel, Fabien Autrel, Marc-Oliver Pahl. The Evolution of Federated Learning-based Intrusion Detection and Mitigation: a Survey

[22] MOHAMED AMINE FERRAG, OTHMANE FRIHA, LEANDROS MAGLARAS, (Senior Member, IEEE), HELGE JANICKE, (Member, IEEE), AND LEI SHU, (Senior Member, IEEE) ] Federated Deep Learning for Cyber Security in the Internet of Things: Concepts Applications, and Experimental Analysis.

[23] PHAM KHANH QUAN , (Graduate Student Member, IEEE), MAJID KUNDROO , (Graduate Student Member, IEEE), AND TAEHONG KIM , (Senior Member, IEEE). Experimental Evaluation and Analysis of Federated Learning in Edge Computing Environments.

[24] Léo Lavaur, Benjamin Costé, Marc-Oliver Pahl, Yann Busnel, Fabien AutrelFederated Learning as enabler for Collaborative Security between not Fully-Trusting Distributed Parties

[25] SAFDAR HUSSAIN JAVED, MAAZ BIN AHMAD, MUHAMMAD ASIF ,WASEEM AKRAM2, KHALID MAHMOOD, (Senior Member, IEEE), ASHOK KUMAR DAS, (Senior Member, IEEE), AND SACHIN SHETTY, (Senior Member, IEEE). APT Adversarial Defence Mechanism for Industrial IoT Enabled Cyber-Physical System.

[26] Dapeng Man , Fanyi Zeng, Wu Yang, Miao Yu , Jiguang Lv, and Yijing Wang. Intelligent Intrusion Detection Based on Federated Learning for Edge-Assisted Internet of Things

[27] Jos´e L. Hern´andez-Ramos, Georgios Karopoulos, Efstratios Chatzoglou, Vasileios Kouliaridis, Enrique Marmol, Aurora Gonzalez-Vidal and Georgios Kambourakis

Intrusion Detection based on Federated Learning: a systematic review.

[28] Muhammad Asad 1,\* , Ahmed Moustafa 1,2 and Chao Yu . A Critical Evaluation of Privacy and Security Threats in Federated Learning

[29] Viraaji Mothukuri , Graduate Student Member, IEEE, Prachi Khare , Reza M. Parizi , Senior Member, IEEE, Seyedamin Pouriyeh , Associate Member, IEEE, Ali Dehghantanha , Senior Member, IEEE, and Gautam Srivastava , Senior Member, IEEE. Federated-Learning-Based Anomaly Detection for IoT Security Attacks

[30] Milad Nasr, Reza Shokri, Amir Houmansadr. Comprehensive Privacy Analysis of Deep Learning:Passive and Active White-box Inference Attacks against Centralized and Federated Learning