**AMERICAN INTERNATIONALA close up of a sign

Description automatically generated**

**UNIVERSITY-BANGLADESH**

**Faculty of Science and Technology**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Assignment Title: | Cybersecurity in the Internet of Things (IoT) and Industrial IoT (IIoT) | | | |
| Assignment No: | 01 | | Date of Submission: | 11 December 2024 |
| Course Title: | RESEARCH METHODOLOGY | | | |
| Course Code: | 00801 | | Section: | H |
| Semester: | Fall | 2025-26 | Course Teacher: | DR. MD. ABDULLAH - AL - JUBAIR |

**Declaration and Statement of Authorship:**

1. I/we hold a copy of this Assignment/Case-Study, which can be produced if the original is lost/damaged.
2. This Assignment/Case-Study is my/our original work and no part of it has been copied from any other student’s work or from any other source except where due acknowledgement is made.
3. No part of this Assignment/Case-Study has been written for me/us by any other person except where such collaborationhas been authorized by the concerned teacher and is clearly acknowledged in the assignment.
4. I/we have not previously submitted or currently submitting this work for any other course/unit.
5. This work may be reproduced, communicated, compared and archived for the purpose of detecting plagiarism.
6. I/we give permission for a copy of my/our marked work to be retained by the Faculty for review and comparison, including review by external examiners.
7. I/we understand thatPlagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a formofcheatingandisaveryseriousacademicoffencethatmayleadtoexpulsionfromtheUniversity. Plagiarized material can be drawn from, and presented in, written, graphic and visual form, including electronic data, and oral presentations. Plagiarism occurs when the origin of them arterial used is not appropriately cited.
8. I/we also understand that enabling plagiarism is the act of assisting or allowing another person to plagiarize or to copy my/our work.

*\* Student(s) must complete all details except the faculty use part.*

\*\* Please submit all assignments to your course teacher or the office of the concerned teacher.

|  |  |
| --- | --- |
| Group Name/No.: |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Name** | **ID** | **Program** | **Signature** |
| 1 | TAUHID HASAN | 22-46438-1 | BSc [CSE] |  |
| 2 | RIAD AL HASAN | 22-46732-1 | BSc [CSE] |  |
| 3 | MD. MOSTAFIJUR RAHMAN | 22-47161-1 | BSc [CSE] |  |
| 4 | JOTHIRMOY SARKER SHUVO | 22-46473-1 | BSc [CSE] |  |
| 5 | MD YEASIN NEWAZ | 22-46803-1 | BSc [CSE] |  |

|  |  |  |
| --- | --- | --- |
| ***Faculty use only*** | | |
| FACULTYCOMMENTS | **Marks Obtained** |  |
|  |
|  |
|  | **Total Marks** |  |
|  |
|  |

1. **Topic of Interest: Cybersecurity in the Internet of Things (IoT) and Industrial IoT (IIoT)**

The Internet of Things (IoT) and Industrial IoT (IIoT) are revolutionizing industries and everyday life by connecting devices, machines, and systems. However, the rapid adoption of these technologies has introduced substantial cybersecurity risks. Many devices within the IoT and IIoT ecosystems are resource-constrained, leaving them vulnerable to attacks such as Denial of Service (DoS), Man-in-the-Middle (MITM), SQL injection, and ransomware. As IoT continues to expand, securing these devices has become an urgent priority for businesses and consumers. Research indicates that traditional security models, relying on perimeter defenses and centralized authentication, are inadequate for IoT environments due to their distributed and heterogeneous nature. Moreover, the vast amounts of data generated by IoT devices, including sensitive information on user activities, operational processes, and system performance, provide attackers with numerous entry points. The complexity of managing and securing extensive IoT networks has driven the development of new security solutions, including intrusion detection systems (IDS), machine learning-based models, and blockchain technologies. Emerging approaches like Privacy-Preserving Fixed-Length Encoding (PPFLE) aim to enhance threat detection accuracy while maintaining data privacy, making them particularly suited for IoT applications. Despite these advancements, securing IoT and IIoT remains a significant challenge due to the constantly evolving nature of cyberattacks and the rapid deployment of new devices. As a result, ensuring robust cybersecurity for IoT and IIoT ecosystems continues to be a critical research focus, emphasizing AI-based solutions and privacy-preserving techniques to address these challenges effectively.

1. **Problem Statement**

The ongoing expansion of the Internet of Things (IoT) and Industrial IoT (IIoT) across industries like manufacturing, healthcare, and smart cities has introduced unprecedented vulnerabilities and cybersecurity risks. The swift deployment of IoT devices often surpasses the implementation of adequate security measures, leaving critical systems exposed to cyber threats such as Denial of Service (DoS), Man-in-the-Middle (MITM), and ransomware attacks. Traditional security models, designed for centralized and static networks, are ill-equipped to handle the dynamic and distributed nature of IoT and IIoT environments. Compounding the issue is the significant volume of data generated by these devices, much of which is sensitive and susceptible to unauthorized access. Furthermore, the limited computational power of many IoT devices makes them ideal targets for exploitation. Given these challenges, there is a pressing need to develop innovative cybersecurity solutions tailored to the unique characteristics of IoT and IIoT networks. While advancements have been made in machine learning-based intrusion detection systems (IDS) and privacy-preserving methods like Privacy-Preserving Fixed-Length Encoding (PPFLE), these solutions alone are insufficient to address the diverse attack vectors present in IoT ecosystems. The primary challenge lies in designing scalable cybersecurity strategies capable of real-time threat detection, mitigation, and privacy preservation. How can emerging technologies like AI, machine learning, and blockchain be leveraged to secure IoT and IIoT networks against evolving threats?

To address this issue, this study aims to investigate the integration of AI-based models and privacy-preserving methods into robust security frameworks for IoT and IIoT. The focus will be on developing solutions that enhance real-time threat detection while safeguarding sensitive data, ensuring a secure and reliable IoT ecosystem.

1. **Primary Research Question:**
2. How effective are machine learning-based intrusion detection systems in identifying and mitigating common IoT and IIoT attacks, such as Denial of Service (DoS), Man-in-the-Middle (MITM), and SQL injection?
3. What role do privacy-preserving techniques, such as Privacy-Preserving Fixed-Length Encoding (PPFLE), play in securing sensitive data generated by IoT devices without compromising the accuracy of cyberattack classification models?
4. How can blockchain technology be integrated into IoT and IIoT systems to enhance data integrity, trust, and security while preventing unauthorized access to sensitive information?
5. What are the challenges and limitations of applying traditional cybersecurity models (e.g., firewalls, antivirus software) to IoT and IIoT networks, and how can these limitations be overcome using more modern approaches, such as AI and machine learning?
6. How can real-time monitoring and threat detection systems be optimized to handle the diverse nature of IoT/IIoT devices and protocols, ensuring timely responses to network anomalies or attacks?
7. What are the key factors that contribute to the vulnerability of IoT and IIoT devices, and how can security protocols be designed to address these factors across different industries (e.g., healthcare, smart cities, manufacturing)?
8. How can automated security systems be developed to dynamically adapt to the continuously evolving nature of IoT and IIoT attacks, ensuring long-term protection without requiring manual intervention?
9. **Here are two specific aims/objectives related to the primary research question:**
10. To investigate the effectiveness of AI-based models and machine learning algorithms in detecting and mitigating cyberattacks in IoT and IIoT networks, with a particular focus on real-time threat detection and attack classification for common attacks such as DoS, MITM, and SQL injection.
11. To evaluate the role of privacy-preserving techniques, such as Privacy-Preserving Fixed-Length Encoding (PPFLE), in securing sensitive data generated by IoT devices while maintaining the accuracy of cybersecurity models and ensuring compliance with data privacy standards.
12. **Hypothesis:**

The hypothesis posits that AI-based models and machine learning algorithms can significantly enhance the security of IoT (Internet of Things) and IIoT (Industrial Internet of Things) networks by improving real-time threat detection and mitigating cyberattacks. As IoT/IIoT ecosystems grow and become more complex, traditional security systems are often ineffective due to the large number of devices and the diverse nature of threats. AI models, particularly those based on machine learning, are expected to offer advanced capabilities by analyzing vast amounts of data quickly, identifying patterns, and detecting anomalies that indicate cyber threats. These AI models can learn from historical attack data, allowing them to identify both known and unknown attack types with greater accuracy and speed than traditional systems.

In addition to AI’s ability to detect threats, the hypothesis suggests that integrating privacy-preserving techniques, such as Privacy-Preserving Fixed-Length Encoding (PPFLE), will further enhance security by protecting sensitive data. IoT and IIoT devices generate large volumes of sensitive data (e.g., user activity, operational information) that are vulnerable to theft or misuse. PPFLE is a privacy-focused technique that helps ensure data security while maintaining the effectiveness of cyberattack detection systems. This technique is designed to allow for the analysis of encrypted or obfuscated data, which will help prevent unauthorized access while still enabling the model to classify and detect cyber threats accurately.

1. **Literature Review**

Numerous research efforts have established the BERT model as an exceptional starting point for identifying cybersecurity threats. Its applications span a variety of domains, from detecting log anomalies to identifying malicious web requests. Alkhatib et al. [10] demonstrated the potential of using BERT in the context of a Controller Area Network (CAN) by applying a "masked language model" unsupervised training objective. Their proposed CAN-BERT transformer model excelled in anomaly detection within modern automotive systems, showcasing improved accuracy and F1-scores over traditional approaches. This study underscores BERT's ability to handle sequence data effectively, making it highly applicable to cybersecurity tasks.

Similarly, Rahali and Akhloufi [6] introduced MalBERT, an innovative tool designed to perform static analysis on the source code of Android applications. MalBERT leverages BERT to comprehend contextual relationships between code tokens and classify them into representative malware categories. The results from their experiments highlight the model's high performance in accurately identifying malware, emphasizing the transformative potential of transformer-based architectures in addressing malicious software detection challenges.

Chen and Liao [8] developed BERT-Log, a robust anomaly detection framework for large-scale computer systems. This approach treats system logs as natural language sequences, enabling the BERT model to learn semantic representations of both normal and anomalous logs. The authors fine-tuned the pre-trained BERT model with a fully connected neural network, achieving significant improvements in detecting system abnormalities. Their work demonstrates how natural language processing techniques can be effectively extended to cybersecurity domains involving log data.

Seyyar et al. [7] explored the use of BERT in detecting anomalous HTTP requests in web applications, utilizing deep learning techniques to enhance detection accuracy. Their findings highlight the adaptability of BERT in diverse cybersecurity scenarios, particularly in addressing web application vulnerabilities. Similarly, Aghaei et al. [11] proposed SecureBERT, a specialized language model tailored for Cyber Threat Intelligence (CTI) tasks. By transforming natural language CTI into machine-readable formats, SecureBERT minimizes the need for labor-intensive manual analysis. This was achieved through the development of a unique tokenizer and fine-tuning pre-trained weights to optimize performance for both general English and cybersecurity-specific text. However, SecureBERT's scope does not extend to network-based cyber threat detection, highlighting an area for potential improvement.

Ranade et al. [12] introduced CyBERT, a customized version of BERT fine-tuned with extensive cybersecurity datasets. This adaptation enhances the model's ability to process intricate information related to threats, attacks, and vulnerabilities. Their research further solidifies the role of BERT in handling domain-specific data effectively. Yu et al. [13] employed BERT to detect advanced persistent threats (APTs) in the Industrial Internet of Things (IIoT). By addressing the challenges posed by long attack sequences, their approach demonstrated exceptional performance, achieving high accuracy and low false alarm rates.

Breve et al. [14] focused on detecting harmful automation rules within IoT platforms such as If-This-Then-That (IFTTT). Using a BERT-based model, they evaluated over 76,000 rules and achieved significantly higher accuracy compared to traditional information flow analysis techniques. This work highlights BERT's potential in safeguarding user privacy and security in IoT environments. Additionally, Wang et al. [9] proposed BERT-of-Theseus, a lightweight IoT intrusion detection model. Designed for resource-constrained environments, this model employs a Siamese network for feature reduction and a Vision Transformer to train a compact PoolFormer model. By optimizing parameter efficiency without compromising accuracy, BERT-of-Theseus addresses the unique challenges of IoT security.

These studies collectively highlight the adaptability of BERT-based models for various cybersecurity tasks, demonstrating their ability to process textual and sequential data such as code, emails, and log sequences. By leveraging BERT's capability to capture contextual relationships, researchers have achieved precise detection and classification results across different domains. However, challenges persist in real-world scenarios, particularly concerning privacy in training data. For instance, network traffic data often contains sensitive information, raising concerns about data sharing and analysis. SecurityBERT addresses these issues by introducing a lightweight, privacy-preserving architecture tailored for cybersecurity applications.

Table 1 compares recent works on cyber threat detection, focusing on four key parameters:

**D** = Detect: Network-based Cyber Threat Detection

**L** = LLM: Utilization of LLMs

**N** = Network PCAP: Packet data analysis of traffic

**P** = Privacy: Privacy-preserving training data

Research from 2022, such as [6], [10], and [11], highlights the integration of large language models (LLMs) for cybersecurity applications but lacks support for network-based threat detection and packet data analysis. Conversely, studies by Hamouda et al. [1] and Friha et al. [2] focused on detecting cyber threats and analyzing packet data but did not utilize LLMs. More recent efforts in 2023, including [3], [4], and [5], have made significant strides in combining network-based detection with packet data analysis. However, the application of LLMs in these studies remains limited, signaling an opportunity for further advancement in leveraging LLMs for comprehensive cybersecurity frameworks.

Overall, the evolution of BERT-based models in cybersecurity research underscores their transformative impact, with promising directions for improving privacy-preserving mechanisms and integrating network-based threat detection capabilities.

1. **Selection of Design: Mixed Methods**

**Reason for Selection:**

The Mixed Methods research design was selected due to its ability to integrate quantitative and qualitative methodologies, enabling a comprehensive exploration of the research problem. This approach ensured a holistic understanding by combining measurable statistical analyses with nuanced, real-world insights. The research was aimed at assessing the performance of AI-driven models and privacy-preserving strategies for cybersecurity in IoT and IIoT networks. This included the evaluation of their technical efficacy through precise metrics (quantitative) and the investigation of challenges and practical applications of these techniques in diverse environments (qualitative).

* 1. **Quantitative Aspects:**
* Effectiveness Evaluation:The performance of AI-based algorithms in detecting and mitigating cyber threats was measured through statistical analysis. Metrics such as accuracy, precision, recall, and F1 scores were utilized, along with real-time detection capabilities.
* Performance Metrics Analysis: Data related to the number of detected threats, system resource utilization, false positive rates, and success in attack mitigation was analyzed. This provided insights into the effectiveness of techniques like Privacy-Preserving Fixed-Length Encoding (PPFLE) when integrated with machine learning models.
  1. **Qualitative Aspects:**
* **Practical Challenges:** Interviews and expert surveys were conducted to explore the complexities associated with deploying these models in real-world IoT/IIoT networks. This included examining the feasibility of privacy-preserving methods and their alignment with existing cybersecurity frameworks.
* **Contextual Insights:** Perspectives from industry stakeholders (e.g., healthcare and manufacturing) were gathered to highlight how different sectors implemented cybersecurity measures. Barriers, risks, and the applicability of AI-based models in specific contexts were identified.

By adopting this Mixed Methods approach, the research bridged the gap between technical performance metrics and the contextual factors shaping the adoption and impact of these cybersecurity solutions, resulting in a well-rounded perspective.

1. **Research Variables**
2. **Independent Variable:** 
   * + **AI-Based Models and Machine Learning Algorithms:** Advanced cybersecurity techniques, such as neural networks, decision trees, and hybrid models for threat detection and mitigation, were deployed.
     + **Privacy-Preserving Techniques:** Data protection methods, such as PPFLE, were applied to secure sensitive information while maintaining the effectiveness of AI-based threat detection models.
3. **Dependent Variable:**
   1. **Effectiveness of Threat Detection and Mitigation**: The ability to identify and respond to cyberattacks was measured. Key indicators included detection accuracy, false positive rates, response times, and success in neutralizing threats.
   2. **Classification Accuracy of Threat Models**: The ability of cybersecurity models to accurately classify various attack types was evaluated, particularly when privacy-preserving techniques were applied.

Relationship Between Variables: The impact of the independent variables (AI-driven models and privacy-preserving methods) on the dependent variables (detection effectiveness and classification accuracy) was analyzed. The study aimed to determine how these technologies improved the identification and mitigation of threats in IoT/IIoT systems while ensuring data privacy.

1. **Diagram**
2. **Reference**

**[1]** D. Hamouda, M. A. Ferrag, N. Benhamida, and H. Seridi, ‘‘PPSS: A privacy-preserving secure framework using blockchain-enabled federated deep learning for industrial IoTs,’’ Pervas. Mobile Comput., vol. 88, Jan. 2023, Art. no. 101738.

**[2]** O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, K.-K. R. Choo, and M. Nafaa, ‘‘FELIDS: Federated learning-based intrusion detection system for agricultural Internet of Things,’’ J. Parallel Distrib. Comput., vol. 165, pp. 17–31, Jul. 2022.

**[3]** S. Selvarajan, G. Srivastava, A. O. Khadidos, A. O. Khadidos, M. Baza, A. Alshehri, and J. C.-W. Lin, ‘‘An artificial intelligence lightweight blockchain security model for security and privacy in IIoT systems,’’ J. Cloud Comput., vol. 12, no. 1, p. 38, Mar. 2023.

**[4]** M. Douiba, S. Benkirane, A. Guezzaz, and M. Azrour, ‘‘An improved anomaly detection model for IoT security using decision tree and gradient boosting,’’ J. Supercomput., vol. 79, no. 3, pp. 3392–3411, Feb. 2023.

**[5]** X. Wang, C. Fang, M. Yang, X. Wu, H. Zhang, and P. Cheng, ‘‘Resilient distributed classification learning against label flipping attack: An ADMM-based approach,’’ IEEE Internet Things J., vol. 10, no. 17, pp. 15617–15631, Sep. 2023.

**[6]** A. Rahali and M. A. Akhloufi, *"MalBERT: Malware detection using bidirectional encoder representations from transformers,"* in Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), Oct. 2021, pp. 3226–3231.

**[7]** Y. E. Seyyar, A. G. Yavuz, and H. M. Ünver, *"An attack detection framework based on BERT and deep learning,"* IEEE Access, vol. 10, pp. 68633–68644, 2022.

**[8]** S. Chen and H. Liao, *"BERT-log: Anomaly detection for system logs based on pre-trained language model,"* Appl. Artif. Intell., vol. 36, no. 1, Dec. 2022, Art. no. 2145642.

**[9]** Z. Wang, J. Li, S. Yang, X. Luo, D. Li, and S. Mahmoodi, *"A lightweight IoT intrusion detection model based on improved BERT,"* [Details incomplete in provided text].

**[10]** N. Alkhatib, M. Mushtaq, H. Ghauch, and J.-L. Danger, *"CAN-BERT do it? Controller area network intrusion detection system based on BERT language model,"* in Proc. IEEE/ACS 19th Int. Conf. Comput. Syst. Appl. (AICCSA), Dec. 2022, pp. 1–8.

**[11]** E. Aghaei, X. Niu, W. Shadid, and E. Al-Shaer, *"SecureBERT: A domain-specific language model for cybersecurity,"* in Proc. Int. Conf. Secur. Privacy Commun. Syst. Cham, Switzerland: Springer, 2022, pp. 39–56.

**[12]** P. Ranade, A. Piplai, A. Joshi, and T. Finin, *"CyBERT: Contextualized embeddings for the cybersecurity domain,"* in Proc. IEEE Int. Conf. Big Data, Dec. 2021, pp. 3334–3342.

**[13]** K. Yu, L. Tan, S. Mumtaz, S. Al-Rubaye, A. Al-Dulaimi, A. K. Bashir, and F. A. Khan, *"Securing critical infrastructures: Deep-learning-based threat detection in IIoT,"* IEEE Commun. Mag., vol. 59, no. 10, pp. 76–82, Oct. 2021.

**[14]** B. Breve, G. Cimino, and V. Deufemia, *"Identifying security and privacy violation rules in trigger-action IoT platforms with NLP models,"* IEEE Internet Things J., vol. 10, no. 6, pp. 5607–5622, Mar. 2023.