

AI-Driven Cybersecurity for Industrial and Business IoT: Predictive Threat Detection

Rabia Akter Sukhi¹, Mahbub Hasan², Fahmida Akter³, Naim Rahman Ifti⁴

¹Doctor of Information Technology Trine University, USA

²Master of science in information studies Trine University, USA

³Master of Science in Information Technology Washington University of science and technology

⁴Bachelor of Science in Computer Science San Francisco Bay University



DOI : <https://doi.org/10.61796/ipteks.v3i4.646>



Sections Info

Article history:

Submitted: July 20, 2026

Final Revised: August 11, 2026

Accepted: September 16, 2026

Published: October 29, 2026

Keywords:

Cybersecurity threats

Industrial

Business

Network Traffic

MQTT protocol

AI-based

Dataset

ABSTRACT

Objective: The study examines cybersecurity threats in IoT environments used in the industrial and business sector by analyzing characteristics of network traffic and attacks in the MQTT protocol. The objective of the research is to look at the differences between legitimate traffic and various classes of attacks, to potentially support future AI-based threat detection. **Method:** A quantitative exploratory research design was selected, and the dataset was the MQTTEEB D processed, standardized data with 634644 observations equally distributed among legitimate traffic, denial-of-service attack, SlowTe attack, malformed data injection, brute force attack, and MQTT publish flooding. To explore relationships between the variables of the TCP and MQTT communications, descriptive statistics, distribution of the features, correlation analysis, and multivariate visualization techniques were used. **Results:** The results show that some network characteristics have characteristic patterns of behavior for each traffic type. However, TCP packet length and MQTT QoS show meaningful differences in the classes of attacks and legitimate communications, with comparatively small discriminatory capability for TCP time delta. Correlation analysis has been used to identify strong correlations between features of the MQTT protocol and interesting correlations between application layer and packet-level properties. Malicious traffic is also spread across wider communication patterns than legitimate traffic on average, with some patterns that overlap between categories, as seen in the integrated visual assessment. **Novelty:** The results indicate that only a combination of TCP and MQTT features might be useful for threat identification for cybersecurity monitoring. Although the study does not build or test predictive models, it presents a foundation for future research in AI-powered cybersecurity in industrial and business IoT environments and gives an evidence-based understanding of MQTT traffic behaviour.

INTRODUCTION

Background of Study

As a cornerstone of digital transformation, the Internet of Things (IoT) is a network of interconnected sensors, devices, communication systems, and information technology that enables the sharing of data and information between them. The proliferation of connected devices has led to an escalation in cybersecurity worries as well as interconnected environments opening up chances for unauthorised access, data theft, service disruption, and so much more. [1] Industrial and business IoT systems often share operational information via network protocols, requiring reliable communication that ensures service continuity and sound decision-making. Traditional security architectures may not recognize attacks unfamiliar in the IoT context because it involves a variety of devices, varied traffic patterns, and resource-limited components. Given these

circumstances, there is a need to continuously monitor and systematically analyze network behaviour to identify suspicious communication patterns. A study from 2025 underscores the need to explore the vulnerabilities of IoT and bolster protection by implementing appropriate detection and mitigation measures [1]. Knowing what constitutes normal and malicious network traffic can therefore be a good starting point for determining what the potential indicators of cyber threats are. In this work, the characteristics of the MQTT traffic and the types of attacks are discussed to analyze how these patterns could be utilized in the design of more intelligent predictive threat detection techniques in industrial and business IoT.

AI-Driven Cybersecurity in Industrial and Business IoT

AI can be used in industrial and business IoT networks to improve cybersecurity by detecting and analyzing network activity automatically and identifying suspicious activity. Various machine learning and deep learning techniques can assist in the detection of intrusions, recognizing anomalies, analyzing malware and classifying threats in the context of connected devices [2]. These capabilities apply to circumstances where the amount of communication data that needs to be monitored is enormous and the mode of attack is constantly evolving, making continuous monitoring difficult. There are also some practical challenges associated with AI-driven security strategies, such as computational demands, data quality issues, model interpretability, and adversarial attacks [2]. If you have a resource-constrained IoT device, you won't want to drain its resources with a complex detection system or risk the impact on efficiency. Therefore, it is beneficial to know the features and differences of the underlying traffic and to understand which category is an attack when making decisions on appropriate security policies. By examining the relationships between the characteristics of the communication and the observed threats, exploratory analysis can be used to verify evidence for further development of efficient and context-appropriate attack detection systems.

Problem Statement

Internet of Things technologies have made connected devices more vulnerable and exposed to potential security risks, which can have an impact on information security, operational reliability, or public safety. Unlike traditional computers, IoT devices are connected to real-world processes and can have limited hardware capabilities, specific communication interfaces, and various software configurations [3]. All of this is making vulnerability identification and effective risk management more difficult. Current security approaches should be able to deal with the threats arising from digital communication as well as from communication between physical signals and cyber systems [3]. In addition, an attacker can exploit several vulnerabilities at once and isolated security assessments are not enough to grasp the complexity of attacks [3]. In an IoT application with MQTT, malicious activities like denial of service, flooding, brute force attacks, and malformed data injection need to be investigated to understand the communication behaviour. Nevertheless, the determination of the difference between

legitimate and malicious traffic is still significant before developing an effective predictive detection mechanism. This study thus aims to systematically study both the network traffic characteristics and attack categories. Exploratory Data Analysis is used to provide insights into patterns that can be observed in the data that can help inform future threat detection strategies without attempting to train or validate a predictive model.

Research Objectives

1. To analyze network traffic characteristics associated with cyber threats by analysing normal and abnormal traffic, differences in the network traffic volume, connection duration, packet characteristics, protocols, and other features related to networks.

2. Comparing attack types by protocol, service, attack characteristics, and network use patterns to find exploratory patterns that can be used to inform predictive threat detection.

Research Question

What are the characteristics of the IoT network traffic, and what are the patterns of cyberattacks that can be used for exploratory analysis to support predictive threat detection in IoT environments for industrial and business applications?

Significance of the Study

This study is important because it introduces some of the visible traffic patterns of legitimate communication in an MQTT-based IoT environment, along with the various categories of cyberattacks. The study can be used to understand how malicious activities can deviate from typical network activities by analyzing TCP and MQTT communication features. The results could help researchers determine candidates for further study and aid in the design of future predictive threat detection systems. The research could also help cybersecurity practitioners comprehend the nature of an attack that needs to be taken into consideration when conducting network monitoring and risk assessment. It also has an exploratory method that offers a structured framework for traffic distribution interpretation, attack category comparison, and communication variable relationships. However, the results are not for the purpose of prediction but to aid in analytical understanding.

Industrial and Business IoT Security and Emerging Cyber Threats

Industrial Internet of Things (IoT) has revolutionized the way industrial management is conducted by linking physical equipment, sensors, communication networks, and computational systems together to enable automation and data-driven operations. The integration also increases the surface of attack, though as vulnerabilities can arise from connected devices, communication paths, software platforms and industrial control processes, it creates the potential for vulnerabilities. The combination of industrial IoT with edge computing brings in new challenges in cybersecurity on Cyber-Physical systems in architectures [4]. They have found denial-of-service attacks, ransomware, malware, and man-in-the-middle attacks as significant threats to industrial environments. One key aspect of their research is the focus on the interaction between

industrial IoT and edge computing, not in isolation as additional security elements. However, a general taxonomy of threats and protection mechanisms is not sufficient to give insights into which observable network properties are used to separate legitimate traffic from malicious traffic. The difference is important because when making decisions about the security needs of industry, one needs to understand the architecture as well as know what attacks are seen in real communication.

Tsiknas et al. [5] have already studied the correlation between connected industrial systems and cyber risk by carrying out a survey of attacks and countermeasures for industrial IoT systems. Their analysis pinpoints system heterogeneity, an inadequate set of interoperability standards, and the need to rely on various intermediaries as factors that make security more complicated. They also say that link encryption can not be enough when there are threats that can impact various layers of a networked system. Zhukabayeva et al.'s architectural focus is complemented by this view, which emphasizes the need for a more robust security strategy than is used in individual communication channels. The survey approach, however, mainly focuses on the organisation of known attacks and defensive strategies, which is beneficial for categorizing attacks, but less for establishing the strength of the association of specific traffic features with individual attack categories for a given dataset. As an example, if packet length is identified as a threat, it is important to know whether it is always different from other attacks, or if TCP timing or application-level message properties are always different. These relationships can be explored by performing exploratory traffic analysis, but a testbed's findings do not necessarily generalize to an industrial network.

In addition to the classification of attack, the following two aspects of cyber security should be taken into account: industrial management and organisational objectives, and operational continuity. Raimundo and Rosário [6] provide a discussion on IoT cybersecurity in industrial management, and position security in the broader context of connectedness in industrial and business activities. The management mindset is significant since a cyber incident will not just impact the confidentiality of data but also the availability of services, the reliability of processes, and confidence in connected operations. This wide-ranging perspective is not seen in attack-focused surveys, which tend to focus on the impact of poor protection on attacks. Management-level discussions, however, and technical threat catalogues answer other analytical questions. An organisation may identify an attack causing continuity issues without possessing enough evidence to identify the network behaviour that are associated with that attack. Likewise, a technical detection method can detect unusual traffic without telling you why it is abnormal. The connection between technical indicators and the risks they can pose to the operation of industry and business processes, especially communications systems that are used in time-sensitive processes, should therefore be taken into account in the context of effective IoT security.

Alnajim et al. [7] describe a more detailed classification of vulnerabilities and defensive measures for industrial IoT in a survey that considers the type of attacks, risks

associated with attacks, and defense measures in distributed industrial environments. The analysis highlights that the adoption of various hardware, software, and data sources poses a security challenge beyond simple encryption. This is a more technical approach than the management approach used by Raimundo and Rosário, where more attention is given to threats and possible protective responses. However, the scope of a survey also presents a challenge, as the measures suggested could vary in aspects such as implementation requirements, computing resources needed, and applicability for specific network configurations. A suitable solution for an industrial gateway with a significant amount of resources may not be suitable for an IoT device with limitations. In addition, examining the current countermeasures does not tell you whether they are effective under the traffic conditions that exist in a specific dataset. These restrictions apply to an MQTT environment where flooding, brute force, denial of service, and malformed message attacks can have different characteristics in the message communication. By examining labelled traffic and relationships between TCP and MQTT features, it is possible to learn more about these differences, but care should be taken not to make any generalizations beyond the dataset.

MQTT Communication and Cybersecurity Vulnerabilities in IoT Networks

Message Queuing Telemetry Transport (MQTT) is a lightweight messaging protocol, especially popular in the Internet of Things (IoT) application area, due to its publish-subscribe architecture where communication between devices works via a broker. This design minimizes direct communication between individual devices, making MQTT an ideal choice for resource-limited applications. However, the broker is a critical element of trust as the delivery of messages relies on the broker's configuration, access controls, and availability. Lakshminarayana et al. [8] study MQTT security from the application layer point of view and address issues related to authentication, authorization, confidentiality, and securing message exchanges. Their reviews are essential because they establish a relationship between protocol functionality and security flaws, and not because the MQTT protocol is so light that it is secure. However, application layer security is dependent on protocol usage, and a security mechanism does not ensure that it is correctly implemented across devices and consistently used.

Laghari et al. [9] take a wider look at the MQTT system and explore vulnerabilities, mitigation strategies, and future research directions. They've pointed out that risks can arise from weak authentication, insufficient access control, insecure communication, and configuration issues on the broker and/or client. The concerns in these are complementary to the application layer emphasis of Lakshminarayana et al., with the addition of a focus on interactions between multiple ecosystem components needed for MQTT security. For instance, if the authentication of a client is not sufficient, or topic permissions are not set, an unauthorized client may publish a message even if it is protected. Likewise, strict access control can be ineffective at stopping excessive message traffic from disrupting service. The comparison suggests MQTT security needs to be layered, not relying on a single security. Vulnerability surveys, however, mostly help to

determine the potential attack routes and measures to protect against attacks, and the network characteristics observable during attacks need to be investigated independently.

Hintaw et al. [10] analyze MQTT vulnerabilities, attack vectors, and possible solutions, complementing the analysis of the weaknesses in the protocol that can be exploited in IoT communication. They focus on threats to authentication, handling of messages, and availability of the communication services. This threat-based strategy can be helpful to illustrate the reason why lightweight messaging may be a security risk in the absence of deployment practices that take into account the assumptions behind trusted messaging. As opposed to Laghari et al., who focus on the wider MQTT ecosystem and future research avenues, Hintaw et al. offer a more direct link between vulnerabilities and attack methods and countermeasures. However, the classification of an attack vector does not necessarily provide insight into the regularity of its occurrence in network observations. Attack detection might be dependent on packet size, timing, TCP flags, message features, and intensity of communication in the practical world. These are properties that need to be investigated for various categories of attacks, not just assumed to be a sure sign of a specific attack.

Another layer is added to the security aspect of IoT due to the choice of protocols – the efficiency of the communications and the degree of vulnerability need to be considered simultaneously. Kolisnyk [11] explains the vulnerabilities in communication protocols and choosing protocols based on the requirements of IoT systems. This view is not limited to MQTT weaknesses, nor only to the security requirements of the application; it understands that the appropriateness of the protocol is a function of the system context and the security requirements of the application. Although MQTT is appealing for lightweight messaging, its benefits cannot be assumed to be the same for all operational settings. A protocol can satisfy the communication needs but be vulnerable due to weak authentication, inadequate permissions or inadequate monitoring. On the other hand, extra security measures can result in more processing load and complex implementations on limited devices. The focus of Kolisnyk's protocol selection viewpoint then supports reviews focused on MQTT by emphasizing the need to weigh the requirements for communication with the security risks. Moreover, comparing two protocols is not enough to determine which features are the most informative for detecting attacks in a specific data set of MQTT communications.

The big issue for empirical analysis is the linkage between vulnerabilities that have been documented in the literature and communication patterns that can be observed in the network. An MQTT attack can cause various impacts on packet timings, message sizes, connection behavior, and application-level message fields (denial of service, publish flooding, brute force attacks, and malformed data injection cases). However, there can be overlapping differences and some features of the standardized or transformed data sets may have different operational meanings from those they had in their original form. In some cases, a high message rate could be an indication of malicious activity (flooding) or of a high frequency of legitimate messages transmitted, depending

on the application context. Likewise, an abnormal TCP doesn't necessarily indicate that an attack has taken place. Labelled classes can be compared to each other to see if they share distributions, feature relationships, or might indicate a testable feature. It relies on clear pre-processing, a careful analysis of the classification labels, and an understanding that outcomes of a particular MQTT testbed may not translate to all IoT implementations. These constraints are important to differentiate discovery of candidate indicators and validation of an operational detection system.

AI-Driven and Predictive Cyber Threat Detection in IoT Environments

As Internet of Things (IoT) environments become more complex, a greater need has arisen for intelligent cybersecurity mechanisms to detect new threats before they seriously disrupt operations. While signature-based methods are still effective for detecting known attack patterns, relying on some signatures can restrict their ability to detect new and evolving attacks. Dhanushkodi and Thejas [12] highlight the potential of leveraging AI to enhance cyber threat detection by intelligently analysing and automatically mitigating threats. These capabilities, however, can only be effective if the training data is of high quality, the attack patterns are varied, and the ability to discriminate between actual threats and legitimate variations in network behaviour is high. This is especially critical for IoT applications, where devices are heterogeneous and produce different traffic patterns with highly disparate computing capacities. Therefore, the accuracy of detection can only be a partial assessment of security performance since it does not encompass overhead, false alarms, or consequences of delayed detection.

While AI can be useful in many IoT scenarios, it can cause new problems for cybersecurity, depending on the specific application. Abreu et al. explore the use of Artificial Intelligence for cyber threat detection in connected vehicles, which presents unique security challenges due to communication delays, interconnected electronic components, and mobility. They offer a valuable domain-specific perspective, as vehicle security isn't solely about spotting malicious network traffic. The detection mechanisms should also enable prompt response, while not impairing safety-critical functions. However, the suitability of AI-driven detection depends on the techniques' ability to adapt to evolving driving situations, varying vehicle architectures, and novel attack modes not previously observed. Solutions that work well in a controlled environment will not necessarily work well across a diversity of connected vehicles. This leads to doubts over the generalizability of detection techniques and whether the reported performance leads to reliable protection in actual operating scenarios.

Predictive security is a security model that goes beyond just detecting attacks in progress and looks for what might happen and the behavioral patterns that could lead to a security breach. Thayalan et al.'s analysis of consumer applications centers on real-time threat detection and AI-powered predictive security, emphasizing the need for rapid threat identification in today's dynamic digital landscape. The capabilities are applicable to IoT networks, as connected devices are constantly sharing information and can help detect suspicious behaviors before the attacks can take their toll. Prediction, however, is

very different from detection. An effective classification system of what has been observed is not always a reliable predictor of future threats. A careful analysis of the temporal information, the variable distributions of attacks, and the impact of false warnings is important for predictive performance. Furthermore, there is a practical balance between analytical complexity and response time when processing in real time. Advanced threat assessment models can be more complex and require significant amounts of computing power, but might not be feasible on an IoT device that has limited resources.

Another view of the relationship between intelligent detection and advancing cyber threats is adaptive network security. Sunkara [15] explores the integration of AI in the realm of advanced cyber attacks, highlighting the importance of intelligent threat detection and adaptive security solutions. This view supports predictive methods because it acknowledges the need for the detection system to adapt its behavior to the opponents and not rely solely on predetermined decision rules. However, adaptation brings inherent risks as well because legitimate network behavior can be altered and mimic an attack, and attackers can also be able to intentionally alter the information that the detection models use. Unlike domain-specific investigations of connected vehicles, a more general adaptive security approach allows for conceptual flexibility, but may need more contextual information for determining its appropriateness with respect to a given application of the IoT. In all of these methods, false positive rate, latency of detection, computational requirement, and resistance to new types of attacks are all significant evaluation criteria. A comprehensive understanding of these factors is essential to determine if AI-powered detection can offer meaningful, timely, and context-rich security in various IoT contexts.

Literature Gap

While previous studies have investigated the threats from the Industrial IoT, vulnerabilities in MQTT, and AI-assisted attack detection, few studies focus on correlating documented attack mechanisms with observed network traffic features. Additional exploratory analysis is required to investigate variations of labelled MQTT traffic classes within communication features and to look for any clues in the traffic that could be used to detect cybersecurity aspects in industrial IoT environments.

METHODOLOGY

Research Design

This study is conducted in a quantitative, exploratory research method to explore the threat of cybersecurity in industrial and business IoT in the Internet of Things by analyzing the network traffic in MQTT. The research aims to showcase how characteristics of the network and attack categories can be distinguished from each other by observing the differences between legitimate communication and malicious activities. In the approach of the research, descriptive statistics and graphical analysis are used to examine patterns in the chosen data, without building or testing a machine learning

model. This method enables analysis of the distribution of traffic, communication properties and correlations between TCP and MQTT features. The research design used is suitable for the identification of potential indicators of cyber threats and developing evidence-based understanding of network behaviour. The findings are not an assertion of performance in predicting detection and are only considered in the context of the data available.

Data Source and Description

The MQTTEEB-D Dataset Processed Standardized file (mqttteeb-d_dataset_Processed_standardized.csv) is used to analyze the communication behaviour of MQTT-based IoT devices in an exploratory manner [16]. It consists of 634,644 data rows and 13 columns pertaining to communication timestamps, TCP characteristics, MQTT protocol fields, and the target labels. The variables are: timestamp, tcp_flags, tcp_time_delta, tcp_len, mqtt_conack_flags, mqtt_conflag_cleansess, mqtt_conflags, mqtt_dupflag, mqtt_hdrflags, mqtt_kalive, mqtt_msg, mqtt_qos, and Target [16]. The target variable consists of six traffic types: legitimate traffic, denial of service, SlowITe, malformed data injection, brute force attacks, and MQTT publish flooding. The supplied processed file contains 105,774 records for each category. The data set is thus suitable to compare and contrast legitimate and malicious traffic, and to study patterns of communication among attack categories.

Data Preprocessing

Before doing exploratory analysis, it is important to check the structure of the data, variable types, missing observations, and duplicate records in the data; this process is called data preprocessing. The provided data set does not have any missing values, and there are 14 duplicate rows, which should be considered before discarding them. If it is desired to describe temporal patterns, the variable timestamp must be converted to a suitable date and time format. A categorical variable should be used for the target variable, as this will ensure the classes of attacks are represented correctly in visualizations. It is also important to see if there are constant values and limited variation of the numerical variables because if they are there, they could impact the correlation analysis and interpretation. The supplied file contains mqtt_dupflag, which is always 0, and should not be counted in the correlation calculations. Several features are standardized, so they do not necessarily correspond to the actual measurements in the packets without reference to the documentation of the dataset. The following steps help to ensure the analytical consistency and minimize the possibility of misinterpretation.

Exploratory Data Analysis

Exploratory data analysis is performed to find out if there is any relationship between the TCP/MQTT variables, variations in network characteristics, and the distribution of traffic categories. Descriptive summaries are generated to analyze the frequencies of each target class and the central tendency and spread of numerical features. This enables a comparison to be performed between different traffic types and different types of individual attacks to determine if there are any differences in

communication characteristics. Correlation analysis to explore associations between appropriate numerical variables, with the understanding that correlation does not indicate causation or attack mechanisms. Multivariate analysis includes selected communication features and target categories to look for patterns that could be of interest for cybersecurity monitoring. The analysis is directed towards understanding the provided observations, not towards the predictive models or accuracy of the classifications. Potential threat indicators are thus viewed as exploratory evidence which must be investigated to decide whether to incorporate them into an operational detection system.

Visualization Strategy

The visualization approach involves four complementary graphical approaches to the research goals. A bar chart shows the distribution of legitimate traffic and the five types of malicious traffic. Second, boxplots or violin plots are used to compare selected characteristics of TCP and MQTT across the target classes to identify differences in distributions and variabilities. Third, a correlation heatmap investigates relationships between appropriate numeric features, omitting constant features which can't generate valuable correlation estimates. Finally, a multivariate visualization combines selected communication features and attack types to explore combinations of features that could result in different attack patterns. Appropriate axis labels, legends, figure titles and readable scales are used to aid interpretation. Since the supplied data set has an equal number of observations in all six classes, the class distribution chart should be regarded as being derived from a processed data set, and not as proof of attacks occurring at the same rate all the time in the original network.

Ethical Considerations

The study is based on an existing dataset for research and education, instead of gathering new information from people or active attacks on IoT systems. The analysis targets network attributes and labeled traffic categories, and aims are to gain insight into cybersecurity patterns and to enable defensive applications. The documentation of the datasets and the source of the datasets should be cited correctly in the finished paper. Analysis should maintain the integrity of the observations provided, report preprocessing decisions clearly and transparently, and be careful not to present results selectively that may be misleading. If there are any restrictions in representativeness and/or labelling in the dataset, these must be reported. As the analysis is exploratory, potential indicators should not be provided as a validated detection rule or evidence of a specific system's security or vulnerability. If sensitive network information is discovered during further examination of the data, that information should not be included in the research.

Limitations

First, this data is specific to this particular MQTT-based IoT environment, and the results don't necessarily apply to every industry or business network, communication protocol, or configuration. Second, records are balanced or resampled and may not

accurately represent the actual prevalence of attacks in the real world in the records within the six target categories. Thirdly, standardized features do not allow a direct interpretation in the original measurement units unless the transformation procedure is described. Fourth, the variable `mqtt_dupflag` does not offer anything useful for correlation analysis, because its value will always be zero. Exploratory associations also do not conclude causality or show that a feature can reliably predict a future attack. The study is not a training-and-validation study for a machine learning model and therefore cannot claim predictive accuracy, detection rate, or operational effectiveness. To evaluate generalizability and practical performance, further research is needed using different datasets, different preprocessing methods, and independent model validation.

Conceptual Framework

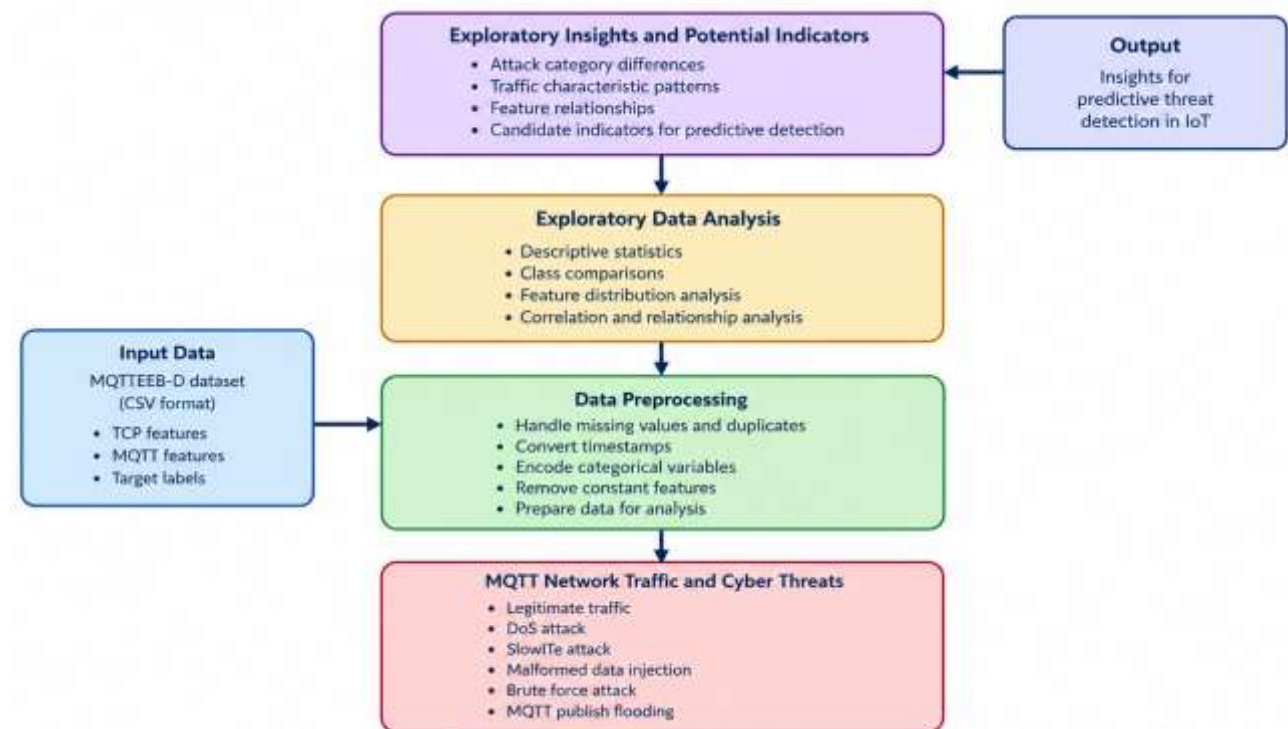


Figure 1. Conceptual Framework Diagram for this Research Paper
(Source: Created by learner)

RESULTS AND FINDINGS

Distribution of Normal and Malicious MQTT Traffic

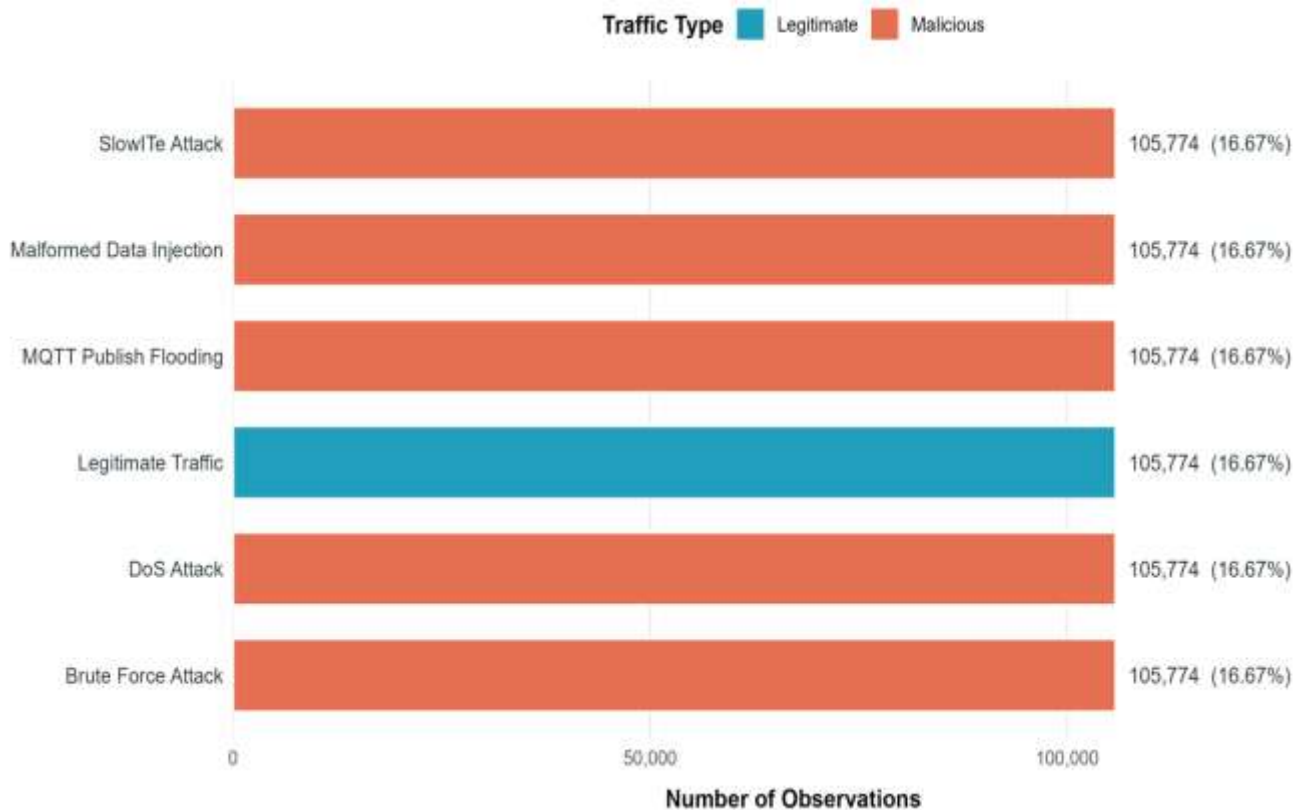


Figure 2. Frequency and Percentage of Observations by Traffic Category
(Source: Created by learner)

All six traffic categories have the same number of observations: each of them has 105774 observations, or 16.67 percent of the total number of observations. This is because there is a balance between legitimate traffic and the five categories of cyberattacks, with none of them dominating the analysis. The visualization suggests that the data has been normalized so that each category is represented evenly. This balanced structure provides a significant level of class balance to minimize class imbalance effects and facilitate the uniformity of assessing traffic patterns across attack types.

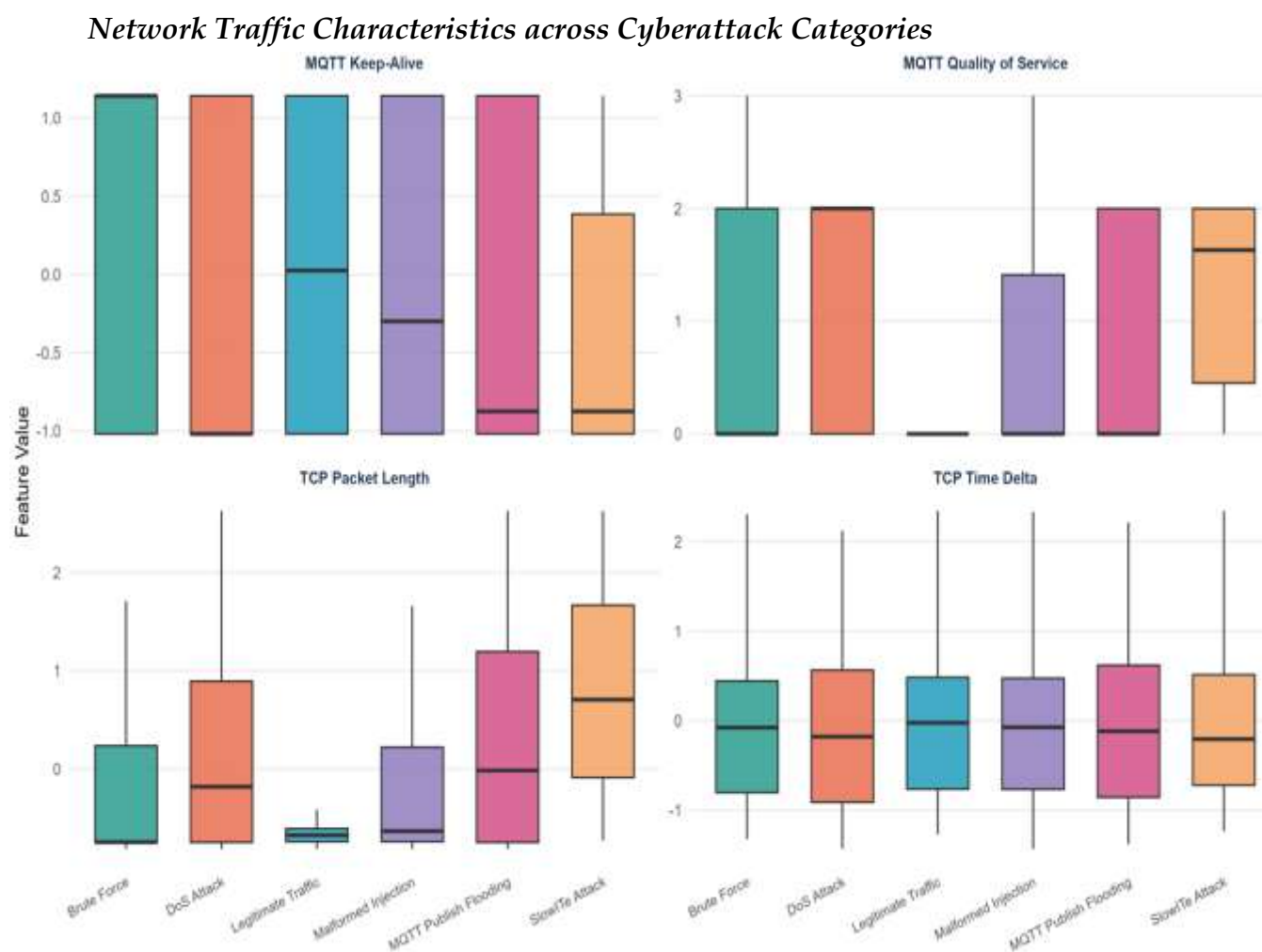


Figure 3. Comparison of TCP and MQTT Feature Distributions
(Source: Created by learner)

There are clear differences in the features of TCP and MQTT communication in the various traffic types. The packet lengths of the TCP packets are higher in the case of slowITe attack records compared to the legitimate traffic. The differences in the MQTT Quality of Service levels for each class are quite significant, with some attacks showing larger median values than the normal communication. The time delta distributions of TCP seem to be somewhat alike among the categories. These differences indicate that the packet-related and protocol-related properties could be good features to differentiate between malicious and legitimate MQTT network traffic.

Relationships between TCP and MQTT Communication Features

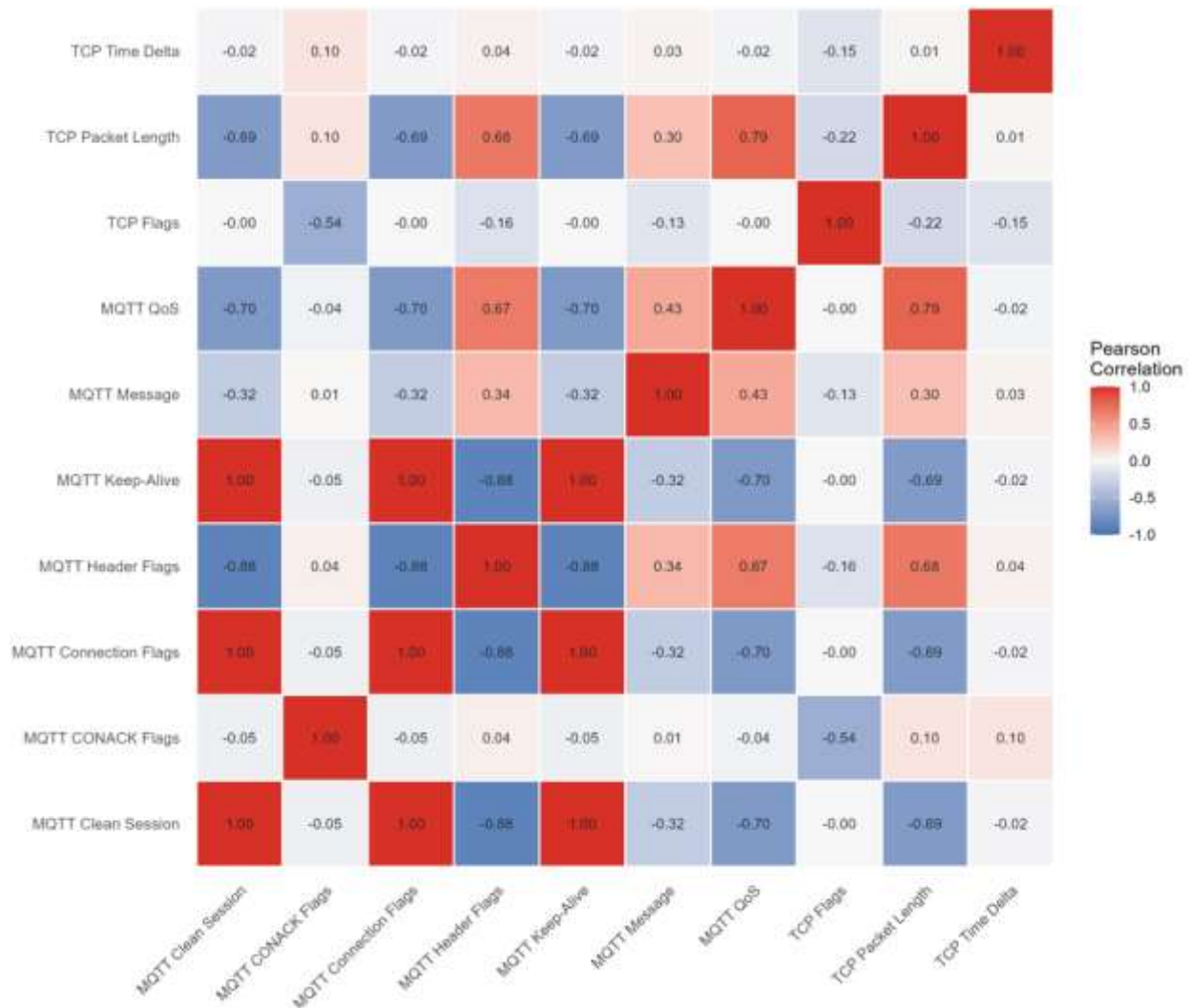


Figure 4. Pearson Correlation Coefficients for Numerical Variables
(Source: Created by learner)

There are several good relationships between MQTT and TCP variables. The positive associations between the variables MQTT Clean Session, MQTT Connection Flags, and MQTT Keep Alive are perfect, which means there is very similar behaviour in the dataset. The variables have strong negative correlations with the MQTT Header Flags and positive correlations with the other variables, TCP Packet Length and MQTT Quality of Service. The correlation between TCP Packet Length and MQTT Quality of Service is high, indicating that there is a strong linkage between the quality of service of the message and the characteristics of the packets. For TCP Time Delta, in comparison, there are very weak correlations with most variables, suggesting that there is not much linear association with communication features.

Integrated MQTT Cyber Threat Patterns and Potential Predictive Indicators

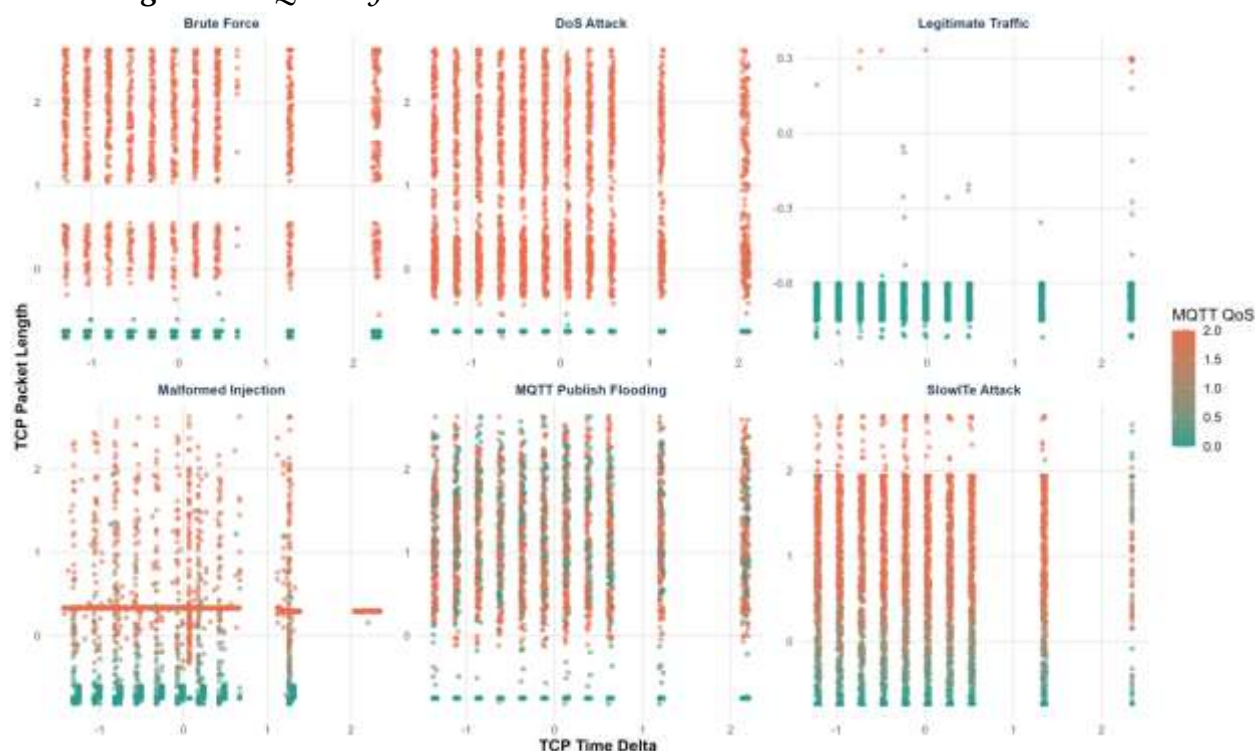


Figure 5. TCP Timing and Packet Length Compared Across Traffic Classes, with colour representing MQTT Quality of Service

(Source: Created by learner)

TCP time delta and TCP packet length, with the MQTT Quality of Service, show the communication behaviour of the traffic classes. Valid traffic is relatively compact and has relatively small packet length values and relatively small variation in Quality-of-Service levels. The larger distributions and more widely spread observations in contrast attacks are more noticeable in the DoS MQTT Publish Flooding and SlowITe attacks. Some of the classes overlap, meaning that no single feature is sufficient to differentiate threats. However, certain packet length or service quality characteristics combinations may be useful to future threat detection research.

Analysis of Results/Findings

The results offer important information about the nature of MQTT network traffic and directly answer the research question related to identifying traffic patterns that could help predict threats in industrial and business IoT environments. The distribution analysis revealed that the number of observations for each traffic category was equal; thus, the legitimate communication and malicious activities were well represented. This balance allowed meaningful comparison of the traffic classes without bias due to dominance of classes. Five different categories of attack can be identified in parallel with the legitimate traffic, which illustrates the variety of communication behaviours within the data set. Therefore, the data set is suitable for investigating the phenomenon of cyber threats from a visible point of view related to the network characteristics, and is suitable

for systematic analysis of the difference between normal and malicious MQTT communication.

Comparisons of network traffic characteristics showed that there are some differences between the various network traffic classes in regards to certain TCP and MQTT features. The length of the TCP packets was found to be very different; for example, attack categories like SlowITe and MQTT Publish Flooding generally resulted in wider distributions than legitimate traffic. Additionally, there were variations in the Quality of Service (QoS) for the messaging, indicating varying communication needs and message delivery patterns in the various attack scenarios. By comparison, TCP time delta had more overlap between categories, suggesting that the timing information alone doesn't seem to be enough to differentiate attacks. The results indicate that the way malicious activities affect network communication patterns is different and that the use of a combination of protocol-level and packet-level features can be more informative than all of them individually. Thus, the findings provide an answer to the research question, showing that attack categories have discernible communication properties that could be used to investigate as cybersecurity issues.

The correlation analysis and the integrated multivariate visualization further confirmed the understanding of the relationships between the communication features and the attack behaviour. The MQTT Keep Alive and MQTT Connection Flags are strongly positively correlated with MQTT Clean Session, suggesting that the same characteristics are present across the dataset, while MQTT Header Flags is strongly negatively correlated with MQTT Clean Session, which suggests contrasting communication behaviour. Furthermore, it is noted that there is a close connection between the MQTT Quality of Service and the TCP Packet Length, which implies that application-level decisions of the messages affect the behaviour of the packet. The integrated visualization showed that there is an overlap between the attack categories as well as an overlap between the legitimate categories and the attack categories, but the overlap is not as large as it might have been. This observation suggests that multiple network attributes are likely to need to be considered, rather than relying on a single attribute to identify a threat. The results show that there are meaningful patterns in MQTT traffic and offer some exploratory evidence that could inform the efforts to develop future AI-based predictive cybersecurity solutions for IoT environments.

Future Work

The exploratory findings can be developed and evaluated into predictive cybersecurity models with the identified MQTT and TCP communication features, which can be used in future research. Machine learning algorithms like Random Forest, Support Vector Machine, Extreme Gradient Boosting and a Neural Networks can be used to see if the detected traffic patterns can be used to accurately classify legitimate and malicious traffic. Further experiments could be used to compare the predictive power of individual features to that of combinations of features to determine which features are most informative. Efficient feature selection and dimensionality reduction techniques can be

investigated to retain the detection effectiveness and increase the efficiency in resource-limited IoT environments.

Extending the results by testing with various IoT datasets from different industrial and business contexts should be explored in further studies. This dataset is specific to an MQTT environment and needs to be validated in different network architectures, communication protocols, and different conditions. Additionally, future research could explore real-time threat detection systems at the edge of IoT networks where low-latency and efficient resource utilization are crucial. Finally, explainable artificial intelligence techniques can be used to enhance transparency and aid security analysts in comprehending why certain traffic patterns are flagged as suspicious. These advancements would help advance more practical, reliable, and scalable cybersecurity solutions for connected industrial systems.

CONCLUSION

Fundamental Finding: This paper examines the nature of IoT network traffic using MQTT and tries to establish patterns for both normal and different kinds of cyberattacks. The research showed that measurable differences exist between Traffic Classes using descriptive analysis, feature distribution comparison, correlation assessment, and multivariate visualization. Individual TCP and MQTT features exhibited different trends between different attack types, and correlations between protocol variables revealed key communication behaviors in the dataset. **Implication:** The findings suggest combinations of network characteristics could be useful indicators to classify malicious activities from normal operations. The outcomes help to create a better understanding of how MQTT communicates and provide useful information to design intelligent threat detection strategies in industrial and business IoT scenarios. **Limitation:** While the study doesn't create or test a predictive model, it certainly sets the groundwork for future cybersecurity research using AI. **Future Research:** While the study doesn't create or test a predictive model, it certainly sets the groundwork for future cybersecurity research using AI.

REFERENCES

- [1] E. Dritsas and M. Trigka, "A Survey on Cybersecurity in IoT," *Future Internet*, vol. 17, no. 1, Art. no. 30, Jan. 2025, doi: 10.3390/fi17010030.
- [2] F. C. Ogenyi, C. N. Ugwu, and O. P. C. Ugwu, "Securing the future: AI-driven cybersecurity in the age of autonomous IoT," *Frontiers in the Internet of Things*, vol. 4, 2025, doi: 10.3389/friot.2025.1658273.
- [3] C. Yan et al., "Cybersecurity vulnerabilities in IoT devices," *Nature Reviews Electrical Engineering*, vol. 3, pp. 381–395, 2026, doi: 10.1038/s44287-026-00296-5.
- [4] T. Zhukabayeva, L. Zholshiyeva, N. Karabayev, S. Khan, and N. Alnazzawi, "Cybersecurity solutions for industrial internet of things–edge computing integration: Challenges, threats, and future directions," *Sensors*, vol. 25, no. 1, Art. no. 213, Jan. 2025, doi: 10.3390/s25010213.

- [5] K. Tsiknas, D. Taketzis, K. Demertzis, and C. Skianis, "Cyber threats to industrial IoT: A survey on attacks and countermeasures," *IoT*, vol. 2, no. 1, pp. 163–186, Mar. 2021, doi: 10.3390/iot2010009.
- [6] R. J. Raimundo and A. T. Rosário, "Cybersecurity in the internet of things in industrial management," *Applied Sciences*, vol. 12, no. 3, Art. no. 1598, Feb. 2022, doi: 10.3390/app12031598.
- [7] A. M. Alnajim, S. Habib, M. Islam, S. M. Thwin, and F. Alotaibi, "A comprehensive survey of cybersecurity threats, attacks, and effective countermeasures in industrial internet of things," *Technologies*, vol. 11, no. 6, Art. no. 161, Nov. 2023, doi: 10.3390/technologies11060161.
- [8] S. Lakshminarayana, A. Praseed, and P. S. Thilagam, "Securing the IoT application layer from an MQTT protocol perspective: Challenges and research prospects," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 4, pp. 2510–2546, 2024, doi: 10.1109/COMST.2024.3372630.
- [9] S. U. A. Laghari, W. Li, S. Manickam, P. Nanda, A. K. Al-Ani, and S. Karuppayah, "Securing MQTT ecosystem: Exploring vulnerabilities, mitigations, and future trajectories," *IEEE Access*, vol. 12, pp. 139273–139289, 2024, doi: 10.1109/ACCESS.2024.3412030.
- [10] A. J. Hintaw, S. Manickam, M. F. Aboalmaaly, and S. Karuppayah, "MQTT vulnerabilities, attack vectors and solutions in the Internet of Things (IoT)," *IETE Journal of Research*, vol. 69, no. 6, pp. 3368–3397, 2023, doi: 10.1080/03772063.2021.1912651.
- [11] M. Kolisnyk, "Vulnerability analysis and method of selection of communication protocols for information transfer in Internet of Things systems," *Radioelectronic and Computer Systems*, no. 1, pp. 133–149, 2021, doi: 10.32620/reks.2021.1.12.
- [12] K. Dhanushkodi and S. Thejas, "AI-enabled threat detection: Leveraging artificial intelligence for advanced security and cyber threat mitigation," *IEEE Access*, vol. 12, pp. 173127–173136, 2024.
- [13] R. Abreu, E. Simão, C. Serôdio, F. Branco, and A. Valente, "Enhancing IoT security in vehicles: A comprehensive review of AI-driven solutions for cyber-threat detection," *AI*, vol. 5, no. 4, pp. 2279–2299, 2024.
- [14] S. Thayalan, N. Radhakrishnan, T. V. Ramana, G. G. Devarajan, M. Karuppiah, and M. M. Al Dabel, "Real-time threat detection and AI-driven predictive security for consumer applications," *IEEE Transactions on Consumer Electronics*, vol. 71, no. 2, pp. 5100–5107, 2025.
- [15] G. Sunkara, "AI-driven cybersecurity: Advancing intelligent threat detection and adaptive network security in the era of sophisticated cyber attacks," *Well Testing Journal*, vol. 31, no. 1, pp. 185–198, 2022.
- [16] A. Aqachtoul, K. Karam, A. Elamrani, M. Najib, N. Rafalia, and M. Bakhouya, *MQTTEEB D A Real World IoT Cybersecurity Dataset for AI Powered Threat Detection in MQTT Networks*, Version 2. Mendeley Data, 2026. doi: 10.17632/jfttfn6tr.2.

Rabia Akter Sukhi

Doctor of Information Technology

Trine University, USA

Email: kazirabia01@gmail.com

Mahbub Hasan

Master of science in information studies Trine University, USA

Email: mahbub0016@gmail.com

Fahmida Akter

Master of Science in Information Technology Washington university of science and technology

Email: a.fahmida42@gmail.com

Naim Rahman Ifti

Bachelor of Science in Computer Science San Francisco Bay University

Email: naimrahmanifti@gmail.com
