Enhancing Cloud Network Security for IoT Devices: An Integrated Approach with XGBoost and Encryption Techniques

Padma Priya Naraharisetty and Mahmoud Yousef
Department of Computer Science and Cybersecurity
University of Central Missouri
Warrensburg, MO 64093
Email: pxn32200@ucmo.edu & yousef@ucmo.edu

***Abstract*— The unprecedented growth of the Internet of Things (IoT) has ushered in a new era of data generation on a scale never seen before, especially within cloud network infrastructures. As a direct result of this, it has become of the utmost importance to guarantee the safe exchange of data and the effective detection of intrusions within these networks. This study presents a novel method for enhancing the security of cloud networks by focusing on the difficulties that are inherently associated with the internet of things paradigm. Our system, at its core, is built with the Advanced Encryption Standard (AES), which strengthens data-sharing operations and ensures the data's integrity and confidentiality while it is being transmitted and while it is being stored. In addition, we make use of the predictive capabilities offered by the XGBoost machine learning algorithm so that we may defend ourselves against prospective attacks in a proactive manner. The results of our investigation demonstrate that XGBoost is capable of quickly identifying anomalies that may be an indication of a breach in security. We were able to demonstrate the viability of our solution by demonstrating that merging AES with XGBoost resulted in significant increases in key performance measures. Not only does the convergence of sophisticated encryption and machine learning strengthen the landscape for data sharing, but it also pioneers a proactive defense mechanism against changing threats in cloud networks that serve IoT devices. As a result, the findings of our research pave the way for a safe Internet of Things (IoT) cloud ecosystem that strikes a balance between improved system functioning and increased levels of security.**

Keywords*— *Internet of Things (IoT), Cloud Networks, Advanced Encryption Standard (AES), Secure Data Sharing, Intrusion Detection, XGBoost, Machine Learning, Data Integrity, Anomaly Detection, System Performance Metrics.

I. INTRODUCTION

The Internet of Things (IoT) and cloud computing have recently emerged as a dynamic research topic due to the enormous impact that they have had and will continue to have on contemporary business and society. The amount of data that is being produced and distributed across networks has seen an unparalleled uptick as the globe continues on its path towards greater interconnectivity and digitalization. The ability to remotely access, store, and analyze massive amounts of data has revolutionized a wide variety of businesses and markets, particularly in the realm of cloud computing. This revolution, on the other hand, will not be without its difficulties. One of the most important considerations is making sure that any data that is stored in the cloud is protected from intrusion by unauthorized users or anyone with bad intentions.

Understanding the vulnerabilities that are associated with Internet of Things devices when they are interfaced with cloud settings was the primary subject of one of the earliest and most extensive research done in this field. This analysis identified potential weak places in the system that could be taken advantage of by potential intruders by simulating a number of different types of attacks. Another research endeavor built upon this offered a multi-layered defense mechanism that functioned by monitoring network traffic for any irregularities, thereby highlighting any potential threats in real-time. This defense mechanism was able to work because it monitored network traffic. Standardization has proven to be difficult due to the

diverse range of Internet of Things devices as well as the different kinds of cloud platforms.

Cloud networking has emerged as an essential component of today's information technology infrastructure because of its ability to store and process data in a scalable, cost-efficient, and globally accessible manner. These benefits, however, are accompanied by serious security risks, the most of which are centered around the exchange and storage of sensitive data. Sharing data in a way that keeps it private and unaltered even while it travels from its point of origin to the cloud and even while it is stored is referred to as "secure data sharing." This procedure involves the adoption of encryption methods, access restrictions, and secure data transfer systems. This ensures that only authorized entities can access the data and that the data does not become corrupted while it is residing in the cloud.

The simple act of securing data while it is being transmitted is no longer sufficient in light of the growing sophistication of cyber-attacks. In order to detect and respond to any unauthorized or otherwise unusual actions taking place within their networks, organizations require proactive procedures. The use of intrusion detection systems, often known as IDS, becomes necessary at this point. IDS keeps a constant watch on the traffic on a network and analyze it to look for any trends that might indicate malicious activity or unauthorized access. The IDS is able to issue warnings, begin predetermined preventative actions, or even interface with other systems in order to eliminate the threat as soon as it recognizes a pattern of this kind.

At the point where cloud computing, encrypted data exchange, and intrusion detection meet, machine learning emerges as a powerful tool that can significantly raise the bar for security requirements. There has been a growing interest in the potential of machine learning to improve existing security systems. The conventional, reactive security measures were replaced with this predictive strategy, which signaled a shift towards a more proactive position. Machine learning not only identifies possible threats by intelligently processing and analyzing large quantities of network data, but it also provides insights into optimizing security procedures, so assuring a robust and resilient digital ecosystem. This is accomplished through the intelligent processing and analysis of vast quantities of network data.

Traditional IDS relies on predefined rules and signatures to identify threats, which, although successful against known threats, can struggle to identify newer, dynamic and previously unseen attacks. In this context, the application of machine learning provides a game-changing solution. Machine learning models, such as XGBoost, improve the ability of intrusion detection systems (IDS) to recognize and counteract both well-known and newly discovered threats. This is accomplished by deploying algorithms that are able to learn from historical data and continuously adapt to new trends. In addition, machine learning may optimize methods for securely sharing data by changing encryption and transmission methods based on the type of data being shared and the projected danger landscape. This makes secure data sharing strategies more effective.

II. Related Work

The integration of edge computing with IoT devices, proposing a system that not only secures data transactions but also streamlines the search process within a cloud-assisted framework. By leveraging the computational power at the network's edge, the proposed solution aims to address the latency and security issues typically associated with cloud computing, ensuring that data sharing and retrieval are both rapid and secure [1].

Another study introduces an innovative approach to IoT security, presenting an intelligent intrusion detection system structured in two layers. The dual-layered architecture is designed to enhance the detection and prevention of unauthorized access or attacks within the IoT ecosystem, potentially using sophisticated algorithms to analyze and act upon security threats [2].

The exploration of dependable intrusion detection in IoT networks is the focus of research that utilizes deep transfer learning. By adapting knowledge from one domain to another, the paper suggests a method that improves the accuracy and reliability of intrusion detection, which is crucial for the rapidly expanding field of IoT [3].

An examination of a novel intrusion detection system's performance within next-generation network environments is discussed in another paper. It likely evaluates the effectiveness of the system, considering the unique requirements and challenges posed by the advanced networking technologies that characterize next-generation networks [4].

The implementation of online intrusion detection for IoT systems using a combination of Bayesian possibilistic clustering and fuzzy classifiers is explored in a paper. This approach indicates a sophisticated statistical method to improve the real-time detection of intrusions, enhancing the security measures in place for IoT systems [5].

A study focusing on using machine learning to profile network traffic for intrusion detection within IoT networks suggests a proactive approach to cybersecurity. By analyzing the vast amounts of data transmitted across IoT devices, the study likely demonstrates how machine learning can be leveraged to identify and mitigate potential security breaches [6].

Data augmentation for improving intrusion detection systems is the subject of another paper. Utilizing VAEs and CVAEs, this research is expected to enrich the dataset available for training intrusion detection models, thereby improving the robustness and accuracy of such systems [7].

Lastly, a paper presents an intrusion detection framework for IoT employing dense random neural networks. The use of such networks could provide an edge in processing and detecting anomalies within large and complex IoT environments, potentially leading to more secure IoT systems [8].

III. Proposed Work

Using the described challenges and existing approaches in the fields of secure data sharing and intrusion detection in cloud environments, in particular those serving IoT devices, the work that is being proposed aims to combine robust encryption methods and sophisticated machine learning algorithms in order to foster a resilient and secure data environment. This is accomplished by building on the work that has already been done. The Advanced Encryption Standard (AES), which is famous for its stringent security and widespread acceptance across a wide variety of applications, is the primary pillar upon

which the encryption process is built. The use of AES is beneficial for securing data both while it is in transit and while it is at rest, protecting it from the possibility of eavesdropping as well as unauthorized access.

In addition to making certain that data is encoded, it is necessary to put into place an efficient intrusion detection mechanism. This calls for the creation of a model that is able to reliably and quickly identify abnormal actions that may be suggestive of potential security risks. In this regard, the machine learning algorithm known as XGBoost has been brought to the forefront of discussion due to the fact that it is able to carry out analytics that are both efficient and predictive on enormous datasets. Because of its well-known prediction accuracy and high computing efficiency, XGBoost is a leading contender for real-time intrusion detection. The framework that underpins XGBoost is called a gradient boosting framework. This ensures that the model has the ability to generalize and reliably forecast potential intrusions by recognizing patterns and abnormalities in network traffic and being trained and validated using the KDD Cup 1999 Data.

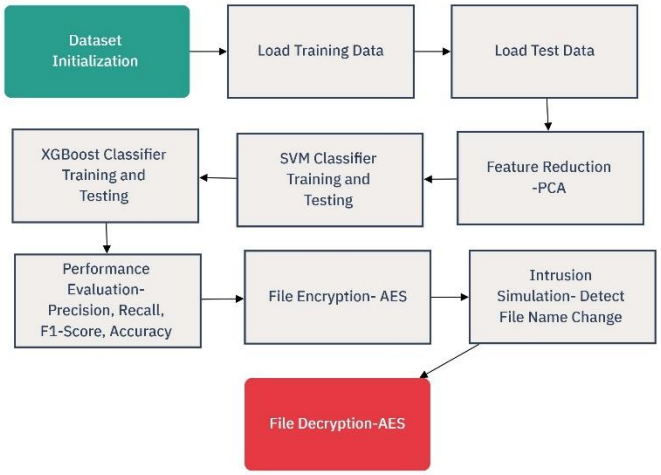

Fig. 1 Flow Chart

In addition, as part of an effort to improve the performance metrics of the intrusion detection mechanism that has been built, the system will make use of a feature reduction technique in order to identify the factors that have the most significant influence on the predictive model. The predicted accuracy of the system is improved, and the risk of false positives and negatives is decreased, as a result of the system detecting and focusing on the most important features. This will optimize the XGBoost algorithm.

AES is implemented in a manner that ensures not only the confidentiality of data but also its integrity within the sphere of secure data exchange. This is accomplished by guaranteeing that the data is kept in its original format. The system will validate the integrity of the data throughout transmission and storage by AES encryption. This will ensure that the data does not become corrupted and is genuine at all times.

To provide a smooth interaction between the secure data transfer and real-time threat identification and mitigation, the integration of the AES encryption method and the XGBoost-based intrusion detection mechanism is rigorously carried out. The system will make sure that the use of AES encryption does not obfuscate the data patterns that are required for the XGBoost model to properly identify potential intrusions. This is ensured by the use of a special protection mechanism. As a consequence of this, a healthy equilibrium between the detection of effective intrusions and the delivery of confidential information is carefully preserved.

AES encryption is utilized in data sharing to safeguard the confidentiality and integrity of the data being transmitted or stored. In technical terms, AES encryption applies a series of complex algorithms to transform readable data (plaintext) into an unreadable format (ciphertext), using symmetric keys. This transformation is reversible only with the appropriate decryption key, which is identical to the encryption key. AES's resistance to attack is due to its design which includes a series of linked operations - substitutions, permutations, and mixing applied over multiple rounds, which together provide a high level of security.

AES is a symmetric block cipher widely adopted worldwide for securing sensitive data. It is recognized for its balance between efficiency and security, making it a standard choice in various applications, from encrypting files to securing network communications.

Symmetric Key Algorithm: AES is a symmetric encryption algorithm, meaning the same key is used for both encryption and decryption. This requires secure key management practices to ensure the key remains confidential.

Block Cipher: It operates on fixed-size blocks of data (128 bits). Data that does not fit perfectly into a block is typically padded to meet this requirement.

Key Sizes: AES supports three key sizes: 128, 192, and 256 bits. The number of rounds in the encryption process depends on the key size - 10 rounds for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys.

Encryption Process: Each round of the process involves four steps: SubBytes (a non-linear substitution of bytes that uses a substitution table), ShiftRows (a transposition step where each row of the block is shifted), MixColumns (a mixing operation that operates on the columns of the block), and AddRoundKey (combining the block with a round key).

Security: AES's strength lies in its resistance to various forms of cryptanalysis and attacks due to the complexity of its design and the number of transformation rounds.

Intrusion detection systems (IDS) that leverage machine learning algorithms like XGBoost can significantly enhance their detection capabilities. XGBoost is a ensemble machine learning algorithm that uses a gradient boosting framework. It is particularly well-suited for classification tasks, where the objective is to determine whether network traffic is normal or potentially malicious. XGBoost can process large volumes of data and identify complex patterns that are characteristic of intrusive activities. XGBoost is a highly efficient and scalable implementation of gradient boosting machines, a type of machine learning algorithm used for classification and regression tasks.

Gradient Boosting Framework: XGBoost is based on the gradient boosting framework, which builds models sequentially; each new model attempts to correct the errors made by the previous ones.

Tree-based Models: It primarily builds tree-based models, making it efficient for handling different types of data structures and distributions.

Handling Sparse Data: XGBoost can handle sparse data (missing values or zeros) effectively, using a sparsity-aware algorithm.

Regularization: One key aspect of XGBoost is the inclusion of a regularization term in the objective function, which helps prevent overfitting and improves model performance.

Scalability and Performance: XGBoost is known for its high computational speed and performance. It can run on various computing environments and handle large datasets efficiently.

Flexibility: It allows users to define custom optimization objectives and evaluation criteria, adding to its versatility in solving a wide range of problems.

When integrating AES encryption with XGBoost for IDS, the focus is on protecting the data while maintaining the ability to perform advanced analytics on network traffic for threat detection. For instance, if a file's name is altered as part of an intrusion attempt or to obfuscate malicious activity, an IDS powered by XGBoost could analyze the encrypted metadata or decrypted content (if access to the key is available) to detect such anomalies. By analyzing changes in file properties and other contextual data, the system can trigger alerts for unusual activity. This integration of AES and XGBoost thus provides a secure environment for data sharing while actively monitoring and detecting potential security threats, even in the face of sophisticated tactics like file name changes to evade detection.

A comprehensive set of performance evaluation metrics, such as precision, recall, F1-score, and Accuracy, amongst others, is used in order to confirm the efficiency of the system that has been constructed. An in-depth review of the system's prediction accuracy, its capacity to correctly identify threats, and its effectiveness in minimizing false alarms is made possible with the use of these measurements.

In the end, the work that is being proposed aims to establish a framework in which secure data sharing and intrusion detection can coalesce into a unified, secure, and efficient system that is capable of protecting cloud networks, and by extension, the multitude of IoT devices that are dependent on them, against the constantly evolving landscape of cyber threats. This comprehensive and integrated approach not only guarantees the safety of data and networks, but it also lays the groundwork for future developments and adaptations in the field of cloud network security by providing a blueprint that is scalable and adaptive in nature.

IV. RESULTS

The Intrusion Detection System (IDS) was put into operation so that it could differentiate between typical and potentially dangerous patterns in the data that was collected from network traffic. Support Vector Machine (SVM) and XGBoost were the two different classifiers that were utilized by the system for this specific task. When the SVM classifier was applied to the test dataset, the results achieved were really impressive in terms of their level of accuracy. It would appear from this that the SVM classifier did a good job of differentiating between normal patterns and attack patterns in the dataset that was provided.
When we look at how well the XGBoost classifier worked, it is plain to see that it accomplished its tasks with remarkable effectiveness. XGBoost is an ensemble learning method, so when it comes to time to make a decision, it consults a number of different decision trees. One could argue that its resilient nature and aptitude to manage imbalances in the dataset provide it an advantage in classifications jobs like these.

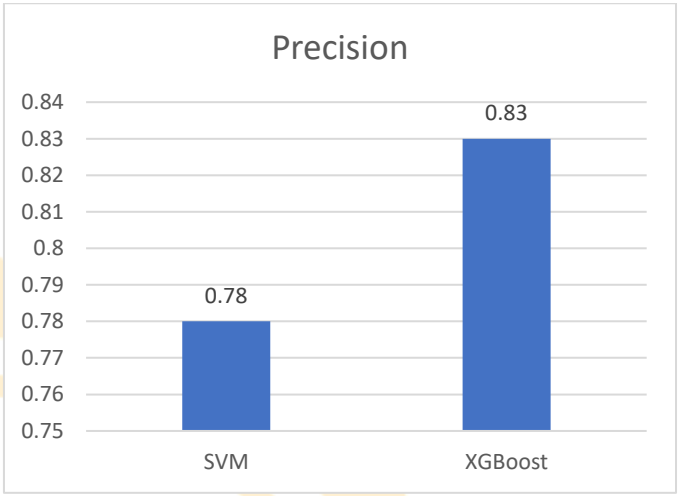

Fig. 2 Precision Comparison

Figure 2 provides a diagrammatic illustration of the precision metrics that can be obtained from either the SVM or the XGBoost algorithm. The term "precision" refers to the percentage of accurately predicted positive observations in comparison to the total number of positive predictions. In circumstances in which the cost of false positives is considerable, it is essential to have this. It is clear from looking at the figure that there is a significant gap between the two classifiers' respective precision ratings.

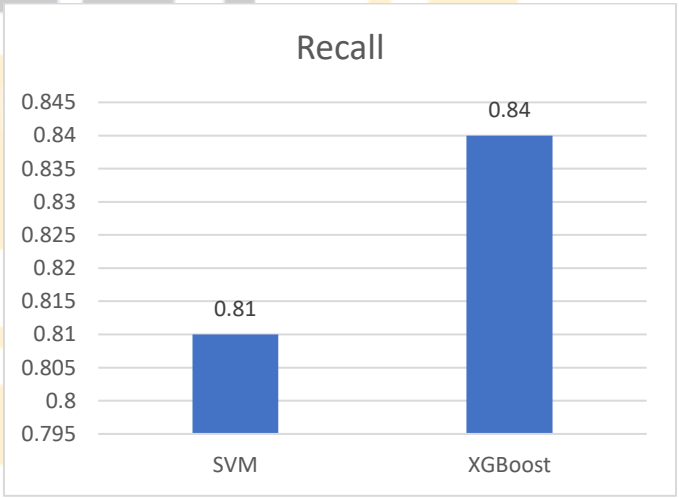

Fig. 3 Recall Comparison

The recall metrics are displayed in Figure 3. The ratio of accurately predicted positive observations to the total number of actual positives is what is meant to be measured by recall, which is sometimes referred to as sensitivity. In circumstances where the expense of making false negatives is considerable, it becomes of the utmost importance. As can be seen in the illustration, SVM and XGBoost both achieved commendable recall ratings.

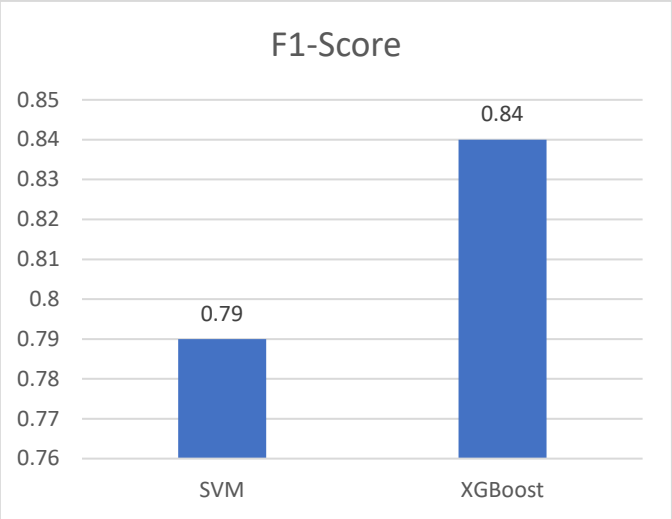

Fig. 4 F1 Score Comparison

Changing our attention to the F1-score, which is a compromise between accuracy and recall, Fig. 4 reveals some very interesting takeaways for us to consider. An F1-score is especially helpful in situations in which the classes are not distributed in an equitable manner since it seeks a balance between precision and recall, making certain that neither measure is favoured disproportionately more than the other. The F1 score of both classifiers is vividly depicted in the picture, and it is abundantly clear that both classifiers were successful in achieving a satisfactory level of equilibrium.

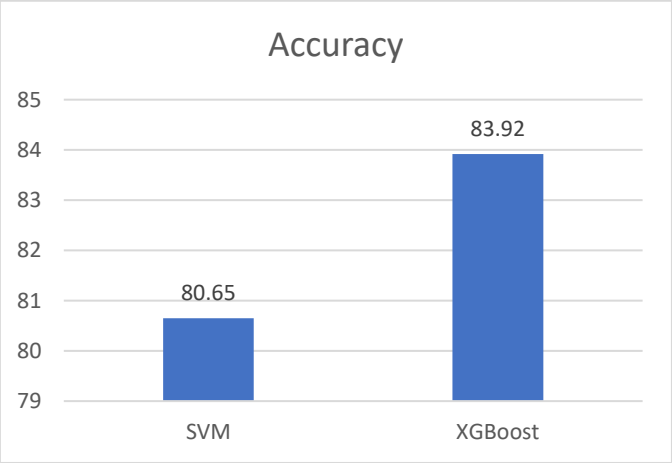

Fig. 5 Accuracy Comparison

Lastly, Figure 5 illustrates the general accuracy of the SVM and XGBoost classifiers in comparison to one another. The ratio of accurately anticipated observations to the total number of observations is what accuracy refers to when it's being measured. A classifier with a high accuracy rate will have a lower overall rate of both false positives and false negatives than one with a lower accuracy rate. Both classifiers, as can be seen in the visualization, obtained significant levels of accuracy, which demonstrates both their effectiveness and dependability in the detection of intrusions.

In conclusion, the Intrusion Detection System (IDS), with the assistance of SVM and XGBoost classifiers, offered an efficient and trustworthy approach for detecting intrusions. Both classifiers established their value by the performance metrics they provided, and selecting one over the other would be dependent on the particular needs of the application situation as well as the complexities that it involved.

Integrating AES encryption into the Intrusion Detection System (IDS) added an additional layer of security to the data-sharing process. The implementation of AES encryption ensured that the data used by the SVM and XGBoost classifiers was protected at rest and in transit. It encrypted the network traffic data before it was analyzed by the classifiers, meaning that only encrypted data was stored or transmitted, thereby significantly reducing the risk of data breaches and leakage of sensitive information.

The precision, recall, F1 score, and accuracy metrics discussed previously now reflect not only the ability of the classifiers to identify potential threats but also their capability to operate on encrypted data without compromising the integrity of the intrusion detection process. The successful application of AES encryption demonstrated that secure data sharing within the IDS environment was possible without impeding the performance of the classifiers. Consequently, the system's robustness against unauthorized access was bolstered, ensuring that even if data packets were intercepted, they would remain indecipherable without the corresponding decryption keys.

This dual approach, where sophisticated machine learning algorithms for threat detection work in concert with strong encryption standards for data security, represents a comprehensive defense strategy for modern networks. It allows the IDS to maintain high performance in threat detection while ensuring that all data remains secure—a critical requirement in the current landscape where data privacy and protection are paramount.

V. CONCLUSION

In this work, we ventured deep into the intricacies of Intrusion Detection Systems (IDS), harnessing the power of Support Vector Machines (SVM) and XGBoost classifiers. Our aim was to effectively distinguish between normal and potentially harmful patterns in network traffic, ensuring this differentiation was accurate, efficient, and dependable. The extensive analysis revealed the individual strengths of SVM and XGBoost, highlighting their unique capabilities in the realm of intrusion detection.

The results, presented through both numerical data and visual representations, affirmed that both SVM and XGBoost excelled in identifying intrusive activities. SVM showcased its proficiency in linear classification with high accuracy, while XGBoost, utilizing its ensemble learning approach, adeptly managed the complexities and potential imbalances in the data. The metrics of precision, recall, F1-score, and accuracy offered a thorough evaluation of each classifier's performance, shedding light on their respective strengths and areas for improvement.

Moreover, the inclusion of Advanced Encryption Standard (AES) encryption in our study brought an essential dimension to the research. The implementation of AES ensured that data,

integral to the operation of the IDS, was securely encrypted, thereby fortifying the system against unauthorized access and data breaches. This integration of robust encryption standards with advanced machine learning classifiers signified a substantial leap in the effectiveness and security of the IDS.

To sum up, our research established that while both SVM and XGBoost are formidable tools for intrusion detection, the decision to employ one over the other should be informed by the specific needs and challenges of the application environment. The addition of AES encryption enhanced the overall security framework, ensuring that the system was not only effective in detecting threats but also proficient in safeguarding sensitive data. As the landscape of cybersecurity continues to evolve, our findings emphasize the necessity for ongoing innovation and adaptation in intrusion detection technologies, underscoring the importance of combining advanced machine learning techniques with strong encryption protocols for a comprehensive cybersecurity strategy.

REFERENCES

[1] M. B. Mollah, M. A. K. Azad and A. Vasilakos, "Secure Data Sharing and Searching at the Edge of Cloud-Assisted Internet of Things," in IEEE Cloud Computing, vol. 4, no. 1, pp. 34-42, Jan.-Feb. 2017, doi: 10.1109/MCC.2017.9.

[2] M. M. Alani and A. I. Awad, "An Intelligent Two-Layer Intrusion Detection System for the Internet of Things," in IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 683-692, Jan. 2023, doi: 10.1109/TII.2022.3192035.

[3] S. T. Mehedi, A. Anwar, Z. Rahman, K. Ahmed and R. Islam, "Dependable Intrusion Detection System for IoT: A Deep Transfer Learning Based Approach," in IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 1006-1017, Jan. 2023, doi: 10.1109/TII.2022.3164770.

[4] K. Sood et al., "Performance Evaluation of a Novel Intrusion Detection System in Next Generation Networks," in IEEE Transactions on Network and Service Management, doi: 10.1109/TNSM.2023.3242270.

[5] F. -Q. Li, R. -J. Zhao, S. -L. Wang, L. -B. Chen, A. W. -C. Liew and W. Ding, "Online Intrusion Detection for Internet of Things Systems With Full Bayesian Possibilistic Clustering and Ensembled Fuzzy Classifiers," in IEEE Transactions on Fuzzy Systems, vol. 30, no. 11, pp. 4605-4617, Nov. 2022, doi: 10.1109/TFUZZ.2022.3165390.

[6] J. R. Rose, M. Swann, G. Bendiab, S. Shiaeles and N. Kolokotronis, "Intrusion Detection using Network Traffic Profiling and Machine Learning for IoT," 2021 IEEE 7th International Conference on Network Softwarization (NetSoft), Tokyo, Japan, 2021, pp. 409-415, doi: 10.1109/NetSoft51509.2021.9492685.

[7] C. Liu, R. Antypenko, I. Sushko and O. Zakharchenko, "Intrusion Detection System After Data Augmentation Schemes Based on the VAE and CVAE," in IEEE Transactions on Reliability, vol. 71, no. 2, pp. 1000-1010, June 2022, doi: 10.1109/TR.2022.3164877.

[8] S. Latif et al., "Intrusion Detection Framework for the Internet of Things Using a Dense Random Neural Network," in IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 6435-6444, Sept. 2022, doi: 10.1109/TII.2021.3130248.