TRUST-BASED SECURITY MODEL FOR DETECTING MISBEHAVING NODES AND ENHANCING ROUTING IN AN INTERNET OF THINGS ENVIRONMENT

*Aswin Vignesh Ramesh** [1], *Dr. E. J. Thomson Fredrick* [2]

ABSTRACT

Authentication in Internet of Things (IoT) is essential due to its unique features which includes sensing, connectivity, intelligence, scalability, self-configuration and an on open dynamic environment. The Internet of Things (IoT) devices must establish a mutual trust without a prior familiarity which might expose them to potential attacks. Therefore, a trust-based solution is crucial for Internet of Things security. This discussion is mainly focused on four different important issues namely Support Vector Machines (SVM), Dynamic Source Routing, Cryptographic Techniques and Bayesian Network Model. This survey first outlines the challenges in Support Vector Machines to detect the malicious nodes. The merits and demerits of support vector machines under this review are also considered. This paper brings a close outlook to detect malicious nodes using support vector machines. Dynamic Source Routing (DSR) is used for an efficient routing from source to destination to enhance trust among IoT nodes. The Cryptographic techniques are used to enhance the security in communication among IoT nodes. The Bayesian networks and random forest to detect the trustworthiness of IoT nodes with a hybrid approach.

I. INTRODUCTION

The Internet of Things (IoT) devices has transformed various industries by enabling connectivity and exchange between devices. The growing number of connected devices introduces notable security challenges. Detecting malicious nodes is essential in IoT networks to preserve network integrity, confidentiality and availability. Support Vector Machines (SVMs) are supervised learning algorithms frequently used for classification and regression problems. IoT networks are vulnerable to different forms of attacks, including Denial of Service (DoS) attacks, unauthorized access and data leaks. Detecting malicious nodes involves analysing the behaviour and traffic patterns of devices to identify anomalies that indicate potential threats.

Dept. of Computer Technology[1],
achuaswinvigneshr@gmail.com[1]
Dept. of Computer Technology[2]
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
thomson.ej@kahedu.edu.in
* Corresponding Author

II. SUPPORT VECTOR MACHINES

SVMs can be instrumental in this process due to its ability to distinguish between normal and malicious behaviour based on multiple features. Devices like sensors, actuators, mobile phones will detect, track and gather variety of information about an individual's social interactions.

Detecting malicious nodes in an Internet of Things (IoT) environment using Support Vector Machines (SVM) involves the approaches to feature extraction, model training and continuous monitoring.

The main goals to achieve this:

(1) Feature selection and extraction.

(2) Data Pre-processing.

(3) Model training and evaluation.

(4) Real-time Monitoring and Detection.

(5) Adaption and improvement.

(6) Scalability and Efficiency.

(7) Integration with Existing Security Measures.

Feature extraction involves creating new features from raw data, often using techniques like PCA which means Principal Component Analysis or using the domain knowledge to derive features. Effective feature selection and extraction help reduce the dimensionality of the dataset by making the model training much faster and much more efficient while improving the accuracy.

Data Cleaning involves handling missing values outlies and noise which can distort the model's understanding of the data. Normalization or scaling might be required to ensure all features are on same scale, especially if the features have different units. Data is split into training, validation and testing sets to evaluate the model's performance effectively.

The model is trained to use a labelled dataset where patterns, anomalies or malicious activities are known. Common algorithms include Supervised Learning Models or Unsupervised Models. Evaluation metrics such as accuracy, precision, recall, ROC-AUC, F1 Score are used to access the model's effectiveness.

Real-Time Monitoring and detection are to deploy the model for continuous monitoring and real-time detection of suspicious activities or anomalies. The system would monitor network traffic, user behaviour or system logs in real time. Real-Time detection requires a low-latency system that can process incoming data quickly and generate alerts when anomalies or malicious nodes are detected.

Adaption and Improvement is to ensure that the system remains effective over time and adapts to new threats or by changing the data patterns. In Dynamic environments like cyber security, threats evolve and new patterns may emerge requiring the model to adapt continuously.

The goal of scalability and efficiency is to design the system to handle large volumes of data without performance degradation. Efficient algorithms and distributed computing frameworks can help to process large datasets quickly.

The integration with existing security measures is to ensure the new system complements and enhance the existing security infrastructure. The model should be integrated with current security tools such as firewalls, Intrusion detection System (IDS) and security information and Event management.

Achieving these goals require a comprehensive approach involving data science, Cyber security and efficient software engineering practices.

SVMs simplify many machines learning tasks, with optimization being key part. The convex optimization problem is especially important on how the SVMs works.

The Advantages of SVM in malicious node detection are:

- Precision in marginal Cases: by maintaining the margin, SVM excels at distinguishing between closely situated data points in the future space.
- Scalable to Dynamic Environments: with proper kernel selection, SVM adapts well to diverse and dynamic network conditions.
- Effective for sparse data: Works well even when the data is sparse or the number of malicious nodes is small compared to benign nodes.

The challenges of SVM in malicious node detection are;

Scalability with large datasets where SVM can become computationally expensive when dealing with very large datasets, especially in dynamic IoT or MANET environments.

Kernel selection is by choosing the right kernel and tuning its parameters can be non-trivial, requiring domain expertise or trial-and-error.

Imbalanced Data:

SVM may struggle with highly imbalanced datasets. Techniques like Synthetic Minority Over sampling Technique – SMOTE can be used to address this issue.

Support Vector Machines (SVM) offer an effective solution for detecting malicious nodes in various network environments. Their ability to model complex decision boundaries makes them particularly well-suited for non linear problems, which are commonly in network anomaly detection.

However, to achieve optimal performance, careful feature selection, kernel tuning and dataset pre-processing are essential. Additionally, integrating SVM into a hybrid framework can further improve its scalability and robustness for real-world applications.

III. RELEVANT WORKS IN SVM IN DETECTING MALICIOUS NODES

In [1], Ravi Pariharet al, 2017 proposed that SVM is used to classify nodes into two categories: normal or malicious. It takes the neighbor trust value as input, which is computed based on control packets and data.

In [2], Jayashree Jha and Leena Ragha, 2013 stated that by applying SVM in Intrusion Detection Systems (IDS) has certain limitations. As a supervised learning method, SVM needs labelled data for an effective training. Classification also requires a prior knowledge, which may not always be accessible.

In [3], Theodoros Evgeniou Massimiliano Pontil and Tomaso A. Poggio, 1999 stated that the Regularization networks and SVMs(Support Vector Machines) are used to address specific learning challenges.

In [4], Yong Shiet al, 2011 stated that the SVMs have been extended to the domain of regression, clustering problems.

In [5], Babacar Gayeet al,2021 stated that based on the scale and characteristics various solution spaces are selected, and the dual problems solution is converted into the classification surface of the original space to boost the algorithms efficiency.

In [6], Simon Tong and Daphane Koller, 2001 proposed a new algorithm that implements pool-based active learning using Support Vector Machines (SVMs).

IV. RELEVANT WORKS IN DYNAMIC SOURCE ROUTING

Each data packets contains a list of intermediate nodes it must pass through to reach the destination, enabling flexible routing without relying on network-wide route discovery.

Dynamic Source Routing (DSR) includes mechanism to detect and avoid routing loops by maintain sequence numbers and route request records.

Dynamic Source Routing (DSR) is well suited for dynamic networks where nodes frequently move and establish temporary connections, such as military operations, disaster recovery and vehicular networks.

DSR provides flexibility, efficiency and reliability in communication which makes it adaptable to diverse dynamic and resource constrained environments. It also offers a lightweight and efficient routing solution forad-hoc mobile networks and wireless sensor networks.

Nodes can monitor the behaviour of neighbouring nodes by keeping track of the number of packets forwarded verses the number received and the changes in the route information provided by the intermediate nodes and any unusual delays in forwarding packets.

Nodes can maintain a trust score or reputation system based on the observed behaviour of their neighbours. Initially all nodes are assigned a neutral trust value and the trust score increases if the nodes behave correctly and decreases the trust value if suspicious activities are detected.

A low trust score can indicate a potentially malicious node, and other nodes may avoid using it in future routing. Nodes in ad-hoc networks often have limited battery and processing power, so detection mechanism must be light weighted.

Nodes can collaborate by sharing suspicious behaviour reports with the neighbours. If multiple nodes detect malicious activity from the same source, they can share their findings to confirm the threat. This consensus approach helps mitigate the risk of false positives and improves detection accuracy. The network can isolate the detected malicious nodes by avoiding them in future route discoveries.

During the route discovery phase, additional checks can be added to verify the authenticity of the route, Nodes can validate the integrity of route information by using cryptographic techniques, such as digital signatures or hash functions. This prevents malicious nodes from altering the route without detection.

The DSR protocol can be integrated with an Intrusion Detection System (IDS) tailored for ad-hoc networks. Th IDS can analyse traffic patterns, monitor for common attacks alert nodes when anomalies are detected. The IDS can use machine learning algorithms to identify deviations from normal behaviour, adapting to new attack patterns over time.

Extending DSR with malicious node detection capabilities can significantly enhance the security of mobile ad-hoc networks. By leveraging behaviour monitoring, trust-based systems and collaborative decisions. The protocol can identify and isolate threats, maintaining the integrity and reliability of the network even in hostile environments.

In [7], David B. Johnson and David A. Maltz, 1999 proposed a protocol for ad-hoc networks that implements dynamic source-based routing, allowing it to rapidly adjust to route changes during frequent host movement.

In [8], Amer Abu Salemet al,2014 stated that route discovery is the process initiated at the packet source to find a path to the destination.

In [9], Salem A. Almazok and Bulent Bilgehan, 2020 stated that the Dynamic Source Routing (DSR), a common routing protocol, uses the minimum hop count to determine the path, neglecting factors such as energy consumption and node energy levels, which can significantly influence the performance of the routing algorithm.

In [10], David A. Maltzet al, 2000 stated that each node operates the Dynamic Source Routing protocol and integrates smoothly with the existing internet infrastructure and the mobile IP Protocol.

In [11], Shams Qaziet al, 2013 proposed that the wormhole attack can block two nodes from finding legitimate routes that are more than two hops away, thereby disrupting network functionality.

In [12], C.V. Nanda Kishore and Venkatesh Kumar, 2023 stated that each wireless sensor networks (WSN) node is equipped with several sensing devices and are governed by a micro controller.

V. RELEVANT WORKS IN CRYPTOGRAPHIC TECHNIQUES

Cryptographic techniques are used to enhance security in communication among IoT nodes. Many IoT devices face resource limitations, such as restricted processing power, memory, and battery life), applying cryptographic requires careful consideration to balance security and performance. Transforms plain text data into ciphertext using cryptographic keys. The main two types are:

- Symmetric Encryption: Employs a single key is used for both encryption and decryption.
- Asymmetric Encryption: employs a pair of keys (public and private) for encryption and decryption.

Cryptography can be applied to IoT (Internet of Things) devices to ensure the security while being efficient enough for resource constrained devices. These algorithms are optimized for low power consumption, small code size and minimal computational overhead.

AES-CCM (Counter with CBC-MAC) is an operational mode for cryptographic algorithms that offers both encryption and authentication. Encryption uses the AES algorithm in counter mode for encryption data. Authentication uses CBC-MAC (Cipher Block Chaining Message Authentication Code) ensures data integrity and authenticity. AES-CCM combines AES encryption and CBC-MAC authentication to offer a secure and efficient way to protect a secure reliable and efficient method for securing data in IoT networks.

In a decentralized network, it can be difficult to manage trust and detect malicious behaviour without a central authority. The solution is to implement a block chain-based trust management system to maintain a tamper proof record of node behaviour.

Nodes record their interactions and trust scores on a distributed ledger and the block chain ensures that records cannot be altered, providing a reliable way to identify and track malicious nodes. Consensus algorithm can be used to validate updates to trust the ledger.

Cryptographic operations, especially public key cryptography can be computationally expensive making them challenging for resource limited. The key management in dynamic and decentralized environments liker MANET securely distributing and managing cryptographic key is complex. The overhead introduced by cryptographic operations and data exchanges can impact network scalability and performance.

Incorporating cryptographic techniques to DSR and other routing protocols can significantly enhance the detection and mitigation of malicious nodes in mobile ad-hoc networks. By leveraging digital signatures, hash chains, MACs and block chain the protocol can ensure the data integrity, authenticate nodes and maintain secure communication in the adversarial nodes.

In [13], Javier Sanchez Guerreroet al, 2017 stated that security must be maintained throughout the device's lifecycle from design to deployment by using credentials that enable secure access.

In [14], Thi Van Than Doanet al, 2023 stated that the calculation results stay encrypted and can only be decrypted by the data owner, ensuring confidentiality. This allows the third parties to work with cipher texts without accessing the decrypted data.

In [15], Armaan Sidhu, 2023 ensures the privacy and security of sensitive information, stressing the importance of ongoing research and innovation in this crucial field. It also outlines the key milestones in cryptography, especially those related to modern ciphers.

In [16], Shalini Subramaniet al, 2023 stated that the Quantum cryptography secures data through encryption, utilizing the principles of quantum physics to guarantee absolute security in communication.

In [17], Alfred Menezeset al, 1996 stated that the Public-Key cryptography techniques are widely used today. Cryptography is used to protect personal privacy, such as in electronic mail, across the financial services industry, public sector and by individuals.

In [18], Miles E. Smid, 2021 stated that the robust cryptographic algorithms are vital for protecting data, both stored and transmitted, worldwide. AES was the outcome of a multi-year collaborative effort between the U.S. government, industry and academia.

VI. RELEVANT WORKS IN BAYESIAN NETWORK MODEL

A Bayesian Network is a type of probabilistic model that employs a direct acrylic graph (DAG) to show the relationships and conditional dependencies among a set of variables.

Bayesian Network and random forest used for predictive modelling and uncertainty quantification. They provide a framework for decision making in uncertain environments by integrating probabilistic reasoning.

Bayesian inference is used to update the belief about a node's malicious status based on observed evidence. Prior probability is the initial belief about a node's behaviour, often set based on historical data or assumed to be neutral if no prior information is available. The posterior probability is an updated probability of a node being malicious after incorporating observed evidence.

In [19], Desmond Onam, 2024 stated that the Bayesian networks, often referred to as belief networks, are graphical model used to capture the probabilistic dependencies among a collection of random variables.

In [20], Judea Pearl, 1998 stated that the Bayesian networks are characterized by their ability to capture conditional dependencies between variables using conditional probabilistic distributions.

In [21], Dapahne Koller and Nir Friedman, 2009 stated that the Bayesian interface algorithms enable the computational posterior probabilities and likelihoods, allowing agents to update their beliefs in the presence of new evidence.

In [22], Bouckaert and Lucas, 2009 stated that the Bayesian network provide a flexible framework for modelling complex systems with uncertain or incomplete information.

In [23], Heckermanet al, 1995 stated that the Bayesian networks exhibit modularity, allowing for the modular construction and modification of complex models by adding or removing nodes and edges.

In [24], David J. Speigelhalteret al,1992 stated that the eventual aim is to produce a program that can handle arbitrarily complex problems by decomposing them into modular components which then can be handled by the software.

VII. RELEVANT WORKS IN RANDOM FOREST

Random forest is a machine learning method that builds numerous decision trees during the training phrase and makes predictions based on the majority vote classification or the average for regression. Random forest can leverage the enhanced feature set and uncertainty estimates from Bayesian networks to improve predictive performance.

Random forest typically provides high classification accuracy due to its ensemble nature, reducing over fitting compared to single decision trees. It can measure the importance of each feature, helping identify the key indicators of malicious behaviour and the model is robust to noise and can handle missing or corrupted data making it effective in dynamic and unreliable network environments.

The random forest is trained using a labelled dataset where nodes are classified as malicious or benign based on the observed features which includes data collection, which gathers network traffic data, including features such as packet forwarding ratio, delay anomalies and route

modification. Feature extraction extract meaningful features from raw data that capture potential malicious behaviours. The model training uses the labelled dataset to train the random forest classifier

In [25], Yanjun Qi, 2012 stated that the random forest is an efficient, interpretable, and non-parametric method suitable for various types of datasets.

In [26], Leo Brieman, 2001 stated that the random forest consists of an ensemble of unpruned classification or regression trees, each constructed from randomly chosen subsets of the training datasets.

In [27], Adele Cutleret al, 2011 stated that random forest is a model that combines multiple decision trees, where each tree is built using randomly sampled data and features, with the same distribution applied across the entire forest

In [28], Gerard Biauet al, 2008 stated that the base classifiers employed for averaging are general and simple and randomized, often using random samples from the data.

In [29], Khaled Fawagrehet al, 2013 an ensemble learning method, Random Forest is used for classification and regression.

In [30], Khaled Fawagrehet al, 2014 proposed a method to increase the diversity of Random Forest by selecting random subspaces, assigning a weight to each based on its predictive ability, and using the weight in majority voting process.

VIII. HYBRID APPROACH IN BAYESIAN NETWORK MODEL AND RANDOM FOREST TO DETECT TRUSTWORTHINESS OF IOT NODES

In [31], Fathy, N et al, 2023 introduces a hybrid trust management model designed to enhance the detection of misbehaving nodes in Internet of Things (IoT) networks, combining both direct trust based on observed behaviours and indirect trust (based on reputation from neighbouring nodes.

By combining Bayesian networks and random forest, we can leverage the probabilistic reasoning and uncertainty quantification capabilities of Bayesian networks with powerful predictive performance of random forests. This hybrid approach can result in more accurate and robust models especially in complex domains with significant uncertainty and feature dependencies.

Use a Bayesian network model the probabilistic dependencies between features and compute an initial trust score for each node. The random forest uses the full set of features and incorporates the initial trust score from the Bayesian network as an additional input feature.

By combing these two approaches, the hybrid model can utilize probabilistic reasoning for initial assessment and ensemble-based classification for accurate detection, achieving a balance between interpretability and performance.

IX. CONCLUSION

Authentication in the Internet of Things (IoT) is vital due to its unique characteristics such as sensing, connectivity, intelligence, scalability, and self-configuration in an open dynamic environment. The necessity for mutual trust between IoT devices, without prior familiarity, makes them susceptible to potential attacks, emphasizing the importance of trust-based security solutions. This survey delves into four critical components—Support Vector Machines (SVM), Dynamic Source Routing (DSR), Cryptographic Techniques, and Bayesian Network Models—to address IoT security challenges. It discusses the strengths and weaknesses of SVM in detecting malicious nodes, the role of DSR in efficient routing to enhance trust, the use of cryptographic techniques for securing communication, and the application of Bayesian networks combined with random forests for evaluating node trustworthiness. Together, these methods provide a comprehensive approach to improving the security and trust mechanisms within IoT systems.

REFERNCES

[1] R. Parihar, A. Jain and U. Singh, "Support vector machine through detecting packet dropping misbehaving nodes in MANET," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2017, pp. 483-488, doi: 10.1109/ ICECA.2017. 8212711.

[2] Jayshree Jha, Leena Ragha,"Research work on Intrusion Detection System using Support Vector Machine." International Conference and workshop on Advanced Computing 2013. ICWAC, 3 (June 2013).

[3] Evgeniou, Theodoros & Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7_12.

[4] Shi, Yong & Tian, Yingjie & Kou, Gang & Peng, Yi & Li, Jianping. (2011). Support Vector Machines for Classification Problems. Advanced Information and Knowledge Processing. 10.1007/978-0-85729-504-0_1.

[5] Gaye, Babacar & Zhang, Dezheng&Wulamu, Aziguli. (2021). Improvement of Support Vector Machine Algorithm in Big Data Background. Mathematical Problems in Engineering. 2021. 1-9. 10.1155/2021/5594899.

[6] Tong, Simon & Koller, Daphne. (2001). Support Vector Machine Active Learning with Applications to Text Classification. The Journal of Machine Learning Research. 2. 45-66. 10.1162/153244302760185243.

[7] Johnson, David & Maltz, David. (1999). Dynamic Source Routing in Ad Hoc Wireless Networks. Mobile Comput.. 353.

[8] Abu Salem, Amer & Samara, Ghassan &Alhmiedat, Tareq. (2014). Performance Analysis of Dynamic Source Routing Protocol. 10.48550/arXiv.1712.04622.

[9] Almazok, Salem & Bilgehan, Bülent. (2020). A novel dynamic source routing (DSR) protocol based on minimum execution time scheduling and moth flame optimization (MET-MFO). EURASIP Journal on Wireless Communications and Networking. 2020. 219. 10.1186/s13638-020-01802-5.

[10] Maltz, David & Broch, Josh & Johnson, David. (2000). Experiences Designing and Building a Multi-Hop Wireless Ad Hoc Network Testbed.

[11] Qazi, Shams & Raad, Raad & Mu, Yi & Susilo, Willy. (2013). Securing DSR against wormhole attacks in multirate ad hoc networks. Journal of Network and Computer Applications. 36. 10.1016/j.jnca.2012.12.019.

[12] Cv, Nanda & H, Venkatesh. (2023). Providing End-to-End QoS over Mobile and Adhoc Networks using Dynamic Source Routing (DSR) Protocol. 1-5. 10.1109/ICONAT57137.2023.10080251.

[13] Paez-Quinde, Maria & Guerrero, Javier & Alban, Robert & Narváez Rios, Magaly &Guachimboza, Marco. (2017). Cryptography Applied to the Internet of Things. 10.1007/978-3-319-73210-7_47.

[14] Doan, Thi& Messai, Mohamed-Lamine & Gavin, Gérald& Darmont, Jérôme. (2023). A survey on implementations of homomorphic encryption schemes. The Journal of Supercomputing. 79. 1-42. 10.1007/s11227-023-05233-z.

[15] Sidhu, Armaan. (2023). Analyzing Modern Cryptography Techniques and Reviewing their Timeline (2023).

[16] Subramani, Shalini &Munuswamy, Selvi & Arputharaj, Kannan & Kumar Svn, Santhosh. (2023). Review of Security Methods Based on Classical Cryptography and Quantum Cryptography. Cybernetics and Systems. 1-19. 10.1080/01969722.2023.2166261.

[17] Menezes, Alfred & Oorschot, Paul & Vanstone, Scott. (1996). Handbook of Applied Cryptography. 10.1201/9780429466335.

[18] Smid, Miles. (2021). Development of the Advanced Encryption Standard. Journal of Research of the National Institute of Standards and Technology. 126. 10.6028/jres.126.024.

[19] Onam, Desmond. (2024). Understanding Bayesian Networks: Significance in Artificial Intelligence.

[20] Judea Pearl. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[21] Koller, Daphne & Friedman, Nir. (2009). Probabilistic Graphical Models: Principles and Techniques.

[22] Bouckaert, R. R., & Lucas, P. J. F. (2009). Bayesian networks: A view on the horizon. Machine Learning, 75(1), 5-9.

[23] Heckerman, David & Geiger, Dan & Chickering, David. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. Machine Learning.

[24]T homas, Andrew & Spiegelhalter, David & Gilks, Wally. (1992). BUGS: a Program to Perform Bayesian Inference using Gibbs Sampling.

[25] Qi, Y. (2012) Random Forest for Bioinformatics. In: Zhang, C. and Ma, Y.Q. Ed., Ensemble Machine Learning, Springer, US, 307-323.

[26] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001).

[27] Cutler, Adele & Cutler, David & Stevens, John. (2011). Random Forests.

[28] Biau, Gerard &Devroye, Luc & Lugosi, Gábor. (2008). Consistency of Random Forests and Other Averaging Classifiers. Journal of Machine Learning Research.

[29] Fawagreh, Khaled & Gaber, Mohamed & Elyan, Eyad. (2014). Diversified Random Forests Using Random Subspaces.

[30] Fawagreh, K., Gaber, M. M., & Elyan, E. (2014). Random forests: from early developments to recent advancements. Systems Science & Control Engineering, 2(1), 602–609.

[31] Fathy, N., Khaled, M. E., & El-Bakry, A. M., A hybrid trust management model for misbehaving node detection in IoT networks. International Journal of Network Management, Vol 33, No.4,2023