A Review on Machine Learning Techniques for Secure IoT Networks

Lohesh M

January 22, 2023

1 INTRODUCTION

IoT is a network of interconnected devices that communicate through wireless or wired medium. These devices have low computational capabilities and use software instructions to send data to cloud servers. The number of connected devices in IoT networks is expected to grow significantly in the coming years, leading to the generation of a large amount of data known as Big Data. This poses a challenge for the security of IoT networks. Traditional security solutions for IoT are not effective due to the complexity and interconnectedness of the network. IoT networks are also vulnerable to physical attacks, wireless connection eavesdropping and interconnection undependability. Machine Learning (ML) can be used to provide security for IoT networks by analysing and detecting abnormal behaviour in the data. ML enables IoT devices to automate and make decisions based on previous knowledge from data. ML can also be used for tasks such as malware detection, computer vision, medical image detection, speech and face recognition, and bio-informatics. However, there are limitations to using ML for IoT security such as scalability and generalization, and further research is needed in this area. In addition, the integration of the network also introduces more security threats and makes IoT surfaces more vulnerable. Therefore, a critical privacy breach may occur if a large amount of valuable data from IoT devices is not analysed and transmitted securely.

2 MOTIVATION OF USING MACHINE LEARNING TO SECURE IOT NETWORKS

IoT networks are vulnerable to security threats due to their heterogeneous nature, massive deployment, and interconnectivity of devices, making it difficult to secure the network. Edge computing can help to improve security, but there is still a need for standard protocols to ensure the confidentiality, authorization, authentication, and integrity of the network and its data.

2.1 Security Threats to IoT Networks

IoT networks are vulnerable to both active and passive threats, such as physical access to devices, attacks on communication links, and lack of computational power. Confidentiality, authorization, authentication, and integrity are all important parameters to consider when designing an IoT network. Confidentiality ensures that sensitive information is protected from predators, authorization ensures that only authorized users have access to the network, authentication establishes the identity of entities in the network, and integrity ensures that data is not modified during transfer and that physical devices are secure. The need for a security solution is paramount in order to protect IoT networks from various security breaches. Additionally, the authentication process in IoT networks can be different for different systems, for example in a medical smart IoT network, both safety and security should be considered, and the authentication scheme is designed based on both properties. Furthermore, the integrity of the system should also constantly monitor the physical devices of a network to ensure security. For example, smart medical wearables and implantable devices need robust integrity monitoring, otherwise, it can affect human lives directly.

2.2 Machine Learning approach to overcome the threats

Machine learning techniques can be used to enhance the security of IoT networks by detecting and preventing attacks. The large amount of data generated by IoT devices can be used to train ML algorithms to establish new patterns and detect anomalies in network behaviour. These trained algorithms can be used to detect threats and prevent attacks on the network by monitoring incoming and outgoing data from registered devices and creating a profile for normal IoT ecosystem behaviour. ML methods such as supervised, unsupervised, semi-supervised, and reinforcement learning can be used to detect and mitigate authentication issues, DDoS attacks, and intrusion.

3 IOT SECURITY AND MACHINE LEARNING

Different machine learning algorithms have different strengths and weaknesses and can be applied to various security problems in IoT networks. Some of the commonly used ML algorithms include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. The choice of algorithm will depend on the specific problem that needs to be solved and the type of data available for training.

3.1 Machine Learning Algorithms

...

1. Supervised Machine Learning:...

    - Support Vector Machines (SVM): Support Vector Machines (SVMs) is a supervised learning algorithm that classify data into two or more classes by finding a hyperplane that maximizes the distance between the closest samples of each class. They are commonly used in security applications such as intrusion detection and attack prevention and can also be used to effectively break cryptographic devices.
    - Bayesian Algorithms: Bayesian algorithms are a type of machine learning algorithm that use the Bayesian theorem to calculate the probability of an event based on previous behaviour. The most common algorithm is the Naive Bayes (NB) algorithm, it is notable for its simplicity and commonly used in network intrusion detection, it calculates the posterior probability of an event and uses the Bayesian theorem to label unlabelled input.
    - k-Nearest Neighbours (KNN): KNN (K-Nearest Neighbours) is a supervised machine learning algorithm that can perform both classification and regression. It categorizes new input based on the votes of its nearest neighbour samples and assigns the class with the majority of votes. It is commonly used in network intrusion detection, and in wireless sensor networks to detect intrusion.
    - Artificial Neural Network (ANN): ANNs are a type of ML algorithm that are inspired by the biological neural network and uses multiple layers of processing elements to classify input data. Researchers have proposed the use of ANNs for securing IoT communications and ongoing research is being done on using neural networks for IoT traffic prediction.
    - Ensemble Learning (EL): Ensemble Learning is a popular ML technique that combines multiple models to improve performance and accuracy. It is often used in various applications and can achieve greater accuracy compared to individual models.

2. Unsupervised Machine Learning:...

    - Principal Component Analysis (PCA): PCA is a feature reduction tool that reduces large datasets without losing important information, commonly used in intrusion detection for real-time IoT systems.
    - k-Means Clustering: k-Means Clustering creates k clusters based on similar features in input data. It is commonly used for anomaly detection and the performance of this technique is not as effective as supervised machine learning models.

3. Semi-supervised Machine Learning: Supervised ML is the most robust class of ML but requires labelled data, which can be costly and time-consuming. Unsupervised ML uses unlabelled data but is less accurate. Semi-supervised ML combines the two by augmenting unlabelled data to train supervised ML algorithms, but its accuracy and robustness are still lower than supervised ML. There is limited research on semi-supervised ML models, but some have been used for network intrusion detection with good performance.

4. Reinforcement Learning (RL) Models: Reinforcement learning (RL) is a type of machine learning that allows an agent to learn from its environment by taking actions and receiving feedback in the form of rewards or punishments. It is based on the idea of trial and error, where the agent learns to select actions that lead to the highest reward over time. RL is often used in the IoT security domain, for example, in anti-jamming projects for autonomous radios, where the algorithm learns to evade jamming frequencies by using data on jamming signals and other interference.

3.2 Machine Learning Applications in IoT Security

...

- Both the supervised and unsupervised ML algorithms are mainly focused on data training but RL is dedicated to decision making and comparison tasks.
- The supervised ML models are currently used in malware analysis intrusion detection, and attack detection.
- It is utilized in IoT networks for channel estimation, adaptive filtration, and spectrum utilization.
- Unsupervised ML mainly focused on dimensionality reduction with clustering.
- It is currently used by researchers in tasks like anomaly and intrusion detection as well as malware detection.
- RL is in its developing stage but widely used in IoT networks.
- It involves creating learning models from the information of the surroundings.
- RL is being used against aggressive jamming in tactical mobile networking and autonomous radios.

4 LIMITATIONS OF MACHINE LEARNING IN IOT NETWORKS

Accurate results in ML training for IoT networks can be achieved by using a large initial dataset. However, in IoT networks, there is a lack of sufficient security-related data and using confidential data for training can be an issue. To address this problem, a crowd-sourcing platform could be developed to create different datasets for various security tasks. These datasets should be updated by continuously monitoring new attack patterns. Additionally, low-level data, which can be noisy or corrupted, should be filtered before being sent to the ML model for training in real-time. This will ensure that the ML model is trained on high-level, clean data.

5 CONCLUSION

IoT networks are vulnerable to security threats due to their heterogeneous devices and information. Traditional security solutions are not feasible at the commercial level, so ML algorithms are being used to enable IoT devices to learn from the environment and input data, and train themselves to secure the IoT networks. This work discussed various ML techniques used to secure IoT networks, the threats to IoT networks and the corresponding ML solutions, and a comparison of different ML techniques and their advantages, disadvantages, and current deployment. Additionally, the basics of different types of learning such as supervised, unsupervised, and reinforcement learning were covered. The real-time applications of each ML model were also described, as well as the limitations of using ML in IoT networks.