Intelligent Deep Learning in IoT Smart Home Networks

Subjects: Computer Science, Artificial Intelligence

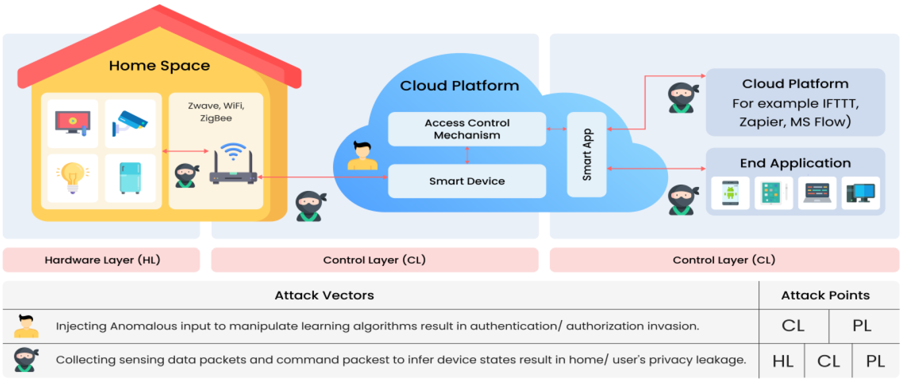
Contributor: Kashif Naseer Qureshi , Nazia Butt , Ana Shahid , Sajjad Haider , Ashraf Osman Ibrahim , Faisal Binzagr , Noman Arshad

The Internet of Things (IoT) is the interconnection of sensors, machines, objects, or other computing devices over the internet to communicate with the least human interference. Specific types of sensors are involved in obtaining information from physical entities, and after analysis, it is stored in local storage, which is then sent to cloud storage, where appropriate action is taken according to the information.

internet of things smart homes machine learning

1. Introduction

The Internet of Things (IoT) is the interconnection of sensors, machines, objects, or other computing devices over the internet to communicate with the least human interference. Specific types of sensors are involved in obtaining information from physical entities, and after analysis, it is stored in local storage, which is then sent to cloud storage, where appropriate action is taken according to the information. The smart home is one example of connected home devices that can be controlled from anywhere at any time [1]. These networks are implemented as a global technology and have gained popularity among users. These networks have suffered from various challenges, among which security is one of the top challenges. Some of the common security attacks on these networks are Denial of Service (DoS), brute force, and ransomware. Intrusion detection is the concept of monitoring the traffic and classifying it as benign or malign. Intrusion detection in IoT networks can be signature-based, anomaly-based, or specification-based. In an anomaly-based intrusion detection system, normal behavior is recorded and stored as patterns and then compared with traffic patterns to see whether noise and other potential intrusions are anomalous or normal [2]. There are multiple techniques for anomaly-based intrusion detection systems (IDSs), such as data mining, statistical models, rule models, payload models, protocol models, and signal processing models. Machine learning (ML) and deep learning (DL) techniques are used for anomaly detection to tackle attacks on a network with significant performance [3][4]. **Figure 1** shows an IDS in a smart home network.



Attack Vectors	Attack Points		
Injecting Anomalous input to manipulate learning algorithms result in authentication/ authorization invasion.	CL	PL	
Collecting sensing data packets and command packest to infer device states result in home/ user's privacy leakage.	HL	CL	PL

Figure 1. IDS in a Smart Home Network.

In signature-based techniques, detection is performed by matching signatures stored in the database with the signatures of the traffic flow. Anomaly-based IDSs with ML techniques are designed for attack detection [5][6]. However, the existing solutions are limited in function due to a lack of real-time dataset usage or an updated dataset, which leads to a degradation in performance, as attacks usually change their patterns and methods. Public datasets are available but not specifically created for smart home IoT networks. Some of the existing solutions suffer from overhead issues or increased time complexity [7]. Moreover, there are various issues observed, such as noise, overfitting, underfitting, complexity, and dimensionality, which lead to carelessness in data cleaning, feature extraction, selection, and normalization techniques. Accuracy is one of the significant parameters for measuring the model performance and needs to be maximized. However, there are differences in the test and training accuracy of many existing models, which suffer from overfitting or underfitting. As a countermeasure to these issues, appropriate methods are selected for data cleaning and hyperparameter selection. Hybrid ML/DL-based classifiers are used on recent real-time datasets for intrusion detection.

2. Intelligent Deep Learning in IoT Smart Home Networks

The authors of [8] used different Support Vector Machine (SVM) techniques, namely, Linear, Quadratic, Fine Gaussian, and Medium Gaussian SVM, on the NSL-KDD dataset. Linear SVM involves a linear kernel and is used when data are linearly separable. If zs and zt are data points, then the kernel in this scenario is Fine Gaussian SVM, which showed a clear difference between classes with the kernel sqrt(P)/4 (P is a predictor). Medium Gaussian showed fewer differences between classes when the kernel was sqrt (P). The analysis was conducted through ROC and a confusion matrix. Fine Gaussian SVM achieved high performance among other SVM techniques with a minimal error rate. However, this technique was not tested on a real-time dataset. A real-time dataset is needed for the future of IoT security in terms of intrusion detection. The authors of [9] used multinomial NB, a continuous dataset with discrete data, whereas, in Bernoulli NB, both discrete and categorical data are used, but the feature vector should be binary. The experimental setup used Gaussian NB, as the aim was to deal with more than two groups of attacks. Moreover, the sklearn library of Python was used in this technique to evaluate all parameters over the KDD dataset. PCA was also used to reduce the attributes and execution time of the KDD dataset over the KDD dataset, which exhibits better performance than the traditional Naïve Bayes. However, if the number of components is increased, then it affects the accuracy, which could be a challenge to overcome in the future.

The authors of [10] used four ensemble learning ML models on the RPL-NIDIDS17 dataset to overcome routing attacks. The ensemble learning models were Boosted Trees, Bagged Trees, Discriminant, and RUS Boosted Trees, and the dataset contained packet traces of Sybil, Clone ID, Black Hole, and Hello Flooding. The preprocessing of data was performed through cleaning, one-hot encoding, and scaling methods. Missing values were handled through cleaning, where one-hot encoding converted the nominal data into numerical form, and scaling was used to scale them between 0 and 1. After preprocessing, the data were converted into training and test samples, and four ensemble learning models were trained on the training set and then tested to see the expected outcomes in terms of attack detection (normal or an attack class), which indicated the good performance of EL ML models. However, lightweight solutions for securing smart nodes in IoT networks will be a target in the future.

The authors of [11] used the CART algorithm based on decision trees (DTs) to split the parent and child nodes based on the Gini index criterion. Ensemble classifiers utilized the results of multiple DTs through voting. This means that multiple classifiers are used for the selection of sample classes through voting rather than a single model. CART Decision Tree is famous for classification and regression. Combinations of three decision trees were used along with the NSL KDD dataset, which resulted in improved performance and accuracy while detecting a variety of attacks, such as DoS, R26, and Probe. However, the time required for modeling was increased due to the combination of trees, which could be ignorable or manageable after further research and testing.

The authors of [12] used KNN and LSTM for protection against illegitimate users in IoT networks. The proposed technique is based on three phases. In preprocessing, the normalization of data is conducted in R [0,1] through the min–max function. After preprocessing, feature selection is completed, through which the best features for intrusion detection are selected. Finally, KNN and LSTM are implemented to detect intrusion. The grouping of instances is managed according to the value of K and the distance measured. LSTM is used to minimize the error rate by calculating the difference between the expected outcome and the original outcome and then adjusting these calculations by varying the values of weights and biases accordingly. Simulations were run in MATLAB, and the BOT-IOT dataset was used. The mean detection time and Kappa stats were evaluated as parameters for the performance check. The detection time is the time needed to recognize the attack, whereas the mean value is used to set and balance TPR and TNR. A comparison of KNN and LSTM was also made, which determined that LSTM was not underfitted or overfitted, so it is a better-performing algorithm in the scenario. More attacks and a high number of instances in real-time IoT scenarios needed to be extended from this work. The authors of [13] improved the feature sets and association rule mining techniques, such as the FP growth algorithm, for the improvement of feature sets through the FP growth algorithm. The CNN model was implemented for Botnet attack detection with higher accuracy than existing features. However, several attacks, a larger sample size, and more ML/DL models with tested thresholds are issues and challenges for future work.

The authors of [14] used Gray Wolf Optimization (GWO) and Particle Swarm Optimization (PSO) for feature extraction and selection. Random Forest (RF) was used as a classifier for intrusion detection through simulations in Python language on KDD 99, NSL-KDD, and CIC IDS 2019 datasets. RF has a high variance and low bias, but with the GWO-PSO-RF problem, the biasing problem was solved. Hence, it showed optimal results, but it needs to be implemented in a real learning environment implementing IoT security to ensure its performance, and a distillation technique needs to be applied to enhance its performance. The authors of [15] used the Q learning model to predict cyberattacks, and the problem of QoS control was managed by the RL learning algorithm. The RL-based model was also compared in terms of accuracy and precision with other DL models with significant performance, and AUC was also improved. However, an increased number of epochs caused a decrease in precision. More DL models with different calculations need to be trained to develop an effective IDS.

The authors of [16] used a protocol-based DID dataset, which reduced the number of features compared with the UNSW-NB15 and BoT-IoT datasets, and LSTM was used as a classifier with promising results. However, the misclassification of DoS and DDoS occurred due to similarities between their features, which needs to be mitigated in the future. The authors of [17] used AAE and GAN to prevent noise in the data and latent representation of the data, whereas KNN was used as a classifier on the IoT 23 dataset. Both techniques showed good performance, but GAN outperformed in terms of accuracy. Instances of minority classes were increased, and the feature resemblance method was used for the detection of new attacks. **Table 1** shows the comparison of the discussed schemes.

Table 1. Comparison of ML/DL schemes for IDS.

Algorithm/Model	Dataset	Classification	Attacks	Performance Metrics	Achievements	Demerits
Nonlinear SVM [8] 2019	UNSW NB 15	Binary, multiclass	Analysis, Backdoor, DoS, Exploits, Fuzzers, Generic, Reconnaissance, Shellcode, worms	ACC, DR, FPR	Performed well in attack classification.	The difference in training and test accuracy reveals low bias and high variance, which may indicate overfitting.
PCA+NB [9] 2019	NSL-KDD	Multiclass	Probe, DoS, U2R, R2L	ACC, confusion matrix	Decreases the execution time by minimizing the number of components.	Degradation in accuracy if the number of components is increased.

Algorithm/Model	Dataset	Classification	Attacks	Performance Metrics	Achievements	Demerits
EL [10] 2019	RPL-NIDDS17	Binary	The sinkhole, Blackhole, Sybil, Clone ID, Selective Forwarding, Hello Flooding, and Local Repair	ACC, AUC	Ensemble learning showed good performance in mitigating routing attacks.	High accuracy.
Hybrid DT [11] 2020	NSL-KDD	Binary	DOS, Probe, U2R, R26	ACC, precision, recall, F1 score	Improved accuracy.	Value of recall is lower than benchmark schemes.
KNN, LSTM [12] 2020	Bot-IOT	Binary	DDoS, DoS, OS, Service Scan, Keylogging, data exfiltration	ACC, DR, Kappa Stats, Geometric Mean	Counters overfitting and underfitting issues and has a faster learning rate with LSTM.	Difficult to select the suitable value of K when using KNN, and LSTM takes more time and power to train.
FP growth algorithm, CNN [13] 2020	N-BaIOT	Binary	IoT Botnet attacks	ACC, precision, recall, F1 score	Method for improvement of the original feature set is given, and ACC is also good.	The number of classes and data size are small. The threshold value could have been tested. Only one attack class is countered.
Linear SVM [5] 2021	NSL-KDD	Multiclass	DoS, Probe, U2R, R26	ACC, error, ROC, confusion matrix	Fine Gaussian SVM performed well with the lowest error rate.	Lack of optimization of SVM.
GWO-PSO-RF [14] 2021	KDDCUP99, NSLKDD99	Binary class, multiclass	DoS, DDoS, Heartbleed, Botnet, Infiltration	ACC, precision, recall, F1 score,	Balanced GWO-PSO-RF reduced biasing	A real-time dataset was not used.

Algorithm/Model	Dataset	Classification	Attacks	Performance Metrics	Achievements	Demerits
				Support, confusion matrix	problem and DR of minority classes.	
MDP [15] 2021	NSL-KDD	Multiclass	DDOS, DOS	ACC, precision, sensitivity, AUC	Gave the best precision and AUC curve.	Low accuracy and high epoch number could cause overfitting.
LSTM [16] 2022	UNSW-NB15, Bot-IoT	Multiclass	Dos, DDos	ACC, confusion matrix	Imbalanced data issues and overfitting were countered.	More methods for noise removal in data are used.
BiGAN + KNN [17] 2022	IOT23	Multiclass	Dos, Botnet Attacks	ACC, precision, recall, F score	Efficient detection of Zero-Day Attack.	Shared features in the data were not analyzed and required more instances of the minority class.

traffic using vector convolutional deep learning approach in fog environment. Future Gener. Comput. Syst. 2020, 113, 255–265.

4. Qureshi, M.A.; Qureshi, K.N.; Jeon, G.; Piccialli, F. Deep learning-based ambient assisted living for self-management of cardiovascular conditions. Neural Comput. Appl. 2021, 34, 10449–10467.
5. IvanCvitic, D.; Gupta, B.B.; Choo, K.-K.R. Boosting-Based DDoS Detection in Internet-of-Things Systems. IEEE Internet Things J. 2021, 9, 2109–2123.
6. Cvitić, I.; Peraković, D.; Periša, M.; Gupta, B. Ensemble machine learning approach for classification of IoT devices in smart home. Int. J. Mach. Learn. Cybern. 2021, 12, 3179–3202.
7. Zaib, M.H.; Bashir, F.; Qureshi, K.N.; Kausar, S.; Rizwan, M.; Jeon, G. Deep learning based cyber bullying early detection using distributed denial of service flow. Multimed. Syst. 2021, 28, 1905–1924.
8. Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J. Survey of intrusion detection systems: Techniques, datasets and challenges. Cybersecurity 2019, 2, 1.
9. Sharmila, B.; Nagapadma, R. Intrusion detection system using Naive Bayes algorithm. In Proceedings of the 2019 IEEE International WIE Conference on Electrical and Computer

Engineering (WIECON-ECE), Bangalore, India, 15–16 November 2019; IEEE: New York, NY, USA, 2019; pp. 1–4.

10. Verma, A.; Ranga, V. ELNIDS: Ensemble learning based network intrusion detection system for RPL based Internet of Things. In 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 18–19 April 2019; IEEE: New York, NY, USA, 2019; pp. 1–6.
11. Taghavinejad, S.M.; Taghavinejad, M.; Shahmiri, L.; Zavvar, M.; Zavvar, M.H. Intrusion detection in IoT-based smart grid using hybrid decision tree. In 2020 6th International Conference on Web Research (ICWR); IEEE: New York, NY, USA, 2020; pp. 152–156.
12. Sugi, S.S.S.; Ratna, S.R. Investigation of machine learning techniques in intrusion detection system for IoT network. In 3rd International Conference on Intelligent Sustainable Systems (ICISS), Coimbatore, India, 3–5 December 2020; IEEE: New York, NY, USA, 2020; pp. 1164–1167.
13. Van Huong, P.; Minh, N.H. Improving the feature set in IoT intrusion detection problem based on FP-Growth Algorithm. In 2020 International Conference on Advanced Technologies for Communications (ATC); IEEE: New York, NY, USA, 2020; pp. 18–23.
14. Keserwani, P.K.; Govil, M.C.; Pilli, E.S.; Govil, P. A smart anomaly-based intrusion detection system for the Internet of Things (IoT) network using GWO–PSO–RF model. J. Reliab. Intell. Environ. 2021, 7, 3–21.
15. Kalnoor, G. Markov Decision Process based Model for Performance Analysis an Intrusion Detection System in IoT Networks. J. Telecommun. Inf. Technol. 2021, 3, 42–49.
16. Zeeshan, M.; Riaz, Q.; Bilal, M.A.; Shahzad, M.K.; Jabeen, H.; Haider, S.A.; Rahim, A. Protocol-Based Deep Intrusion Detection for DoS and DDoS Attacks Using UNSW-NB15 and Bot-IoT Data-Sets. IEEE Access 2021, 10, 2269–2283.
17. Abdalgawad, N.; Sajun, A.; Kaddoura, Y.; Zualkernan, I.A.; Aloul, F. Generative Deep Learning to Detect Cyberattacks for the IoT-23 Dataset. IEEE Access 2021, 10, 6430–6441.

Retrieved from https://encyclopedia.pub/entry/history/show/89001