IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Cyber Attacks Detection In IOT Based Smart City Applications Using Machine Learning Technique

DILEEP SOMALA
Department of Electronics and Computer Engineering
Presidency University, Bangalore, India

VAISHNAVI U
Department of Electronics and Computer Engineering
Presidency University, Bangalore, India

KOTESWARARAO PANGA
Department of Electronics and Computer Engineering
Presidency University, Bangalore, India

SYED MUJEEB
Department of Electronics and Computer Engineering

VENKATESH EDIGA
Department of Electronics and Computer Engineering
Presidency University, Bangalore, India

Dr. RADHA RAM MOHAN S
Professor
School of Computer Science & Engineering
Presidency University, Bangalore, India

ABSTRACT

The spread of IoT applications in recent years has supported the development of smart cities. A smart city leverages Internet of Things-enabled technologies, communications, and applications to maximize operational efficiency and improve the quality of services provided by his providers as a result of the health and well-being of people. However, as smart city networks grow, so does the risk of cyber threats and attacks. In a smart city network, sensors connected to massive cloud servers are connected to his IoT devices and exposed to threats and malicious attacks. Therefore, it is imperative to protect the Internet of Things from failures and thwart attacks. This study explores machine learning (LR, DT, RF, and MLP) based attacks and detection methods to mitigate and counter IOT cyber security risks in smart cities. The proposed technique is highly effective in detecting cyber-attacks, and the Stacking Ensembles model outperforms comparable models in terms of accuracy, precision, recall, and F1 score, making it a promising candidate in this field. It further proves the sex.

KEY WORDS:

Smart city; Internet of Things, cyber security, cyber-attacks, machine learning, Random Forest, Linear Regression, Multi layer Perception, Decision Tree.

1. INTRODUCTION

The Internet of Things (IoT) is a network of connected devices that facilitate the transfer of information between gadgets (smart home sensors, environmental sensors, automotive and roadside sensors, medical equipment, industrial robots, surveillance equipment, etc.). . IoT has recently grown dramatically in popularity with communities and services around the world, with 27 billion connected IOT devices (as of 2018). Smart city applications face many security challenges. First, zero-day attacks are possible by exploiting vulnerabilities in many protocols used in smart city applications. Second, network-based hacks can be intelligently detected before they impact smart city functionality. Third, IOT systems used in smart cities often have resource limitations (such as RAM) and transfer collected data to cloud servers for processing. It also has limited onboard capacity for security operations.

The need for ensemble models of classifiers was spurred by the fact that a single classifier is often insufficient to provide efficient IDS. The ensemble approach considers many models and then merges them into one model. Studies show that ensemble models perform better than single classifiers. However, ensemble methods can perform well if many aspects are carefully considered (such as feature selection and base classifier). Bagging, boosting, and stacking are ensemble strategies that work best. In this study, we apply both single classifier and ensemble strategies to improve the performance of IDS on various evaluation metrics such as accuracy, precision, recall and F1 score.

2. LITERATURE SURVEY

This section will deal with all the previous information related to cyber attacks detection in IOT.. Literature survey is the most important step in software development process. For any software or application development, this step plays a very crucial role by determining the several factors like time, money, effort, lines of code and company strength. Once all these several factors are satisfied, then we need to determine which operating system and language used for developing the application. Once the programmers start building the application, they will first observe what are the pre-defined inventions that are done on same concept and then they will try to design the task in some innovated manner.

MOTIVATION

Cyber-attacks detection in IOT based application using ANN by MdMahedi Hassan and Tasadduq Imam Widespread adoption of Internet of Things (IoT) apps has recently helped drive the growth of smart cities that use intelligent applications to maximize operational efficiency, quality of service and well-being of citizens. increase. In this study, we present machine learning-based attack and anomaly detection techniques to mitigate his IoTcyber security risks in smart cities. There are many other machine learning (ML) techniques, including computationally intensive deep learning networks, but I chose to use artificial neural networks (ANNs).The Internet of Things (IoT) and cloud-based systems are opening up new areas of development.They have numerous applications such as smart grids, smart cities, smart farming, smart homes, and more. IoT sensors are so close to people and critical infrastructure that there are concerns about privacy and security. IoT network security is a popular research area and of great importance. Various types of intrusion detection systems (IDS) have been developed to detect and stop unauthorized network intrusions.

IDS Based on Ensemble Techniques

In the literature, several ensemble methods based IDSs are proposed to enhance accuracy over base classifiers. In [32], ANN and Bayesian net based ensemble method was proposed where they used gain ratio (GR) feature selection technique and performance was evaluated on KDD'99 and NSL-KDD datasets where ensemble methods achieved 99.42% and 98.07% accuracy, respectively.

In Haq et al. proposed an ensemble method that combines Naive Bayes, Bayesian Net and decision tree classifier. They extracted the common features by using Best First Search, Genetic and Rank Search feature selection techniques. The ensemble technique produced 98% true positive rate when tested with 10-fold cross validation method. Gaikwad et al. introduced a bagging ensemble method where they used REPTree as a base classifier. Their model achieved 81.29% accuracy on NSL-KDD dataset. In Jabbar et al. proposed an ensemble method comprising alternating decision tree (ADTree) and KNN, and the performance evaluation demonstrated that the proposed ensemble achieved better detection rate (~99.8%) compared to the existing techniques.

3. EXISTING METHODOLOGY

To detect assaults and anomalies, existing works typically use signature-based techniques, however these techniques have substantial overheads and are susceptible to known threats.

LIMITATION OF EXISTING SYSTEM

1. Low Accuracy
2. It is less Efficient
3. Signature based techniques are un-susceptible

4. PROBLEM STATEMENT

The majority of the published work to far uses signature-based methods to spot attacks and anomalies. Both the overhead costs and vulnerability of these methods are well-known. This study analyses the viability of ensemble based learning versus single model classifiers for detecting cyber-attacks in IOT-based smart city applications. Additionally, we contrast a multi-class classification configuration with the binary class prediction employed in the bulk of important articles. In addition to this, we take into account the integration of feature selection and cross validation, as even traditional machine learning approaches have not been sufficiently covered in the body of literature that is already accessible in this subject. Considering the integration above, extended research shows that a collection of machine learning-based classifiers outperforms a single classifier in correctly classifying attacks and their categories

5. PROPOSED SYSTEM

Our proposed model tracks the network traffic flowing through each fog node. Because fog nodes are closer to IOT sensors than cloud centers, IOT sensors can detect cyber-attacks better than cloud centers. This instantly identifies threats and alerts network and IOT administrators so they can review and update their systems

ADVANTAGES OF PROPOSED SYSTEM:

The following are the benefits of the proposed system. They are:

1. High accuracy
2. Low complexity

6. IMPLEMENTATION PHASE

The step of implementation is when the theoretical design is translated into a programmatically-based approach. The application will be divided into a number of components at this point and then programmed for deployment. Modules:

1. Data Collection
2. Data preprocessing
3. Training
4. Predicting
5. Result analysis

1) Data Collection Module

UNSW-NB15: a comprehensive data set for network intrusion detection systems**,** These features are described in UNSW-NB15_features.csv file. A partition from this dataset is configured as a training set and testing set, namely, UNSW_NB15_trai ning-set.csv and UNSW_NB15_testing-set.csv respectively.The number of records in the training set is 175,341 records and the testing set is 82,332 records from the different types, attack and normal. Figure 1 and 2 show the test bed configuration dataset and the method of the feature creation of the UNSW-NB15, respectively. The additional features are as described in UNSW-NB15_features.csv file.

2) Data preprocessing Module

i. Null value handling
ii. Categorical handling
iii. Scaling

In this data preprocessing module for the given data set, there are few values missing in the data set, the missing values could lead to false prediction results. The missing values are removed using the python. It is also important to remove the replicated data from the data set to fetch the accurate results. Normalization is used to scale the data to a specific range.

3. Training Module

A. Model training with classifier

B. Accuracy

C. Save model

It is important to split X and y as training set and testing set. Here, we will split the original data as 70% training set and 30% testing set. But the partition action from this dataset was pre-configured as a training set and testing set,

namely, UNSW_NB15_training-set.csv and UNSW_NB15_testing-set.csv respectively.The number of records in the training set is 175,341 records and the testing set is 82,332 records from the different types, attack and normal.

4. Predicting Module

A. Input test data

B. Preprocess

C. Load model

D. Recognition

5. Result analysis Module

A. View result

7. MACHINE LEARNING ALGORITHMS

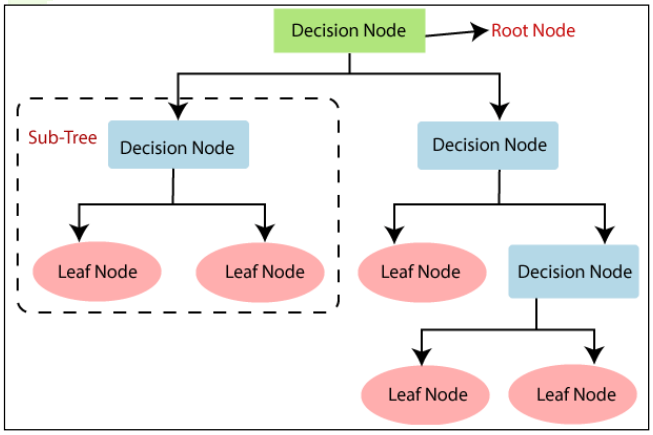
1) Logistic Regression:

Predictive analytics and classification frequently use this kind of machine learning regression model, also referred to as a logit model. Depending on the given dataset of independent features, the logistic regression model calculates the probability that an event will occur, such as voting or not voting. Given that the result is a probability of happening an event, the dependent feature's range is 0 to 1.

In the logistic regression model, the odds of winning the probability of success of an event divided by the probability of failure-are transformed using the logit formula.

2) Decision Tree:

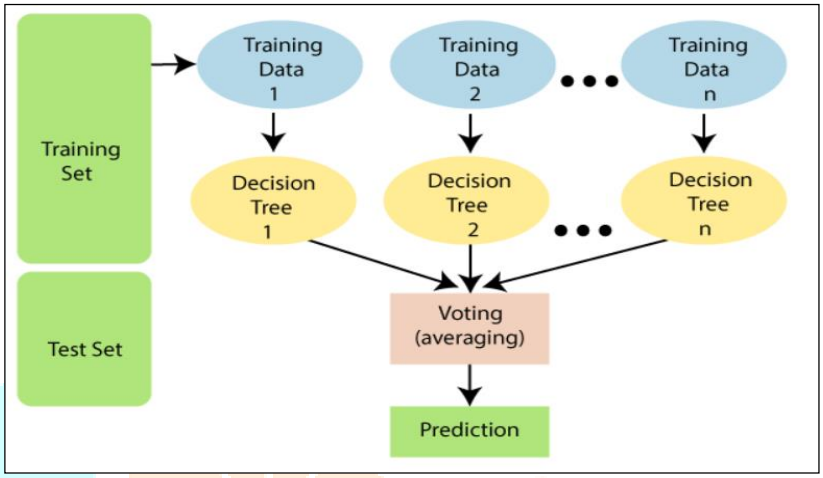
Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset.



3) Random Forest Classifier:

Random forests are for supervised machine learning, where there is a labeled target variable.

Random forests can be used for solving regression (numeric target variable) and classification (categorical target variable) problems. Random forests are an ensemble method, meaning they combine predictions from other models.Each of the smaller models in the random forest ensemble is a decision tree.



8. EXPERIMETAL REPORTS

In this proposed application, we try to use google collab or Jupitor as working platform and try to show the performance of our proposed application.

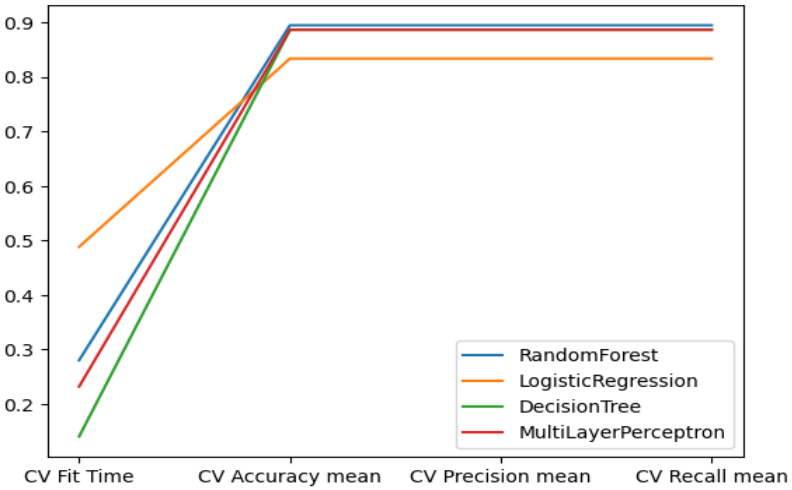
1) CONFUSION MATRIX

```
Confusion Matrix:
 [[    0     0     5   190    24   189   267     2     0     0]
 [    0     0     5   104    44   189   221    20     0     0]
 [    0     0    61  3004   242   260   321   201     0     0]
 [   16     0    46  8443  1102   437   784   304     0     0]
 [    0     0    16   708  3180   403  1651   104     0     0]
 [    0     0     9  1421   222 16507   641    71     0     0]
 [  166     0    45  3708  9592   293 22347   831    18     0]
 [    0     0     6  1373   404    10    97  1606     0     0]
 [    0     0     0   125    92     0    14   147     0     0]
 [    0     0     0    23    13     6     1     1     0     0]]
```

2) CLASSIFICATION REPORT

```
Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00       677
           1       0.00      0.00      0.00       583
           2       0.32      0.01      0.03      4089
           3       0.44      0.76      0.56     11132
           4       0.21      0.52      0.30      6062
           5       0.90      0.87      0.89     18871
           6       0.85      0.60      0.71     37000
           7       0.49      0.46      0.47      3496
           8       0.00      0.00      0.00       378
           9       0.00      0.00      0.00        44
```



9. CONCLUSION

We are investigating the feasibility of using learning ensembles with single model classifiers to detect cyber attacks in IOT based smart city applications in this research. Our ensemble strategy, particularly the stacking strategy, has been shown to be superior to single models in detecting attacks from benign samples in experiments with the most recent IOT attack datasets. In terms of accuracy, precision, recall and F1 score metrics, our combined approach with a stack performs better than any single or other combination model to identify the attack types. In order to detect IOT attacks, we will be looking at Deep Learning techniques for the next research.

10. REFERENCES

1) Chowdhury A, G. Karmakar and J. Kamruzzaman, "The Co-Evolution of Cloud and IoT Applications: Recent and Future Trends", Handbook of Research on the IoT Cloud Computing and Wireless Network Optimization, pp. 213-234, 2019.
2) J. Howell, "Number of connected iot devices will surge to 125 billion by 2030", ihsmarkit says - ihs technology, 06 2020, [online] Available:
3) M.-O. Pahl, F.-X. Aubet and S. Liebald, "Graph-based IoTmicroservice security", Proceedings of the NOMS 2018–2018 IEEE/IFIP Network Operations and Management Symposium, pp. 1-320, 2018.
4) F. Restuccia, S. D'Oro and T. Melodia, "Securing the internet of things: New perspectives and research challenges", Journal of IEEE Internet of Things Journal, vol. 1, no. 1, pp. 1-14, 2018.
5) M Yar and K.F. teinmetz, Cybercrime and Society, Thousand Oaks, CA, USA:SAGE Publications Limited, 2019.
6) M Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric et al., "Understanding the mirai botnet", USENIX Security Symposium, pp. 1092-1110, 2017.