International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Detection and Prevention of Brute Force Attacks Using Machine Learning

P Adithya Vardhan Reddy[1], P Yoganandha Reddy[2], S Nivetha[3]

[1,2]Bachelor's Student, CSE AI ML, Sathyabama University
[3]Assistant Professor, CSE AI ML, Sathyabama University

Abstract
With the development in remote correspondence, there are numerous security dangers over the web. The interruption discovery framework (IDS) assists with tracking down the assaults on the framework and the interlopers are identified. Already different AI (ML) procedures are applied on the IDS and attempted to work on the outcomes on the discovery of intruders and to improve the IDS's accuracy. The paper has suggested method for using the principal component analysis to create effective IDS (PCA) as well as the random forest classification method Where the PCA can assist arrange the dataset by lessening the dimensionality of the dataset and the arbitrary forest will make classification easier. According to the obtained results, the proposed method performs more effectively and accurately than other methods such as Decision Tree, Naive Bayes, and SVM. The outcomes acquired by proposed strategy are having the qualities for execution time (min) is 3.24 minutes, Exactness rate (%) is 96.78 %, and the Mistake rate (%) is 0.21 %.

Keywords: Intrusion, Detection, Machine Learning, PCA

1. Introduction

Hackers try to take advantage of or hack laptop devices. The act of jeopardizing the availability, confidentiality, or integrity of data or computer assets is called intrusion. Attackers attempt to circumvent authentication or authorization procedures by taking advantage of device architecture defects or capacity limitations. Network security is more important than ever because of the growth of community services and the information they contain. Using a Network Intrusion Detection System (NIDS) is one way to deal with this inconvenience. NIDS keeps an eye on attacks and finds a lot of network sports. Consequently, it is crucial to detect an assault in such buildings; this is highly accurate, examines quickly, and has the fewest false positives possible. By identifying harmful intrusions, intrusion detection systems (IDS) protect networks. Consequently, IDS has emerged as a significant issue in computer networks. Two requirements for an IDS are alertness and agility. Safety is paramount in all efforts to prevent loss. An IDS's primary duties include alerting network directors, blocking questionable connections and echo/monitoring, and providing documentation of unusual activities. Furthermore, an intrusion detection system (IDS) can differentiate between attacks that originate externally (by using hackers) and internally (from certain workers, users, or other assets). Intrusion detection systems (IDS) come in two common varieties: host-based (HIDS) and network (Network IDS). The foundation of network intrusion detection systems is the ability to identify unlawful, irregular, and Unlawful behaviour among community visitors.

2. Literature Survey

2.1 Feature Selection and Dimensionality Reduction

Principal Component Analysis (PCA) has been widely used to reduce dataset dimensionality while retaining essential features. Studies show that PCA helps improve the efficiency of Intrusion Detection Systems (IDS) by eliminating redundant data and focusing on critical attributes for attack detection.

2.2 Machine Learning-Based Intrusion Detection Systems

Several machine learning models have been proposed to detect brute force attacks effectively:

- Random Forest (RF): Known for its high accuracy and robustness in classification problems.
- Decision Tree (DT): Simple and interpretable but prone to overfitting.
- Naïve Bayes (NB): Works well with probabilistic models but has lower accuracy for complex attack patterns.

2.3 IP Reputation and Geolocation-Based Restrictions

Using machine learning to analyze IP reputation and block suspicious activities from high-risk regions can mitigate brute force attempts.

3. Existing System

Iftekhar Ahmed et al investigated diverse gadget mastering calculations for interruption identification framework. They as looked at various procedures which incorporate SVM, Outrageous Learning Machine and Arbitrary Backwoods. The creators of the results proposed that the concentrated device concentrating on strategy achieved very well when contrasted with different calculations. B. Rias et al., worked here to work on the best of realities feed in an interruption location framework. They utilized a standard fundamentally based choice component to upgrade the statistics set. They used the KDD dataset and, as an end result, confirmed a dynamic increase in IDS outcomes.

Disadvantages of this system is as mentioned below:

Internet working systems are at risk of diverse malicious activities. The predominant hassle to be addressed on this regard is the penetration of the data dispersal system. The present effects suggest that some improvements may be made in phrases of accuracy, detection price and fake alarm. Other strategies may be changed through some previously used methods inclusive of SVM and Naive Bayes. In addition, the study indicates that some of the techniques inside the dataset might be improved. Increase the exceptional of input within the proposed system.

4. System Modules

4.1 Data Collection

This is the primary actual step to studying actual engine development. Sampling, statistics series. This is a crucial step and relies upon on how accurate it is the better the model, the better the information, the better our model can be to finish. There are many facts collection methods like internet scraping, guide facts series. Intervention, etc. This intrusion detection system dataset is taken from the dataset kdd. Link: http://kdd.Ics.Uci.Edu/databases/kddcup99/kddcup99. Html.

4.2 Dataset and Preparation

Let's share the information. by getting rid of several columns and missing records. Let's start by making a list of the column names that we must store or shop for.

After that, we remove or delete every column except the ones we must keep.

Lastly, we remove from the dataset any rows that have missing values.

Training and assessment are not the same thing.

4.3 Model Selection

One method for lowering the dimensionality of a statistics collection is principal issue evaluation. One of the most accurate and environmentally friendly methods for reducing the dimensionality of data and the desired results is principal factor assessment.

By using this technique, the measures' capacities are reduced to an optimally broad range of attributes, referred to as directors. The elements of the dataset may be extraordinarily large because this process uses all of the input reality as a dataset, which includes a vast number of characteristics. By placing the measurement factors on the same pivot, this method reduces the amount of information. Information factors become essential components and are converted to an axis. ATP is able to

This may be fulfilled using those way :

- Take the given dimension with all confines d.
- Add the suggest vector for each dimension d.
- Add the covariance matrix for the entire data set.
- Add the eigenvectors(e1, e2, e three. Ed) and eigenvalues(v1, v2, v3, Vd).
- Sort the eigenvalues in descending order and pick out the use of n eigenvectors
- Sum the eigenvalues to get the matrix d * n = M.
- Use this M to produce a brand new model area.
- The performing durations are top.

4.3 Saving the trained model

If you are assured that you could take the template prepared for a manufacturing surroundings, the first step is to save it to a .H5 or .Pkl document the use of the .H5 or .Pkl library. Make positive ALEX is mounted for your environment. Then we import the module and replica it to a .Pkl file.

5. Methodology

5.1 System Architecture

There are two primary components to the suggested architecture:

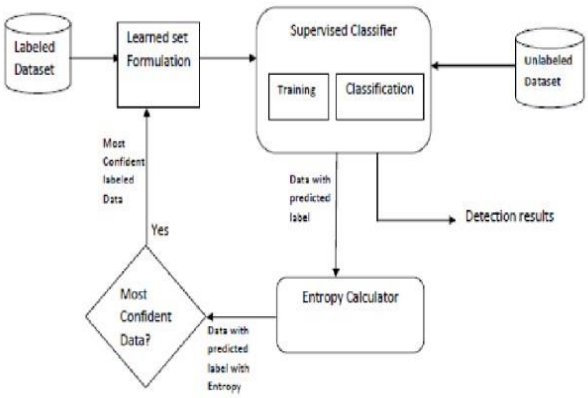

Figure 1 System Architecture

5.1.1 Random Forest

Irregular lush region is a well-known gadget acquiring information on set of decides that has a place with managed acquiring information on strategy. It tends to be utilized for both sort and relapse commitments in ML. It is principally based at the idea of troupe getting to be aware, that is a multi-type technique for

fixing an issue. A more perplexing difficulty and better acting model. An irregular lush region calculation comprises of different choice trees. The backwoods is created by means of irregular lush region set of rules utilizing bootstrap binning or bunching. Pressing it an outfit calculation is a metacalculation that works on the precision of framework learning calculations. As the name indicates, "Random Forest is a classifier that has many selection trees in different sets. Given a facts set, it takes an algorithm to enhance the predictive accuracy of that statistics set. Instead of relying on a unmarried decision tree, a random woodland takes one prediction from every tree and relies on numerous. He foretells the voices and foretells the final occasion.

5.1.2 Principal Component Analysis

Principal thing evaluation is a fuzzy learning algorithm used to reduce dimensionality in gadget getting to know. It is a statistical system that transforms observations of correlated features into a hard and fast of complex features using an orthogonal transformation. These new transformed features are called foremost additives. It is one of the maximum popular tools used for exploratory evaluation and predictive modelling. It is a way of extracting valid styles from information with the aid of minimizing variables. PCA works via searching on the variance of every characteristic due to the fact a high attribute indicates a terrific separation among instructions, for that reason lowering dimensionality. Some real packages of PCA are picture processing, film advice gadget, optimized electricity distribution over various conversation channels.

6. Conclusion

In conclusion, we've used system learning techniques and community simulation using NS2 to remedy the important problem of intrusion detection in dispensed IoT networks. Our proposed method successfully detects intrusions and outperforms conventional algorithms together with SVM, naive base, and decision timber. Our method using the Knowledge Discovery dataset showed a processing time of 3.24 minutes, an accuracy of 96.78%, and a blunders price of 0.21%. These results highlight the robustness and efficiency of our approach in detecting intrusions in complex IoT environments. By integrating ML and NS2 simulation, we've got created a reliable and scalable intrusion detection system that complements the safety of IoT networks. Our findings spotlight the potential of advanced device studying strategies blended with community simulation to deal with the rising security challenges of IoT networks.

References

1. Tesfahun A., Bhaskari D., "Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction," 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, IEEE, 2013, 978-0-4799-2235-2/13.
2. Park K., Song Y., Cheong Y., "Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm," 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (Big Data Service), 2018.
3. Le T.-T.-H., Kang H., Kim H., "The Impact of PCA-Scale Improving GRU Performance for Intrusion Detection," 2019 International Conference on Platform Technology and Service (PlatCon), IEEE, 2019, DOI:10.1109/platcon.2019.8668960.
4. Ahmad I., Basheri M., Iqbal M. J., Rahim A., "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," IEEE Access, 2018, 6,

33789–33795.

5. Ge M., Fu X., Syed N., Baig Z., Teo G., Robles-Kelly A., "Deep Learning-Based Intrusion Detection for IoT Networks," 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC), 2019, pp. 256–265, Japan.