Semi-Supervised Machine Learning Approach For DDOS Detection

GUDIE:
Mrs. R. Jagadeeswari
CSE
BIHER

TEAM MEMBER-1:
A.Yashwanth Reddy
CSE
BIHER

TEAM MEMBER-2:
Ch. Bhargava Siva Kumar Reddy
CSE
BIHER

TEAM MEMBER-3:
D. Dinesh Goud
CSE
BIHER

TEAM MEMBER-4:
R. Mani Raghavendra
CSE
BIHER

ABSTRACT

The fast propagation of computer networks has changed the viewpoint of network security. An easy access condition causes computer networks to be susceptible to several threats from hackers.

Threats to networks are numerous and potentially devastating. Up until now, researchers have developed intrusion detection systems (IDS) capable of detecting attacks in several available environments. A boundless number of methods for misuse detection as well as anomaly detection have been applied.

Many of the technologies proposed are complementary to each other since, for different kinds of environments, some approaches perform better than others. This project presents new intrusion detection systems that are then used to survey and classify them. The taxonomy consists of the detection principle and certain operational aspects of the intrusion detection system. In our project, we have used algorithms like Nave Bayes (NB) and Random Forest (RF). All are measured in terms of accuracy.

OBJECTIVE

- To minimize data loss.
- More throughputs.
- To reduce time consumption.
- Continuous energy checkup of all data to avoid communication failure.
- To find the intruder at the early stage.

INTRODUCTION

In the statistical context, Machine Learning is defined as an application of artificial intelligence where available information is used through algorithms to process or assist the processing of statistical data. While Machine Learning involves concepts of automation, it requires human guidance. Machine Learning involves a high level of generalization in order to get a system that performs well on yet unseen data instances. Machine learning is a relativelynew discipline within Computer Science that provides a collection of data analysis techniques. Some of these techniques are based on well established statistical methods (e.g. logistic regression and principal component analysis) while manyothers are not.

The fast propagation of computer networks has changed the viewpoint of network security. An easy accessibility conditions cause computer network as susceptible against several threats from hackers.
Threats to networks are numerous andpotentially devastating.

In the statistical context, machine learning is defined as an application of artificial intelligence where available information is used through algorithms to process or assist the processing of statistical data. While machine learning involves concepts of automation, it requires human guidance. Machine learning involves a high level of generalisation in order to get a system that performs well on yet unseen data instances. Machine learning is a relatively new discipline within computer science that provides a collection of data analysis techniques. Some of these techniques are based on well-established statistical methods (e.g., logistic regression and principal component analysis), while many others are not.

The fast propagation of computer networks has changed the viewpoint of network security. An easy access condition causes computer networks to be susceptible to several threats from hackers.

Threats to networks are numerous and potentially devastating. Up until now, researchers have developed intrusion detection systems (IDS) capable of detecting attacks in several available environments. A boundless number of methods for misuse detection as well as anomaly detection have been applied. Many of the technologies proposed are complementary to each other since, for different kinds of environments, some approaches perform better than others. This project presents a new

intrusion detection systems that are then used to survey and classify them. The taxonomy consists of the detection principle and certain operational aspects of the intrusion detection system.

In our project, we have used algorithms like Dismissal of Conquered Movements,

LITERATURE REVIEW

NL-IDS: Trust Based Intrusion Detection System for Network layer in Wireless Sensor Networks Umashankar Ghugar, Jayaram Pradhan IEEE 2021.

- we have proposed a trustbased intrusion detection system (NL-IDS)for network layer in WSN to detect the Black hole attackers in the network. The sensor node trust is calculated as per the deviation of keyfactor at the network layer based on the Black hole attack.

- We use the watchdog technique where a sensor node continuouslymonitors the neighbor node by calculating a periodic trust value.
- Finally, the overall trust value of the sensor node is evaluated by the gathered values of trust metrics of the network layer (past and previous trustvalues).

This NL-IDS scheme is efficient toidentify the malicious node with respect to Black hole attack at the network layer.

Analyzing the Vulnerability of Wireless Sensor Networks to a Malicious Matched Protocol Attack George D. O'Mahon, Philip J. Harris,Colin C. Murphy IEEE 2021.

Lexicographic Game Method,Collaborative Game Method, Repeated Game Approach, Stochastic Random Process, Petri Net Process, Artificial Neural Network, and Convolutional Neural Network, respectively.

- In this paper, a specific vulnerability of WSNs is explored, termed here thematched protocol attack.
- This malicious attack uses protocol-specific structures to compromise anetwork using that protocol.
- Through attack exploration, this paper provides evidence that traditional spectral techniques are notsufficient to detect an intrusion usingthis style of attack.
- Furthermore, a ZigBee cluster head network, which co-exists with ISM band services, consisting of XBee COTS devices is utilized, along with a real time spectrum analyzer, to experimentally evaluate the effect ofmatched protocol interference on a realistic network model.
- Results of this evaluation are provided in terms of device errorsand spectrum use.

Secure Knowledge and Cluster-Based Intrusion Detection Mechanism for Smart Wireless Sensor Networks Houbing Song IEEE 2021.

- This paper proposes a knowledge- based context-aware approach for handling the intrusions generated bymalicious nodes.
- The system operates on a knowledge base, located at the base station, whichis used to store the events generated bythe nodes inside the network.
- The events are categorized and the cluster heads (CHs) are acknowledged to block maliciously repeated activities generated.
- The CHs can also get informational records about the maliciousness of intruder nodes by using their inference engines.
- The mechanism of events logging and analysis by the base station greatly affects the performance of nodes in the network by reducing the extra security-related load on them.

A Light-Weight Countermeasure to Forwarding Misbehavior in Wireless Sensor Networks: Design, Analysis, andEvaluation Cong Pu, Sunho Lim IEEE 2021.

- In this paper, we propose a light- weight countermeasure to a selective forwarding attack, called SCAD, where a randomly selected single checkpoint node is deployed to detect the forwarding misbehavior of malicious node.
- The proposed countermeasure is integrated with timeout and hop-by-hop retransmission techniques to quickly recover unexpected packet losses due to the forwarding misbehavior or bad channel quality.
- We also present a simple analytical model and its numerical result in termsof false detection rate.

Energy Efficient Detection-Removal Algorithm for Selective ForwardingAttack In Wireless Sensor Networks
T.R Sreelakshmi, G.S Binu IEEE 2021.

- Network layer attacks are more severesince if the routing information is

disregarded, disturbances may bring about routing loops, changing of routesetc.

- Selective forwarding attack is a type ofactive attack affecting network layers that selectively drops or refuses to forward the data packets.
- This paper discusses about an energy efficient detection-removal algorithm for effective detection of selective forwarding attack in a clustered WSNscenario.
- The impact of the malicious node innetwork parameters like packet delivery ratio, throughput, residual energy of network and end to end delay are analyzed.

EXISTING SYSTEM

Naive Bayes (NB):

In statistics, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong

(Nave) independence assumptions exist between the features.

They are among the simplest Bayesian network models, but coupled With kernel density estimating, they can achieve higher accuracy levels.

Nave Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features or predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation, as used for many other types of classifiers.

In the statistics and computer science literature, naive Bayes models are known under a variety of names, including simple

Bayes and independence Bayes. All these names refer to the use of Bayes' theorem in the classifier's decision rule, but Nave Bayes is not (necessarily) a Bayesian method.

DISADVANTAGE OFEXISTING SYSTEM

- Less Accuracy
- Attack can happen at any time
- More timeconsuming process
- Higher Computational Cost
- Lack of standards
- Cannot be implemented in alldatasets

PROBLEM STATEMENT

- Not Economical
- More timeconsuming process.
- Early prediction of attack is notpossible.
- Less accuracy rate
- Cannot be implemented in alldatasets.

PROPOSED SYSTEM

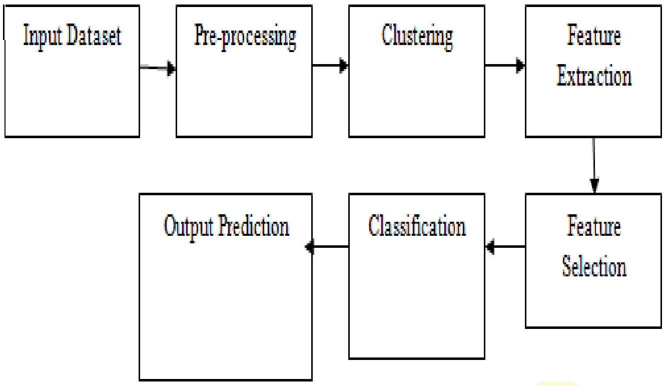
Random Forest:

- Random forests or random decision forests are an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual trees.
- Random decision forests correct for decision trees' habit of overfitting their training.
- Random forests generally outperform decision trees, but their accuracy is lower than that of gradient-boosted trees. However, data characteristics can affect.

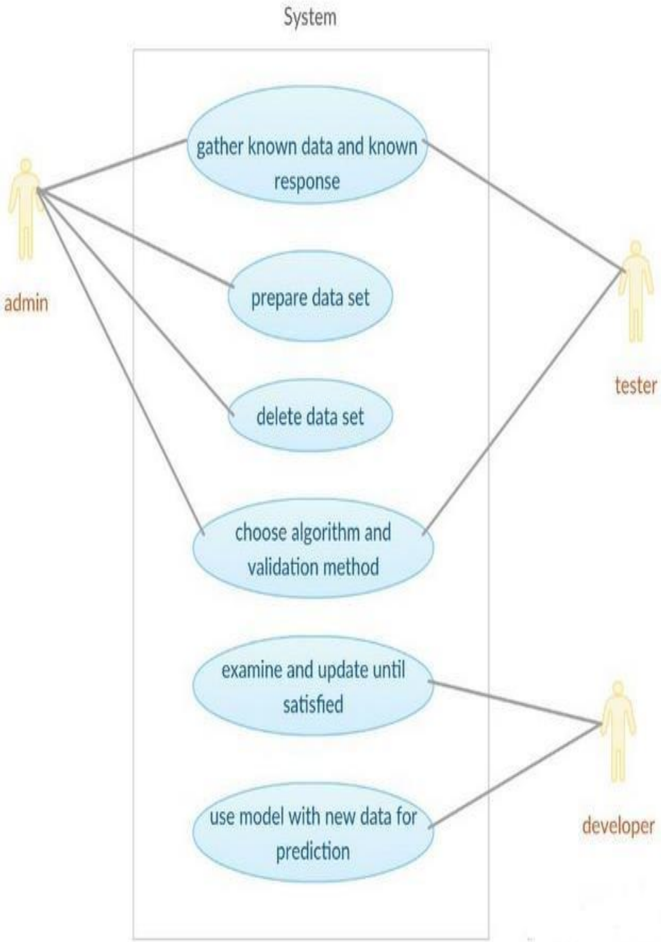
PROPOSED SYSTEMADVANTAGES

- High output efficiency.
- User friendly.
- Less time consumption.
- Can be implemented in all datasets.
- Early prediction of attack is possible

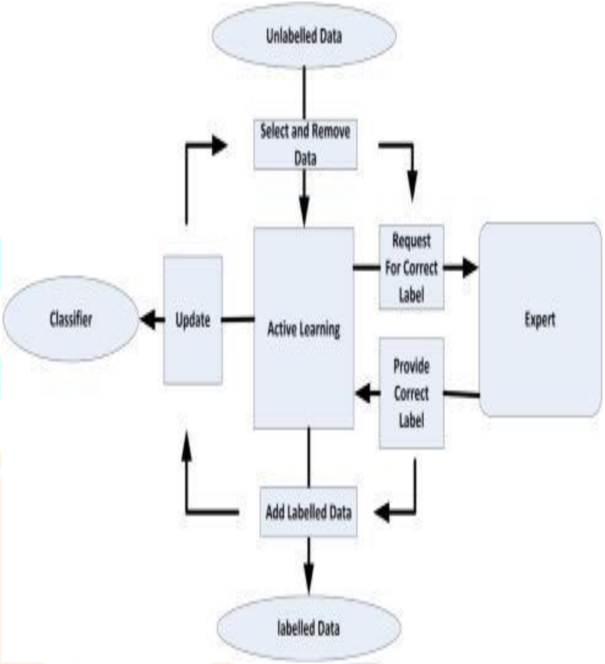
ARCHITECTURE DIAGRAM



USECASE DIAGRAM


System
gather known data and known response
prepare data set
delete data set
choose algorithm and validation method
examine and update until satisfied
use model with new data for prediction
admin
tester
developer

UML DIAGRAM



MODULES DESCRIPTION

- **Incoming packet:** In our project, we are directly connected to the network (we capture the online packet) and transfer the packet
- **Packet Capture:** In real-time networking, many packets are transferred from source to destination.
  - In this module, live packets from the network are captured and passed to the pre-processor.
  - According to the concept of machine learning, the data is grouped or clustered based on its features or characteristics, e.g., the protocols used by the packets.
- **Packet scanner:** After catching the packet, we use a packet scanner for the purpose of scanning the packet. Packet scanning is an important part of our
- **Packet analyzer: A packet** analyzer is also known as a packet sniffer. As data streams flow across the network, sniffers capture each packet and, if needed, decode it, producing raw data showing the values of various fields in the packet and analysing its content according.
- **Labelling:** Labelling is used for defining the corresponding
- **Training model: The training** model contains a set of trained data sets that are used to detect attacks in the
- **Prediction:** Based on the training model, we make the prediction that the packet is a normal packet or abnormal.
- **Output:** Based on the prediction (i.e., normal or abnormal), we generate output in the form of an abnormal packet.
-

HARDWARE REQUIREMENTS

- Processor: Core I5Processor.
- Ram: 4 GB RAM
- Hard Disk: 500 G.B HardDisk
- 14inch monitor

SOFTWARE REQUIREMENTS

- Technology: Python
- IDE: Python IDE
- Web Server Jupiter/Anaconda/Panda
- Database: My SQL

CONCLUSION

Intrusion detection is currently attracting interest from both the research community and commercial companies. We have given background on the current state-of-the-art of IDS, based on a proposed taxonomy illustrated with examples of past and current projects. This

Taxonomy also highlights the recent work and covers the past and current developments adequately. Each of its techniques has its own advantages and disadvantages.

We believe that no single criterion can be used to completely defend against computer network intrusion. There is no single version of it that can be used as a standard solution against all possible attacks. It is both technically difficult and economically costly to build and maintain computer systems and networks that are not susceptible to attacks. The technique to be selected depends on the specifications of the types of anomalies that the system is supposed to face, the type and behaviour of the data, the environment in which the system is working, the cost and computation limitations, and the security level required.

REFERENCES

[1] I. F. Akyildiz et al., “Wireless Sensor Networks: A Survey, “Elsevier Comp. Networks, vol. 3, no. 2, 2019, pp. 393–422

[2] G.Li, J.He, Y. Fu. “Group-based intrusion detection system in wireless sensor networks” Computer Communications, Volume 31, Issue 18(December 2019)

[3] Michael Brownfield, “Wireless Sensor Network Denial of Sleep Attack”, Proceedings of the 2019 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY.

[4] FarooqAnjum,DhanantSubhadrabandhu, SaswatiSarkar *, Rahul Shetty, “On Optimal Placement of Intrusion Detection Modules in Sensor Networks”, Proceedings of the First International Conference on Broadband Networks (BROADNETS19).

[5] Parveen Sadotra et al, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.9, September- 20