IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 7, July 2022

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107145 |

Intrusion Attacks Detection and Alerting System using Machine Learning

Prajwalgouda Patil[1], Rajesh S[2], Sanjay Y [3], Vinay P M [4], Prof. Kamala R$_{MTech}$ [5]

B.E. Student, Department of Information Science & Engineering, Bapuji Institute of Engineering & Technology, Davangere, Bangalore, Karnataka, India [1]

B.E. Student, Department of Information Science & Engineering, Bapuji Institute of Engineering & Technology, Davangere, Bangalore, Karnataka, India [2].

B.E. Student, Department of Information Science & Engineering, Bapuji Institute of Engineering & Technology, Davangere, Bangalore, Karnataka, India [3].

B.E. Student, Department of Information Science & Engineering, Bapuji Institute of Engineering & Technology, Davangere, Bangalore, Karnataka, India [4].

Assistant Professor, Department of Information Science & Engineering, Bapuji Institute of Engineering & Technology, Davangere, Bangalore, Karnataka, India [5].

ABSTRACT: Cyber-crime is proliferating everywhere exploiting every kind of vulnerability to computing environment. Ethical Hackers pay more attention towards assessing vulnerabilities and recommending mitigation methodologies. The development of effective techniques has been an urgent demand in the field of the cyber security community. Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks.Machine Learning algorithms can be used to train and detect if there has been a cyberattack. As soon as the attack is detected, an email notification can be sent to the security engineers or others. Any classification algorithm can be used to categorize different types of attack. Our main goal is that the task of finding attacks is fundamentally different from these other applications, making it significantly harder for the intrusion detection community to employ machine learning effectively.

.

KEYWORDS: Random Forest Algorithm, Logistic regression, Decision tree, Support vector machine.

I. INTRODUCTION

Intrusion detection is the process of dynamically monitoring events occurring in a computer system or network, analysing them for signs of possible incidents and often interdicting the unauthorized access. This is typically accomplished by automatically collecting information from a variety of systems and network sources, and then analysing the information for possible security problems.Traditional intrusion detection and prevention techniques, like firewalls, access control mechanisms, and encryptions, have several limitations in fully protecting networks and systems from increasingly sophisticated attacks like denial of service. Moreover, most systems built based on such techniques suffer from high false positive and false negative detection rates and the lack of continuously adapting to changing malicious behaviours. In the past decade, however, several Machine Learning (ML) techniques have been applied to the problem of intrusion detection with the hope of improving detection rates and adaptability. These techniques are often used to keep the attack knowledge bases up-to-date and comprehensive.In recent days, cyber-security and protection against numerous cyber-attacksare becoming a burning question. The main reason behind that is the tremendous growth of computer networks and the vast number of relevant applications used by individuals or groups for either personal or commercial use, especially after the acceptance of the Internet of Things (IoT). The cyber-attacks cause severe damage and severe financial losses in large-scale networks.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107145 |

The existing solutions like hardware and software firewalls, user's authentication, and data encryption methods are not sufficient to meet the challenge of upcoming demand, and unfortunately, not able to protect the computer network's several cyber-threats. These conventional security structures are not sufficient as safeguard due to the faster rigorous evolution of intrusion systems. Firewall only controls every access from network to network, which means prevent access between networks. But it does not provide any signal in case of an internal attack. So, it is obvious to develop accurate defence techniques such as machine learning-based intrusion detection system (IDS) for the system's security.

In general, an intrusion detection system (IDS) is a system or software that detects infectious activities and violations of policy in a network or system. An IDS identifies the inconsistencies and abnormal behaviour on a network during the functioning of daily activities in a network or system used to detect risks or attacks related to network security, like denial-of-service (DoS). An intrusion detection system also helps to locate, decide, and control unauthorized system behaviour such as unauthorized access, or modification and destruction. There are different types of intrusion detection systems based on the user perspective. For instance, they are host-based and network-based IDS.

II. LITERATURE REVIEW

In this paper, we employ various popular machine learning classification algorithms, namely Bayesian Network, Naive Bayes classifier, Decision Tree, Random Decision Forest, Random Tree, Decision Table, and Artificial Neural Network, to detect intrusions due to provide intelligent services in the domain of cyber-security. Finally, we test the effectiveness of various experiments on cyber-security datasets having several categories of cyber-attacks and evaluate the effectiveness of the performance metrics, precision, recall, f1-score, and accuracy.In general, an intrusion detection system (IDS) is a system or software that detects infectious activities and violations of policy in a network or system. An IDS identifies the inconsistencies and abnormal behaviour on a network during the functioning of daily activities in a network or system used to detect risks or attacks related to network security, like denial-of-service (DoS). An intrusion detection system also helps to locate, decide, and control unauthorized system behaviour such as unauthorized access, or modification and destruction. There are different types of intrusion detection systems based on the user perspective. For instance, they are host-based and network-based IDS.

These are in the scope of single computers to large networks to some extent. In a host-based intrusion detection system (HIDS), it lies on an individual system and keeps track of operating system files for inconsistency and abnormalities in the activity. In contrast, the network intrusion detection system (NIDS) investigates and scans connections in the network for unwanted traffic. On the other hand, there are two approaches based on detection, one is signature-based, and another one is anomaly-based detection. Signature-based IDS explores the byte patterns in the path of the network. One can treat it as malicious instruction sequences used by malware. It arises from antivirus software referred to the groups or patterns as signatures detected in it. Signature-based IDS cannot detect attacks, for which there is no pattern available before. An anomaly-based IDS, it examines the behaviour of the network and finds patterns, automatically creates a data-driven model for profiling the expected behaviour, and thus detects deviations in the case of any anomalies. The merit of this anomaly-based IDS is to trace current, latest, and unseen inconsistencies or cyber-attacks like denial-of-services.

III. PROBLEM STATEMENT

Within the ever-growing and quickly increasing field of cyber security, it is nearly impossible to quantify or justify the explanations why cyber security has such an outsized impact. We are able to detect behaviour that is not normal for average network usage. While it's good to be able to detect normal network usage, the disadvantage is that the system can create large number of false alarms.

PROPOSED SOLUTION

Machine Learning algorithms can be used to train and detect if there has been a cyber- attack. As soon as the attack is detected, an email notification can be sent to the security engineers or users.The classification algorithm can be used to categorize if it is adenial-of-service attack or not. Example algorithm is Support Vector Machine which is a supervised learning method that analysis data and recognizes pattern our best part is early detection which will help mitigate the risk of repairable damage.

International Journal of Innovative Research in Science, Engineering and Technology

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107145 |

SYSTEM DESIGN

- The System Design consists of

1. Initially the network packets will be captured from the Organisational network within firewall.
2. Then the captured packets are converted into a file with formats like win cap, Jpcap.
3. Then the features of the file will be extracted.
4. Finely our machine learning model which is trained with the Attack packets will analyse these data and estimates the Attack packets based on Attack vectors.
5. Then this report will be automatically mailed to the Network Administrators and also a telegram notification will be sent.

IV. METHODOLOGY

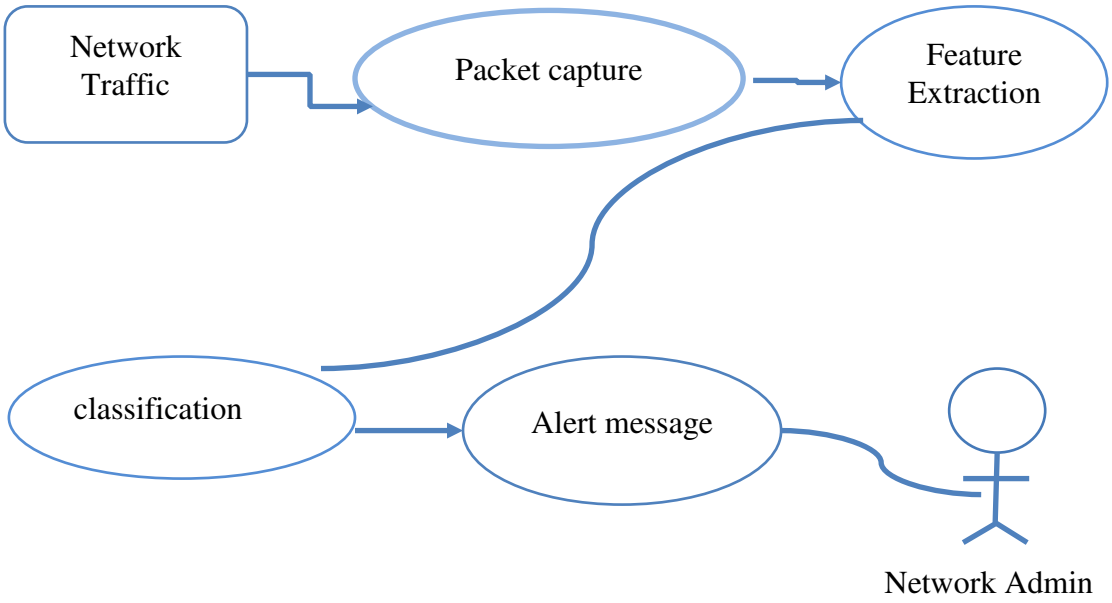

Fig 1: Data Flow Diagram

1. A data flow diagram (DFD) is a graphic representation of the "flow" of data through an information system. A data flow diagram can also be used for the visualization of data processing (structured design). It is common practice for a designer to draw a context level DFD first which shows the interaction between the system and outside entities. DFD's show the flow of data from external entities into the system, how the data moves from one process to another, as well as its logical storage.

2. There are only four symbols:Squares representing external entities, which are sources and destinations of information entering and leaving the system.

3. Rounded rectangles representing processes, in other methodologies, may be called 'Activities', 'Actions', 'Procedures', 'Subsystems' etc. which take data as input, do processing to it, and output it.

4. Arrows representing the data flows, which can either, be electronic data or physical items. It is impossible for data to flow from data store to data store except via a process, and external entities are not allowed to access data stores directly. The flat three-sided rectangle is representing data stores should both receive information for storing and provide it for further processing.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107145 |

PSECUDO CODE

1. Decision Tree

```
from sklearn.tree import DecisionTreeClassifier

model2 = DecisionTreeClassifier(criterion="entropy", max_depth = 4)

start_time = time.time()
model2.fit(X_train, Y_train.values.ravel())
end_time = time.time()

print("Training time: ",end_time-start_time)

Training time:  0.9031133651733398
start_time = time.time()
Y_test_pred2 = model2.predict(X_test)
end_time = time.time()

print("Testing time: ",end_time-start_time)

Testing time:  0.030696392059326172
print("Train score is:", model2.score(X_train, Y_train))
print("Test score is:",model2.score(X_test,Y_test))

Train score is: 0.9905829108684749
Test score is: 0.9905230421954646
```

2.Random Forest

```
from sklearn.ensemble import RandomForestClassifier

model3 = RandomForestClassifier(n_estimators=30)

start_time = time.time()
model3.fit(X_train, Y_train.values.ravel())
end_time = time.time()

print("Training time: ",end_time-start_time)

Training time:  6.59112811088562
start_time = time.time()
Y_test_pred3 = model3.predict(X_test)
end_time = time.time()

print("Testing time: ",end_time-start_time)

Testing time:  0.6176023483276367
print("Train score is:", model3.score(X_train, Y_train))
print("Test score is:",model3.score(X_test,Y_test))

Train score is: 0.9999728091747887
Test score is: 0.9996749004766081
```

3.Random Forest Algorithm

Step 1: In random forest N number of random records are taken from the dataset having K number of records.
Step 2: Individual decision tress are constructed for each sample.
Step 3: Each decision tree will generate an output.
Step 4: Final output is considered based on majority voting for classification and regression respectively.

International Journal of Innovative Research in Science, Engineering and Technology

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107145 |

V. RESULTS AND DISCUSSION

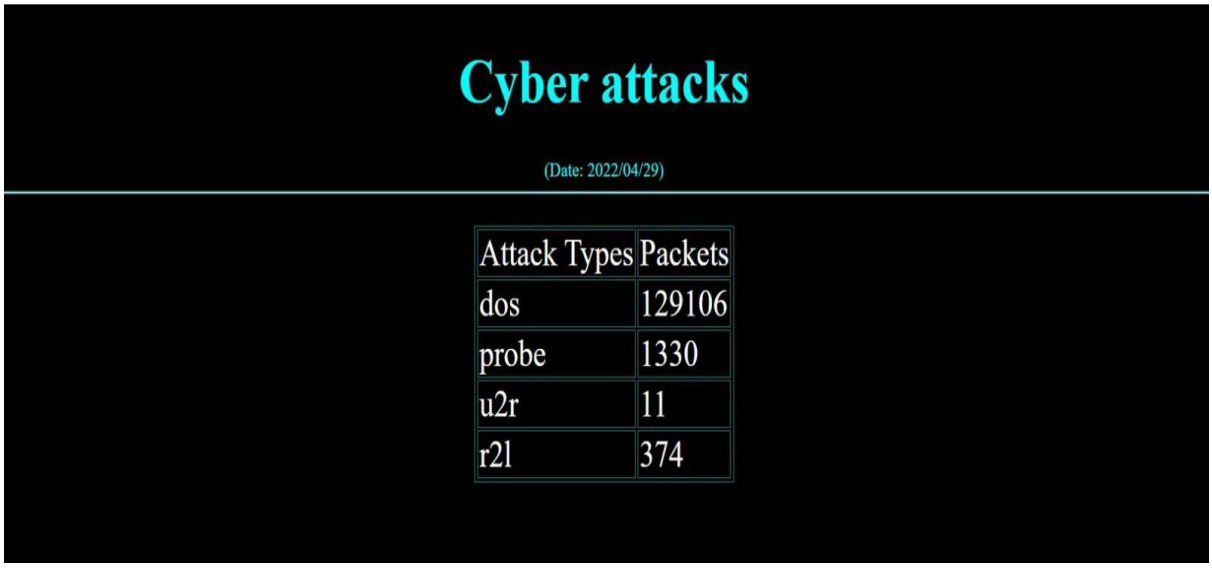

Fig 2: Cyber Attack

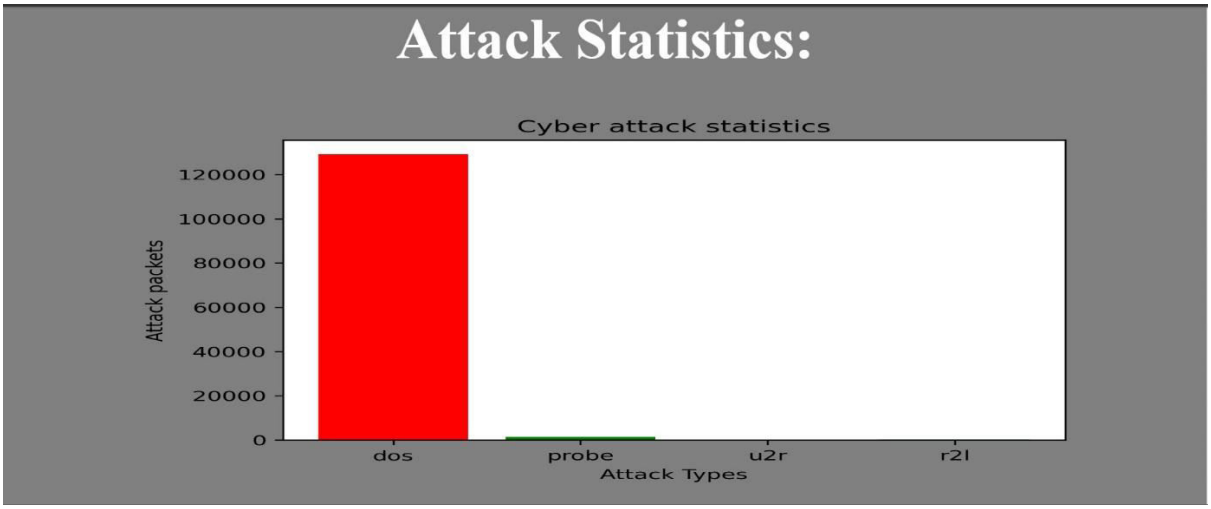

Fig 3: Attack Statistics

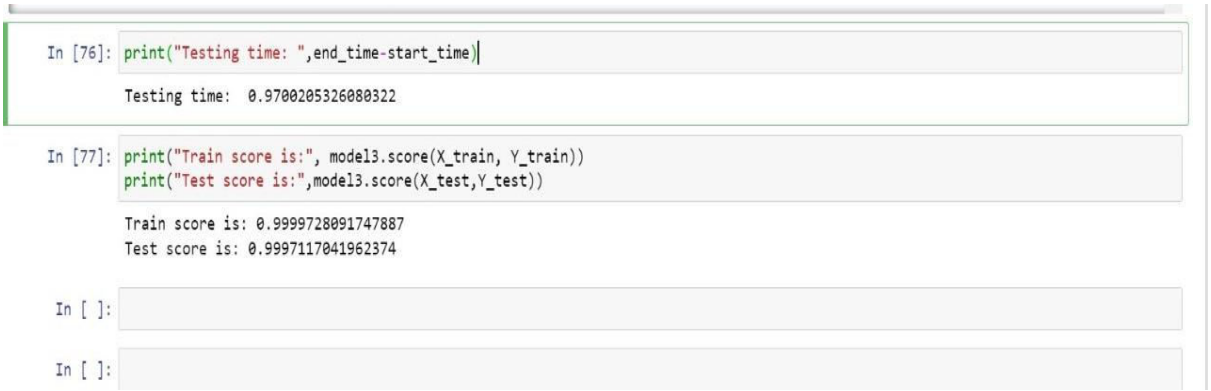

Fig 4: Accuracy of Random Forest

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 7, July 2022||

| DOI:10.15680/IJIRSET.2022.1107145 |

VI. CONCLUSION

` Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Hence, efficient adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. We reviewed several influential algorithms for intrusion detection based on various machine learning techniques. Characteristics of ML techniques makes it possible to design IDS that have high detection rates and low false positive rates while the system quickly adapts itself to changing malicious behaviour. IDS using many Machine Learning Techniques like Random Forest, Decision tree and logistic regression to perform better in various metrics. The IDS should provide the most effective solutions based on the requirements. One thing is sure, any company failing to adopt these techniques now or in the immediate future risk compromising data.

REFERENCES

[1] *Dipankar Dasgupta. Immunity-based intrusion detection system*: A general frame-work. In Proceedings of the22nd National Information Systems Security Conference (NISSC). Arlington, Virginia, USA, 1999.
[2] *Jonatan Gomez and Dipankar Dasgupta*. Evolving fuzzy classier for intrusiondetection. In Proceedings of the 2002 IEEE Workshop on Information Assurance,West Point, NY, USA, 2002.
[3*] Steven A. Hofmeyr, Stephanie Forrest, and Anil Somayaji*. Intrusion detectionusing sequences of system calls. Journal of Computer Security, 6(3):151{180,August 1998.
[4] *Peter Mell Karen Scarfone*. Guide to intrusion detection and prevention systems(IDPS). National Institute of Standards and Technology, NIST SP - 800-94, 2007.
[5] *Jungwon Kim, Peter J. Bentley, Uwe Aickelin, Julie Greensmith, Gianni Tedesco,and Jamie Twycross*. Immune system approaches to intrusion detection a review.Natural Computing, 6(4):413{466, December 2007}.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details