TE-A

Group No: - 14

Title of Project: Network Intrusion Detection System using Machine Learning Technique.

Adrian Dsouza A-630

Vedant Lanjewar A-660

Abhishek Mahakal A-663

1. Technical Papers.

- L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446–452, 2015.
- Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection
- MACHINE M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016.

2. Analysis

In order to examine malicious activity that occurs in a network or a system, intrusion detection system is used. Intrusion Detection is software or a device that scans a system or a network for a distrustful activity. Due to the growing connectivity between computers, intrusion detection becomes vital to perform network security. Various machine learning techniques and statistical methodologies have been used to build different types of Intrusion Detection Systems to protect the networks. Performance of an Intrusion Detection mainly depends on accuracy. Accuracy for Intrusion detection must be enhanced to reduce false alarms and to increase the detection rate. In order to improve the performance, different techniques have been used in recent works. Analyzing huge network traffic data is the main work of intrusion detection system. A well-organized classification methodology is required to overcome this issue. This issue is taken in proposed approach. Machine learning techniques like Support Vector Machine (SVM) and Naïve Bayes are applied. These techniques are well-known to solve the classification problems. For evaluation of intrusion detection system, NSL– KDD knowledge

discovery Dataset is taken. The outcomes show that SVM works better than Naïve Bayes. To perform comparative analysis, effective classification methods like Support Vector Machine and Naive Bayes are taken, their accuracy and misclassification rate get calculated.

3. Literature Review

The chapter presents an overview of computer attacks and some of the techniques employed against intrusion. IDS architectures, models and implementations are also discussed. An intrusion is a successful violation of a network's security policy. An Intrusion Detection System (IDS) is an ad hoc security solution to protect flawed computer systems. The system sends a notification or takes an action should an intruder attempt to go past the security mechanism such as authentication, or firewall. Intrusion Detection Systems (IDS) are those systems that have the ability to detect both internal and external attacks on a computer system and undertake some measures to eliminate them. Intrusion Detection Systems not only detect intrusion but also identify failed intrusion attempts, providing necessary information for precautionary measures and sometimes counter intrusions.

Intrusion detection systems will detect an intrusion or intrusion attempt by examining features such as:

1. Network traffic,
2. CPU and Input/Output(I/O) utilization,
3. File activity or even user location for signs of attacks.

Apart from sending a notification in the form of an electronic mail or by sounding an alarm, the system could also be configured to terminate a TCP (Transport Control Protocol) session and replace the connection with the administrator's connection in case of breaches to the security policy.

On the other hand, the systems could log suspicious activity for subsequent review. Normally, for the security of networks, IDS's will have monitoring agents or sensors on local networks, between subnets or on links to remote networks such as the Internet. The agents analyse traffic for any signs of attack and send this information to a management console. The IDS could be placed just behind a network firewall or behind a perimeter router. Data sources used to detect intrusion could also be from audit trails produced by the operating

system, network traffic flowing between systems, application logs or data collected from system probes for example the file system alteration monitors.

4. Existing system

To protect target systems and networks against malicious activities anomaly-based network IDS is a valuable technology. Despite the variety of anomaly-based network intrusion detection techniques described in the literature in recent years, anomaly detection functionalities enabled security tools are just beginning to appear, and some important problems remain to be solved. Several anomaly-based techniques have been proposed including Linear Regression, Support Vector Machines (SVM), Genetic Algorithm, Gaussian mixture model, knearest neighbour algorithm, Naive Bayes classifier, Decision Tree. Among them the most widely used learning algorithm is SVM as it has already established itself on different types of problem. One major issue on anomaly-based detection is though all these proposed techniques can detect novel attacks but they all suffer a high false alarm rate in general. The cause behind is the complexity of generating profiles of practical normal behaviour by learning from the training data sets.

5. Proposed system

The system proposed is composed of feature selection and learning algorithm. Feature selection component are responsible to extract most relevant features or attributes to identify the instance to a particular group or class. The learning algorithm component builds the necessary intelligence or knowledge using the result found from the feature selection component. Using the training dataset, the model gets trained and builds its intelligence. Then the learned intelligences are applied to the testing dataset to measure the accuracy of home much the model correctly classified on unseen data.

Based on the best features found in the feature selection process, learning models are developed. To develop the learning model, machine learning algorithm is used. Training dataset is used to train the algorithm with the selected features. In supervised machine learning, each instance in the training dataset has the class it belongs to. The algorithm builds the learning model based on which machine learning algorithm is being used.

6. System Architecture

Three basic IDS architectures that have been proposed for intrusion detection systems and they are:

1. Host-based IDS

2. Network-based IDS

3. Distributed IDS

- **Host-based Intrusion detection System (HIDS)**

  A piece of software is loaded onto a system to detect intrusion. The software uses log files or system auditing agents, which look at communication traffic.

  1. Log file analysers - analyses log files for patterns that indicate intrusion
  2. File system monitor - monitors the system to check for integrity of files and directories
  3. Connection analyzers - to monitor connection attempts
  4. Kernel based analyzer- to detect malicious activity on a kernel

  The hosts within the private network will have an intrusion detection system that will send alerts to the agent console from where they are analyzed.

  For private small scale network we use HIDS.

- **Network-based Intrusion Detection System (NIDS)**

  Network Intrusion Detection systems (NIDS) monitor the network by capturing network packets. They parse the packets, analyse them and extract useful information from them. All the analysis of the different packets are done without changing or inserting any data on the

network. A sensor is used to monitor packets traveling on that particular segment. The IDS determine if the traffic matches any known signatures. Examples of these known signatures may include:

1. String Signatures— This represents a text string that may indicate a possible network intrusion.
2. Port Signatures—Watches out for connection attempts to well-known ports
3. Header Signatures represent dangerous header combinations that could characterize intrusion.

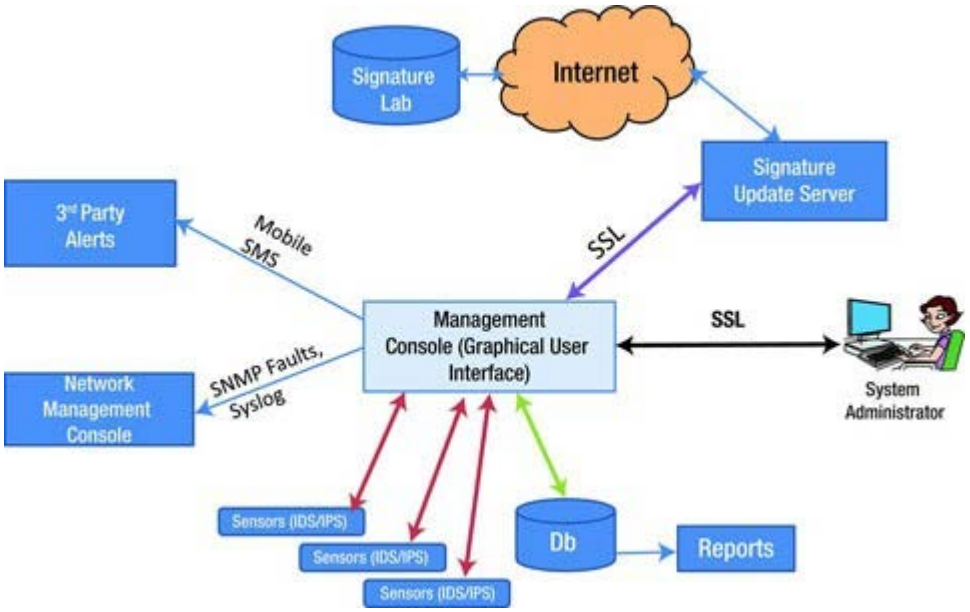
For small to medium scale we use NIDS.

- **Distributed Intrusion Detection Systems.**

Despite their advantages, HIDS and NIDS still have a few limitations arising from the fact they all have to collect data either from audit trails or by monitoring packets in the network to a centralized location where they are analyzed. And the problems associated with these are, because of having a centralized analyzer, a single point of failure is introduced. Should the attacker fail the central point, then intrusion detection will be ineffective.

Therefore, for large enterprises we use Distributed IDS.

We are therefore selecting the Network Intrusion Detection System.



---

7. Required Algorithms

- SUPPORT VECTOR MACHINES
- NAIVE BAYES ALGORITHM
- ARTIFICIAL NEURAL NETWORKS

8. Hardware and software requirements.

Hardware:

As of now a single computer hardware with following is sufficient.

Minimum Requirements

- 64-bit 2.6 GHz Intel core i5 CPU
- 8 GB RAM
- Windows 7 environment
- limited network traffic instances.

Later on, we can simulate it further by making use of IOT devices for collecting data, simulating and defending an attack using IDS we generate.

Software:

- Weka 3.8/3.9
- Jupiter Lab (Python)
- R studio (R)