*“In Pursuit of Technical Excellence”*

**INTRUSION DETECTION SYSTEM USING ML**

For the Degree of

**Bachelor of Engineering**

In

**Information Technology**

Submitted By

**ANAND INGLE BE16S06F006**

**AKSHAY CHACHARKAR BE15F06F008**

**HITESH TAPDIYA BE15F06F054**

**VISHAL POTGANTE BE15F06F043**

Under the Guidance of

**Prof. S. M. Chavan**



Department of Information Technology

**Government College of Engineering, Aurangabad**

Maharashtra State, India

**(An autonomous Institute of Government of Maharashtra)**

(2018-19)

**CERTIFICATE**

This is to certify that the thesis entitled **“Intrusion Detection System Using ML”,** which is being submitted herewith for the award of the **‘Degree of Bachelor of Engineering’** in **‘Information Technology’** of Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. This is the result of the research work and contribution by **‘Anand M. Ingle, Akshay B. Chacharkar, Hitesh A. Tapdiya, Vishal B. Potgante’** under my supervision and guidance.

Place: Aurangabad.

Date :

**Prof. S. M. Chavan Prof. C. M. Gaikwad**

Project Guide Head of Department

(Information Technology Department) (Information Technology Department)

**Dr. P. B. Murnal**

Principal

Government College of Engineering,

Aurangabad (M.S) – 431005

**ABSTRACT**

An Intrusion Detection System (IDS) is a software that monitors a single or a network of computers for malicious activities (attacks) that are aimed at stealing or censoring information or corrupting network protocols. 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. It attempts to identify intrusions, which we define to be unauthorized uses, misuses, or abuses of computer systems by either authorized users or external perpetrators. Some IDSs monitor a single computer, while others monitor several computers connected by a network. IDSs detect intrusions by analyzing information about user activity from sources such as audit records, system tables, and network traffic summaries.

Intrusion detection is the process of dynamically monitoring events occurring in a computer system or network, analyzing 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 analyzing 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 behaviors. 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.

**INDEX**

**NAME Page No**

**LIST OF ABBREVIATIONS I**

**LIST OF FIGURES** **II**

**LIST OF TABLES III**

**1 INTRODUCTION** **1**

1.1 The Introduction of an IDS 1

1.2 Need of Project 1

1.3 Types of IDS 2

1.3.1 Network Intrusion Detection System (NIDS) 2

1.3.2 The Host Intrusion Detection System (HIDS) 2

1.4Intrusion Detection Methods 3

1.4.1 Signature based intrusion detection system (S-IDS) 3

1.4.2 Anomaly based intrusion detection system (A-IDS)

**2 LITERATURE SURVEY 4**

2.1 Intrusion Detection System using AI and Machine Learning Algorithm 4

2.2 Attacks Classification 4

2.3 External Abnormal Behavior 5

2.4 Internal Abnormal Behavior 5

2.5 A Look at Firewall 5

2.6 Intrusion Detection System (IDS) 6

2.6.1 Why do we need IDS? 6

2.6.2 Anomaly detection 7

2.6.3 Efficiency of IDS 8

2.6.3.1 Accuracy 8

2.6.3.2 Performance 8

2.6.3.3 Completeness 8

2.6.3.4 Fault Tolerance 8

2.6.3.5 Timeliness 9

2.6.4 Intrusion Detection Tools 10

2.6.4.1 Arp 10

2.6.4.2 Ping 10

2.6.4.3 Netstat 10

2.6.4.4 Telnet 10

2.6.4.5 Nmap 10

2.6.4.6 Nessus 11

2.6.4.7 Tracert 11

2.6.4.8 RealSecure 11

2.6.5 Types of IDS 12

2.6.5.1 Network Based 12

2.6.5.2 Host Based 12

2.7 Detection Types 13

2.7.1 Signature-Based Detection 13

2.7.2 Anomaly-Based Detection 13

2.7.3 Stateful Protocol Inspection 13

2.8 Issues and Challenges in IDS 14

2.9 Machine Learning 16

2.9.1 Supervised Machine Learning 16

2.9.2 Unsupervised Machine Learning 17

* + 1. Evaluating ML for an IDS 17

**3 SYSTEM DEVELOPMENT 21**

3.1 Proposed Model 21

3.1.1 Technology Stack 22

3.1.2 Program Execution 22

3.1.3 Structure 22

3.1.4 Datasets 22

3.1.5 Libcap 23

3.2 States of Machine Development 23

* 1. Input Processing 27

3.4 Data Cleaning and Pre-processing 32

3.5 Model 34

**4 PERFORMANCE ANALYSIS 35**

4.1 Introduction to Testing and Evaluation of IDS 35

4.2 Existing Tools and Methodologies 35

4.3 Testing Methodology 35

4.3.1 General Methodology 35

4.3.2 DARPA approach 36

4.3.3 LARIAT 36

4.4 Tools for Testing IDS 36

4.5 Issues in Testing of an IDS 37

4.6 Present Evaluation Environments 37

4.6.1 DARPA Environment 37

* + 1. Vendor Independent Testing Lab 37

4.7 Using the Test Results 38

**5 CONCLUSION 39**

5.1 Conclusion 39

5.2 Future Work 39

5.2.1 Implement feature extraction in the inline mode 39

5.2.2 Use a wider dataset 39

5.2.3 Use pre-selected API’s 39

**REFERENCES 40**

**ACKNOWLEDGEMENT 41**

**LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
| IDS | Intrusion Detection System |
| DOS | Denial of Service |
| DDOS | Distributed Denial of Service |
| IP | Internet Protocol |
| TCP | Transmission Control Protocol |
| DARPA | Defense Advanced Research Projects Agency |
| ML | Machine Learning |
| API | Application Programming Interface |
| NSS | National Security Strategy |
| LAN | Local Area Network |
| WAN | Wide Area Network |
| UDP | User Datagram Protocol |
| ICMP | Internet Control Message Protocol |

I

**LIST OF FIGURES**

|  |  |
| --- | --- |
| **Name** | Page No. |
| Figure 1. basic types of IDS | 7 |
| Figure 2: Various data sources in IDS | 9 |
| Figure 3: Types of Intrusion Detection System | 18 |
| Figure 4. proposed model | 21 |
| Figure 5. Iterative Development Model | 22 |
| Figure 6: States of Machine Development | 23 |
| Figure 7. Selecting the dataset to scan it. | 24 |
| Figure 8. uploading the dataset to scan | 24 |
| Figure 9. scanning the dataset | 25 |
| Figure 10. plotting the outlier’s graph | 25 |
| Figure 11. showing the ip addresses with anomalous behavior | 26 |
| Figure 12. About the project developers | 26 |
| Figure 13. simple-typical IDS | 27 |
| Figure 14: Flow of Machine Learning | 29 |
| Figure 15: Training labels and data | 33 |

II

**LIST OF TABLES**

|  |  |
| --- | --- |
| **Name** | Page No. |
| Table 1: Comparison of various intrusion detection techniques | 19 |
| Table 2: NSL KDD Dataset Description | 30 |
| Table 3: Analysis of Attacks Classes and types | 31 |
| Table 4: Number of Instances In The Test Set | 32 |

III

**1 INTRODUCTION**

**1.1 The Introduction of an IDS**

There are many types of dangers on the internet, including malware and DDOS attacks. A network can be protected against such attacks using an intrusion detection system. An IDS system can detect intrusions and intrusion de-generates an alert when it detects an intrusion. This intrusion detection system in a network analyses all traffic. For large datacentres this is a difficult task. There's an enormous amount of data through the network of a datacentre. Standard intrusion systems cannot then all traffic completely. A way to fix this is by IP flows is regeneration of packet data. Using IP flows ensures that an intrusion detection system can check all traffic. Intrusion detection systems also require a lot of maintenance. This depends on course and also involves high cost. Sensitive data is also increasingly being stored digitally. All these new services could contain security flaws which could leak private data, such as passwords or other sensitive data. This means that security flaws become more and more important since they can cause so much damage.

**1.2 Need of this Project**

It is not just the leaking of sensitive data that is an issue, but also protecting a computer or network against malware is important. Considering this, it becomes more important to be able to detect and prevent attacks on network systems. Intrusion detection systems are used for this purpose. An intrusion detection system can alert administrators of malicious behaviour. In order to have good performance, most intrusion detection systems need a lot of manual maintenance. This thesis tries to find out whether an intrusion detection system can work out-of-the-box with an acceptable performance. This is done by using machine learning algorithms. These are algorithms which can learn and find patterns in input. Machine learning algorithms seem promising for the problem of automatic intrusion detection. This thesis tries to view therefore that an intrusion detection system out-of-the-box may have a good performance. This is done via machine learning algorithms. These are algorithms that can learn from data and patterns. This seems well applicable to the problem of intrusion detection, this will also view this thesis, as well as the algorithms may or may not work.

The term intrusion refers to any unauthorized access that attempt to compromise confidentiality, integrity and availability of information resources. In general, we can say that any malicious use or misuse of any entity in a network can be referred as an intrusion. The intruder tries to find the vulnerability in the security system and then prepare for attack. Intrusion detection is the process of fast detection of unwanted violation in system’s normal behaviour due to attacks performed by the malicious user. Intrusion detection system (IDS) deals with detection of such type of attacks in just in time or real time and report, alert or countermeasure on the attack to the administrator. Now a days, artificial Intelligence, data mining and machine learning algorithms have been subjected to extensive research in intrusion detection with emphasis on improving the accuracy of detection and make an immune model