# Chapter 1

# INTRODUCTION

Computer networks are telecommunication networks which allow computers to exchange data. Network security is implemented on a variety of computer networks to secure daily transactions and communications among businesses, governments and individuals.

## 1.1 Overview

The current network security technologies (such as Firewall, IDS/ IPS) do provide security to networks but have limitations in terms of detecting only the known attack patterns. The system becomes prone to unknown vulnerabilities until the signatures of such attacks are found and configured in the signature databases of network security solutions. The moment attack pattern changes, the signature based solutions fail.

## 1.2 Network Security

Network security[1] consists of the provisions and policies adopted by a network administrator to prevent and monitor unauthorized access, misuse, modification, or denial of a computer network and network-accessible resources. Network security involves the authorization of access to data in a network, which is controlled by the network administrator.

### 1.2.1 Intrusion Detection Systems

An Intrusion Detection System (IDS)[2] is a device or software application that monitors network or system activities for malicious activities or policy violations and produces reports to a management station. Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts.

In a passive system, the intrusion detection system (IDS) sensor detects a potential security breach, logs the information and signals an alert on the console and/or owner. In a reactive system, also known as an intrusion prevention system (IPS), the IPS auto-responds to the suspicious activity by resetting the connection or by reprogramming the firewall to block network traffic from the suspected malicious source. The term IDPS is commonly used

where this can happen automatically or at the command of an operator; systems that both "detect (alert)"and "prevent".

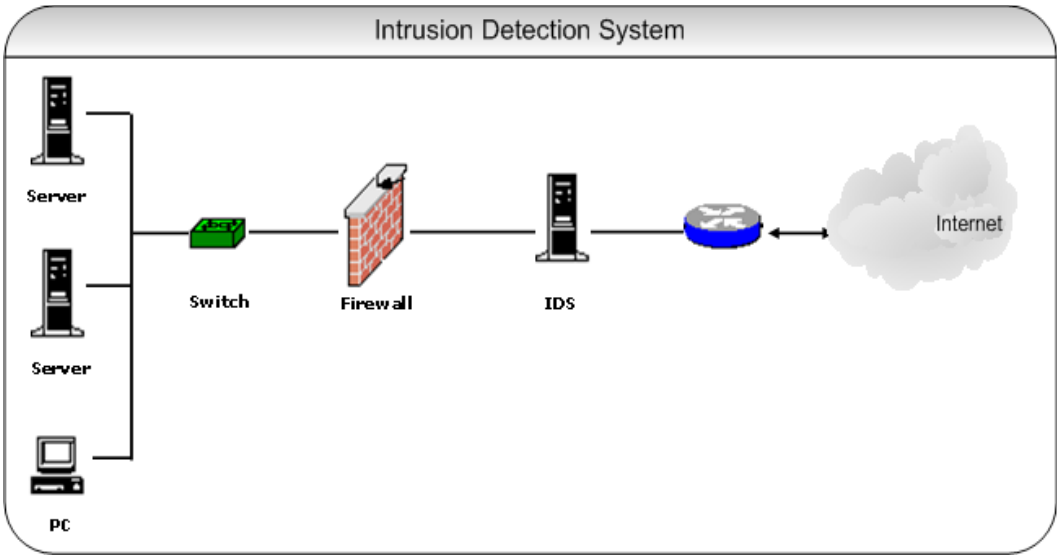


**Figure 1.1 Intrusion Detection Systems**

### 1.2.2 Attack Classification

In computer networks, an attack is any attempt to destroy, expose, alter, disable, steal or gain unauthorized access to or make unauthorized use of an asset.[3]

| Attack Class | Attack Type |
|---|---|
| Probe | portsweep, ipsweep, quesosatan, msscan, ntinfoscan, lsdomain, illegal-sniffer. |
| DoS | apache2, smurf, neptune, tosnuke, land, pod, back, teardrop, tcpreset, syslogd, crashiis, arppoison, mailbomb, selfping, processtable, udpstorm, warezclient. |
| R2L | dict, netcat, sendmail, imap, ncftp, xlock, xsnoop, sshtrojan, framespoof, ppmacro, guest, netbus, snmpget, ftpwrite, hhtptunnel, phf, named. |
| U2R | sechole, xterm, eject, ps, nukepw, secret, perl, yaga, fdformat, ffbconfig, casesen, ntfsdos, ppmacro, loadmodule, sqlattack. |

**Table 1.1 Attacks present in DARPA 1999 Dataset**

## 1.3 Machine Learning and its Applications

Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data. For example, a machine learning system could be trained on network traffic to learn to distinguish between anomalous and normal traffic. After learning, it can used to classify the incoming traffic into anomalous and normal traffic.

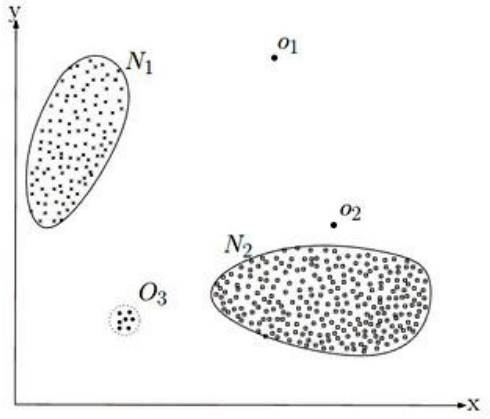


**Figure 1.2 Example of Anomalies**

Machine Learning is used in a variety of fields, a few of which are mentioned below.

i. Network Anomaly Detection: Statistical machine learning is one class of statistical approaches used in the field of Network Anomaly Detection. A choice of a learning scenario depends on the information available in measurements. For example, a frequent scenario is that there are only raw network measurements available and thus unsupervised learning methods are used. If additional information is available, e.g. from network operators, known anomalies or normal network behaviors, learning can be done with supervision. A choice of a mapping depends on the availability of a model, the amount and the type of measurements, and complexity of learning algorithms[4].

ii. Evaluating Learned Knowledge: Rules induced from training data are not necessarily of high quality. The performance of knowledge acquired in this way is an empirical question that must be answered before that knowledge can be used on a regular basis. One standard approach to evaluation involves dividing the data into two sets, training on the first set, and testing the induced knowledge on the second. One can repeat this process a number of times with different splits, and then average the results to estimate the rules' performance on completely new problems. Kibler and Langley (1988) experimental

methods of this sort for a broad class of learning algorithms. An important part of the evaluation process is experts' examination of the learned knowledge. Evans and Fisher (1994) encourage an iterative process in developing a fielded application.[5]