

Adaptive Network Intrusion Detection and Mitigation Model using Clustering and bayesian Algorithm in a Dynamic Environment

Musyimi S. Muthama
Institute of Computing (TUM)

Waweru Mwangi
Professor
School of Computing & IT
(JKUAT)

Otieno Calvin, PhD
School of Computing & IT
(JKUAT)

ABSTRACT

Today, there is a serious challenge facing Network Security, especially Networking Intrusion Detection and prevention attacks of Denial of Service. Denial of Service (DoS) has the most devastating effects on Networking and Information security. It has also put tremendous pressure over the security experts lately, in bringing out effective defense solutions. The adversary methods are ever changing day night, the complexity and sophistication of attacks and vulnerability methods continue to rise yearly, and the potential impact to the bottom line is significant organization information systems. These attacks could be implemented diversely with a variety of tools and techniques. Since there is no single solution for DoS, this attack has managed to prevail on internet for nearly a decade. The task of uncovering previous unseen DOS attacks and new attacks in dynamic Intranet networks quickly becomes unmanageable. Network Intrusion Detection Systems (NIDS) have become a necessity in information security model because of the increase DOS and malicious activities. Therefore, the research proposes a Data mining technique enhanced with Artificial intelligent that is classification, clustering, and Behavior profiling networks algorithm to categorize a network process as either normal or abnormal. Prevention of DOS attacks using techniques as Quarantine IP address, Blackholing, Sinkholing. The proposed model Adaptive Network Intrusion Detection Model (ANIDM) based on the combination of techniques artificial intelligent and anomaly detection. Which is self-learning and adaptive in presence of DOS attacks and malicious activities under minimal human intervention. The proposed model is based on clustering and classification through K-NN algorithm. The results show that these mitigation approaches improve the ability to separate between unknown abnormalities in the dataset and the legitimate traffic structure. The ANIDM can effectively detects existing and new attacks focusing on Denial of service attack (DOS), while providing continuous service even under attacks. The model is based on speed, crypt-isolation, log-audit and preemptively restored on a regular basis. The model can be applied to Dynamic environment learning institution and business enterprise. Finding the primary challenges to Network intrusion detection are the problem of misjudgment, previous unseen attacks, overwhelming volume of false alarms, misdetection, new attacks tactics. Therefore, the paper concludes that variation ANIDM can be considered for detection and mitigate of DoS and DDOS attacks.

General Terms

Denial-of-service attack, network traffic characterization, Flooding, Network Security

Keywords

Network Security Smurf, Neptune, Ping-Of-Death (POD), ICMP, Back, LAND (Local Area Network Denial &KDD Cup 99 Dataset, Dedicate to my family and all A.I, Datamining, Hackathons and knowledge discovery all of me for all of us.

1. INTRODUCTION

The Internet connects hundreds of millions of computers across the world running on multiple hardware and software platforms providing communication and commercial services. However, this interconnectivity among computers also enables malicious users to misuse resources and mount Internet attacks. The continuously growing Internet attacks pose severe challenges to develop a flexible, adaptive security-oriented methods. Unlike a firewall and antivirus that filters “bad” traffic, an IDS monitor and analyzes packets to detect malicious attack attempts[1]. The attacker and intruder’s techniques are ever changing overnight using techniques as Key-lockers, trojan installation, spywares, machine generated malware, viruses, and multiplicative worms, which are primary threats to information security in a dynamic environment. Information security is to protect three goals of a secure environment, these are confidentiality, integrity and availability[2, 3].

1.1 Back ground information

Denial of service (DOS) attacks have emerged as one of the most severe networks intrusive behaviors and have posed serious threats to the infrastructures of computer networks and various network-based services. These attacks can be launched by deliberately exploiting system vulnerabilities of a victim. DoS attack aims at disrupting the authorized use of networks, systems, or applications. By sending messages which exhaust service provider’s resources (network bandwidth, system resources, application resources).

Increased network complexity, greater access, and a growing emphasis on the Internet, have made network security a major concern for organizations [2, 4]. IDS is to detect intrusions and intrusion attempts within our network, allowing a savvy admin to take appropriate mitigation and remediation steps [5],[6].

Traditional techniques for IDS focus on technical solutions in various forms details such as firewall, encryptions, cryptography, intrusion detection systems, data backup, remote disaster recovery, backup power, system recovery and web application security[7]. As Internet devices and applications continue to grow, it becomes increasingly

important to understand network behavior for efficient network management and security monitoring [8].

In this research discuss uncover previous unseen attacks from the traces of the traffic flow. Using data mining algorithms called K-NN and clustering via Naive Bayes classification for anomaly-based network intrusion detection.

Data mining is the set of activities used to find new, hidden, or unexpected patterns in data. These techniques are often called knowledge data discovery (KDD), and include statistical analysis, neural or fuzzy logic, intelligent agents or data visualization.

1.2 Problem Statement

Insecurity incidents and evolving threats are on the rise and are increasing exponentially. Therefore, Intrusion detection is an important component of a modern information technology protection from unauthorized users.

The propagation of malware, trojan horse in a Network has presented a thoughtful vulnerability to Information security of transmitted information. The current tactics of Intrusion Detection i.e. Firewall, proxy servers, anti-virus and sensors systems fails to detect new, already inconspicuous malicious executables.

The adversary is ever dynamic as software, hardware and Networks Technology is dynamic new tactics for intrusion and more sophisticated are been launched. Most of the existing research work of NIDS are based on signature, normally, Host based and Network based but all not self-learning and adaptive

Intranet monitoring is crucial in providing reliable, efficient networked services in distributed network, Internet service provider, are placed in geographically diverse locations. The task of uncovering previous unseen attacks and new attack pattern in organization networks quickly becomes unmanageable.

Tactics, tools, and algorithms used by hackers are ever changing day night; the complexity of malicious attacks and penetration methods continues to rise yearly. The potential effect is critical to association data security.

Security incidents and evolving threats are on the rise and are increasing exponentially. Therefore, Intrusion detection is an important component of a modern information technology protection from unauthorized users. It detects and treats anomalies efficiently, because they affect the quality of services provided, resulting in degradation of network, performance and even in operations' interruption [9]. The researcher proposes artificial network intrusion detection and immunization system enabling firms to reduce undetected intrusion and new attackers. Artificial Intrusion Detection System is one of the most important security systems to detect intrusions in a variety of networks in a dynamic environment [10].

Current techniques i.e. anti-virus systems fail to detect polymorphic/metamorphic, new attack patterns and previously unseen attacks. This is intended to "System and Methodology for Unseen DOS attack " describes the state's overall requirements regarding the acquisition and implementation of intrusion prevention and detection with intelligence [11, 12].

1.3 Research contribution

- i. To uncover previously unseen DOS attacks in existing Network intrusion detection model and mitigate the attacks in a dynamic environment.
- ii. Visualizing Denial of service attacks in intranet dynamic environment.

1.4 Research Scope

The paper covers enhancement of Network intrusion detection and mitigate of DOS. Integrating artificial intelligent and K-NN classification & clustering monitoring computer network behavior profiling, in a dynamic environment based on Data mining.

1.5 What is Dynamic Environment Network in traffic

Network dynamics environment is a research field for the study of networks whose status changes in time. Traffic flow, number of connections, services on the Intranet, Hardware platform, number of users, applications and bandwidth usage. Thus the reason, it is necessary for the model to visualize and uncover previously unseen DOS attacks. Network dataset includes both normal and anomalous traffic.

2. LITERATURE REVIEW

Kusumah, P., S. Sutikno, et al.[13] To uncover previous unseen attacks and search for a new attacker pattern which compromises with Intranet performance, based on the results of the initial studies that have been conducted, many information security problems occur in many organizations. The basic vulnerability of security can software, hardware and network. That problem can be overcome by the application of specific information security solutions directly or partially it information security achievement of enterprise goals. Therefore, components of the IT-related organizations must be managed properly.

2.1 Anomaly-based Detection

Anomaly-based detection mechanism shows promising results in detecting zero-day attacks that exploit previously unknown system vulnerability, and it has less dependency on domain knowledge. Recent work on DoS attack detection primarily adopts this concept. Techniques used in these anomaly-based detection systems can be divided into two categories, namely Data mining and machine learning[14].

2.1.1 A theoretical Review of Data Mining ANIDM.

The paper has listed many advantages of using AI based on Datamining techniques over other conventional approach [15, 16]. The major advantages include Flexibility (vs. threshold definition of conventional technique); Adaptability (vs. specific rules of conventional technique); Pattern recognition (and detection of new patterns); Fast computing (faster than humans, actually) In the area of computer security, Intrusion Detection (ID) is a mechanism that attempts to discover abnormal access to computers by analyzing various interactions. The use of AIS in ANIDM is an appealing concept in current techniques [10]. With the exponential growth of network bandwidth, this task slowly demanded substantial improvements in both speed and accuracy.

[4] Statistical analysis techniques have been employed to conduct investigation into attributes of network traffic packets and to determine a rationale threshold for discriminating attacks from the legitimate user.

2.1.2 Apriori

Apriori is an algorithm that uses data mining to find frequent item sets and is a method for discovering the interesting relationship between data variables. The apriori principle states that if an item or collection of items is frequent, then all of its subsets must also be considered frequent. The algorithm bases itself on identifying frequent items set k , before using the frequent itemset of k to find the frequent itemset of $k + 1$ [17].

History-based IP Filtering

History-based IP Filtering suggested using a database of previously seen legitimate IP addresses to counter DDoS attacks. This method bases itself on the common assumption that normal traffic differentiates itself from traffic under an attack. The idea is that the network should learn from previous network connections, before under an attack, the learned behavior is used to characterize incoming packets as normal or abnormal. This mechanism known as Under an attack, only IP addresses from the IAD are allowed to access the network or service [18].

2.1.3 Adaptive History IP filtering

The approach relies on observing data and using the Bayesian theory to derive optimal IP networks rules from this data, which then decides which packet to accept and drop based on the dynamically assigned threshold [18].

Pattern recognition is used to derive the normal traffic distribution based on previously observed data. Based on the observed data the proposed approach creates a binary tree with access control lists containing IP networks to accept. [19] Anomaly detection assumes that intrusions are anomalies that necessarily differ from normal behavior. Basically, anomaly detection establishes a profile for normal operation and marks the activities that deviate significantly from the profile as attacks. The main advantage of anomaly detection is that it can detect unknown attacks. However, this advantage is paid for in terms of a high false positive rate because, in practice, anomalies are not necessarily intrusive. Moreover, anomaly detection cannot detect the attacks that do not obviously deviate from normal activities. It has a relatively high detection rate for new types of intrusion.

2.1.4 Statistics

The most mature data mining technologies but are often not applicable because they need clean data. In addition, many statistical procedures assume linear relationships, which limits their use.

2.1.5 Neural networks, fuzzy logic

These technologies are able to work with complicated and imprecise data. Their broad applicability has made them popular in the field. High computational time.

2.1.6 Decision tree

These technologies are conceptually simple and have gained in popularity as better tree growing software was introduced. Because of the way they are used, they are perhaps better called “classification” trees.

2.2 DOS & Ddos attack techniques

Over the last couple of years, DDoS attacks have focused on attacking the infrastructure and application layer, either by overloading the bandwidth capacity or focusing on depleting some limited network resource. Attackers often successfully exploit the open internet infrastructure into launching DoS or DDoS attacks.

2.2.1 Network layer attacks

Network layer attacks, are DDoS attacks that target the network layer by attempting to overwhelm the bandwidth and routing infrastructure. A secure network is a web application's first line of defense against malicious attacks. It is the gateway to the servers where your application resides. Securing the network layer is the only way to ensure your application is not flooded with attacks which could be easily blocked at that outermost layer.

2.2.2 DNS reflection attack

The basic technique in DNS reflection attack is to request a large zone file with the source IP address spoofed to be the intended victim. The attacker's request is only a fraction of what the DNS server will respond with, efficiently amplifying the attack to many times the size of the available bandwidth.

2.2.3 Broadcast reflection attack

The ICMP protocol is a diagnostic tool that can be used to test the reachability of different computer systems. A host can send an ICMP (Internet Control Message Protocol) echo request message to a computer system. When the receiving system gets this message, the system will respond by sending an ICMP echo reply message back to the sender.

2.2.4 Specialized Classification of DOS Attacks

An explicit attempt by attackers to prevent legitimate users of a service from using that service. Consumption of network connectivity and/or bandwidth. Consumption of other resources, e.g. queues, CPU. Destruction or alternation of configuration information. Malformed packets confusing an application, cause it to freeze. Physical destruction or alternation of network components.

- i. Flooding Attacks: - A SYN flood is a form of denial-of-service attack in which an attacker sends a succession of SYN requests to a target's system in an attempt to consume enough server resources to make the system unresponsive to legitimate traffic.
- ii. Smurf DoS Attack: - The Smurf attack is a distributed denial-of-service attack in which large numbers of Internet Control Message Protocol (ICMP) packets with the intended victim's spoofed source IP are broadcast to a computer network using an IP broadcast address.
- iii. DNS server Attack:- A DNS attack is an exploit in which an attacker takes advantage of vulnerabilities in the domain name system (DNS). DNS is a protocol that translates a user-friendly domain name, like WhatIs.com, into the computer-friendly IP address.
- iv. DHCP Starvation Attack: -A DHCP starvation attack works by broadcasting DHCP requests with spoofed MAC addresses. If enough requests are sent, the network attacker can exhaust the address space available to the DHCP servers for a period of time. This is a simple resource starvation attack just like a synchronization (SYN) flood attack.
- v. Ping of Death:- On the Internet, ping of death is a denial of service (DoS) attack caused by an attacker deliberately sending an IP packet larger than the 65,536 bytes allowed by the IP protocol. One of the features of TCP/IP is fragmentation; it allows a single IP packet to be broken down into smaller segments.

2.3. Data Mining ANIDM for denial of service attack

Data mining is becoming one of the popular techniques for detecting intrusion. IDS can be classified on the basis of their

strategy of detection. There are two categories under this classification. Networks are easy targets for any type of attack like viruses, worms, any kind of malicious program, etc. This is because the infection can spread easily across the network with high speed. Therefore, to prevent this, networks need to be designed and equipped with the sophisticated intelligence to diagnose and mitigate threats in real-time. Network Intrusion Prevention provides self-defending solutions that offer network wide protection and mitigation techniques. It has the intelligence to accurately detect, analyze, classify, and mitigate malicious traffic in real-time, offering comprehensive protection for a wide range of network intrusions and attacks [20].

Example, dropping a packet that was regarded to be malicious along with the possible blocking of all further traffic from that IP address or port) [21]. Classification Data Mining Intrusion Detection Models Different data mining approaches are frequently used to analyze network data to gain intrusion related knowledge.

2.4. Classification and Clustering

Classification is similar to clustering in that it also partitions Datasets records into distinct segments called classes. Unlike clustering, classification analysis requires that the end-user/analyst know ahead of time how classes are defined. Given a set of records, where one of the features is the class label (i.e., the concept), classification algorithms can compute a model that uses the most discriminating feature values to describe each concept [2, 5]. Consider the problem of clustering n sequences of characters. First, each of the $(n) \times (n-1)/2$ pairs of possible merges is evaluated, and the two clusters that have maximum value of the criterion function are merged. After performing m merging steps, each of the $(n-m) \times (n-m-1)/2$ pairs possible merges is evaluated. This process continues until there are only k clusters left [22].

2.4.1 Integrating Data Mining in ANIDS

The central theme of intrusion detection using data mining approach is to detect the security violations in an information system. Data mining can process a large amount of data and it discovers hidden and ignored information [9, 23].

Classification is a data mining technique that assigns objects to one of several predefined categories. Algorithms (like ID3 and C4.5) require that the target attribute will have only discrete values. As decision trees use the “divide and conquer” method, they tend to perform well if a few highly relevant attributes exist, but less so if many complex interactions are present. The greedy characteristic of decision trees leads to another disadvantage that should be pointed out [24].

Stampar, M. and K. Fertalj et al, [17] A Bayesian Network (BN) is a model that encodes probabilistic relationships among variables of interest. This technique is generally used for intrusion detection in combination with statistical schemes. It has several advantages including the capability of encoding interdependencies between variables and of predicting events, as well as the ability to incorporate both prior knowledge and data. Main issue is the lag between a new threat and the signature for detecting that threat being applied to the IDS. During that lag period IDS is unable to detect the threat. An anomaly-based IDS monitors behavior and compares its characteristics against an established baseline. The baseline will identify what is “normal” for that subject and alert when anomalous behavior is detected, or significantly different than the baseline. The main issue is the higher false positive rate.

Reddy, R.R., Y. Ramadevi, et al [25] Support vector machine (SVM) is a technique used for solving a variety of learning, classification and prediction problems. Support vectors are a subset of experimental dataset used to define the boundary between the two classes. In situations, where SVM cannot separate two classes, it solves this problem by mapping input data into high-dimensional feature spaces using a kernel function. In high-dimensional space, it is possible to create a hyper plane that allows linear separation (which corresponds to a curved surface in the lower-dimensional input space).

2.5. Network Intrusion Prevention Techniques

Sou, S.I. and C.S. Lin, et al [16] An Intrusion Prevention System is a network security device that monitors network and system activities for malicious or unwanted behavior and can react, in real-time. IPS make access control decisions based on application content, rather than IP address or ports as traditional firewalls had done. These systems are proactive defenses mechanisms designed to detect malicious packets within normal network traffic and stop intrusions deals, blocking the offending traffic automatically before it does any damage rather than simply raising an alert as, or after, the malicious payload has been delivered.

2.5.1 Quarantine IP address

Use the IP address and port area of the quarantine rules page for the network IPS appliance to block events that are occurring on the source and target IP addresses or ports. Network Access Control (NAC) is a set of technologies and defined processes, which its aim is to control access to the network. NAC is a valid technology that should play a key role in internal network security. A common criterion for NAC does not exist and therefore the definition of what does a NAC solution should (and/or must) contain varies from one vendor to another. The distributed mode of NFV can be seen in interconnecting data centers, regionally or globally distributed CDNs, LAN, and location-specific services in campus networks. If the element does not comply with the defined security policy, the NAC solution must restrict the element's access to the network [26] [27].

2.5.2 Blackholing of DOS and DDOS

Blackhole is a common defense strategy used by Internet Service Providers (ISP) to stop DDoS attacks by blocking incoming traffic and redirecting it into a “black hole” or null route. It is a place where packets enter, are analyzed and discarded so they do not come out. A firewall has black hole capabilities based on its configuration, and so does any computer/host that does not provide ICMP Destination. Attempts to mitigate the impact of an attack, redirects traffic from attacked DNS or IP address to a “black hole” Then all traffic will be dropped, must know IP address of attacker or else legitimate traffic will be dropped as well.

2.5.3 IP address Blocking (Shun)

Cisco IPS Sensor software supports dynamic response action by blocking the offending traffic during an attack. The Attack Response Controller (ARC) function in the sensor software is responsible for managing network devices to respond to suspicious events by blocking (shunning) network access from attacking hosts and networks [26].

2.5.4 Sink holing of DOS and DDOS

Sink holing is a technique for manipulating data flow in a network; you redirect traffic from its intended destination to

the server of your choosing. It can be used maliciously, to steer legitimate traffic away from its intended recipient, but security professionals more commonly use sink holing as a tool for research and reacting to attacks.

2.6. Discuss of the critiques of the existing literature

Jyothsna, V., V.R. Prasad, et al [28] In contrast to the previously proposed data mining-based IDS, we employ random forests for anomaly intrusion detection. Random forests algorithm is more accurate and efficient on large dataset like network traffic. We also use the data mining techniques to select features and handle imbalanced intrusion problem. k-NN method was used as a supporter method for multi-class classification.

Midzic, A., Z. Avdagic, et al [12] IDS and IPS construction

requires high-quality training and testing data set and that is difficult to ensure, because network traffic can contain elements which could have impact on the personal privacy or company security policies. Marking activity needs some expert knowledge. Moreover, even when data classification and marking is done, those Data packets soon will become “outdated”. Researcher proposes Anomaly-based networks are based on Automatic Update.

- i. Analysis the KDD 99 dataset for selecting relevant features. They proposed that some features or attributes are not related to any attack. They have taken only 10% of the whole dataset and performed their analysis. Proposed heuristic rules for R2L attack detection by taking KDD cup99 dataset. They derive some heuristic rules for various R2L attacks by applying the decision tree algorithm and applying the statistical method [29]. The researcher proposes the service and traffic of hosts are different, in the ideal situation, each host should have its own behavior patterns database, but this would increase the data processing expenses[9]
- ii. [30] presented Intrusion detection technique by using k-means, Fuzzy neural network and SVM classifier for attaining high detection rate. Has described Agent based intrusion detection system using data mining approaches? The result shows that the frequent patterns mined from the audit data could be used as reliable agents, which outperformed from traditional signature-based NIDS [29]. The researcher proposes behavior profiling AI dynamic model which based on NIDS system and behavior of [2].
- iii. [31] The reason is these technologies cannot deal with a large number of security attributes due to the problem of knowledge discovering the unknown pattern or behaviors from high dimensionality security state spaces. These attributes can be represented as stream datasets contain collections of data objects that are altered in time due to continuously executed, time-dependent data updates. The researcher proposing proposes a dynamic model Intelligent Intrusion Detection System, based on specific AI approach for intrusion detection.

2.6.1 Summary of proposed model

The paper emphasis on denial of service attack for Network Intrusion Detection. Evaluating a number of attributes which quantify denial of service attack identifying as an intrusion and its prevention techniques (ANIDM). Which is based on Data mining using combined algorithm classification, clustering and knn designed to monitor network traffic. Identifying existing attacks and new attacks by analyzing strange behavior from normal behaviors. Looking the ability of the system to intelligently detect new previously unseen threats and react to them in such a way that minimizes the damage and potentially removes the threat altogether.

2.6.2 Research gap

From the literature above other researchers have focused on traditional intrusion detection methods such as signature and misuse detection. However, these methods are known to detect existing attacks, misjudgment and high false positive alerts. This research focuses on uncovering previously unseen attacks and prevention of attacks mainly DOS. The model seeks to develop an IDS with the ability of artificial intelligent to detect previously unseen attacks and react to them in such a way that will minimize the damage and potentially removes the threat altogether.

2.6.3 Research Findings

The primary challenges of Network intrusion detection models are the problem of misjudgment, misdetection, hidden unseen attacks. Network Intrusion Detection Systems (NIDS) have become a necessity in Information security model because of the increase in attackers, malicious activities and intruders in the network.

3. METHODOLOGY

This chapter explains in details the research design and methodology, target population, population sample, data collection techniques and methods used in data analysis then, it describes the K-Nearest Neighbor (KNN) algorithm and presents a comparative study of KNN-based IDS.

3.1 Research Design

In this research, the researcher applies one of the most efficient data mining algorithms called K-Nearest Neighbor via Euclidean clustering for anomaly-based network intrusion detection.

3.1.1 k-means clustering algorithm

k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster[32]. Each dimension of the feature space represents one attribute (feature) of the sample. The space is partitioned into regions by the classes (labels) of the points. An unknown point is classified to the class whose labels are most frequent among the K nearest samples [33].

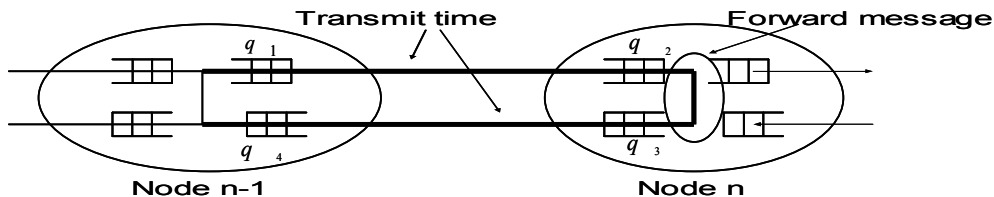


Fig 1: packets transmission from source node to end node (Node n)

$$rtt = q_1 + \left(lat + \frac{packet_size}{bw} \right) + q_2 + forward + q_3 + \left(lat + \frac{message_size}{bw} \right) + q_4$$

3.2 Detection and Prevention

Methodology.

3.2.1 Training Data Set

Data required for experimental dataset the unsupervised anomaly-based intrusion detection system is taken from the 'NSL dataset'. It is assumed that the relative amount of attacks in the training data is very small compared with normal data, a reasonable assumption that may or may not hold in the real-world context for which it is applied. If this assumption holds, anomalies and attacks may be detected based on cluster sizes. Large clusters correspond to normal data, and small clusters.

3.2.2 Test Datasets

Test Datasets: - in order to estimate how well your model has been trained (that is dependent upon the size of your data, the value you would like to predict, input etc) and to estimate model properties (mean error for numeric predictors, classification errors for classifiers, recall and precision for IR-models etc.)

3.2.3 Application of ANIDM

Application phase: - now you apply your freshly-developed model to the real-world data and get the results. Since you normally don't have any reference value in this type of data (otherwise, why would you need your model?), you can only speculate about the quality of your model output using the results of your validation phase.

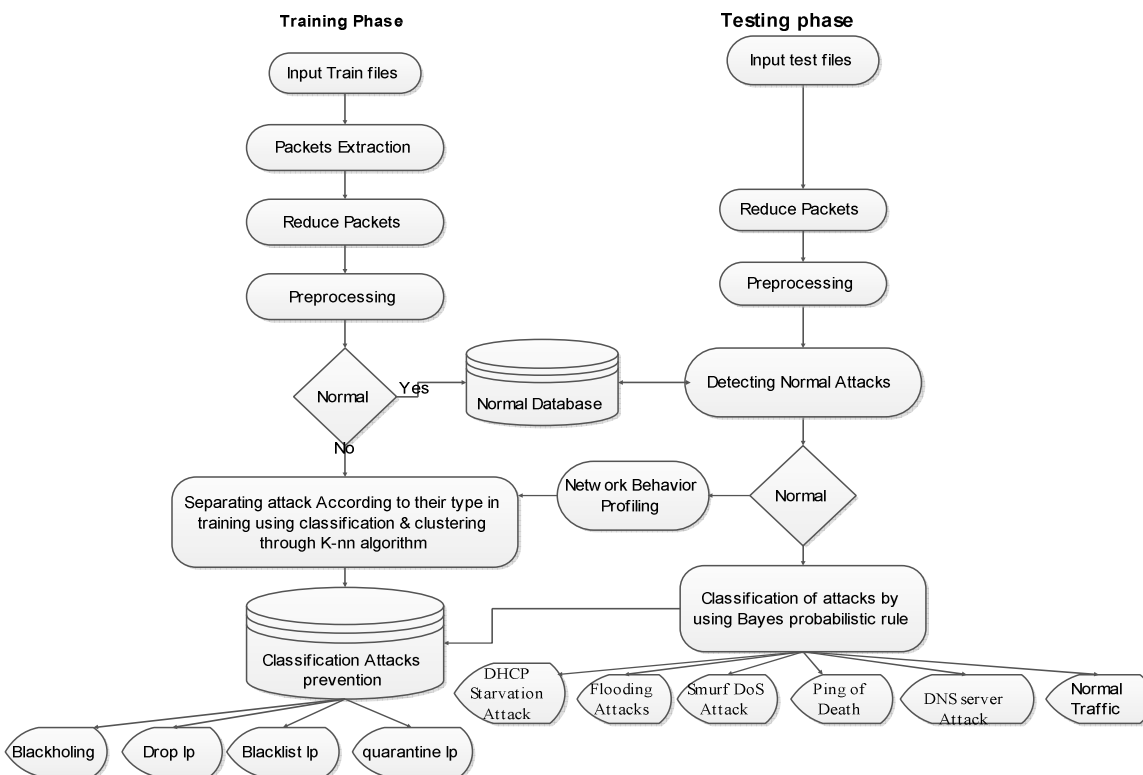


Fig 2: Proposed Model Architectural of Anidm Using K-NN And Mitigation

The detection and mitigation mechanisms designed here are effective for small network topologies and can also be extended to analogous large domains.

Table 1: Attributes Selection For Modelling Nidpm

Feature name	Description	Type	Normal	Abnormally
Duration	length (number of seconds) of the connection	Continuous	Definite 100mbps Time exceeded Timestamp request Timestamp reply	Indefinite, loop Destination unreachable Host unreachable Port unreachable.
Protocol type	type of the protocol, e.g. tcp, udp, etc.	Discrete	Many protocols are defined for use at the Application layer, such as HTTP, FTP, SMTP, and Telnet.	All protocols not defined at application layers for security
Src_bytes	number of data bytes from source to destination	Continuous	Srcbytes the number of source bytes to copy. Defined at begging of transmission	Not defined before transmission begging's
dst_bytes	Number of data bytes from destination to source	Continuous	proto/dst_port, combinations ordered by total dst_bytes defined.	Not defined before transmission begging's
Count	1 if connection is from/to the same host/port; 0 otherwise	discrete	number of connections to the same host as the current connection in the past two seconds	Any connection outside defined frame work
Num_failed_logins	number of failed login attempts	continuous	Only 3 defined attempts	Greater then 5attempts
Wrong_fragment	number of ``wrong" fragments	continuous	a fragment identifier is a short string of characters that refers to a resource that is subordinate to another define by user	maximum offset and minimum ≥ 7 bytes server to allocate a buffer of size > 65535 bytes. Ethernet is 1500 bytes.
urgent	number of urgent packets	continuous	Normal status of the connection.	error status of the connection.

Attribute = packet byte. Time (duration)128, 256 possible values, 48 attributes (packet bytes), 20 bytes of IP header, 20 bytes of TCP header, 8 bytes of payload.

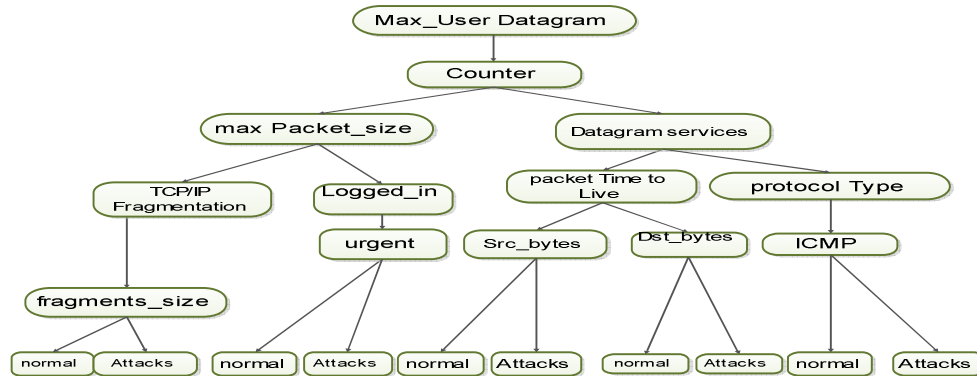


Fig 3: Attributes relationship (Duration time been the root)

Table 2: Scaling of attributes

Attributes	Values def
Duration	[0, 58329],
Packet Time to Live (TTL)	[0, 2 ⁸ = 256],
packet size	[0, 65535],
Max Data set	[0, 65507],
ICMP in bytes	[0, 1500],
Protocol type	[0, 10],

IP Fragmentation	[0, 65535],
Service	[0, 500],
wrong_fragment	[0, 3],
urgent	[0, 14],
num_failed_logins	[0, 5],
srv_bytes	[0, 255],
dst_bytes	[0, 255],

Anomaly detection

If X has unknown system call then **abnormal**
If X is the same as any Dj then **normal**

K-nearest neighbor

Calculate sim_avg for k-nn
If sim_avg > threshold then **normal**
Else **abnormal**

End.

3.3 Implementation of K-mean clustering

Exploratory of datasets analysis techniques, implements nonhierarchical methods of grouping objects together. Determining the mean using Euclidean distance evaluation. Grouping the datasets based on minimum distance of the attributes.

3.3.1 Weka Data Format

Weka permits the input data set to be in numerous file formats like CSV (comma separated values: *.csv), Binary Serialized Instances (*.bsi) and (arff). Data mining supports automated analysis and interpretations of the data and events collected from comparison of trained network traffic and network traffic detection.

3.3.2 The pseudo code adapted k-nn algorithm below

1. Choose random k data points as initial Clusters Mean (cluster center)
 2. Repeat, 3. for each data point x from D
 4. Computer the distance x
 5. Assign x to the nearest cluster, 6. End for
 7. Re-compute the mean for current cluster collections.
 8. Until reaching stable cluster
 9. Use these centroid for normal and anomaly traffic.
 10. Calculate distance of centroid from normal and anomaly centroid points.
 11. If distance(X, Dj) >= 5
 12. Then anomaly found ; exit
 13. Else then X is normal;
- */ end of pseudo code.

The pseudo code for the adapted k-nn algorithm

- 1: Recording the traffic
- 2: Making cluster out of recorded values as normal and anomalous
- 3: Calculating mean values of both clusters
- 4: Read new traffic data
- 5: Calculate distance from new data point to mean of both clusters
- 6: Assign data point to the nearest cluster
- 7: Update mean value of cluster
- 8: Repeat from step 4
- 9: Based on traffic rate cluster will be declared normal or anomalous
- 10: Obtained value along with class label is appended in training dataset.

4. SIMULATION AND MODELING

This segment presents two ANIDM that utilizes diverse varieties of the K-nn classification algorithm. This paper depicts the identification techniques and anticipation strategies arrangement Modeling of Adaptive Network Intrusion framework for dynamic condition.

4.1 Detection Results of DOS Attacks in KDD Dataset Using Anipm KNN-Based IDS

This section presents two NIDM that utilizes diverse varieties of the K-nn classification algorithm as a Data mining. This paper depicts the identification techniques and anticipation strategies arrangement Modeling of Adaptive Network Intrusion framework for dynamic condition.

4.1.1 Unsupervised KNN-based NIDS

The KNN algorithm can be utilized as a part of unsupervised, learning as a clustering system presenting the grouping KNN algorithm in interruption detection where it regards meddling practices as anomalies to the general information. Accepting that the assaults are few focuses that lie in an inadequate area of the feature space, the grouping KNN calculation marks interruptions in light of the total of KNN separations of all test focuses.

4.1.2 KNN Classifier and clustering-based IDS

In classification and clustering, the information comprises of a preparation set and a test set. The preparation set is an arrangement of N includes vectors and their class names; and a learning calculation is utilized to prepare a classifier utilizing the preparation set. The test set is an arrangement of highlight vectors to which the classifier must relegate names. A natural method to choose how to arrange an unlabeled test thing is to take a gander at the preparation information focuses adjacent and make the classification as indicated by the classes of those close by named information focuses.

4.1.3 ANIDM Detection and Prevention System

To address the core problem of a network intrusion detection system (NIDS), the accuracy of the detection results, researcher proposes ANIDM K-NN-based NIDS that possesses highly accurate intrusion detection capability. The current work excludes application information from network context and limits itself to operating system, prevention and alerts system. A model that relies on contextual information of both the network and the exploited vulnerability. The model is designed using three selected network parameters: average packet time, number of packets sent and that received. The researcher therefore seeks to develop a model a self-learning and adaptive model in presence DOS and other malicious activities.

4.2 Data Source NSL-KDD Dataset

The NSL-KDD dataset is a typical benchmark for assessment of interruption location methods. This is the informational collection utilized for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with NSL-KDD dataset. The opposition errand was to fabricate a system interruption locator, a prescient model equipped for recognizing ``bad" associations, called interruptions or assaults, and ``good" ordinary associations [19]. Analysis and extraction data from NSL KDD dataset using WEKA tool datamining.
https://github.com/defcom17/NSL_KDD,

4.2.1 DOS in KNN classification and clustering.

Test and Training must be networks with a similar number of segments. Since the researcher have worked on DoS attacks only and therefore, give a brief background about Smurf, Teardrop, Neptune, PoD, Back and LAND attacks for testing our detection Model:

- i. Smurf: - Smurf attacks in KDD dataset use ICMP echo request packets directed to IP broadcast addresses from remote locations to create DoS

attack. It can be identified by watching large number of Echo requests and replies from the victim machine. The column 'count' values from the dataset can be read and analyzed for Smurf attacks. In our experimentation the searcher has considered 280790 Smurf attacks in total.

- ii. Neptune: - To initiate Neptune attack a large number of SYN packets are sent to target machine to exhaust its buffer. Neptune attack never establishes the TCP session resulting in many zero packets in each connection attempt. In our experimentation the researcher has considered 107201 Neptune attacks in altogether.
- iii. Ping-Of-Death (POD):- POD affects older Operating Systems. It uses oversized IP packets to crash, freeze or reboot the system. During the experimentation POD affected none of the victim systems. ICMP packets longer than 64000 bytes can be due to POD attack. In our experimentation the researcher has considered 264 POD attacks in total.
- iv. Back: - In Back attacks the wrong IP addresses are used by the attacker in the source IP address of the IP packet header. As a result, the receiver fails to determine the real attacking node. Since the attacker node cannot be located therefore the attacker cannot be stopped from sending illegitimate packets. The receiver thus gets inundated by unwanted packets and fails to provide service to regular users thus causing denial of service. the researcher has considered 2023 in total.
- v. LAND (Local Area Network Denial): - A LAND attack is a DoS attack under which spoofed packets are sent to the target computer to bring it down. In this attack large number of TCP/SYN packets are sent to the target machine. These attacks are different from the SYN flooding attacks because in these attacks the spoofed IP packets have both the source as well as the destination IP addresses as the target machines IP address. The machine therefore fails to provide service to legitimate users. the researcher has considered 21 LAND attacks in total.

4.3 Anomaly IDS training using KNN Algorithm

Researcher use KNN classifier to classify the incoming requests as anomaly or normal. Based on the obtained values the users sending anomaly requests are sent warning. Training a KNN classifier when controller starts Counting the number hosts connected per 10 seconds. Then obtained value is classified by KNN based on the distance and this value along with class label is appended in training dataset.

```

if packet is classified as anomaly then
    warning is given to the user else
        entry is made in the table of a switch and sent
            through the port to a destination
end if.
    
```

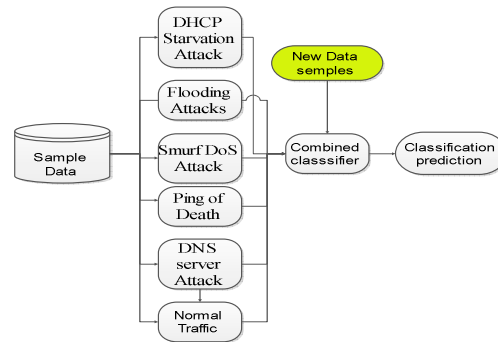


Fig 4: Testing the model training, using new Dataset.

4.3.1 Data Preprocessing

The researcher has performed reduction of dimensionality of the NSL-KDD dataset. It is an important step, not only to reduce the complexity of the training process but also to gain an insight as to which network connection features are significant for the process of any network intrusion detection. Having done that, these features are real numbers on different scales collected. TCPDUMP files are converted to arff and csv format file.

4.3.2 Testing Data Set

The training data consists of around 50,000 normal connection records extracted randomly from the NSL- dataset and all the attack connection records with a maximum of 1% per attack type. Thus, a training data file with around 50,000 normal instances and 5000 attack instances is obtained. The testing data file is created in the same manner from the 'corrected' file. Thus, the study is training the anomaly detection system on some number of attacks and testing it to check whether it is able to detect DOS unseen attacks which exist in the testing file.

4.3.3 Experiment & Result tabulation

Training and Testing Data Set

Data required for training the unsupervised anomaly-based intrusion detection system is taken from the 'NSL dataset'. It is assumed that the relative amount of attacks in the training data is very small compared with normal data, a reasonable assumption that may or may not hold in the real-world context for which it is applied. If this assumption holds, anomalies and attacks may be detected based on cluster sizes as shown below

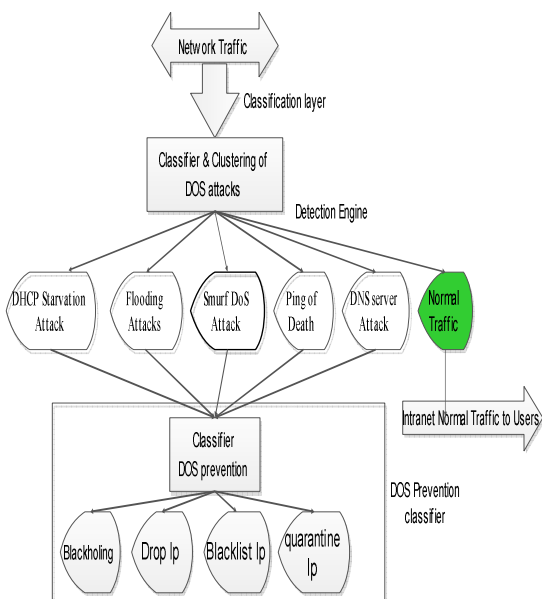


Fig 5: Architecture classification model.

4.4. Data Set Training

- i. Learning storing all training instances
- ii. Classification an instance gets a classification equal to the classification of the nearest instances to the instance.
- iii. Instance-based learning is simple, efficient and accurate approach to concept learning and classification.
- iv. Given a training set, study can use KNN to predict the class of a previously unseen instance by comparing it to other points in the space.
- v. In the k-nearest neighbour algorithm, k is a user defined parameter, where k is a positive integer, usually small. Let's choose k=1. In this case study refer to 1NN as the nearest neighbour algorithm. Choosing a suitable k is both an art and a science. Both low and high values of k have their own advantages.

4.4.1 Preprocessing

The researcher has performed reduction of dimensionality of the NSL-KDD dataset. It is an important step, not only to reduce the complexity of the training process but also to gain an insight as to which network connection features are significant for the process of any network intrusion detection. Having done that, these features are real numbers on different scales collected. TCPDUMP files are converted to arff and csv format file.

Table 3: DOS Attacks in Dataset

DOS attacks	Counter
Smurf	280790
Neptune	107201
Ping-Of-Death (POD)	264
Back	2023
LAND	21 connection fail

4.4.2 K-means algorithm

K-means algorithm is useful for undirected knowledge discovery and is relatively simple. K-means has found wide spread usage in lot of fields, ranging from unsupervised

learning of neural network, Pattern recognitions, Classification analysis, Artificial intelligence,

Table 4: Train and Test Datasets for Detection of DOS

Experiment Analysis			
training dataset	normal	9041	670
	anomaly	4714	8119
Test dataset	normal	63106	4237
	anomaly	7834	50796

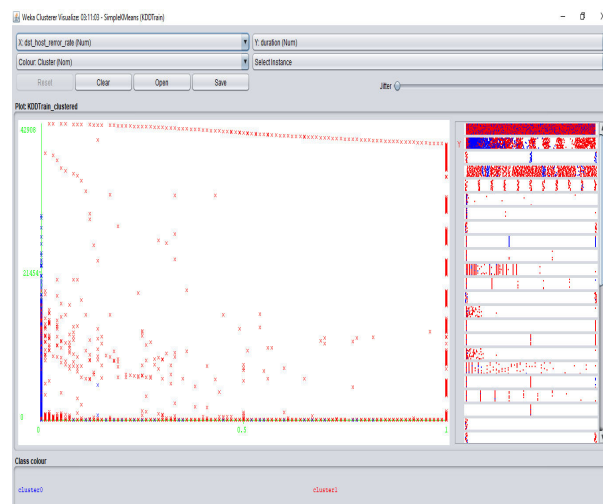


Fig6: Dst-host rate via Duration Train Dataset and Test.

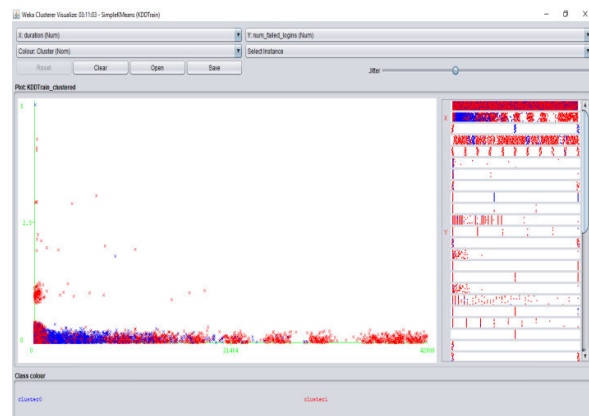


Fig 7: Testing failed login Duration via Num-failed login

4.5 Comparison of Classifiers

```

Tester: weka.experiment.PairedCorrectedTTTester -G 4,5,6 -
D 1 -R 2 -S 0.05 -result-matrix
"weka.experiment.ResultMatrixPlainText -mean-prec 2 -
stddev-prec 2 -col-name-width 0 -row-name-width 25 -mean-
width 0 -stddev-width 0 -sig-width 0 -count-width 5 -print-
col-names -print-row-names -enum-col-names"
Analyzing: Percent correct
Datasets: 1
Resultsets: 3
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 7/10/18 10:33 AM
Dataset (1) rules.On | (2) rules (3) bayes
    
```

kdd_cup_1999 (100) 99.94 | 75.98 * 99.86
(v/*) | (0/0/1) (0/0/1)

Key:

- (1) rules.OneR '-B 6' -3459427003147861443
- (2) rules.ZeroR " 48055541465867954
- (3) bytes.NaiveBayes " 5995231201785697655

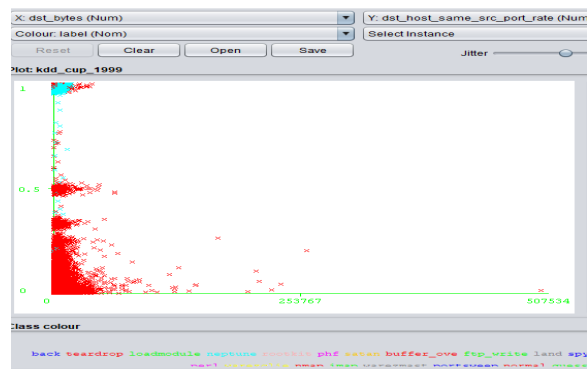


Fig 8: Scatter plots for duration against Dst bytes

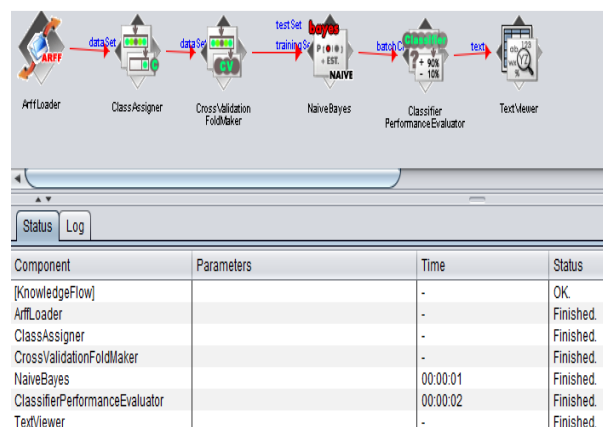


Fig 9: Knowledge flow

=== Evaluation result ===

Scheme: NaiveBayes

Relation: kdd_cup_1999

Correctly Classified Instances	15389	99.8572 %
Incorrectly Classified Instances	22	0.1428 %
Kappa statistic	0.9961	
Mean absolute error	0.0001	
Root mean squared error	0.0111	
Relative absolute error	0.39 %	
Root relative squared error	8.8425 %	
Total Number of Instances	15411	

4.5.1 KNN cluster plots

Unsupervised learning means that there is no outcome to be predicted, and the algorithm just tries to find patterns in the data. In k means clustering, we have to specify the number of clusters we want the data to be grouped into. The algorithm randomly assigns each observation to a cluster, and finds the centroid of each cluster.

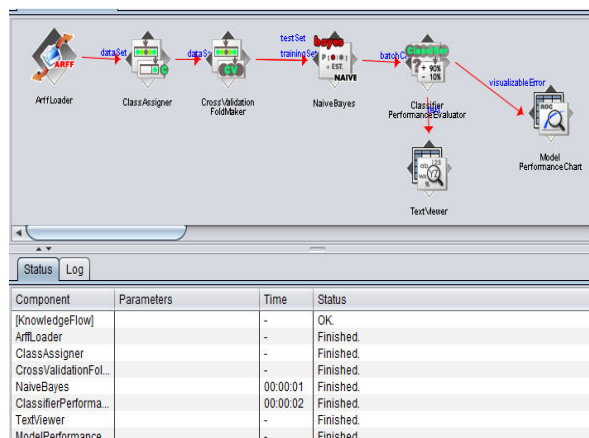


Fig 10: Predicted of normal and abnormal traffic flow

4.5.2 Distance measured by Euclidean distance

$$\text{sim}(X, D_j) = \frac{\sum_{t_i \in (X \cap D_j)} x_i \times d_{ij}}{\|X\|_2 \times \|D_j\|_2}$$

$$\text{dist} = \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$$

$$\sqrt{(\text{Duration} - \text{scr bytes})^2 + (\text{dstbytes} - \text{wrongfragments})^2}$$

$$\sqrt{(\text{Airtargent} - \text{Numfailedlogin})^2 + (\text{login} - \text{wrongfragments})^2}$$

$$\sqrt{(\text{count} - \text{srvcoun})^2 + (\text{dsthostcount} - \text{dsthostsrv})^2}$$

Time taken to build model (full training data) : 1.93 seconds

- Relation: kdd_cup_1999
- Correctly Classified Instances 15389
99.8572 %
- Incorrectly Classified Instances 22 0.1428 %

X – test Packets set, D_j – jth training Packets set,

t_i – attributes of DOS

X and D_j , x_i – weight of attribute t_i in X .

D_{ij} – weight of attributes t_i in D_j .

4.5.3 Euclidean calculation clusters in next column

Table 5: Euclidean centroids and clusters centers

Final cluster centroids:	Cluster#		
Attribute	Full Data	Cluster 0	Cluster 1
	125973	24039	101934
duration	287.1447	316.9924	280.1057
protocol_type	tcp	udp	tcp
service	http	domain	http
flag	SF	SF	SF
src_bytes	45566.743	12788.22	53296.87
dst_bytes	19779.1144	3497.02	3497.02
wrong_fragment	1	0.0227	0
urgent	1	0.0001	0.0001
num_failed_logins	1	0.0012	0.0012
logged_in	3	1	0
count	84.1076	92.4781	82.1335
srv_count	27.7379	105.1082	9.4917
srv_serror_rate	1	0.2825	0.3491
dst_host_count	182.1489	172.2258	184.4891
dst_host_srv_count	115.653	138.3971	110.2893
class	normal	normal	normal
Time taken to build model (full training data) : 1.93 seconds			
Dataset bayes.Naiv			
kdd_cup_1999	100	99.86	
bayes.NaiveBayes		"	
5995231201785697655			

5. CONCLUSION AND FUTURE WORK

In this paper the study Trains & test Datasets using Datamining and Artificial Intelligent to detect and mitigate DOS and DDOS attack in Dynamic Intranet Environment. From this simulation and model, it is found that, attributes packet Time to Live is dependent to Counter and Duration in the Classifier via Clustering methods. The overall accuracy of the proposed model is 99.79% with true positive rate of 99.95%, true negative rate of 99.13%, false positive

rate of 0.87% and false negative rate of 0.05%.

Therefore, the paper concludes that variation ANIDP can be considered for detection and mitigate of DoS and DDOS attacks. Previous, clustering algorithms within anomaly intrusion detection focuses on clustering based on known attributes, and not based on geographical regions. This paper argues, that classification & clustering techniques based on these common attributes, can give an incorrect view of the hidden data structure. One of the most central and unique features to base clustering on, is the source IP address. Anomaly based mitigation techniques often use this feature in clustering. These attributes include features from the TCP protocol and IP protocol. The attributes can differ in everything from source address and destination address to packet size, TTL and flags.

In the past, as the present, DOS & DDoS has been more a nuisance activity conducted by cyber vandals than an activity with specific socioeconomic aims. In the future, DOS and DDoS may be used as a disruptive force, with broad destabilization as its aim instead of the targeting of specific targets. DoS and DDOS attacks are *genuine threats* to many Internet users i.e. Annoying, Debilitating, losses. Level of loss is related to motivation as well shielding attempts from the defender. Knowledge flow in Datamining.

6. REFERENCES

- [1] Verma, K., H. Hasbullah, and A. Kumar. An efficient defense method against UDP spoofed flooding traffic of denial of service (DoS) attacks in VANET. in 2013 3rd IEEE International Advance Computing Conference (IACC). 2013.
- [2] Singh, S., et al. Intrusion Detection Based On Artificial Intelligence Techniques. in International Conference Of Advance Research And Innovation (Icari-2014). 2014.
- [3] Xu, K., Z.-L. Zhang, and S. Bhattacharyya, Internet traffic behavior profiling for network security monitoring. IEEE/ACM Transactions On Networking, 2008. 16(6): p. 1241-1252.
- [4] Tan, Z., et al., Detection of Denial-of-Service Attacks Based on Computer Vision Techniques. IEEE Transactions on Computers, 2015. 64(9): p. 2519-2533.
- [5] Dadhich, A. and S.K. Yadav, Evolutionary Algorithms, Fuzzy Logic and Artificial Immune Systems applied to Cryptography and Cryptanalysis: State-of-the-art review. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2014. 3(6).
- [6] Chang, S. and T.E. Daniels. Correlation based node behavior profiling for enterprise network security. in 2009 Third International Conference on Emerging Security Information, Systems and Technologies. 2009. IEEE.
- [7] Sharma, S.K., et al. An improved network intrusion detection technique based on k-means clustering via Naive bayes classification. in Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on. 2012.
- [8] Xu, K., F. Wang, and L. Gu. Network-aware behavior clustering of Internet end hosts. in INFOCOM, 2011 Proceedings IEEE. 2011. IEEE.
- [9] Zhao, Y. Network intrusion detection system model based on data mining. in 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). 2016.
- [10] Yang, H., et al., A survey of artificial immune system based intrusion detection. The Scientific World Journal, 2014. 2014.
- [11] Murugan, S. and K. Kuppasamy. System and methodology for unknown Malware attack. in Sustainable Energy and Intelligent Systems (SEISCON 2011), International Conference on. 2011.
- [12] Midzic, A., Z. Avdagic, and S. Omanovic. Intrusion detection system modeling based on neural networks and fuzzy logic. in 2016 IEEE 20th Jubilee International Conference on Intelligent Engineering Systems (INES). 2016.

- [13] Kusumah, P., S. Sutikno, and Y. Rosmansyah. Model design of information security governance assessment with collaborative integration of COBIT 5 and ITIL (case study: INTRAC). in *ICT For Smart Society (ICISS)*, 2014 International Conference on. 2014. IEEE.
- [14] Yan, Q. and F.R. Yu, Distributed denial of service attacks in software-defined networking with cloud computing. *IEEE Communications Magazine*, 2015. 53(4): p. 52-59.
- [15] Buczak, A.L. and E. Guven, A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys & Tutorials*, 2016. 18(2): p. 1153-1176.
- [16] Sou, S.I. and C.S. Lin, Random Packet Inspection Scheme for Network Intrusion Prevention in LTE Core Networks. *IEEE Transactions on Vehicular Technology*, 2017. 66(9): p. 8385-8397.
- [17] Stampar, M. and K. Fertalj. Artificial intelligence in network intrusion detection. in *Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015 38th International Convention on. 2015.
- [18] Goldstein, M., et al. Bayes Optimal DDoS Mitigation by Adaptive History-Based IP Filtering. in *Seventh International Conference on Networking (icn 2008)*. 2008.
- [19] Farid, D.M. and M.Z. Rahman, Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm. *JCP*, 2010. 5(1): p. 23-31.
- [20] Degeler, V., R. French, and K. Jones. Self-Healing Intrusion Detection System Concept. in *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity)*, *IEEE International Conference on High Performance and Smart Computing (HPSC)*, and *IEEE International Conference on Intelligent Data and Security (IDS)*. 2016.
- [21] Lorandel, J., J.C. Prévotet, and M. Hélar. Efficient modelling of FPGA-based IP blocks using neural networks. in *2016 International Symposium on Wireless Communication Systems (ISWCS)*. 2016.
- [22] Kwon, D., et al., A survey of deep learning-based network anomaly detection. *Cluster Computing*, 2017.
- [23] Frias-Martinez, V., et al. A network access control mechanism based on behavior profiles. in *Computer Security Applications Conference, 2009. ACSAC'09. Annual. 2009. IEEE*.
- [24] Kumar, M., M. Hanumanthappa, and T.V.S. Kumar. Intrusion Detection System using decision tree algorithm. in *Communication Technology (ICCT)*, 2012 IEEE 14th International Conference on. 2012.
- [25] Reddy, R.R., Y. Ramadevi, and K.V.N. Sunitha. Enhanced anomaly detection using ensemble support vector machine. in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*. 2017.
- [26] Gharaibeh, M., et al. Assessing co-locality of IP blocks. in *2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHP)*. 2016.
- [27] Schinagl, S., R. Paans, and K. Schoon. The Revival of Ancient Information Security Models, Insight in Risks and Selection of Measures. in *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 2016. IEEE.
- [28] Jyothsna, V., V.R. Prasad, and K.M. Prasad, A review of anomaly based intrusion detection systems. *International Journal of Computer Applications*, 2011. 28(7): p. 26-35.
- [29] Sahu, S.K., S. Sarangi, and S.K. Jena. A detail analysis on intrusion detection datasets. in *Advance Computing Conference (IACC)*, 2014 IEEE International. 2014.
- [30] Chandrasekhar, A. and K. Raghuvver. Intrusion detection technique by using k-means, fuzzy neural network and SVM classifiers. in *Computer Communication and Informatics (ICCCI)*, 2013 International Conference on. 2013. IEEE.
- [31] Cui, X., J. Beaver, and T. Potok. Swarm-Based Knowledge Discovery for Intrusion Behavior Discovering. in *2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. 2010.
- [32] Varuna, S. and P. Natesan. An integration of k-means clustering and naive bayes classifier for Intrusion Detection. in *Signal Processing, Communication and Networking (ICSCN)*, 2015 3rd International Conference on. 2015.
- [33] He, D., et al., Software-Defined-Networking-Enabled Traffic Anomaly Detection and Mitigation. *IEEE Internet of Things Journal*, 2017. PP(99): p. 1-1.