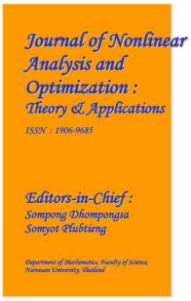
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1 : 2024
ISSN : **1906-9685**



INTRUSION DETECTION SYSTEM IN COMPUTER NETWORKS BASED ON MACHINE LEARNING ALGORITHMS

Dr Chandrasekharan Dinesh, Associate Professor, Department Of Information Technology, VIIT(A), Visakhapatnam, AP, India
Smruti Samantra, K. Sivasai, G. Uday Kiran, F.Sharief, S. Sai Ram Kiran, Student, Department Of Information Technology, VIIT(A), Visakhapatnam, AP, India

1.ABSTRACT
Technology for network security has become essential for protecting government and meticulously calculated structures. Complex requirements are faced by ultramodern intrusion finding activities, which call for dependability, extensibility, ease of management, and minimal conservation costs. In recent years, intrusion detection systems based on machine literacy have proven to be very delicate, adept at identifying novel forms of intrusion, and able to adapt to shifting conditions. In order to provide a stopgap for future intrusion discovery system establishment, this work compares the efficacy various machine literacy types in intrusion discovery systems, including artificial neural networks and support vector machines. In contrast to other related workshops on machine literacy-based intrusion sensors, we suggest utilizing slice varying rates of normal data for each dimension to determine the mean value.

2.PREFACE
One essential resource that helps associations thrive in their responsibility is the information contained in IT products or systems. Additionally, individuality has a legitimate hope that any personal data they may have in IT systems or products would be confidential, be accessible to them upon request, and not be altered by unauthorized parties.
IT systems and products should carry out their intended purposes while maintaining appropriate control over the data to ensure that it is protected against risks like unauthorized or undesired distribution, alteration, or loss. The prevention and reduction of these and similar risks are encompassed under the umbrella of IT security (1). It' crucial that a system's security features be built to prevent unwanted access to its resources and data. However, completely preventing security breaches from occurring.
Attempts at subversion can be dealt with in two ways. Building a totally safe system is one method to aid subversion itself (4). For example, the network director may require all drug users to identify and verify themselves; the director could protect data using vibrant cryptographic designs and extremely stringent access control methods. However, this isn't actually feasible for the following reasons: 1. In reality, creating a completely safe system is impossible as error-free software is still a pipe dream and nobody wants to go to the work of trying to create software that is equivalent. In addition to security concerns with their email software, drug users also feel as though they aren't receiving their money's worth when they purchase software. Consequently, creating and implementing an entirely secure system.

Advantages:

1. It can determine whether the host has been more directly encroached upon. It can more accurately assess network assaults or host intrusions since its data is derived from system inspection records and host system logs, which it compares with network-grounded intrusion discovery systems
2. It has the ability to detect assaults on translated network landscape. The data is unaffected since it originates from system files and is transferred across a network that translates it before hosts decode it.

3. A new tackle is not necessary. All that is needed, in lieu of new equipment, is the deployment of the monitoring system on the assigned hosts.
4. Increased expense. A monitoring system must be installed on each host, and since each host is different, so too are the inspection files and log patterns, resulting in a unique intrusion detection system for each host. It may have an impact on covered hosts' systems' efficacy. Host system sources may be entangled by an intrusion detection system in a covered state. Predicted on networks, IDS (7) Similar to Internet packets, the majority of its data is gathered by a general network aqueduct that travels through a network corridor. And the following are its benefits and drawbacks:
5. Affordable prices All Ethernet assaults can only be detected by network-based intrusion detection systems (IDS), and they only expense the device itself.
6. Like denial of customer service, it can detect assaults that host-predicated intrusion detection systems are unable to detect.

Disadvantages:

1. Due to the huge flux, it may not be able to decode each signal in the connection and may lose some.
2. To analyse bulk information across a massive the internet, additional memory capacity and a CPU with a faster firing rate are needed.
3. It is unable to handle repeated packets and might not allow attack information included in such packets.

b) type based on an alternative analytical system Abuse Investigation (7) Another term for it is "hand-predicated discovery", which has the ability to convert attack symptom or policy-defying information Already treated case information should be examined to the palm of handed database in order to determine the likelihood that it is an attack. Cases that match attacking hand data can be classified as attacks. Its advantages include a high rate of discovery and a low rate of false alarms for known assaults; but, its capacity for discovering unknown discovery styles is minimal, and the attack databases has to be regularly refreshed. Anomaly Finding When discovery is performed, the lives are compared with factual stoners' data; in the event the neutralize is below threshold value, the user's behavior can be considered normal and it has no intention of attacking; if the neutralize is above threshold value, the user's behavior can be considered malicious. It may also establish a life expectancy for normal stoner behavior based on statistics data of stoners from previous periods.

disadvantage of one another, similar to MINDS (8). IDS functions similarly to an internet monitoring and alarm system, monitoring and analyzing potential online threats, promptly sounding the alert before assaults pose a risk, implementing appropriate countermeasures, and minimizing the likelihood of larger losses. Additionally, certain techniques rely on pattern checking, which has a low mistake rate. However, the pattern-based technologies need to be updated often, as comparable systems lack the capacity to detect new and emerging attack methods. Nowadays, machine learning is heavily used in many fields, such as handwriting, professional opinion, physiological hand insulation, please quest machines, and pronunciation identity, etc. Modern machine learning and data mining methods have been deployed to intrusion detection in a multitude of investigations lately. These algorithms are capable of analyzing large amounts of data, and similar technologies have a higher capacity for discovering undiscovered assaults. Even if there have been some dissertation accomplishments recorded, there is still much room for growth.

In same circumstances with identical settings, how successful are the various machine learning approaches when used for intrusion detection? What more styles exist except the ones mentioned above? Therefore, in order to provide potential suggestions for improvement and serve as a guide for building intrusion detection systems, this dissertation compares the efficacy of two well-known machine learning techniques—artificial neural networks and support vector machines—when applied to intrusion discovery.

.

3.COMBINED WORK

Here are several methods for solving intrusion detection issues. By using frequent occasion methods and association rules to system inspection data, Lee et al. (2009) created an intrusion discovery model. Item restrictions in the form of axis particularity(s) are only utilized to decipher suitable patterns; low frequency patterns are discovered semi-automatically using an iterative, level-wise approximation

mining technique. The NIDES system uses statistical techniques to carry out anomaly finding (10). It creates lives via the use of statistical metrics that factor in subject effort and profile creation. There are four main categories of statistical measures: ordinary, classification, assessment data distribution, and endurance intensity.Learning to recognize intrusion systems is done by neural networks. The construction of an n-caste network and the definition of abstract instructions in terms of sequencing
There exist several approaches to address intrusion detection problems. Lee et al. (2009) developed an intrusion finding model based on system examination information using numerous occasion algorithms and correlation criteria. Only appropriate patterns may be deduced from item limits in the form of axis particularity(s); low frequency patterns are found semi-automatically by a recursive, level-wise estimation mining process. The NIDES system provides anomaly discovery by statistical methods (10). Through the application of statistical measurements that take into account respondent engagement and picture creation, it builds lives. Statistical measurements fall into four primary classes: ordinary, classification, persistence quantity and and assessment distribution of information.The neural networks are used for recognizing intrusion solutions. The formulation of abstract commands and the building of an n-caste network .

Intrusion Dataset:

Details 1998 DARPA(KDD-mug collection)(18) intrusion discovery assessment study involved setting up alandscape in order to spoof a normal US Air Force LAN and get raw TCP/IP dump data for a network. Although the LAN was run like a real landscape, it was under constant attack. Forty-one vibrant numeric and qualitative elements were removed for every TCP/IP connection. The training sample of 494014 entries from the said database was employed; around 20 of these records correspond to typical patterns (Table 1). In fact, 311029 informational records made up the test set (see Table 2). The following are the four distinct assault pattern orders (19). It is noteworthy to remark that we have shown the recommended literacy to be capable in this paper. (see Table 2). The following are the four distinct assault pattern orders (19). It's crucial to note that we have shown in this research how the recommended literacy system may identify aberrant behaviors through regular behaviors.

Probing:

A type of attack called as probing occurs when a bushwhacker searches the system for known weaknesses or knowledge. A graphic representation of available computers as well as features on a network can be used by a bushwhacker to find vulnerabilities. Exams come in a variety of forms; some misuse a machine's legitimate functions, while others employ social engineering methods.The most common kind of assault, this one involves very little technical maneuvering.

Denial of service(DOS) attacks:

Attacks known as denial of service (DoS) occur when a hacker causes a computer or memory resource to become excessively busy or full to process legitimate requests, preventing legitimate drug users from using the device. DoS attacks can be initiated in a variety of methods, such as by misusing the computer's authorized functions, focusing on execution vulnerabilities, or taking advantage of system errors. DoS assaults are categorized according to the services that a bushwhacker makes inaccessible to lawful drug dealers.

Attacks from stoner to root (U2R):

Stoner to root exploits are a type of attack in which a bushwhacker initially gains access to a regular stoner account on the system, which they may then use to target vulnerabilities and take control of the system as root. The most prevalent vulnerabilities in this type of attacks are routine buffer overflows, which result from frequent programming errors and fictional terrain.

Remote to stoner(R2L) attacks:

A remote to stoner attack is a type of cyberattack in which a hacker uses a network to deliver packets to a system and then takes advantage of that machine's vulnerability to obtain initial access as a stoner in an unethical manner. While there are many various kinds of R2U assaults, social engineering is the most often used attack method in this class.

Table 1 Training data set:

Class	Class name	No. of cases	%
0	Normal	97271	19.6
1	inquiry	4107	0.83
2	DOS	391458	79.2
3	U2R	59	0.01
4	R2L	1119	0.22

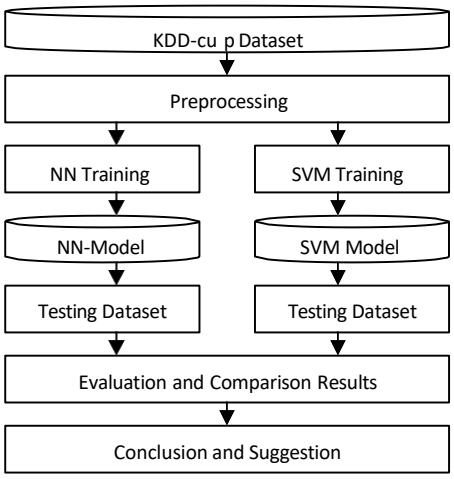
Table 2 Test data set:

class	Class name	No.of cases	%
0	Normal	60593	19.4
1	Inquiry	4166	1.33
2	DOS	231455	74.4
3	U2R	88	0.02
4	R2L	14727	4.73

System Structure:

The KDD dataset is used in this study to evaluate two different machine learning algorithms: support vector machines (SVM) and neural networks (NN). Next is a summary of the testing dataset. Fig. 1 illustrates the exploration's ongoing input.

1.The suggested system's architecture



(Fig.1)

Neural Networks:

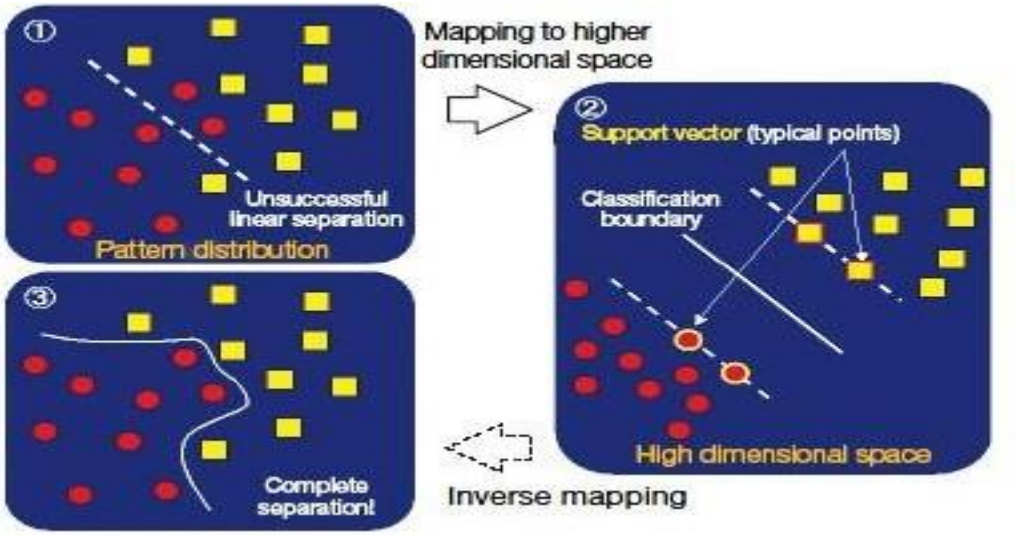
The Multilayer Perceptrons (MLP)(20) neural networks have demonstrated remarkable effectiveness in a range of tasks, yielding outcomes that are competitive or better than existing computational literacy models. As long as they have enough hidden units, they can approach any nonstop function with arbitrary delicacy. Accordingly, comparable models can operate as non-linear discriminant functions by forming any range decision limit in point space. Every part of the coordinate vector has one input knot when the NN is utilized for the structure of the bracket. For every class, there is typically one affair knot that can receive a point. The retired bumps allow the NN to create an internal version of the information collected during literacy.

MSEi=n1∑i=1/n (yi−y^i)2

The network's real affair should gradually approaches the specified affair by lowering the value of this mistake in order for it to learn effectively. In addition to calculating the error for a specific input, the reverse propagation rule backpropagates its damage in type subcaste to the primary one. To decrease error and help the network gain knowledge, connections and weights between the bumps are adjusted in accordance on the reverse-propagated damage.

Machines that provide support: Strong supervised neural network models for both classification and regression applications are support vector machines (SVMs). SVMs work especially well in high-dimensional environments in situations when there are more features than inputs in a categorization

Fig2:-Linear svm



When it comes to machine literacy problems requiring bracket and retrogression, support vector machines (21, 22) are becoming less and less common. The SVMs exhibit attractive and vibrant characteristics along with strong generalization abilities when compared to other classifiers. In fact, the armature doesn't need to be set up experimentally, and there are a good number of free settings to adjust.

4.ANALYSIS AND EVALUTION

This study compares neural networks as well as support vector machine learning based on their efficacy with the KDD-mug dataset. The entire set has over a million colorful, inversely spread data points. As a result, training and test datasets will be sampled in this exploratory effort. Indeed, based on the typical percentage, we select 10,000 groups of data for training and test datasets, with the regular proportion being 10, 20, and 90. We then use the remaining data to create and test vector attacks. Training and testing can start after the data has undergone pre-position correction. The following generalization may be made regarding the discovery and identification of attack and non-attack actions:

a) True positive (TP), which is the amount of assault identified when an actual attack occurs.
b) True negative (TN), which is the quantum of normal that is identified when it is normal in reality.
c) False positive (FP): A false alarm occurs when an assault is identified when it is truly normal.
d) False Negative (FN): An attack is really identified as a quantum of normal, except for certain attacks that an intrusion detection system can identify.

We evaluate delicacy, discovery rate, and false alarm rate since intrusion detection systems have high discovery rates and low false alarm rates. The comparative results of colorful assaults are then shown. Given the conditions of TP and TN, delicacy may be defined as follows: the percentage of data that is accurately identified within the total data

delicacy = (TP TN/ TP TN FP FN) * 100(5)

Table 3 summarizes the results measured by original class marker bracket.

Table 3: delicacy results of NN and SVM classifiers against KDD- mug dataset

Chance of normal data	NN(in %)	SVM(in %)
10	42.3	38.0
20	44.5	42.1
40	58.6	56.2
50	65.0	64.5
60	74.3	73.8
70	80.6	82.9
80	87.1	89.2
90	93.7	95.3
Average	66.6	65.7

The angles displayed in Fig. 4 are generated in order to evaluate and analyze the gesture of NN and SVM classifiers in terms of the delicacy criteria and throughout the complete range of normal data values. Table 1 shows that there isn't much of a difference in the two approaches' levels of delicacy, although when the amount of normal information is little, NN may perform more delicately than SVM. However, SVM outperforms NN for values greater than 70, and the NN and SVM delicacy is nearly similar when the fraction of normal information is about 50. SVMs are marginally inferior than NN classifiers in terms of overall delicacy. Discovery rate, which is another helpful metric, is the percentage of attacks found in all attack data

Discovery Rate (TP/ TP FN) * 100(6)

show the findings of the discovery rate measurements based on NN and SVM classifiers. e 4 The discovery rate findings for various normal data chances for the NN and SVM classifiers versus KDD. Results from both NN and SVM classifiers decrease in discovery rate as the likelihood of normal data increases. SVMs perform better overall than NNs, and they outperform NNs by about in terms of average value. The fraction of regular data that is mistakenly identified as an attack vector, or FP, is known as the false alarm rate. For this reason, the false alarm rate is defined as follows. explain the comparative findings of the four distinct attacks—inquiry, Dos, U2R, and R2L—that were based on NNs and SVMs in terms of delicacy, which is defined as the percentage that the kind of data is corrected categorized

a) SVM performs better than NN for inquiry attack delicacy in other situations, but NN is superior when the percentage of normal information is less than 50. The average delicacy of the SVM and NN classifiers for this kind of assault is 83.2 and 82.5, respectively.

b) SVM classifiers beat NN equivalents in every scenario for the Dos attack, with the exception of the scenario when the percentage of normal data is 70. The average delicacy of the SVM and NN classifiers for this kind of assault is 62.5 and 58.6, respectively.

C) Regarding U2R assault In general, SVM performs better than NN in terms of delicacy when the percentage of normal information is less than 60. In other situations, however, SVM is superior to NN. In this kind of assault, the typical delicacy of

Regarding R2L assault Based on the average value, the delicacy of these two types is similar. SVM is superior to NN when the percentage of normal data is 10, 40, 50, and 70. The average delicacy of the SVM and NN classifiers for this kind of assault is 14.7 and 14.6, respectively. Ultimately, Table 7 displays the comparison between the average results obtained from this effort and the results obtained from the KDD Cup 99 winner. As may be observed, KDD Winner's delicacy in Dos attacks is genuinely great, but in U2R and R2L, it performs much worse than NN and SV

Table 4: Discovery rate average results for colorful attacks through KDD Winner.

	Inquiry	Dos	U2R	R2L
KDD WINNER	83.3	97.1	13.2	8.4
NN	82.5	58.6	65.4	14.6
SVM	83.2	62.5	65.5	14.7

5.CONCLUSION

The exploratory work contrasts other attackers' delicacy, false alarm rate, discovery rate, and delicacy under various normal information proportions. Although the KDD Cup 99 dataset is now the industry standard for intrusion detection, errors may occur if only one set is utilized because of its uneven data distribution. In contrast, the investigation uses distinct normal data proportions for training and testing, finally obtaining a single average value, and hopes to obtain further objective outcomes. Comparing the performance of NN and SVM, we find that SVM is better than NN in terms of discovery, false alarm rate, and delicacy for inquiry, Dos, U2R, and R2L assaults, whereas NN is better than SVM in terms of delicacy alone.

6.UNBORN WORK

This study uses the KDD Cup 99 dataset, which is widely used as a standard dataset in several exploratory workshops. However, since 1999, there have been significant changes in network technology and attack methods, thus this dataset might not accurately represent the state of networks today. Consequently, if more recent data more accurately captures the state of the network. According to our tests and comparisons, NN is more sophisticated than SVM in terms of delicacy; yet, SVM has a higher false alarm and discovery rate. If we combine the two approaches, however, total delicacy may be significantly boosted. This investigation in Slice makes the assumption that the distribution of attack data, which differs from normal data, is in fact unreliable and will not always produce the best outcomes. As such, this should be improved and verified.

7.REFERENCES

1.The Common Criteria for Information Technology Security (CCIMB) version 2.1, specifically Part 1 covering the Introduction and General Model, is documented in CCIMB-99-031, published in Evaluation 1999.

1 R. DeMara and A. Rocke proposed a method for migrating network tempering using the dynamic dispatch of mobile agents in their 2004 paper published in Computers and Security.

3. Y. Fyodor introduced "SNORTNET," a distributed intrusion detection system, in 2000. The details are available at http://snornet.scorpions.net/snortnet.pdf.

4 J. Viega and G. McGraw discussed how to avoid security problems in software development in their book "Building Secure Software: How to avoid Security Problems the Right Way," published by Addison Wesley in 2002.

5 S. Cho described the incorporation of soft computing techniques into a probabilistic intrusion detection system in a 2002 article in IEEE Transactions on Systems, Man, and Cybernetics.

6 G. Shipley's chapter on "Intrusion Detection Systems (IDSs)" in the book "Maximum Security: A Hacker's Guide to Protecting Your Internet Site and Network," edited by Shelley Johnston Markunday and published by Sams Publication in 2001, provides insights into IDSs.