International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 1 | ISSN : 2456-3307

Machine learning algorithm for Cyber Security - A Review

Mohammad Asif, Pratap M. Mohite, Prof. P. D. Satya
Computer Science and Engineering, S.Y.C.E.T. Aurangabad, BATU, Lonere, Maharashtra, India

ABSTRACT

The computer networks are exposed to increasingly safety threats. With new kinds of attacks appearing usually, growing flexible and adaptive protection-oriented strategies is a severe undertaking. In this context, anomaly-primarily based community intrusion detection techniques are a precious era to guard target structures and networks in opposition to malicious sports. Threats the internets are posing higher threat on IDS safety of statistics. The primary concept is to utilize auditing programs to extract an in-depth set of capabilities that describe each network connection or host session and practice statistics mining applications to learn rules that correctly capture the behavior of intrusions and normal activities. Now Intrusion Detection has end up the priority and on the crucial assignment of statistics protection administrators. A device deployed in a network is at risk of numerous assaults and desires to be blanketed towards assaults. Intrusion detection machine is a necessity of these days' information safety area. It performs a vital function in detection of anomalous site visitors in a community and indicators the network administrators to manage such visitors. The painting supplied in this thesis is an attempt to locate such visitor's anomalies in the networks through generating and reading the site visitors float information.

Keywords : IDS (Intrusion Detection System), HIDS (Host Based Intrusion Detection System), ML (Machine Learning), NIDS(Network Based Intrusion Detection System)

I. INTRODUCTION

This IDS supplied in this thesis implements the k-means method of data mining for intrusion detection and the outlier detection technique the usage of community outlier factor to locate the anomalies present within the site visitors glide. The performance of these two processes is compared through numerous confusion matrix and overall performance metrics like fake superb charge, sensitivity, specificity, type charge and precision, and an evaluation is accomplished to find out that which one of the tactics is higher for use for intrusion detection using site visitor's waft. Intrusion detection is the hassle of identifying unauthorized use, misuse, and abuse of computer systems by both machine insiders and outside penetrators. Namely, the accelerated connectivity of laptop systems gives greater get admission to outsiders and makes it less complicated for intruders to avoid detection. IDS are based on the perception that an intruder's conduct can be exceptionally extraordinary from that of a legitimate consumer. In view that many modern IDSs are built by using manual encoding of expert expertise, modifications to IDSs are high priced and sluggish. in this paper, we describe a records mining framework for adaptively building Intrusion Detection (ID) fashions. This record attempts a comprehensive compilation and categorization of available intrusion detection gadget (IDS) products. lengthy studied and prototyped in academic and authority's circles, IDS have most effective within the last few years all started to emerge as a feasible and useful commercial alternative. The primary industrial

IDS product becomes launched in 1991, with a relative handful emerging within the subsequent half of dozen years. Then, inside the remaining couple of years, the sphere underwent explosive increase. Even after the obvious failure of numerous early releases, there nevertheless remains at the least seventeen extant products that claim to provide powerful intrusion detection in a networked environment. Given this recent growth, and the reported improved utility of corporate assets to these products [1, 2, and 3], the time has come for a complete assessment of the subject.

Intrusion detection structures try to discover laptop misuse. Misuse is the overall performance of an motion that isn't always favored by the machine proprietor; one that doesn't comply with the device's ideal use and/or security coverage. IDSs mechanically examine online person hobby for forbidden (i.e., invalid) and anomalous (i.e., strange, inconsistent) conduct. they're primarily based on the speculation that monitoring and reading community transmissions, machine audit information, software audit data, device configuration, facts files, and different information can locate misuse. This record encompasses widespread portions of information, effective analysis calls for detection1 specialized and constantly honed expertise, and at the least close to real-time of misuse is frequently important.

In [4], the subsequent characteristics are identified as suited for IDS:

- ✓ It needs to run constantly with minimal human supervision.
- ✓ It has to be fault tolerant in the sense that it has to be capable of recover from gadget crashes, both unintended or because of malicious activity. Upon startup, the IDS have to be capable of recovering its previous country and resume its operation unaffected.
- ✓ It ought to face up to subversion. The IDS need to be able to screen itself and stumble on if it's been modified by way of an attacker.
- ✓ It must impose a minimal overhead at the system wherein it is walking; with a view to no longer intervene with its ordinary operation.
- ✓ It needs to be capable of being configured in step with the security policies of the gadget this is being monitored.
- ✓ It ought to be able to adapt to changes in device and person conduct over the years (e.g., new applications being mounted, customers converting from one hobby to another or new assets being to be had those purpose modifications in gadget aid utilization patterns).
- ✓ Anomalies are deviations from everyday consumer behavior. Misuses, alternatively, are recognized patterns of attack [5]. at the same time as misuse styles are frequently easier to process and locate, it is frequently the paradox patterns with a purpose to help to discover problems. As misuses are recognized styles of assault, the detection machine tends to fail whilst novel assault techniques are implemented. Detection of anomaly patterns is computationally pricey due to the overhead of keeping track of, and possibly updating several machine profile metrics, because it need to be tailor-made gadget to device, and every now and then even used to a person, due to the reality conduct styles and gadget usage vary significantly.
- ✓ Different IDSs had been designed to do dispensed collection and analysis of records. A hierarchical device is defined in [6], and [7] describes a cooperative gadget without a central authority. these structures clear up maximum of the problems mentioned except for the reconfiguration or adding skills to the IDS, which aren't described in either of the 2 designs.
- ✓ The boom of the net has added excellent benefits to society at the identical time the growing attacks on the IT Infrastructure are

getting an increasing number of serious issues and wishes to be addressed. Together with the growth of the internet, attacks also are growing in parallel.

- ✓ Host-based intrusion detection system pursuits at detecting the intrusions at the host stage. They function in my opinion at each host of the community. Consequently, they are able to operate on distinctive type and quantity of records on the equal time. The host-based machine is dependent on the host operating system. Any vulnerability inside the host-primarily based OS can weaken the integrity of the host-primarily based.

An outsider can exploit this vulnerability to release an assault on the way to be hard to be detected through the HIDS. as a result, a robust HIDS have to be supported with the non-susceptible host OS. The bodily deployment of HIDS in a community may be visible the traffic enters from the internet via the router or switch or firewall to the local network. It is the responsibility of the HIDS in the host to detect possible intrusions in the traffic flow.

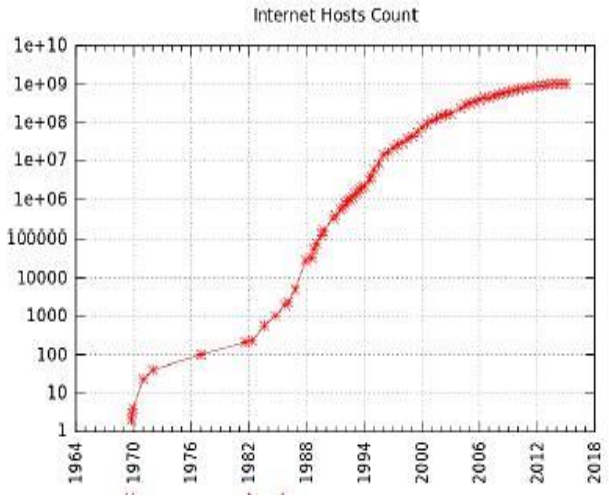

Figure 1 : Host based intrusion detection system

II. RELATED WORK

The growth of Internet has brought great benefits to the society at the same time the growing attacks on the IT Infrastructure are becoming an increasingly serious issue and needs to be addressed. Along with the growth of Internet attacks are also growing in parallel. In earlier days, the attacker should have a good knowledge about the target infrastructure and knowledge on the Network, Operating Systems & Applications. Whereas today there are lots of open tools available in the Internet which can trigger automated attacks.

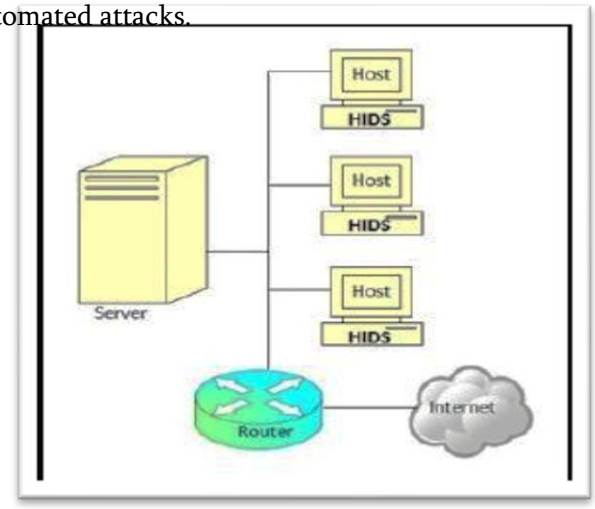

Figure 2. Growth of Internet in terms of Host Count

WannaCry attack: In might also 2017, the WannaCry Ransomware spread via the net, the use of a make the most vector named EternalBlue. The ransomware attack infected greater than 230,000 computer systems in over a hundred and fifty countries the usage of 20 specific languages to demand cash from users the usage of Bitcoin crypto currency. WannaCry demanded US$300 in line with pc. [9]

Petya attack: Petya computer virus spread in the course of April 2016, this malware infected the grasp boot record of the laptop through encrypting the report tables of the NTFS file system. as soon as infected on the next boot expects a ransom is paid. again in the month of June 2017, a modified model

of Petya the usage of Eternal Blue exploit and this was aimed to create disruption alternatively to generate income. [9]

We located that there are many benefits of C4.5algorithms for special attacks occurs on your dataset and C4.5 will detect the R2L and U2R attacks and the neural network is the use of for detecting the DOS and Probe attacks and many greater. In this, we are the usage of four algorithms. The primary one is okay-manner clustering and the second steps are fuzzy good judgment 1/3 steps are SVM and the remaining and very last step is C4.5. The blessings of all this set of rules are to detect the assaults from the datasets. After that, it will examine it with the SVM and C4.5 classifiers to find out how many attacks are coming about in the dataset at the same time as transferring the dataset from supply to destination. [8]

This center also evolved a subsequent-era mechanism which includes audit profiles of person's and may monitor the modern-day reputation of the consumer, if any change takes place with person's hobby as compared with audit profile of person then it will generate an alarm.

Haystack [10] later developed a framework to estimate an intrusion detection approach based totally on person and anomaly strategies. Six varieties of intrusion have been detected and people include the masquerade assaults, malicious use, leakage, carrier denial, the unauthorized consumer's wreck-ins try, and get admission to control of a protection system. The source fire developed suggests a community-based intrusion detection and prevention mechanism referred to as snicker machine that is an open supply. Forrest [11] in 1996 created an ordinary profile based on studying the call sequences between intrusion detection and protection in opposition to a human machine. An assault on this device is taken into consideration as the collection deviation from normal profile sequence. thus, this device works offline the use of previously accrued information and implements the view desk set of rules for gaining knowledge of software profiles significantly. Duan et al. [12] have concentrated on identifying compromised machines which can be recruited to hit upon junk mail zombies. An method SPOT is proposed to experiment sequentially outgoing messages by means of enforcing SPRT (Sequential possibility Ratio check). This method quick estimates whether a number is compromised or not. figuring out compromised machines using malware contamination device is said via Bot hunter [13].

This machine has massive no of steps that allow intrusion detection alarms correlation caused using inbound visitors with outgoing message trade sample consequences. Bot Sniffer [14] explained in his paintings approximately compromised device traits which might be a uniform temporal-spatial conduct for detecting zombies. This approach identifies zombies by using combining flows based totally on server connections and searching flows with comparable conduct respectively. Kumar and Goyal [15] have explained implements genetic algorithms in dataset training to categories the labels that are smurf attacked and achieves a low false high-quality ratio of 0.2%. in addition, paintings were completed through Abdullah [16] and co-people elaborated intrusion detection class regulations using genetic algorithms. Intrusion detection policies the usage of genetic algorithms becomes also the have a look at made through Ojugo et al. [17]. This approach uses health characteristic for estimating the rules. system learning strategies are also applied to hit upon the intrusion. current device mastering strategies (artificial Neural Networks - ANN) for intrusion detection was described by way of Roshani group [18]. Gaikwad et al [19] added a technique based on fuzzy clustering and ANN approach.

PROBLEM STATEMENT :- A few of the existing network- and host-based totally IDSs [20, 21] perform records collection and analysis centrally using a monolithic architecture. by using this, we mean that the facts is accrued through a single host, both from audit trails or via tracking packets in a network, and analyzed by means of a unmarried module the usage of one of a kind techniques. other IDSs [22, 23] perform dispensed facts collection (and some preprocessing) with the aid of the use of modules disbursed inside the hosts which are being monitored, however the accumulated statistics remains shipped to a crucial vicinity in which it is analyzed through a monolithic engine. an awesome overview of systems that take both methods is presented in [24]. There are a number of issues with those architectures:

- ✓ The imperative analyzer is a single factor of failure. If an outsider can by some means save you it from operating (for instance, by crashing or slowing down the host wherein it runs), the complete community is without safety.
- ✓ Scalability is constrained. Processing all of the statistics at a unmarried host implies a limit on the dimensions of the network that can be monitored. After that limit the vital analyzer turns into not able to preserve up with the town of statistics. Disbursed data collection can also motive problems with excessive facts visitors within the network.
- ✓ It's far hard to reconfigure or add abilities to the IDS. Changes and additions are generally carried out by way of editing a configuration report, including an entry to a desk or putting in a brand new module. The IDS commonly must be restarted to make the modifications take impact.
- ✓ Analysis of network facts can be awed. As shown in [20], appearing a set of network information in a number other than the only to which the statistics is destined can offer the attacker the possibility of acting Insertion and Evasion assaults. those assaults employ mismatched assumptions inside the community protocol stacks of different hosts to hide the assaults or create a denial of- provider attacks. different IDSs had been designed to do dispensed collection and analysis of facts. A hierarchical device is defined in [25], and [26] describes a cooperative gadget without a government. those systems resolve most of the problems cited besides for the reconfiguration or including abilities to the IDS, which aren't described in either of the 2 designs.
- ✓ Make a larger framework to resource unsupervised ML: The feature choice" step within the framework currently requires labeled datasets to discover the most discriminative talents. Destiny art work will dispose of this framework requirement to manual unsupervised ML with unlabeled records. big effort in this thesis went into ensuring experiments have been representative of actual-worldwide situations. This ensured our outcomes have been considerable in modern-day networks our datasets have been snapshots of community site visitors which may be analyzed opine in batch mode. For the detectors to artwork on a live network.

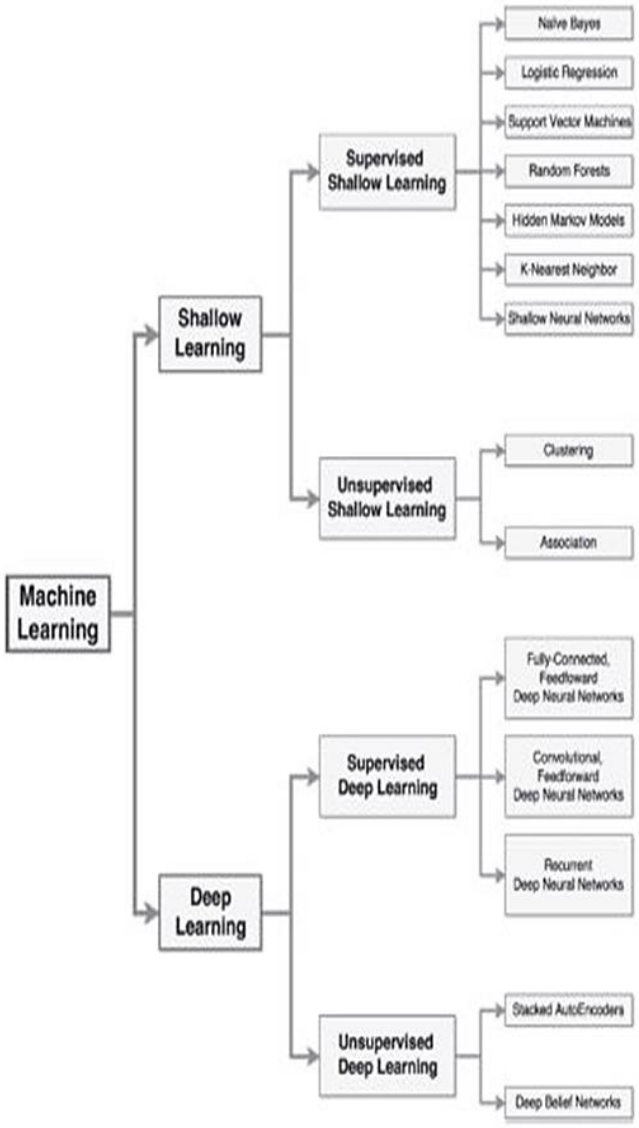

Figure 3 : Classification of ml algorithms for cyber security applications [28]

Machine learning algorithms fragmented as supervised learning and unsupervised learning.

Behave of the scheme as to be defined as, supervised learning which to be deal with more sophistication rather than unsupervised learning but it has too costly and complexity.

But in case of unsupervised learning it takes a huge amount of input but produce low amount of desirable output. The supervised scheme work under the expert system.

To overcome the limitation of this, we can go with semi supervised learning scheme. This semi supervised scheme implementing basis of K-means and KDD CUP algorithms.

Logs documents need to be tested to understand any compromised bills, originating IP's, and all sources accessed via using the attacker. All related sports want to be collected and examined several weeks or even months in advance than the detected event. capability areas of future paintings are computerized correlation and assessment of the log facts from cyber-attacks. additional machine mastering algorithm.

PROPOSED SYSTEM: -According to base paper the machine learning deal with two concept i.e. supervised learning and unsupervised learning. Both concept have some limitations, to overcome this limitation we can concern with the concept of semi supervised approach. In semi-supervised approach concern two scenarios. In this approach front-end to be implementing using of K-means Algorithm and back-end to implementing using of KDDCUP 99. In k-means algorithm concern to secure to be user interface and KDD CUP 99 concern as to centralized data storages on the system or host.

There are numerous procedures and methods utilized in id. each method has deserved and demerits. therefore, this paper highlights the similar distribution of attacks nature with the aid of using ok-way and also the powerful accuracy of the Random forest set of rules in detecting intrusions. This paper describes complete sample popularity and machine learning set of rules overall performance for the four attack categories, which include Denial-of-service (DoS) attacks (deny legitimate request to a system), Probing attacks (statistics gathering attacks), consumer-to-root (U2R) assaults (unauthorized access to nearby splendid-consumer), and faraway-to-neighborhood (R2L) attacks (unauthorized neighborhood get admission to from a faraway device) proven inside the KDD Cup 99 intrusion detection dataset.

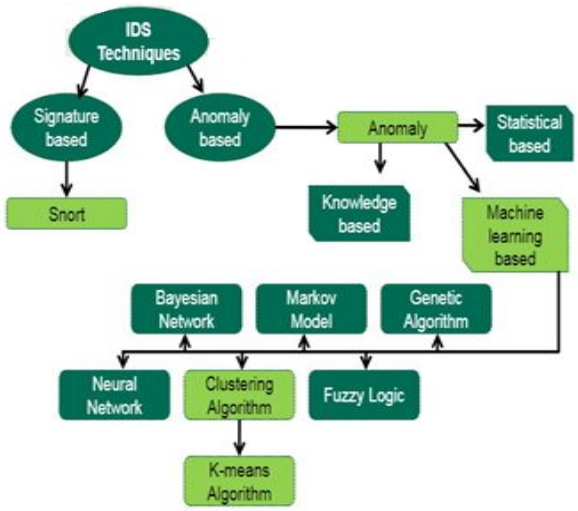

Figure 4 : - Overview of IDS

III. METHODOLOGY

This project consists of the communication of the 1 algorithm of data mining class methods. Those is k-means

K-means Clustering Algorithm: -

Clustering, primarily based on distance measurements carried out on items, and classifying gadgets (invasions) into clusters. not like type, classification because there may be no information about the label of studying statistics is an unattended getting to know system. For anomalous detection, we will use welding and in-intensity evaluation to guide the identity model. Dimension of distance or similarity performs a critical function in collecting observations into homogeneous corporations.

Jacquard affinity size, the longest not unusual order scale (LCS), is vital that the occasion is to evoke the scale to decide if regular or odd. Euclidean distance is about two vectors X and Y in space Euclidean n-dimensions, the size of the distance widely used for vector area. Euclidean distance can be defined because the rectangular root of the whole difference of the same vector dimension. Subsequently, grouping and category algorithms want to be channeled effectively, vastly, it feasible to deal with size of network facts and heterogeneity [29].

In this project, we use k-means algorithm to cluster dataset connections. The k-means set of rules is one of the widely recognized clustering tools.
K-means agencies the information according to with their characteristic values right into a person-particular variety of ok wonderful clusters. Facts categorized into the same cluster have identical characteristic values. k, the fantastic integer denoting the number of clusters, desires to be furnished earlier. the steps involved in a k-means algorithm are given consequently: [30]

1. K points denoting the data to be clustered are placed into the space. These points denote the primary group centroids.
2. The data are assigned to the group that is adjacent to
   the centroid.
3. The positions of all the K centroids are recalculated
   as soon as all the data are assigned.
4. Repeat steps 2 and 3 until the centroid unchanged.

This consequence within the partition of facts into groups. The preprocessed dataset partition is achieved the usage of the ok-means set of rules with k value as 5. due to the fact we have the dataset that contains regular and 4 assault categories which include DoS, Probe, U2R, R2L.

Algorithm 1: K-means algorithm

Input: set of connection records $X = \{x_1, x_2, \dots, x_n\}$
Number of clusters calculated using similarity method: k
Limit of iterations: MaxIteration
Output: set of centroids : $C = \{c_1, c_2, \dots, c_k\}$
Set of labels of C: $L = \{l_1, l_2, \dots, l_k\}$
For $c_i \in C$ such as $i \in \{1 \dots k\}$ **do**
  $c_i \leftarrow x_i \in X$
End For
For $x_i \in X$ such as $i \in \{1 \dots n\}$ **do**
  $d_j \leftarrow \text{Distance}(x_i, c_j)\ j \in \{1, \dots, k\}$
End For
Changed ← false;
Iter ← 0;
Repeat
  For $c_i \in C$ such as $i \in \{1 \dots k\}$ **do**
    UpdateCluster(c_i);
  End For
  For $x_i \in X$ such as $i \in \{1 \dots n\}$ **do**
    minDist ← $\text{Distance}(x_i, c_j)\ j \in \{1, \dots, k\}$
    If minDist $\neq$ d_j **then**
      $d_j \leftarrow$ minDist;
      Changed ← true;
    End If
  End For
  Iter ++;
Until changed = true and iter ≤ MaxIteration
For $c_i \in C$ such as $i \in \{1 \dots k\}$ **do**
  $l_i \leftarrow$ Labeling(c_i);
End for

KDD Cup 99 Dataset: - The assessment of any intrusion detection algorithm on actual network data is extraordinarily tough particularly because of the high fee of acquiring proper labeling of community connections. due to the actual pattern Table be gotten for intrusion detection, the KDDCup'99 datasets are used as the sample to confirm the overall performance of the misuse detection model. The KDDCup'99 datasets, referred by way of Columbia college, became arranged from intrusions simulated in military community surroundings on the DARPA in 1998.
It includes network connections obtained from a sniffer that facts all network visitors the use of the TCP unload layout. The overall simulated length is seven weeks. It was carried out in the MIT Lincoln Labs after which announced at the UCI KDD Cup 1999 Archive [32].

KDDCup'99 dataset has variations of education dataset; one is a complete education set having 5 million connections and the opposite is 10% of this schooling set having 494021 connections. since the whole dataset is huge, the test has been completed on its smaller amount of dataset this is 10% of KDD. moreover, the KDDCup'99 dataset consists of many assault behaviors, categorized into 4 companies: Probe, Denial of provider (DoS), user to Root (U2R), and far-flung to local (R2L) [33]. these may be visible in desk I. normal connections are created to profile than predicted in a military network. The detailed information of the two versions of schooling dataset may be seen in table 1.

The KDDCUP 99 was simulated in a army community environment and used for The 0.33 global understanding Discovery and information Mining equipment opposition, which became held alongside KDDCUP 99 The 5th international convention on knowledge Discovery and facts Mining. The competition challenge changed into to study a predictive model or a classifier able to distinguishing among valid and illegitimate connections in a computer community. This dataset includes one form of normal records and 24 distinctive sorts of assaults which might be categorized into four kinds such:

Table 1: Attack's categories

Attack types	List of attacks
DoS	Back, land, neptune, pod, smurf, teardrop
U2R	Buffer overflow, load module, Perl, root kit
R2L	Ftp_write, guess_pwd, imap, multihop, phf, spy, warezclient, warezmaster
Probes	Satan, ipsweep, nmap, portsweep

IV. CONCLUSION AND FUTURE SCOPE

In this project we can sophisticate more on K-means rather than KDDCUP 99, because it valuable to aspect of semi supervisor concept. This scheme facility to labeled and unlabeled dataset as efficiency.

This paper provides a comparative evaluation hybrid system getting to know approach to detect the Denial of service (DoS) assaults, Probing (Probe) assaults, consumer-to-Root (U2R) attacks, and far off-to-neighborhood (R2L) attacks. we can understand the same nature of assault organization by the use of ok-manner algorithm. And then we use to categories ordinary and attack connections. The experiments display that KDD Cup 99 dataset can be carried out as an effective benchmark dataset to help researchers evaluate distinctive intrusion detection fashions. future paintings consist of analyzing with other records mining algorithms to categories attack categories and how it could hit upon on different real-time environment dataset.

We implement this project as more feasible as the basis of K-means algorithm.

V. REFERENCES

[1] Adaptive Network Security: Solutions for Managing Risk in an Interconnected World, AberdeenGroup, Vol. 11, No. 5, January 1998.

[2] Hacker Stoppers? -- Companies Bought $65 Million Worth Of Network-Intrusion Tools Last Year, But Capabilities Still Lag What's Promised, Information Week, April 1998 http://www.techweb.com/se/directlink.cgi?IWK19980420S0066

[3] Adaptive Network Security Management: Intrusion Detection and Security Assessment Come of Age, The Yankee Group Data Communications Report, Vol. 13, No, 10, June 1998.

[4] Mark Crosbie and Gene Spafford. Active defense of a computer system using autonomous agents. Technical Report 95-008, COAST Group, Departmentof Computer Sciences, Purdue University,West Lafayette, IN 47907-1398, Feb 1995.

[5] Kumar S, Spafford EH (1994) An Application ofPattern Matching in Intrusion Detection. TechnicalReport CSD-TR-94-013. Purdue University.

[6] S. Staniford-Chen, S. Cheung, R. Crawford,M. Dilger, J. Frank, J. Hoagland, K. Levitt,C. Wee, R. Yip, and D. Zerkle. GrIDS: A graphbased intrusion detection system for large networks.In Proceedings of the 19th National InformationSystems Security Conference, volume 1, pages 361-370. National Institute of Standardsand Technology, October 1996.

[7] Gregory B. White, Eric A. Fisch, and Udo W.Pooch. Cooperating security managers: A peer basedintrusion detection system. IEEE Network, pages 20-23, January/February 1996.

[8] Intrusion Detection System by using K-Means Clustering, C 4.5, FNN, SVM Classifier Akshay Takke1, Ravikumar Gujjul2, Mikhil Ghag3 , Vivek Pawar4, Vivek Pandey5 Page no:-636

[9] International Journal of Advanced Research in Computer Science REVIEW PAPER Available Online at www.ijarcs.info© 2015-19, IJARCS All Rights Reserved 356 ISSN No. 0976-5697INTRUSION DETECTION SYSTEMS: A REVIEW D. Ashok Kumar,S. R. Venugopalan Page no:- [356-357]

[10] Patcha, A. and Park, J. M. An overview of anomaly detection techniques: Existing solutions and latest technological trends.Computer Networks, 51(12);2007; 3448–3470.

[11] Forrest, S., Hofmeyr, S. A. , Somayaji, A. and Longstaff, T. A. A Sense of Self for Unix Processes, IEEE Symposium onResearch in Security and Privacy, Oakland, CA, USA, 1996;120--128.

[12] Duan, Z., Chen, P., Sanchez, F., Dong, Y., Stephenson, M. and J. M. Barker, M. (2012). Detecting spam zombies bymonitoring outgoing messages, IEEE Trans. Dependable and Secure Computing, Apr 2012; 9(2):198–210

[13] Gu, G., Porras, P., Yegneswaran V., Fong, M. and Lee, W. BotHunter: detecting malware infection through IDS-drivenialog correlation, Proc. of 16th USENIX Security Symp. (SS '07), Aug. 2007; 12:1–12:16.

[14] Gu, G., Zhang, J. and Lee, W. (2008). BotSniffer: detecting botnet command and control channels in network traffic, Proc. Of15th Ann. Network and Distributed Sytem Security Symp. (NDSS '08),

[15] Goyal, A. and Kumar, C. .GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System, Electrical Engineering and Computer Science, North West University, Technical Report;2008.Feb. 2008.

[16] Abdullah, B., Abd-algafar I., Salama G. I. and Abd-alhafez A. Performance Evaluation of a Genetic Algorithm BasedApproach to Network Intrusion Detection System, Proceedings of 13th International Conference on Aerospace Sciences andAviation Technology (ASAT-13), Military Technical College, Cairo, Egypt, 2009;1-5.

[17] Ojugo, A. A., Eboka, A. O., Okanta, O. E., Yora, R. E. and Aghware, F. O.Genetic Algorithm Rule-Based IntrusionDetection System (GAIDS), Journal of Emerging Trends in Computing and Information Sciences, 3(8);2012; 1182 – 1194.

[18] Roshani Gaidhane, Vaidya, C. and Raghuwanshi, M. Survey.Learning Techniques for Intrusion Detection System (IDS),International Journal of Advance Foundation and Research in Computer (IJAFRC) Feb 2014. ISSN 2348 – 4853, 2014;1(2).

[19] Gaikwad, Sonali Jagtap, D.P. Kunal Thakare and Vaishali Budhawant. Anomaly Based Intrusion Detection System UsingArtificial Neural Network and fuzzy clustering., International Journal of Engineering Research & Technology (IJERT), ISSN:2278-0181, November- 2012; 1(9).

[20] L. Heberlein, G. Dias, K. Levitt, B. Mukherjee, J. Wood, and D. Wolber. A Network SecurityMonitor. In Proceedings of the IEEE Symposium on Research in Security and Privacy, May 1990.

[21] R. Heady, G. Luger, A. Maccabe, and M. Servilla. The Architecture of a Network Level Intrusion Detection System. Technical report, University of New Mexico, Department of Computer Science, August 1990.

[22] Judith Hochberg, Kathleen Jackson, Cathy Stallings, J. F. McClary, David DuBois, and Josephine Ford. NADIR: An automated system for detecting network intrusion and misuse. Computers and Security, 12(3):235-248, May 1993.

[23] S. R. Snapp, J. Brentano, G. V. Dias, T. L. Goan, L. T. Heberlein, C. Ho, K. N. Levitt, B. Mukherjee, S. E. Smaha, T. Grance, D. M. Teal, and D. Mansur. DIDS (Distributed Intrusion Detection System) - Motivation, Architecture, and an early Prototype. In Proceedings of the 14th National Computer Security Conference, pages 167-176, October 1991.

[24] Biswanath Mukherjee, Todd L. Heberlein, and Karl N. Levitt. Network intrusion detection. IEEE Network, 8(3):26-41, May/June 1994.

[25] Thomas H. Ptacek and Timothy N. Newsham. Insertion, evasion, and denial of service: Eluding network intrusion detection. Technical report, Secure Networks, Inc., January 1998.

[26] S. Staniford-Chen, S. Cheung, R. Crawford, M. Dilger, J. Frank, J. Hoagland, K. Levitt, C. Wee, R. Yip, and D. Zerkle. GrIDS: A graph based intrusion detection system for large networks. In Proceedings of the 19th National Information Systems Security Conference, volume 1, pages 361-370. National Institute of Standards and Technology, October 1996.

[27] Gregory B. White, Eric A. Fisch, and Udo W. Pooch. Cooperating security managers: A peer based intrusion detection system. IEEE Network, pages 20-23, January/February 1996.

[28] Review Paper on Shallow Learning and Deep Learning Methods for Network security Afzal Ahmad1*, Mohammad Asif2, Shaikh Rohan Ali 3 (page: -52) 2018

[29] Youssef Ahmed and Ahmed Emam, "Network Intrusion Detection Using Data Mining and Network Behavior Analysis", International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011.

[30] X. Wu, V.Kumar, Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg, "Top 10 algorithms in data mining", Survey Paper(2008).

[31] Intrusion Detection Based On Clustering Algorithm Nadya El MOUSSAID 1, Ahmed TOUMANARI 2, Maryam ELAZHARI 3 Page-1062

[32] P. S. Rath, M. Hohanty, S. Acharya and M. Aich, “Optimization of IDS Algorithms Using Data Mining Technique”, Proceeding of 53rd IRF International Conference, Pune, India,ISBN 978-93-86083-01-2, 2016.

[33] L.S. Parihar and A. Tiwari, “Survey on Intrusion Detection Usingn Data Mining Methods”,IJSART, , Volume-2 Issue-1 ISSN (online: 2395-1052) January-2016.

Cite this article as :

Mohammad Asif, Pratap M. Mohite, Prof. P. D. Satya, "Machine learning algorithm for Cyber Security - A Review", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 1, pp. 535-545, January-February 2019. Journal URL : http://ijsrcseit.com/CSEIT1951141

Authors Profile

Mohammad Asif received his B.E. degree from NMU JALGAON INDIA. Currently pursuing M.Tech.in Computer Engineering from S.Y.C.E.T. Aurangabad Affiliated to BATU Lonere.

Prof. Pratap Mohite B.Tech CSE M.E.(Software Engineerng), Assistant Professor Department of computer Science and Engineering Shreeyash College of Engineering and Technology, Satara Parisar,Beed Bypass Road Aurangabad.

Prof. Satya P.D. B.Tech. CSE, M.Tech. and P.hD.(Pursuing), Assistant Professor Department of computer Science and Engineering Shreeyash College of Engineering and Technology, Satara Parisar,Beed Bypass Road Aurangabad.