International Journal of Research in Advent Technology, Vol.7, No.4, April 2019
E-ISSN: 2321-9637
Available online atp www.ijrat.org

Intrusion Detection System Using Data Mining Techniques – A Survey

Ch. Kiran Kumar[1] ,M.Govindarajan[2]
[1]Research Scholar, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar – 608002, Tamil Nadu, India.
[2]Assistant Professor, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar – 608002, Tamil Nadu, India.
kirankumar.chanumolu@gmail.com[1] , govind_aucse@yahoo.com[2]

Abstract: Security in information mining and Artificial Intelligence has been a critical zone of research amid the most recent couple of years. Today malicious attack is a security risk. These vindictive executables are made at the rate of thousands consistently and make genuine security issues. Intrusion Detection System (IDS) is utilized to access unapproved and vindictive attacks over the system. Information mining procedures that can be connected to IDS to distinguish ordinary and irregular personal conduct standards. An intrusion recognition framework investigates network exercises and distinguishes suspicious action in the system to improve exactness and security and recognize peculiarity attacks. IDS play a pivotal guideline in this period where systems achieved practically any part of action. Shockingly, IDS is a long way from perfection. Thusly, scientists dug constantly more profound to improve them. In this specific situation, information mining systems have been much abused for Intrusion location. In this paper, we present a relative investigation of data mining procedures for intrusion detection. In particular, we consider the general exhibitions of those strategies just as the effect of preparing information measure on their outcomes.
Keywords: Data Mining, Intrusion detection, Network attacks, data security, network failure.

1. INTRODUCTION

Intrusion identification was characterized as "the way toward checking the occasions happening in a PC framework and investigating them for indications of Intrusions, characterized as endeavors to bargain the secrecy, trustworthiness, accessibility, or to sidestep the security instruments of a PC or system". Contingent upon its location component, an Intrusion Detection System (IDS) is an abuse based or peculiarity based [9].

Abuse recognition depends with respect to a lot of marks portraying known attacks. Then again, peculiarity depends with respect to a prebuilt model of typical conduct and take any deviation that demonstrate as an Intrusion attempt. The two methodologies have their points of interest and disadvantages.

The subsequent cautions from abuse discovery based IDSs are truly dependable, due to their low false alert rate. Be that as it may, they are frail against obscure attacks. Then again, oddity based location can deal with obscure attacks that include a deviation from the ordinary conduct, however they trigger a lot of false alerts.

Among different elements, computerized reasoning techniques were left because of their powerlessness to adapt to the expansion of system traffic. Numerous works are discovered that consolidate and change information mining strategies for Intrusion recognition just as works that assess and think about information mining procedures with regards to Intrusion discovery. In any case, there is a space between proposition works and near examinations. The proposition works regularly utilize generally new techniques, while near investigations target established ones, for example, Random Forest, Support Vector Machine, Artificial Neural Networks and C4.5 choice tree.

Fundamental objective of Intrusion identification is to spare information from extortion client. There are gigantic measure of information accumulation in different organizations like records, archives, pictures, recordings, logical information and numerous new information designs identified with human life.

Intrusion recognition is utilized to discover obscure examples, substantial examples and connections in vast informational collections. It likewise breaks down and foresee suspicious exercises and utilize a different kind of parameters to inspect the information and incorporate investigation, characterization, grouping techniques. Intrusion identification framework is a procedure to distinguish unapproved utilization of PC or a media transmission organization. It has capacity to abridge the Intrusion or risk to an association. IDS distinguish three kinds of PC attacks to be specific:

1. Denial of service (Dos) attacks

2. Filtering attacks

3. Access attacks

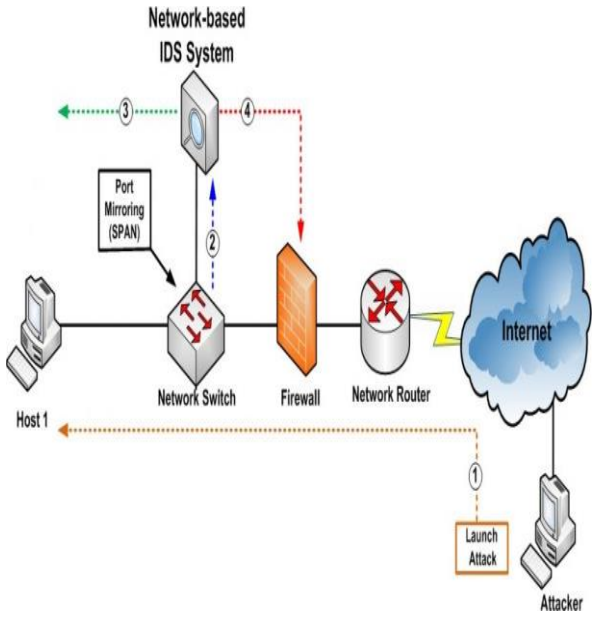

Fig-1: Intrusion Detection System

At whatever point IDS discovers security risk in the framework then it produces an alarm to demonstrate the presence of Intrusion. IDS were first presented by James Anderson in the year 1980. It has turned into a most need and testing assignments for system directors and security specialists. We will likely plan and construct a scanner that filter malevolent example and unapproved information. IDS can identify and end unapproved attack on the system.

1.1 Distinctive Attacks

The attacks can be inactive or dynamic. The dynamic attack is described by the attacker endeavoring to break into the framework. Amid a functioning attack, the interloper will bring information into the framework just as conceivably change information inside the framework. The sorts of dynamic attacks are dispersed DOS, session replay and masquerade. Bug, Trojan are the case of dynamic attacks. The inactive attack endeavors to learn or utilize data from the framework yet doesn't influence framework assets. Encryption, Scanning are a few kinds of uninvolved attacks. An attack can likewise be unleashed by a outsider or an insider of the organization.

1.1.1 DOS (Denial of Service):

DOS is an attack where the culprit tries to make a machine or system asset inaccessible to its expected clients by briefly or uncertainly disturbing administration of a host associated with the web. Flooding in the system, disturbing the associations, keeping the entrance of people are a few instances of DOS attacks.

1.1.2 Filtering attacks (Eavesdropping Attack):

Eavesdropping is an electronic attack where computerized communication is delayed by a person whom they are not planned. Man in the middle attack is the best case of listening in attack. Straightforwardly tuning in to computerized or simple voice communication and moving of information identifying with any type of correspondence are two primary kinds of listening in attack.

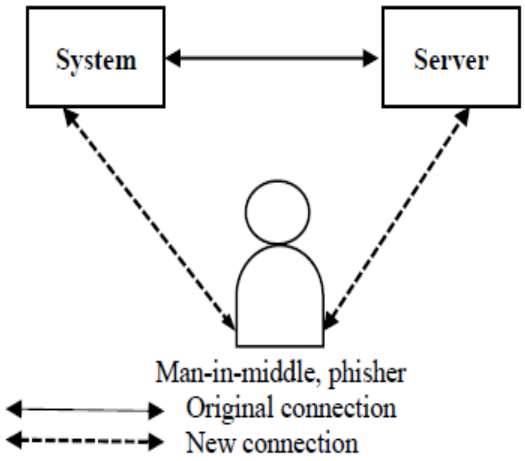

Fig-2: Man in Middle Attack

1.1.3 Access attacks (Phishing Attack):

Phishing is an attack to acquire sensitive data for pernicious reasons. These data's incorporates usernames, passwords is the limit. spear phishing, clone phishing, whaling are the sorts of phishing attacks.

Spear phishing:

Spear phishing is a sort of email mocking which focuses on an individual or an association to get to its secret data.

Clone phishing:

Clone phishing is a sort of phishing where the beneficiary's location is copied for making an indistinguishable email having distinctive substance.

Whaling phishing:

Whaling phishing is a sort of phishing which focuses on the prominent gathering that incorporates senior officials, famous people, businessmen's, lawmakers and so forth. Specialized help tricks, contaminated connections, internet based life misuses, extortion tricks are the instances of phishing.

2. LITERATURE SURVEY

A Hidden Markow display based IDS is produced for programming characterized organizing. The system can help screen the general security of a framework by breaking down the web as an opening and settling on decisions to protect the system dependent on the information from the whole system it incorporates employments of ANN IDS. This procedure permits more noteworthy unique control of a systems administration condition. The paper comprises of the favorable circumstances like expanded in the scope of exercises and furthermore is the expansion of security application.

Huawen Liu et al [1] proposed a genetic calculation based method to deal with system Intrusion identification. The genetic calculation is utilized to infer a lot of arrangement rules from system review information, and the certainty structure is used as wellness capacity to pass decision on the nature of each standard.

Ahmad M et al [2] built up an AI based IDS in which a genetic calculation alongside a support vector machine was utilized for the programmed recognizable proof of the proper arrangement of features. Weiwei Chen et al [8] analyzed and audited arrange Intrusion discovery utilizing information mining.

Ying Gao et al [3] refined the standard definition system with a versatile base-bolster calculation to mine ordinary traffic records. Distinctive node characteristic qualities apply diverse edges. Mehdi Ezzarii et al [13] fabricated a model framework by consolidating the two discovery frameworks, yet they work freely without collaborations. We think about close collaboration between the two subsystems. Numerous scientists have examined directed inconsistency recognition frameworks via preparing over attack free traffic.

Manickam M et al [4] proposed a way to deal with screening the system or host bundles to discover malignant exercises and unapproved people. This method comprises information mining procedures to execute IDS to distinguish both known and obscure attacks. Bing Zhang et al [5] assessed information mining methods for Intrusion location and anticipation to keep up the classification, respectability or accessibility of an resource. To improve precision and security programs are broke down. Calculation comprises sensor, finder and information store to recognize attacks and sort of attacks on database.

Hamidreza Sadreazami et al [6] explained about pernicious executables and program investigation in three distinct dimensions, parallel dimension, machine language level and abnormal state language level. This exploration work introduced an information mining way to deal with concentrate suspicious examples from program and use ordered show for obscure bug identification. It thinks about various classifiers, their outcomes and accuracy.

Mohammed Anbar et al [7] checked on to refresh Intrusion location framework. Our plan to build up a computerized methodology in structure IDS to assemble review information to catch precise conduct. It shows a system to register abuse and irregularity location display. Anna L et al[9] utilized a combinative strategy for k-means grouping in addition to C4.5 to identify Intrusion.

Manoj S et al [10] utilized two generally utilized IDSs that depend on the abuse display. Different endeavors to unravel the Intrusion discovery and reaction issue can be found in [12]. Intrusion location must be intended to screen the association features at the system, transport, and application layers. The MIT/LL IDS assessment informational index and revealed IDS execution results were broke down in [15]. We utilize this attack informational index with blended foundation traffic to test the adequacy of HIDS.

Thabet Kacem et al [11] proposed the Anomaly-based Network Intrusion Detection Systems (A-NIDS). In this work, the authors feature the impediment of existing A-NIDS as they produce an uncommon main part of cautions that can be changed with false-positive alerts. Such tremendous volumes of false alerts stay away from exact acknowledgment of system attacks that would influence adversely the moment response of IDS. Consequently, as an approach to conquer this issue, the creators have presented a technique for sifting such false-positive alerts of A-NIDS.

Shengyi Pan et al [12] clarified that it likely structure and assemble a scanner that precisely filters suspicious examples and unapproved data. Compare location models with customary models (signature based) and copies the discovery rate for new noxious executables.

Panagiotis I et al [14] recommended a system to determine persistent view for inconsistency discovery against typical traffic profiles. They built up a level wise information mining calculation for network structure.

From one viewpoint, Support Vector Machine and Random Forest are notable and utilized AI strategies, while A-NIDS executes the recursive segment chief, a similar guideline whereupon is based ID3, C4.5, C5.0, and RF. The utilized bundle for each examined technique is represented in table 1. The remainder of this area incorporates the definite pre-preparing and testing methodologies that pursued the measurements that is used, lastly, the investigations' outcomes and exchanges.

Table-1: Utilized data bundles

Training Data	OCSVM	ELM	RPART	SVM	RF
20%training	10.07%	81.86%	95.46%	99.29%	99.55%
40%training	9.92%	75.70%	95.91%	99.37%	99.65%
60%training	10.01%	82.71%	95.91%	99.37%	99.67%
80%training	10.11%	86.92%	95.91%	99.38%	99.71%
100%training	10.18%	87.00%	95.91%	99.40%	99.76%
Average	10.06%	82.84%	95.82%	99.36%	99.67%

Despite the fact that the location rate is a decent measure to survey the execution of an identification demonstration, it isn't adequate all alone. Notwithstanding a high discovery rate, a great model needs to keep its fake caution rate as low as would be prudent. The Training data set is represented in table2.

Table-2: Training data set

Training Data	OCSVM	ELM	RPART	SVM	RF
20%	10.02%	5.34%	2.07%	0.52%	0.18%
40%	9.92%	3.78%	1.30%	0.42%	0.14%
60%	9.87%	5.51%	1.30%	0.40%	0.11%
80%	9.88%	5.44%	1.30%	0.40%	0.09%
100%	9.85%	5.14%	1.30%	0.40%	0.08%
Average	9.91%	5.04%	1.46%	0.43%	0.12%

To demonstrate the effect of the preparation information estimation on the exactness of every technique. We plot in table3 the exactnesses against the preparation datasets.

Table-3; Exactness on preparation dataset

Training Data	OCSVM	ELM	RPART	SVM	RF
20%	76.21%	92.45%	97.50%	99.44%	99.78%
40%	76.27%	92.68%	98.22%	99.54%	99.83%
60%	76.33%	92.46%	98.22%	99.56%	99.85%
80%	76.34%	93.24%	98.22%	99.56%	99.88%
100%	76.37%	93.51%	98.22%	99.57%	99.89%
Average	76.30%	92.87%	98.07%	99.53%	99.85%

3. PROPOSED METHOD

In this exploration work, a smart specialist based security demonstration for Database systems should be proposed and executed. For this, various significant commitments are made in this exploration work. The main significant commitment of this work is the proposition of a compositional system for giving security to organize database. The second commitment of this work is the proposition and usage of another anomaly location framework with new calculations. The third commitment of this work is the plan and execution of a security administrator and a system database director which are equipped for organizing with the other real parts of this structure. The fourth and last commitment of this work is the proposition of a specialist based spatio-successive access control framework.

4. CONCLUSION

Intruder may take data and put away in different databases and documents. Individuals presently understand that to deal with alarm based repression, grows new thoughts and guideline to battle with them. One of the difficulties to verifying databases is the principle issue. Intrusion Detection System is utilized to access unapproved and vindictive attacks over the system. According to the concentrated of methods proposed by different creators, the manners in which it can recognize the intruder are displayed here. So, when structuring another IDS, these attributes can be utilized progressively framework to recognize the inside interlopers and their malevolent practices. This will be a substantial IDS which will distinguish the interior gatecrasher's precisely. In this manuscript, different techniques for identifying intrusions are discussed.

REFERENCES

[1].Huawen Liu, Xuelong Li, Jiuyong Li, and Shichao Zhang, “Efficient Outlier Detection for High-Dimensional Data”, IEEE Transactions On Systems, Man, And Cybernetics: Systems, 2017.

[2].Ahmad, M. Basheri, M.J. Iqbal, A. Raheem “Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection”, IEEE. Translations and content mining, 2018.

[3]. Ying Gao, Yu Liu, Yaqia Jin, Juequan Chen, Hongrui Wu, “A Novel Semi-Supervised Learning Approach for Network Intrusion Detection on Cloud-Based Robotic System”, IEEE. Translations and content mining, 2018.

[4].Manickam M, S. P. Rajagopalan, “A hybrid multi-layer intrusion detection system in cloud,” Cluster Computing, 2018.

[5].Bing Zhang , Zhiyang Liu , Yanguo Jia ,Jiadong Ren,and Xiaolin Zhao, “Network Intrusion Detection Method Based on PCA and Bayes Algorithm”, Security and Communication Networks, 2018.

[6].Hamidreza Sadreazami, Arash Mohammadi , Amir Asif , Konstantinos N. Plataniotis z, 2017, “Distributed Graph-based Statistical Approach for Intrusion Detection in Cyber-Physical Systems”, IEEE Transactions on Signal and Information Processing over Networks, DOI 10.1109/TSIPN.2017.2749976.

[7].Mohammed Anbar, Rosni Abdulah, Izan H. Hasbullah, Yung- Wey Chong; Omar E. Elejla, “Comparative Performance Analysis of classification algorithm for Internal Intrusion Detection”, 14th Annual Conference on Privacy Security and Trust (PCT), Dec 12-14,2016, Penang, Malaysia.

[8].Weiwei Chen, Fangang Kong, Feng Mei, GuiginYuan, Bo Li, “a novel unsupervised Anamoly detection Approach for Intrusion Detection System”, IEEE 3rd International Conference on big data security on cloud, May 16-18,2017, Zhejiang, China.

[9].Anna L. Buczak, Erha n Guven, "A Survey of Data Mining and Machine Learning methods for cybersecurity intrusion detection", IEEE communication surveys and tutorials, 18(2) ,2016.

[10]. Manoj S. Koli, Manik K. Chavan, “An Advanced method for detection of botnet traffic using Internal Intrusion Detection”, International Conference on (ICICCT), March 10-11, 2017, Sangli, India.

[11]. Thabet Kacem, Duminda Wijesekera, Paulo Costa, Alexander Barreto, “An ADS-B Intrusion Detection System”, IEEE on ISPA, 2016, Fairfax, Virginia.

[12]. Shengyi Pan, Thomas Morris, Uttam Adhikari, "Developing a Hybrid Intrusion Detection System using Data Mining for power system", IEEE Transactions, 6(6), 2015.

[13]. Mehdi Ezzarii, Hamid Elghazi, Hassan El Ghazi, Tayeb Sadiki, "Epigenetic Algorithm for performing Intrusion Detection System", International Conference on ACOSIS, Oct17-19,2016, Rabat, Morocco.

[14]. Panagiotis I. Radogloa-Grammatikis; Panagiotis G. Sarigannidis, "Flow anamoly based Intrusion Detection System for Android Mobile Devices", 6th International Conference on MOCAST, May 4-6, 2017, Kazani, Greece.

[15]. Trae Hurley, Jorge E. Perdomo, Alexander Perez-pons, "HMMBased Intrusion Detection System for software-defined networking", 15th IEEE Conference on Machine Learning and Application, Dec 18-20, 2016, Miami, Florida.

[16]. Mariem Belhor, Farah Jemili, "Intrusion Detection based on genetic fuzzy classification system", IEEE 13th International Conference on Computer Systems and Application (AICCSA), Nov 29 2016-Dec 2, 2016, Sousse, Tunisia.

[17]. Sharad Awatade, Shweta Joshi. "Improved EAACK: Develop Secure Intrusion Detection System for MANETS using hybrid cryptography", International Conference on computing communication control and automation (ICCUBEA), Aug 12-13, 2016, Maharashtra, India.

[18]. Mayank Agarwal, Sanketh Purwar, Santosh Biswas, Sukumar Nandi, "Internal Detection System for PS-Poll DOS attack in 802.11 networks using real-time discrete event system",IEEE, 4(4), 2017.

[19]. Md Zahangir Alom, Tarek m. Taha, "Network Intrusion Detection for cybersecurity on neuromorphic computing system", International Joint Conference on Neural Networks (IJCNN), May 14-15, 2017, USA.

[20]. Aasia Abdullah and Khaleda Afroaz," Data Mining Approaches on Network Data: Intrusion Detection System", International Journal of Advanced Research in Computer Science, 8(1), 2017.

[21]. Nutan Farah Haq, Musharrat Rafni, Abdur Rahman Onik, Faisal Muhammad Shah, Md. Avishek Khan Hridoy and Dewan Md. Farid, " Application of machine Learning Approaches in Intrusion Detection System : A Survey", International Journal of Advanced Research in Artificial Intelligence, 4(3), 2015.

[22]. Dr. D. Aruna Kumari, N. Tejaswani, G. Sravani and R. Phani Krishna, "Intrusion Detection Using Data Mining Technique (Classification)", International Journal of Computer Science and Information Technologies, 6(2), 2015, pp.1750-1754.

[23]. David, "5.6 million fingerprints stolen in U.S. personnel data hack: government", Reuters, 2015.

[24]. E. Kabir, J. Hu, H. Wang, and G. Zhuo, "A novel statistical technique for intrusion detection systems", Futur. Gener. Comput. Syst., 2017.

[25]. S. Detrow, "Obama On Russian Hacking: 'We Need To Take Action. And We Will," *Npr*, Dec-2016. [Online]. Available: http://www.npr.org/2016/12/15/505775550/obama-on-russianhacking- we-need-to-take-action-and-we-will. [Accessed: 01-Jan-2017].

[26]. J. Zhang and M. L. Huang, "Density approach: a new model for BigData analysis and visualization," Concurr. Comput. Pract. Exp., 28(3), 2016, pp. 661–673.

[27]. R. Singh, H. Kumar, and R. K. Singla, "An intrusion detection system using network traffic profiling and online sequential extreme learning machine," Expert Syst. Appl., 42(22), 2015, pp. 8609– 8624.

[28]. W. Shang, P. Zeng, M. Wan, L. Li, and P. An, "Intrusion detection algorithm based on OCSVM in industrial control system," Secur. Commun. Networks, 9(10), 2016, pp. 1040–1049.

[29]. W. Wang, J. Liu, G. Pitsilis, and X. Zhang, "Abstracting massive data for lightweight intrusion detection in computer networks", Inf. Sci. (Ny), 2016.

[30]. R. C. Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria. 2013.