INTRUSION DETECTION ON CLOUD USING HYBRID MACHINE LEARNING TECHNIQUES

Hem Durgapal
Department of Computer Science and Engineering,
Integral University, Lucknow.
Afsaruddin
Assistant Professor Department of Computer Science and Engineering
Integral University, Lucknow.

Abstract

Cloud computing delivers ubiquitous and pay-per-use services, and as a result of these features, it attracts more consumers to utilize its services. Despite the many benefits of cloud computing, there are a few drawbacks. There is a risk of cloud-based attacks that compromise security features such as confidentiality, availability, and integrity. In order to identify attacks and improve cloud security, security solutions are required for both cloud users and cloud service providers. Cloud computing's primary concern is security. In the cloud, there are a lot of incursions. In order to identify intrusions, a variety of methodologies have been devised. However, no one method can accurately categories all types of assaults. The hybrid machine learning-based Intrusion Detection System (IDS) proposed in this study is a blend of supervised and unsupervised machine learning algorithms. There is a categorization of machine learning and deep learning technologies that are used to identify harmful user assaults in the network. Various researchers' techniques to detecting suspicious activity in the NSL KDD data utilizing tools are highlighted. We've also organized the papers' publishing throughout the last 19 years by year of publication and database source. In the last part of the research, we run an experiment to identify assaults in the dataset NSL-KDD. To identify the assault in the dataset, a machine learning intrusion detection model is developed using effective classifiers.

Index Terms-- Cloud security, Intrusion detection, Machine learning, Hybrid machine learning techniques.

I. INTRODUCTION

Cloud computing, which is commonly done through the internet, provides services, servers, storage, databases, networking, software, analytics, and more with minimum administration effort. Cloud offers three types of services (IAS, PAS, and SAS) as well as three deployment modes (Private, Public, and Hybrid) to its customers [1]. As a result of these features, the majority of cloud users keep sensitive data on the cloud. On the other hand, information security is a concern for both cloud consumers and cloud service providers. Because there is a risk of cloud-based attacks that compromise security features such as confidentiality, availability, and integrity. Intrusion Detection Systems (IDS) are used to improve the system's security and resilience to both internal and external threats. The basic objective of an intrusion detection system is to identify an intrusion and, if required or practicable, to take steps to eliminate it. There are primarily two approaches for detecting intrusions [3].

1. Anomaly-based intrusion detection system: Due to the fast growth of malware, anomaly-based intrusion detection systems were primarily created to identify unknown threats. Behavior-based intrusion detection system is another name for this technology. The core idea is to utilize machine learning to build an anomaly detection model, then compare new behavior to that model.
2. Signature-based intrusion detection system: This method, also known as misuse detection, entails scanning for certain signatures when they occur; nevertheless, it is difficult to detect new assaults using this method.

Host IDS (HIDS), Network IDS (NIDS), and Virtual Machine Monitor IDS are the intrusion detection systems (VMMIDS).

II. LITERATURE REVIEW

We did a thorough literature review of existing intrusion detection strategies for this study. This examination covers the years 1998 through in recent time. We've covered intrusion detection systems, machine learning, deep learning, and block chain technology at a high level. Then, for each of the three strategies outlined below, a full examination of applications is given. In addition, the system's limits and problems in the field of intrusion detection are described. We looked at the current level of block chain technology in terms of identifying cyber-attacks.

There is a categorization of machine learning and data mining technologies that are used to identify harmful user assaults in the network. Various researchers' techniques to detecting suspicious activity in the NSL KDD data utilizing tools are highlighted. Although block chain technology appears to be a potential contribution to intrusion detection, it does not give a means to evaluate its performance to machine learning methods, nor does it supply the dataset needed to develop an effective intrusion detection system. We recommended the following strategy for the reasons stated above. In the last part of the research, we run an experiment to identify assaults in the dataset NSL-KDD. To identify the assault in the dataset, a machine learning intrusion prevention model is developed using three classifiers. Now we study every learning concept in this section, as follow to discuss about something:

1. Supervised Learning: This approach involves predicting a target / result (or dependent variable) from a set of predictor variables (independent variables). We create a function that maps inputs to target values using this collection of variables. The model is trained until it reaches the appropriate degree of precision on the learning algorithm. Regression, Decision Tree, Random Forest, KNN, Logistic Regression, and others are examples of supervised learning.

2. Unsupervised Learning: There is no objective or result variable to predict or estimate in this technique. It's utilized to divide people into separate categories. Unsupervised Learning Examples: K-means is an Apriori algorithm.

3. Reinforcement Learning: The machine is taught to make certain judgments using this method. It works like this: the machine is placed in an environment where it must constantly teach itself via trial and error. This computer learns from its previous experiences and attempts to acquire the most relevant information in order to make appropriate business decisions. Markov Decision Process is an example of Reinforcement Learning.

III. OBJECTIVE OF THE PROJECT

2 Tier Machine Learning Approach: Using deep learning and machine learning methods, it presented a 2-tier architecture for network intrusion detection in this research. They used the Weak data mining programme to run simulations on the KDD data set and found that running the data through two classifiers increases system security. Hybrid machine learning techniques: In this paper, the author [13] presents a design and implementation to detect known attacks using supervised learning and unknown attacks using unsupervised learning. There are seven hybrid variants in the design. A supervised learning method is used in the first layer, while an unsupervised learning approach is used in the second layer. The purpose of this method is to improve network intrusion detection.

IV. PROBLEM DEFINITION

The highlighted KDD 99 data's intrinsic flaws, such as duplicate records and difficulties accessing the data owing to biased findings, and produced a new data set to address these concerns. As shown in Fig 1, the assaults in the data are split into four types [13]. Used several machine learning techniques such as J48, SVM, and Naive Bayes to detect threats in the National Security Lab Knowledge Discovery and Data mining (NSL KDD) dataset [16]. The Weak tool was used to forecast the accuracy rate of the normal and attack categories.

V. PROPOSED METHODOLOGY

This research offers a hybrid model system [Figure-1] that uses two machine learning techniques. In order to provide the highest level of intrusion detection in the cloud, a hybrid method was used. The proposed hybrid model employs the supervised learning algorithm ANN and the unsupervised learning algorithm K-Means, with supervised learning detecting known assaults and unsupervised learning detecting novel attacks. The Principal

Component Analysis (PCA) technique is used to choose features.

Method:

The researchers detail the suggested model as well as the tools and techniques employed in the proposed method in this section. The Spark-Chi-SVM model is shown in Figure suggested model's steps may be stated as follows:

1. Load dataset and export it into Resilient Distributed Datasets (RDD) and Data Frame in Apache Spark.
2. Data preprocessing.
3. Feature selection.
4. Train Spark-Chi-SVM with the training dataset.
5. Test and evaluate the model with the KDD dataset.

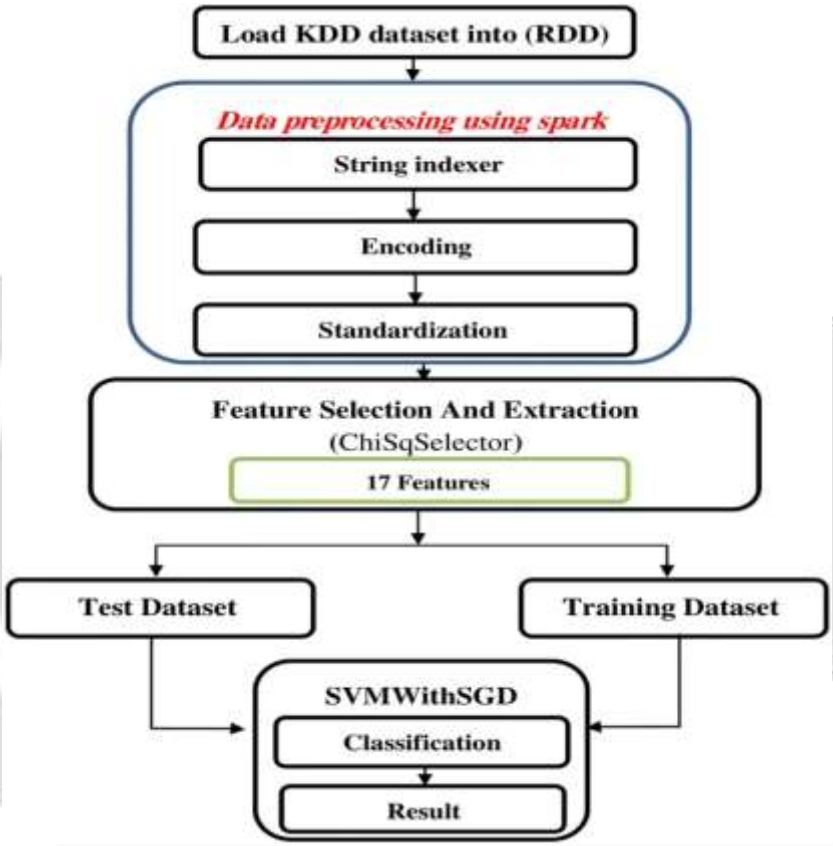

Fig-1: Spark-Chi-SVM model. The sequence of steps that in Spark-Chi-SVM model.

Spark is a fast and general-purpose cluster computing system for large-scale in-memory data processing. Spark has a similar programming model to Map Reduce but extends it with a data-sharing abstraction called Resilient Distributed Datasets or RDD. A Spark was designed to be fast for iterative algorithms, support for in-memory storage and efficient fault recovery. Spark Core consists of two APIs which are the unstructured and structured APIs. The unstructured API is RDDs, Accumulators, and Broadcast variables.

Processing: Large-scale datasets are frequently noisy, duplicated, and contain a variety of data kinds, posing significant hurdles to knowledge discovery and data modelling. In general, intrusion detection algorithms work with one or more forms of raw input data, such as the SVM algorithm, which exclusively works with numerical data. As a result, we prepare the data and transform the dataset's categorical data to numerical data.

VI. REFERENCE

[1] Nikolaos Alexopoulos, Emmanouil Vasilomanolakis, Natalia Reka Ivanko, and Max Muhlhauser.

Towards blockchain-based collaborative intrusion detection systems. In Gregorio D'Agostino and Antonio Scala, editors, Critical Information Infrastructures Security, pages 107–118, Cham, 2018. Springer International Publishing.

[2] Yang Xin, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixia Hou, and Chunhua Wang. Machine learning and deep learning methods for cybersecurity. IEEE Access, 6:35365–35381, 2018.

[3] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. An overview of blockchain technology: Architecture, consensus, and future trends. In Big Data (BigData Congress), 2017 IEEE International Congress on, pages 557–564. IEEE, 2017.

[4] Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, and Wei-Yang Lin. Intrusion detection by machine learning: A review. Expert Systems with Applications, 36(10):11994–12000, 2009.

[5] Bo Dong and Xue Wang. Comparison deep learning method to traditional methods using for network intrusion detection. In 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN), pages 581–585. IEEE, 2016.

[6] Ahmad Javaid, Quamar Niyaz, Weiqing Sun, and Mansoor Alam. A deep learning approach for network intrusion detection system. In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (Formerly BIONETICS), BICT'15, pages 21–26, ICST, Brussels, Belgium, Belgium, 2016. ICST (Institute forComputer Sciences, SocialInformatics and Telecommunications Engineering). 43

[7] Zheng Wang. Deep learning-based intrusion detection with adversaries. IEEE Access, 6:38367–38384, 2018.

[8] Uzair Bashir and Manzoor Chachoo. Intrusion detection and prevention system: Challenges & opportunities. In Computing for Sustainable Global Development (INDIACom), 2014 International Conference on, pages 806–809. IEEE, 2014.

[9] D.H. Lakshminarayana J. P. Philips, N. Tabrizi. A survey of intrusion detection techniques. IEEE International Conference on Machine Learning Application, 175(542):7–9, 2019.

[10] Karen Scarfone and Peter Mell. Guide to intrusion detection and prevention systems (idps). NIST special publication, 800(2007):94, 2007.

[11] Steven R Snapp, James Brentano, Gihan V Dias, Terrance L Goan, L Todd Heberlein, Che-Lin Ho, Karl N Levitt, Biswanath Mukherjee, Stephen E Smaha, Tim Grance, et al. Dids (distributed intrusion detection system)-motivation, architecture, and an early prototype. In Proceedings of the 14th national computer security conference, volume 1, pages 167–176. Washington, DC, 1991.

[12] Vinod Yegneswaran, Paul Barford, and Somesh Jha. Global intrusion detection in the domino overlay system. In NDSS, 2004.

[13] Ryan Huebsch, Brent Chun, Joseph M Hellerstein, Boon Thau Loo, Petros Maniatis, Timothy Roscoe, Scott Shenker, Ion Stoica, and Aydan R Yumerefendi. The architecture of pier: an internet-scale query processor. 2005.

[14] Paul Brutch and Calvin Ko. Challenges in intrusion detection for wireless ad-hoc networks. In null, page 368. IEEE, 2003.

[15] Shiyong Yin, Jinsong Bao, Yiming Zhang, and Xiaodi Huang. M2m security technology of cps based on blockchains. Symmetry, 9(9):193, 2017.

[16] Marko Vukoli´c. Rethinking permissioned blockchains. In Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts, pages 3–7. ACM, 2017.

[17] [18] Hadi Sarvari and Mohammad Mehdi Keikha. Improving the accuracy of intrusion detection systems by using the combination of machine learning approaches. In 2010 international conference of soft computing and pattern recognition, pages 334–337. IEEE, 2010.

[18] Phurivit Sangkatsanee, Naruemon Wattanapongsakorn, and Chalermpol Charnsripinyo. Practical real-time intrusion detection using machine learning approaches. Computer Communications, 34(18):2227–2235, 2011. 44

[19] T Poongothai and K Duraiswamy. Intrusion detection in mobile adhoc networks using machine learning approach. In International Conference on Information Communication and Embedded Systems (ICICES2014), pages 1–5. IEEE, 2014.

[20] David Endler. Intrusion detection. applying machine learning to solaris audit data. In Proceedings 14th Annual Computer Security Applications Conference (Cat. No. 98EX217), pages 268–279. IEEE, 1998.

[21] Arkadiusz Warzyn´ski and Grzegorz Ko~laczek. Intrusion detection systems vulnerability on

adversarial examples. In 2018 Innovations in Intelligent Systems and Applications (INISTA), pages 1–4. IEEE, 2018.

[22] Deyban Perez, Miguel A Astor, David Perez Abreu, and Eugenio Scalise. Intrusion detection in computer networks using hybrid machine learning techniques. In 2017 XLIII Latin American Computer Conference (CLEI), pages 1–10. IEEE, 2017.

[23] Bisyron Wahyudi, Kalamullah Ramli, and Hendri Murfi. Implementation and analysis of combined machine learning method for intrusion detection system. International Journal of Communication Networks and Information Security, 10(2):295–304, 2018.

[24] MA Jabbar, Rajanikanth Aluvalu, and S Sai Satyanarayana Reddy. Intrusion detection system using bayesian network and feature subset selection. In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pages 1–5. IEEE, 2017.

[25] Tahir Mehmood and Helmi B Md Rais. Machine learning algorithms in context of intrusion detection. In 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), pages 369–373. IEEE, 2016.

[26] Kwangjo Kim and Muhamad Erza Aminanto. Deep learning in intrusion detection perspective: Overview and further challenges. In 2017 International Workshop on Big Data and Information Security (IWBIS), pages 5–10. IEEE, 2017.

[27] Sidharth Behera, Ayush Pradhan, and Ratnakar Dash. Deep neural network architecture for anomaly based intrusion detection system. In 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), pages 270–274. IEEE, 2018.

[28] Chuanlong Yin, Yuefei Zhu, Jinlong Fei, and Xinzheng He. A deep learning approach for intrusion detection using recurrent neural networks. Ieee Access, 5:21954–21961, 2017. 45

[29] Li Deng. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing, 3, 2014.

[30] Tuan A Tang, Lotfi Mhamdi, Des McLernon, Syed Ali Raza Zaidi, and Mounir Ghogho. Deep learning approach for network intrusion detection in software defined networking. In 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM), pages 258–263. IEEE, 2016.

[31] Weizhi Meng, Elmar Wolfgang Tischhauser, Qingju Wang, Yu Wang, and Jinguang Han. When intrusion detection meets blockchain technology: a review. Ieee Access, 6:10179–10188, 2018.