10.48047/jocaa.2024.32.01.37

Machine Learning for Cloud-Based Privilege Escalation Attack Detection and Mitigation with CATBOOST

Mrs.B.Laxmi Kalpana[1], Chavalla deekshitha [2], Samreen Begum [3], Ch.Swetha[4]
[1] Assistant Professor, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India
Email: kalpana.laxmi@gmail.com
[2,3,4,]B.Tech Student, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India

ABSTRACT:

The exponential growth in attack frequency and complexity in recent years has made cybersecurity a major concern due to the development of smart devices. Despite the revolutionary impact cloud computing has had on businesses, dispersed services such as security systems are hindered by the cloud's centralization. There is a significant risk of unintentional or malicious disclosure of sensitive information owing to the large amount of data sent between companies and cloud service providers. Since the hostile insider has more access and opportunities to do substantial harm, they pose a critical danger to the company. others on the inside have exclusive and authorised access to resources and information that others on the outside do not. This paper presents a machine learning-based method for classifying insider threats and identifying unusual events that may indicate security issues related to privilege escalation. The system takes a systematic approach to finding these anomalies. Machine learning results and prediction performance are both improved by ensemble learning, which combines many models. The topic of finding security defects or threats involving privilege escalation in network systems via the detection of abnormalities and vulnerabilities has been discussed in several research. However, the assaults cannot be accurately identified in these research. Machine learning (ML) ensembles are suggested and assessed in this research. The purpose of this work is to classify insider assaults using machine learning methods. A dataset that has been tailored from several files inside the CERT dataset is used. The dataset is subjected to the analysis of four machine learning algorithms: LightGBM, XGBoost, Adaboost, and Random Forest (RF). The best overall performer was LightGBM. On the other hand, RF or AdaBoost are two algorithms that could be superior at fending off internal assaults (such as Behavioural Biometrics attacks). Hence, it is possible to get a better categorization in various

10.48047/jocaaa.2024.32.01.37

internal assaults by combining many machine learning algorithms. With an accuracy of 97%, the LightGBM method outperforms the other suggested algorithms; RF comes in at 86%, AdaBoost at 88%, and XGBoost at 88.27%.

INTRODUCTION

One novel approach to enabling and providing services over the Internet is noisy computing. Data storage, processing, and presentation have all undergone major overhauls under the present Cloud Model due to the present financial crisis and the increasing computing needs. By using cloud infrastructure, cloud computing helps individuals save a lot of money on equipment upkeep and purchasing. Providers of cloud storage use encryption, access controls, and authentication as foundational security measures for their infrastructure and the data they manage. The almost limitless capacity of the cloud to store any kind of data in various cloud data storage structures is dependent on the accessibility, speed, and frequency of data access. Businesses and cloud service providers exchange large amounts of data, which increases the risk of sensitive data breaches—whether intentional or not. Businesses have it tougher to stop unauthorised access to their online services because of the same features that make them convenient to use for employees and IT systems. Businesses that rely on cloud services are particularly susceptible to two emerging security threats: authentication and open interfaces. Hackers with specialised knowledge are able to breach Cloud systems by using their expertise. In order to overcome the security barrier and improve data management, machine learning makes use of a number of different methodologies and algorithms. It is not possible to disclose many datasets because of privacy issues or because they lack important statistical features. Regulators are responding to privacy and security concerns brought forth by the Cloud industry's meteoric ascent. Even if an employee moves responsibilities or positions within the Cloud Company, their access credentials may not change. Consequently, sensitive information is inadvertently stolen or damaged due to the misuse of outdated rights. There is a certain amount of power associated with any account that can connect with a computer. Databases, sensitive information, and other services on servers are often only accessible to authorised users. Gaining access to a higher-level user account

10.48047/jocaaa.2024.32.01.37

and either abusing its privileges or increasing them allows a malevolent attacker to access a sensitive system. Attackers may achieve their goals by gaining control of additional systems horizontally or by gaining access to administrative and root privileges vertically. Obtaining the rights of another user at the same level of access is called horizontal privilege escalation. A malicious actor may get access to data that isn't directly related to them by exploiting a vulnerability in the system. Web applications may be vulnerable to attacks that compromise users' personal information if they are not well-designed. After successfully executing a horizontal elevation of privileges exploit, the attacker has the ability to view, modify, and copy sensitive data. Organisational entities were the targets of a horizontal privilege escalation assault, as shown in Figure 1. This kind of attack often requires an in-depth understanding of the vulnerabilities affecting certain operating systems and the use of malicious software. Giving a user, piece of software, or other asset additional privileges or privileged access than they currently have is known as privilege elevation assault. The main goal of the attacker is to increase the level of special access from a modest degree of privileged access. In order to circumvent security measures and obtain vertical access control, the attacker could have to resort to a number of tricks. To achieve business goals like least privilege and job separation, vertical privileges controls use granular versions of security models. One common tactic is for an attacker to impersonate a regular registered user in order to get administrative or root access to a network. By using behavioural analytics, suspicious activity on company systems or user accounts may be identified. This might indicate an incursion or an increase in privileges.

RELATED WORK

Machine learning and cloud-based email phishing assault

"Cloud computing" means that resources like data storage and processing power are made available to clients on demand via personal computer systems, without requiring any kind of input from the customer themselves. Data sent and received for individuals or organisations is typically done via email. Oftentimes, sensitive information like financial records, credit histories, and the like is sent via the Internet. Phishing is a method that fraudsters try to trick consumers into giving up sensitive information by making themselves seem to be from reputable

10.48047/jocaaa.2024.32.01.37

sources. By misleading you in a phishing email, the sender might trick you into divulging sensitive information. When sending and receiving emails, the biggest concern is phishing attempts. If you open and read an attacker's spam email, they will get your personal information. It has been an enormous issue for all parties involved in the last several years. In order to identify fresh emails, this study employs a variety of criteria and algorithms for categorization, and it utilises data sizes that vary between legal and phishing sources. The current methods are evaluated, and then a revised dataset is generated. We used the SVM, NB, and LSTM algorithms to transform comma-separated values (CSV) files into feature extraction files and label files. The detection of a phishing email is seen as a classification problem in this experiment. Results from tests and comparisons show that LSTM, SVM, and NB are the most effective and efficient methods for detecting email phishing attempts. The best accuracy rates for email attack categorization were 99.62% with SVM, 97% with NB, and 98% with LSTM classifiers.

Modelling and Detection of Insider Threats using Machine Learning

As of late, one of the gravest dangers facing businesses and governments has been malevolent insider assaults. An insider threat detection system that is user-cantered and uses machine learning at many levels of data granularity is proposed in this study. We publish and discuss the findings of insider scenario specific analyses and system assessments on both benign and malevolent insiders, as well as on individual data instances. Our findings demonstrate that the detection method based on machine learning may effectively learn from sparse ground truth and identify new malevolent insiders.

An Overview of Cloud Computing Security Risks and Their Solutions

Cloud computing has the ability to do away with the need for expensive computer infrastructure to be set up in order for the industry to leverage IT-based solutions and services. It says it will create a malleable IT infrastructure that can be accessed online by small, portable devices. Because of this, both the current and future software might have their capacities multiplied by a factor of several. Everything is stored on a shared

10.48047/jocaaa.2024.32.01.37

network in a cloud computing environment, and users may access it via their virtual PCs. Data centres may be located anywhere in the globe, out of the control of users. This poses a multitude of security and privacy concerns that must be addressed. Additionally, it is impossible to rule out the potential of a server outage, which has happened very often recently. When using the cloud, there are a number of concerns about data privacy and security that must be addressed. Aiming to detail and analyse the many outstanding difficulties hindering the acceptance and proliferation of cloud computing, this broad study report covers a wide range of stakeholders.

Platform for cloud computing: Evaluation of well-known cryptographic algorithms

Thanks to scientific and technological breakthroughs, cloud computing is set to become the industry standard in the near future. Data security may be achieved by the use of cloud cryptography, which employs encryption techniques. Because of the many benefits of cloud storage, including accessibility, reduced hardware requirements, minimal security risks, and cheap maintenance costs, every organisation is moving its operations to the cloud. Information may be protected against unauthorised access by encoding it using encryption. Protecting data from potential threats whether it is saved on a computer or sent over the internet is a top priority these days. They are responsive, private, bandwidth-dependent, and integrity-dependent; these factors determine the cryptographic approach. Ensuring client data is securely stored in the cloud is another important aspect of cloud computing. Various cryptographic methods are compared in this study report based on their efficiency, consumption, and utility. The results of the evaluation show which algorithms work better in different environments and with different kinds of data.

An Overview of Cloud Security Risks and Countermeasures

Cloud computing refers to the practice of making computer framework resources available on demand. Particularly the ability to store and process data, even in the absence of individualised client administration. On a unified platform accessible from anywhere in the world, it has given its users access to both public and private cloud computing and data storage. In addition, there are a number of

10.48047/jocaaa.2024.32.01.37

security concerns that can delay the widespread use of cloud computing. This article covers the challenges, risks, tactics, and solutions related to cloud computing security. A large number of respondents voiced worries about safety in an earlier poll. The cloud computing architecture paradigm is the subject of yet another assessment, with many articles discussing the difficulties and solutions to cloud security. Everything you need to know about security—the issues, challenges, strategies, and solutions—is right here in this article.

METHODOLOGY

We have developed the following modules to carry out this project.

1) The first step is to **upload the CERT database** to the programme. After the database is uploaded, the application will read all of the numbers and create a graph showing both regular and insider assaults.

2) Pre-process & Split Dataset: This module will eliminate missing values, shuffle and normalise the values, and divide the dataset into a "train" and "test" set. The application will use the former 90% of the time and the latter 20%.

3) Execute Random Forest: The random forest algorithm will be fed 80% of the training data in order to train a model, and then it will be tested on 20% of the data in order to determine the accuracy of the predictions.

4. Execute ADABOOST: The ADABOOST algorithm will be fed 80% of the training data in order to train a model, and then the model will be tested on 20% of the data in order to determine the accuracy of the predictions.

5) Execute XGBOOST: The XGBOOST algorithm will be fed 80% of the training data in order to build a model, and then it will be tested on 20% of the data in order to determine the accuracy of the predictions.

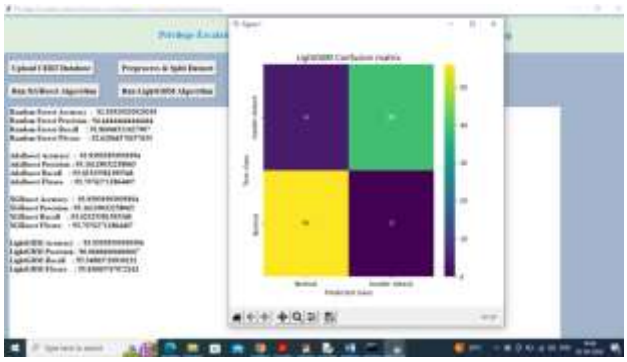
6) Execute LIGHTGBM: Train a model using 80% of the training data fed into the algorithm, and then apply it to 20% of the test data to determine the prediction accuracy. 7) The seventh step is to **execute the CATBOOST extension**, which involves feeding the algorithm 80% of the training data in order to build a model and then

10.48047/jocaaa.2024.32.01.37

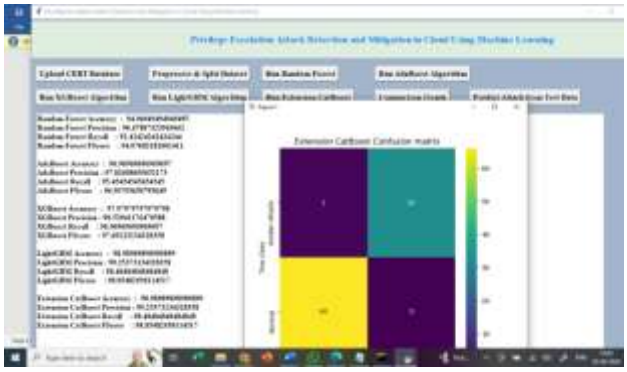
applying the model to 20% of the test data in order to determine the prediction accuracy.

8) Algorithm Comparison: will display an algorithm-by-algorithm graph.

RESULT AND DISCUSSION



Running the CATBOOST algorithm yielded the following results after LIGHTGBM achieved 95% accuracy (see above screen).



By clicking the "Comparison Graph" button, you can see that CATBOOST achieved an accuracy rate of 97% in the previous screen extension.

The names of the algorithms are on the x-axis, while various coloured bars on the y-axis show their accuracy and other metrics. It's clear that all of the algorithms that were extended had excellent performance. Afterwards, choose "Predict Attack from Test Data."

We projected the value from the test data as normal or Insider Attack utilising extension techniques, as seen in the above screen.

CONCLUSION

Since the hostile insider has more access and opportunities to do substantial harm, they

10.48047/jocaaa.2024.32.01.37

pose a critical danger to the company. others on the inside have exclusive and authorised access to resources and information that others on the outside do not. In order to identify and categorise insider attacks, this article presented machine learning techniques. This study makes use of a dataset that has been customised from several files inside the CERT dataset. Applying four machine learning methods to the dataset improved the results. The algorithms in question include LightGBM, XGBoost, Random Forest, and AdaBoost. The paper's experimental findings showed improved classification accuracy when using these supervised machine learning techniques. Of the algorithms that have been suggested, LightGBM has the best accuracy at 97%, followed by RF at 86%, AdaBoost at 88%, and XGBoost at 88.27%. Adding more variables to the dataset and keeping up with emerging insider threat patterns might help the suggested models perform better in the future. There may be new avenues for investigation into the detection and classification of insider assaults in various organisational domains as a result of this. Credible business choices are made by firms using machine learning models, and better judgements are made as a consequence of enhanced model outcomes. Errors may have a significant impact, but they can be mitigated by making the models more accurate. Research based on ML allows people to feed computer algorithms large volumes of data, and the algorithms then utilise that data to make decisions, recommendations, and evaluations.

REFERENCES

[1] U. A. Butt, R. Amin, H. Aldabbas, S. Mohan, B. Alouffi, and A. Ahmadian, ‘‘Cloud-based email phishing attack using machine and deep learning algorithm,’’ Complex Intell. Syst., pp. 1–28, Jun. 2022.

[2] D. C. Le and A. N. Zincir-Heywood, ‘‘Machine learning based insider threat modelling and detection,’’ in Proc. IFIP/IEEE Symp. Integr. Netw. Service Manag. (IM), Apr. 2019, pp. 1–6.

[3] P. Oberoi, ‘‘Survey of various security attacks in clouds based environments,’’ Int. J. Adv. Res. Comput. Sci., vol. 8, no. 9, pp. 405–410, Sep. 2017.

[4] A. Ajmal, S. Ibrar, and R. Amin, ‘‘Cloud computing platform: Performance analysis of prominent cryptographic algorithms,’’ Concurrency Comput., Pract. Exper., vol. 34, no. 15, p. e6938, Jul. 2022.

10.48047/jocaaa.2024.32.01.37

[5] U. A. Butt, R. Amin, M. Mehmood, H. Aldabbas, M. T. Alharbi, and N. Albaqami, ‘‘Cloud security threats and solutions: A survey,’’ Wireless Pers. Commun., vol. 128, no. 1, pp. 387–413, Jan. 2023.

[6] H. Touqeer, S. Zaman, R. Amin, M. Hussain, F. Al-Turjman, and M. Bilal, ‘‘Smart home security: Challenges, issues and solutions at different IoT layers,’’ J. Supercomput., vol. 77, no. 12, pp. 14053–14089, Dec. 2021.

[7] S. Zou, H. Sun, G. Xu, and R. Quan, ‘‘Ensemble strategy for insider threat detection from user activity logs,’’ Comput., Mater. Continua, vol. 65, no. 2, pp. 1321–1334, 2020.

[8] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido, and M. Marchetti, ‘‘On the effectiveness of machine and deep learning for cyber security,’’ in Proc. 10th Int. Conf. Cyber Conflict (CyCon), May 2018, pp. 371–390.

[9] D. C. Le, N. Zincir-Heywood, and M. I. Heywood, ‘‘Analyzing data granularity levels for insider threat detection using machine learning,’’ IEEE Trans. Netw. Service Manag., vol. 17, no. 1, pp. 30–44, Mar. 2020.

[10] F. Janjua, A. Masood, H. Abbas, and I. Rashid, ‘‘Handling insider threat through supervised machine learning techniques,’’ Proc. Comput. Sci., vol. 177, pp. 64–71, Jan. 2020.

[11] R. Kumar, K. Sethi, N. Prajapati, R. R. Rout, and P. Bera, ‘‘Machine learning based malware detection in cloud environment using clustering approach,’’ in Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2020, pp. 1–7.

[12] D. Tripathy, R. Gohil, and T. Halabi, ‘‘Detecting SQL injection attacks in cloud SaaS using machine learning,’’ in Proc. IEEE 6th Int. Conf. Big Data Secur. Cloud (BigDataSecurity), Int. Conf. High Perform. Smart Comput., (HPSC), IEEE Int. Conf. Intell. Data Secur. (IDS), May 2020, pp. 145–150.

[13] X. Sun, Y. Wang, and Z. Shi, ‘‘Insider threat detection using an unsupervised learning method: COPOD,’’ in Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE), May 2021, pp. 749–754.

[14] J. Kim, M. Park, H. Kim, S. Cho, and P. Kang, ‘‘Insider threat detection based on user behaviormodeling and anomaly detection algorithms,’’ Appl. Sci., vol. 9, no. 19, p. 4018, Sep. 2019.

10.48047/jocaaa.2024.32.01.37

[15] L. Liu, O. de Vel, Q.-L. Han, J. Zhang, and Y. Xiang, ‘‘Detecting and preventing cyber insider threats: A survey,’’ IEEE Commun. Surveys Tuts., vol. 20, no. 2, pp. 1397–1417, 2nd Quart., 2018.

[16] P. Chattopadhyay, L. Wang, and Y.-P. Tan, ‘‘Scenario-based insider threat detection from cyber activities,’’ IEEE Trans. Computat. Social Syst., vol. 5, no. 3, pp. 660–675, Sep. 2018.

[17] G. Ravikumar and M. Govindarasu, ‘‘Anomaly detection and mitigation for wide-area damping control using machine learning,’’ IEEE Trans. Smart Grid, early access, May 18, 2020, doi: 10.1109/TSG.2020.2995313.

[18] M. I. Tariq, N. A. Memon, S. Ahmed, S. Tayyaba, M. T. Mushtaq, N. A. Mian, M. Imran, and M. W. Ashraf, ‘‘A review of deep learning security and privacy defensive techniques,’’ Mobile Inf. Syst., vol. 2020, pp. 1–18, Apr. 2020.

[19] D. S. Berman, A. L. Buczak, J. S. Chavis, and C. L. Corbett, ‘‘A survey of deep learning methods for cyber security,’’ Information, vol. 10, no. 4, p. 122, 2019.

[20] N. T. Van and T. N. Thinh, ‘‘An anomaly-based network intrusion detection system using deep learning,’’ in Proc. Int. Conf. Syst. Sci. Eng. (ICSSE), 2017, pp. 210–214.

[21] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, ‘‘Deep learning for anomaly detection: A review,’’ ACM Comput. Surv., vol. 54, no. 2, pp. 1–38, Mar. 2021.

[22] A. Arora, A. Khanna, A. Rastogi, and A. Agarwal, ‘‘Cloud security ecosystem for data security and privacy,’’ in Proc. 7th Int. Conf. Cloud Comput., Data Sci. Eng., Jan. 2017, pp. 288–292.

[23] L. Coppolino, S. D’Antonio, G. Mazzeo, and L. Romano, ‘‘Cloud security: Emerging threats and current solutions,’’ Comput. Electr. Eng., vol. 59, pp. 126–140, Apr. 2017.

[24] M. Abdelsalam, R. Krishnan, Y. Huang, and R. Sandhu, ‘‘Malware detection in cloud infrastructures using convolutional neural networks,’’ in Proc. IEEE 11th Int. Conf. Cloud Comput. (CLOUD), Jul. 2018, pp. 162–169.

[25] F. Jaafar, G. Nicolescu, and C. Richard, ‘‘A systematic approach for privilege escalation prevention,’’ in Proc. IEEE Int. Conf. Softw. Quality, Rel. Secur. Companion (QRS-C), Aug. 2016, pp. 101–108

10.48047/jocaaa.2024.32.01.37