

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

Dr.AR.Sivakumaran^[1], G.Sreeja^[2],Ch.Poojitha^[3], I.Ramya^[4]

^[1]Associate Professor, Department of Information and Technology, Malla Reddy Engineering College for Women,
Autonomous,Hyderabad

^{[2],[3],[4]} Student, Department of Information and Technology, Malla Reddy Engineering College for Women,
Autonomous, Hyderabad, ramyachinnam77@gmail.com, dhodahimavarsha2021@gmail.com,
bogidewarlxmi@gmail.com

ABSTRACT:

The exponential growth in attack frequency and complexity in the past few years has made cybersecurity a major concern with the advent of smart devices. Cloud computing has changed the way businesses operate, but users may find it more challenging to use dispersed services, such as security systems, due to their centralization. Organizations and cloud service suppliers exchange massive amounts of data, which poses a significant risk of accidental or intentional disclosure of sensitive information. Because of their increased access and potential to do substantial harm, an antagonistic insider poses a serious threat to the company. Only approved individuals within the organization have access to sensitive data and assets. This research details a machine learning-based strategy for classifying insider threats and finding out-of-the-ordinary events that can indicate privilege escalation security issues. The system uses a systematic approach to detect these irregularities. Machine learning and prediction accuracy are both enhanced by ensemble learning, which considers several models simultaneously. Using anomaly and weakness detection, some studies have attempted to identify security issues or hazards associated with privilege delegation in network systems. However, the assaults cannot be definitely identified from these research. Ensembles for machine learning (ML) are suggested and assessed in this research. The objective of this endeavor is to classify insider assaults using machine learning approaches. The dataset it uses has been modified from many files beneath the

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

CERT dataset. The dataset is subjected to four machine learning techniques: Light GBM, XG Boost, Ada boost, and three Random Forest (RF) methods. In terms of overall performance, light was superior. In contrast, RF and AdaBoost are two algorithms that may be better at preventing assaults from inside, such as attacks using behavioral biometrics. Consequently, it is possible that various internal threats may be better classified by combining various machine learning algorithms. With a 97% dependability rate, the Light GBM method outperforms the other suggested techniques; RF, AdaBoost, and XG Boost all have 88% accuracy rates.

INTRODUCTION

Noiseless computing is a novel approach to enabling and providing activities via the Internet. Major modifications in data processing, presentation, and storage have been brought about by the current economic downturn and expanding computing needs under the contemporary Cloud Model. Customers may save a ton of money on equipment purchases and upkeep by using cloud computing. Providers of cloud storage rely on authentication, access limitations, and encryption as their primary security measures for both their on-site facilities and the information that they manage. The almost limitless capacity of the cloud to store any kind of data in various cloud storage structures is limited only by the accessibility, speed, and frequency of data access. Businesses exchange more data with their cloud service providers, which increases the risk of confidential information breaches (deliberate or not). Businesses have a harder time preventing unauthorized access to internet services due to the same attributes that make them accessible to workers and IT systems. Businesses relying on cloud services are particularly vulnerable to two emerging security threats: authentication and open interfaces. It is possible for hackers with specialized expertise to breach cloud systems. To improve data management and overcome security hurdles, machine learning makes use of a wide variety of techniques and algorithms. Due to privacy concerns, many datasets are not made publicly accessible because they are either too tiny or do not have significant statistical features. Regulators are addressing security and privacy concerns brought about by the exponential expansion of the computer sector. No matter what a worker does for the Cloud Company, their login credentials may remain the same.

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Confidential information is inadvertently leaked or stolen due to the misuse of outdated permissions. An assigned quantity of power is associated with each computer-connected account. Databases, sensitive information, and other functions of servers are often accessible only to authorized users. An adversarial hacker might potentially compromise a sensitive system by obtaining access to it via a subscriber account with elevated privileges and then abusing or expanding those rights. In order to achieve their goals, attackers may leverage either horizontal control of various systems or vertical access to root and executive powers. Taking on another user's rights at the same level of access is called horizontal privilege escalation. An evildoer can potentially gain information that isn't personally relevant to them by exploiting a security hole in the system. Security breaches involving users' personal information might occur in web programs that are not well-built. Once a horizontal elevation of privilege exploit is successfully executed, the attacker has the ability to view, change, and delete sensitive data. An assault employing horizontal privilege escalation was launched against organizational domains, as seen in Figure 1. This kind of attack often requires malicious software and an in-depth understanding of the issues impacting certain computer operating systems. Giving a person, piece of code, or other asset more limited access or rights than they already have is known as a privilege elevation attack. From a modest degree of privileged access, the attacker aims to increase the amount of elevated privilege. The attacker could have to use a plethora of tactics to bypass security systems and get control over the entrance from above. Business goals like separation of duties and least privilege are achieved via the deployment of more complicated security models in vertical privilege constraints. One method an attacker can use to get administrative or root access to a networking is to pose as a valid registered member. Behavioural analytics may help identify suspicious activity on company systems or personal accounts. This can indicate a potential expansion of rights or an invasion of privacy.

RELATED WORK

Cloud-based phishing attacks and machine learning

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

"Cloud computing" refers to the method by which resources like data storage and processing power are made available to users via their personal computers on demand, without requiring any kind of interaction from the users themselves. Most individuals and businesses use email to deliver and receive data. Credit reports, bank records, and other such sensitive information frequently go online. Scammers employ phishing to trick victims into giving over sensitive information by making their communications seem to come from reputable sources. The sender of a phishing email may try to trick you into giving over sensitive information. Phishing attempts are the biggest concern while sending or receiving emails. If you fall for a scam email and click on the link, an attacker will have access to your personal information. It has been a huge issue for everyone involved throughout the last several years. In order to discover fresh emails, this study employs a number of categorization criteria and algorithms and makes use of varying amounts of data from both genuine and phishing sources. Following an evaluation of the current methods, updated data is created. Employing the SVM, NB, and LSTM algorithms, we generated characteristic extracting documents including label files from CSV files with value pairs delimited by commas. The experiment treats the problem of identifying malicious emails as an issue of classification. Research and testing have shown that LSTM, SVM, and NB are the best methods for detecting email phishing attempts. The best accuracy rates for email attack classification were 99.62% with SVM, 97% with NB, and 98% with LSTM classifiers.

Using Machine Learning to Model and Detect Insider Threats

One of the most pressing problems facing governments and businesses today is the rise of malicious insider assaults. Using several levels of data granularity, this article suggests an insider threat identification system that is user-centric and based on machine learning. In both general and particular information instances, we publish and discuss the findings of our system assessments and scenario-specific investigations on trustworthy sources, both good and negative. Our findings demonstrate that the detection strategy based on machine learning can identify new

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

malevolent insiders and adapt to sparse ground truth.

An Overview of Cloud Security Threats

Computing and Their Remedies

Cloud computing has the potential to help the sector save money on computer infrastructure setup costs while still taking use of IT-based solutions and products. This flexible IT infrastructure will allow small, portable devices to access the internet, as stated in the announcement. So, it's possible to double the ability of both the current and future software. Within the cloud computing system, users have access to a shared network from which they may access any resource. Users have no say over the location or security of the computing facilities that they utilize. A lot of issues with security and privacy arise from this, and they require fixing. Additionally, it is difficult to totally exclude the notion due to the frequency of server disruptions recently. The use of the cloud raises several concerns about the privacy and security of user data. With the goal of describing and assessing the many unsolved challenges preventing the acceptance and expansion of cloud computing, this lengthy study paper targets a broad range of stakeholders.

Cloud computing platform evaluation: renowned cryptographic algorithms

As technology and science continue to grow, cloud computing is predicted to progressively replace on-premises systems. To ensure the safety of data, cloud cryptography may be used. Due to the many benefits of cloud storage—including accessibility, cheap maintenance costs, less security concerns, and lower hardware needs—every firm is moving its operations over the cloud. Information may be protected against unauthorized access by using encryption. Keeping information secure either in transit or at rest on a computer is a top priority in the modern day. It is these traits—private, bandwidth-dependent, integrity-dependent, and responsive—that dictate the method of encryption. Ensuring the security of consumer data on the internet is another

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

essential aspect of cloud computing. Various cryptographic systems are compared in this study paper based on their efficacy, consumption, and practicality. The results show which algorithms work better with different kinds of data and in different environments.

A Synopsis of the Cloud's Security Risk and Remedials

"Cloud computing" refers to the practice of making available framework resources for computer systems on an as-needed basis. In particular, the ability to manage and save data despite tailored client administration. With it, its clients have gained access to a unified platform for private and public cloud computing, data storage, and global accessibility. Additionally, cloud computing could not be extensively adopted due to certain security concerns. Problems, threats, solutions, and tactics related to cloud computing security are covered in this article. A number of respondents voiced worries about security in an earlier poll. There have been a number of articles discussing the problems with and potential solutions to cloud security, and another study of the use of cloud computing architecture paradigm is in the works. Everything you need to know about security, from its challenges and issues to its methods and answers, is right here on this page.

METHODOLOGY

Here are the components that we have developed to finish this project:

- 1) The first stage is to upload the CERT data to the program. The software will examine the data when the database is uploaded and generate a graph showing both normal and insider assaults.
- 2) The second step is to pre-process the dataset by removing missing values, shuffling and normalizing the values, and splitting the dataset into a "train" and "test" set. The software will make 90% use of the first and 20% use of the second.

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

3) Run Random Forest: The algorithm for random forests receives 80% of the data that was used to build a model. Its predictive accuracy is then tested on a 20% subset of the data.

4) Implement the ADABOOST algorithm as follows: after feeding the algorithm 80% of the training data, train a model on top of it, then evaluate its prediction accuracy employing the other twenty percent of the information.

5) In order to generate a model, the XGBOOST algorithm is fed 80% of the raw data. After that, it checks the forecasts' accuracy using 20% of the data.

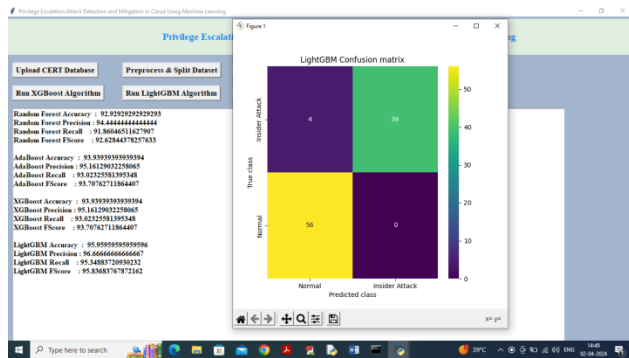
6) Apply LIGHTGBM: utilize a model trained on 80% of the original data set to evaluate prediction accuracy, then utilize 20% of the experimental data to refine the model.

7) In the seventh step, another CATBOOST extension is activated. One way to do this is to use eighty percent of the training data to train the method, and then use twenty percent with the test data to determine the model's prediction accuracy.

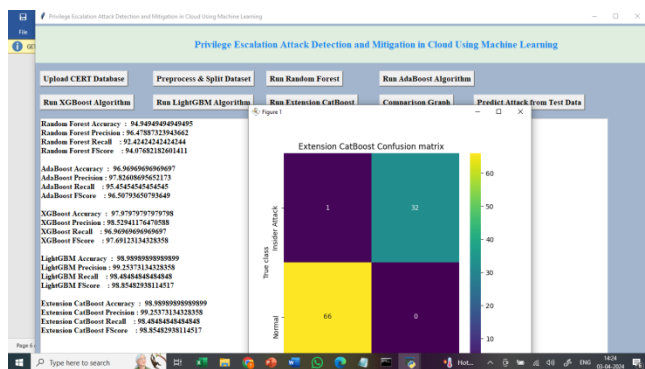
8) Graphs dissecting every algorithm will be demonstrated for the sake of algorithm comparison.

RESULT AND DISCUSSION

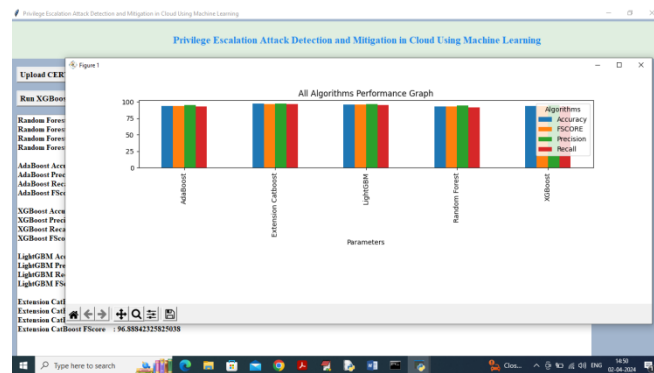
Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).



After achieving 95% accuracy with LIGHTGBM, the CATBOOST algorithm produced the following results (see screen above).



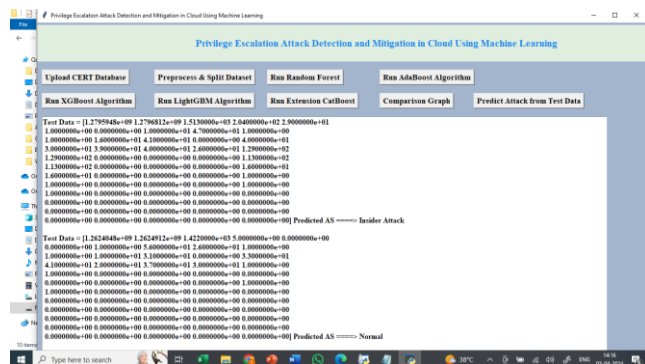
By clicking the "Comparison Graph" button, you can see that CATBOOST achieved an accuracy



rate of 97% in the previous screen extension.

On the y-axis, different coloured bars display the algorithms' names and other metrics, while on the x-axis, the algorithms' names are shown. It is evident that every algorithm that was expanded performed very well. After that, choose "Determine Attack using Test Data."

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).



Using extension methods, we projected the result from the experimental data as either normal or an insider attack, as seen in the screen above.

CONCLUSION

Because of their increased access and potential to do substantial harm, an antagonistic insider poses a serious threat to the company. Personnel with proper authorization are the only ones who can access information and records that are not accessible to the general public. The use of machine learning techniques for the detection and categorization of insider attacks was presented in this work. Several files from the computer emergency response team dataset were used to create a customized dataset for this study. Applying four machine learning approaches improved the dataset's performance. The techniques that are being discussed are Light GBM, XG Boost, Random Forest, and AdaBoost. The experimental findings of the article showed that classification accuracy was improved by applying these regulated machine learning methodologies. With an accuracy of 97%, Light GBM surpasses all other algorithms that have been suggested. RF comes in second with 86%, AdaBoost in 88%, and XG Boost with 88.27%. Keeping up with evolving trends in insider threats and increasing the size of the dataset could lead to future improvements in the suggested models' performance. This can spark fresh ideas for studies that aim to classify and identify insider assaults in various parts of organizations. Machine learning models help businesses make better judgments by using trustworthy information and improving the model's outputs. Improving the accuracy of the models may help

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

reduce the impact of errors, even if they may still be significant. The goal of machine learning is to enable people to input massive volumes of statistical data into computer algorithms, which can then utilize that data to make recommendations, evaluations, and judgments.

REFERENCES

- [1] U. A. Butt, R. Amin, H. Alibis, S. Mohan, B. Aloft, and A. Ahmadi an, “Cloud-based email phishing attack using machine and deep learning algorithm,” *Complex Intel. Syst.*, pp. 1–28, Jun. 2022.
- [2] D. C. Le and A. N. Zincin-Heywood, “Machine learning based insider threat modelling and detection,” in *Proc. IFIP/IEEE Sump. Integer. Newt. Service Manga. (IM)*, Apr. 2019, pp. 1–6.
- [3] P. Oberoi, “Survey of various security attacks in clouds-based environments,” *Int. J. Adv. Res. Compute. Sci.*, vol. 8, no. 9, pp. 405–410, Sep. 2017.
- [4] A. Ajmal, S. Ibra, and R. Amin, “Cloud computing platform: Performance analysis of prominent cryptographic algorithms,” *Concurrency Compute., Pact. Expert.*, vol. 34, no. 15, p. e6938, Jul. 2022.
- [5] U. A. Butt, R. Amin, M. Mehmood, H. albas, M. T. Aldhabi, and N. Alabama, “Cloud security threats and solutions: A survey,” *Wireless Pers. Common.*, vol. 128, no. 1, pp. 387–413, Jan. 2023.
- [6] H. Tourer, S. Zaman, R. Amin, M. Hussain, F. Al-Tudjman, and M. Bilal, “Smart home security: Challenges, issues and solutions at different IoT layers,” *J. Supercomputer.*, 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,” *Compute., Mater. Continua*, vol. 65, no. 2, pp. 1321–1334, 2020.

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

- [8] G. Apprizes, M. Calacanis, 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 (Cyc on), May 2018, pp. 371–390.
- [9] D. C. Le, N. Zincin-Heywood, and M. I. Heywood, “Analysing data granularity levels for insider threat detection using machine learning,” IEEE Trans. Newt. Service Manga., 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. Compute. Sci., vol. 177, pp. 64–71, Jan. 2020.
- [11] R. Kumar, K. Seth, N. Prajapati, R. R. Rout, and P. Brea, “Machine learning based malware detection in cloud environment using clustering approach,” in Proc. 11th Int. Conf. Compute., Common. Newt. Technol. (ICCCNT), Jul. 2020, pp. 1–7.
- [12] D. Tripathi, R. Gohil, and T. Halaby, “Detecting SQL injection attacks in cloud SaaS using machine learning,” in Proc. IEEE 6th Int. Conf. Big Data Secure. Cloud (Bigdata Security), Int. Conf. High Perform. Smart Compute., (HPSC), IEEE Int. Conf. Intel. Data Secure. (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. Common., Inf. Syst. Compute. 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 behaviour modelling and anomaly detection algorithms,” Appl. Sci., vol. 9, no. 19, p. 4018, Sep. 2019.
- [15] L. Liu, O. de Vela, Q.-L. Han, J. Zhang, and Y. Xiang, “Detecting and preventing cyber insider threats: A survey,” IEEE Common. Surveys Tuts., vol. 20, no. 2, pp. 1397–1417, 2nd Quart., 2018.

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

- [16] P. Chattopadhyay, L. Wang, and Y.-P. Tan, “Scenario-based insider threat detection from cyber activities,” *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 3, pp. 660–675, Sep. 2018.
- [17] G. Ravikumar and M. Govind Arius, “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. Memo, S. Ahmed, S. Tayyaba, M. T. Mushtaq, N. A. Mina, 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. Buzau, 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. Thanh, “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. Hengelo, “Deep learning for anomaly detection: A review,” *ACM Comput. Surf.*, 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 Compute., Data Sci. Eng.*, Jan. 2017, pp. 288–292.
- [23] L. Cippolini, S. Antonio, G. Mazzei, and L. Romano, “Cloud security: Emerging threats and current solutions,” *Compute. Electra. Eng.*, vol. 59, pp. 126–140, Apr. 2017.
- [24] M. Abdulsalam, R. Krishnan, Y. Huang, and R. Sandhu, “Malware detection in cloud infrastructures using convolutional neural networks,” in *Proc. IEEE 11th Int. Conf. Cloud Compute. (CLOUD)*, Jul. 2018, pp. 162–169.

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

[25] F. Jaafar, G. Niculescu, and C. Richard, “A systematic approach for privilege escalation prevention,” in Proc. IEEE Int. Conf. Soft. Quality, Rel. Secure. Companion (QRS-C), Aug. 2016, pp. 101–108



Dr. AR. SIVAKUMARAN

Dr.AR.SIVAKUMARAN, has been working as a Associate Professor in Department of Information Technology, Malla Reddy Engineering College for Women, Secunderabad, Telangana, India, since 2019. He received his Doctorate Degree from Anna University, Chennai, Tamil Nadu. He received MTech(CSE) Degree from Motilal Nehru National Institute of Technology (NIT), Allahabad, Uttar Pradesh. He has a Good Academic and Research Experience of more than 23 years. His current area of research includes Web Mining, AI, NLP, Deep Learning and Machine Learning. He has published many papers in Scopus, UGC Care List and reputed International Journals. He has five patent publications.

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).