

Original Article

Machine Learning-Based Detection of Privileged Access Anomalies in Cloud Infrastructure

Uday Allala

Lead SRE Engineer, USA.

Received: 03-02-2019

Revised: 03-24-2019

Accepted: 04-01-2019

Published: 04-03-2019

ABSTRACT

Cloud infrastructure adoption has made privileged account management and identification of any rogue or irregular access more complex. Cloud infrastructure adoption has made privileged account management and identification of any rogue or irregular access more complex. Privileged users have higher permission levels that can give them access to sensitive cloud resources, which makes them a desirable target for credential compromise, privilege abuse and insider attacks. Security mechanisms based on rules tend to be inefficient at detecting behaviorally obscure or new anomalies in large-scale cloud deployments. In this paper, a machine learning based framework for identifying privileged access anomalies through the analysis of access logs, user activity, authentication behavior, resource usage, and temporal behavior is proposed. The suggested approach consists of the following five steps: Data preprocessing, Feature extraction, Feature selection, Behavioral profiling, Anomaly detection using machine learning. The proposed framework is evaluated using representative cloud access data, and multiple machine learning models used to determine the effectiveness of detecting anomalous privileged activities. The accuracy, precision, recall, F1 score, receiver operating characteristic area under the curve (ROC-AUC) and false-positive rate are used to evaluate performance. The experimental analysis is targeted towards the identification of abnormal login activity, unusual resource usage, privilege escalation trends, and abnormal user activity. The effectiveness of the proposed framework is compared with the conventional and baseline machine learning methods. The outcomes illustrate the efficacy of machine learning in enhancing the detection of privileged access anomalies and enable scalable, intelligent monitoring for cloud security. The proposed framework offers a pragmatic basis to improve cloud access security and identify potentially malicious privileged activities early.

KEYWORDS

Privileged Access, Cloud Security, Anomaly Detection, Machine Learning, User Behavior Analytics, Cloud Infrastructure, Cybersecurity.

1. INTRODUCTION

1.1. Background and Motivation

Cloud computing has emerged as a key infrastructure for organizations today that enables scalable access to computing resources, applications, databases and services. In these environments, privileged users like cloud administrators, security administrators and service accounts have broad rights to access sensitive resources and conduct critical operations. [1] Therefore, the compromise or mis-use of privileged accounts may lead to access to unauthorized data, privilege escalation, service disruption and large-scale breaches of security. Most traditional approaches to access control and monitoring rely on a set of policies and patterns of attacks, and are not very effective in cases where these patterns are absent or are more complex and new. The growing amount and complexity of access logs on clouds, therefore, signal the need for more intelligent and adaptive tools that continuously monitor privileged user activities. Machine learning (ML), which learns the behavioral patterns of historical access activities and detects deviations that could signify suspicious or malicious behavior, is an effective solution.

1.2. Security Challenges and Research Problem

Detecting anomalous privileged access to cloud infrastructure can be difficult because normal cloud admin usage may differ greatly between users, resources, locations, devices, and time. [2] Once attackers gain privileged access to credentials, they can try to mimic regular user activity, making it hard to use traditional methods of detecting attacks by signatures. Moreover, large scale cloud environments produce access logs that include authentication events, requests for resources, changes in permissions, network data and temporal activity patterns, and they are heterogeneous and voluminous. The research proposed in this study aims to solve the issue of accurately detecting abnormal privileged access behavior from this complex cloud activity data while minimizing false alarms. The research aims to create machine learning based detection framework that learns meaningful behavioral characteristics, identifies patterns related to normal and abnormal behavior, and classifies suspicious privileged access events to enable early detection and reaction to security incidents.

1.3. Research Objectives and Contributions

The main goal of this research is to create an intelligent machine learning based framework to detect privileged access anomalies in cloud infrastructure. The proposed method includes access-log preprocessing and feature extraction to capture meaningful access-log behaviors of privileged users and anomaly detection and risk assessment using machine learning techniques. Many machine learning models may be explored and compared against each other with the help of suitable performance metrics such as precision, recall, F1-score, accuracy, ROC-AUC and false-positive rate. The most significant aspect of this research is the design and implementation of a behavior based detection mechanism, which can detect unusual privileged access patterns, and can be used for scalable cloud security monitoring. Moreover, systematic analysis of machine learning models is provided and the effectiveness of the models in the detection of the different types of privileged

access anomalies is analyzed, offering practical insights for the improvement of cloud infrastructure security.

2. RELATED WORK

Table 1: Comparison of Existing Privileged Access Anomaly Detection Approaches

Approach	Technique	Data/Features Used	Detection Type	Major Limitation
Rule-Based Detection	Signature/Rule Matching	Authentication and access logs	Known anomalies	Limited detection of unknown attacks
User Behavior Analytics	Statistical/Behavioral Analysis	User and session behavior	Behavioral anomalies	Sensitive to changing user behavior
Random Forest	Supervised ML	Access and activity features	Classification	Requires labeled training data
Support Vector Machine	Supervised ML	Authentication and access features	Classification	Computationally expensive for large datasets
XGBoost	Ensemble ML	Behavioral and access features	Classification	Requires parameter optimization
Isolation Forest	Unsupervised ML	Behavioral feature vectors	Anomaly detection	May produce false positives
Proposed Framework	ML-Based Behavioral Detection	Privileged access and behavioral features	Anomaly detection and risk scoring	Requires continuous model adaptation

2.1. Privileged Access Management and Machine Learning-Based Anomaly Detection

The objective is to give an overview of Privileged Access Management and Machine learning Anomaly detection. Privileged accounts have a wide range of access to important resources, services, and sensitive information, making Privileged Access Management (PAM) an important element of cloud security. [3] The main methods for mitigating the risk of unauthorized privileged operations in existing PAM approaches include: authentication controls, role based access control, multi factor authentication, least privilege policies, and activity monitoring. But existing rule-based and signature-based approaches typically rely on pre-established rules and known attack patterns, making them vulnerable to novel or behaviorally subtle attacks. To overcome these drawbacks, the machine learning techniques have been explored to detect cloud anomalies through studying authentication records, access logs, resource requests, network activities, and user behavior records. The known malicious activities can be classified by supervised algorithms like Random Forest, Support Vector Machine, and Gradient Boosting methods, while unsupervised and semi-supervised algorithms like Isolation Forest, One-Class SVM, and Autoencoders can detect deviations without necessitating large amounts of labelled attack data. The research and studies show how a combination of privileged access monitoring and machine learning can be used to enhance the security of the cloud in more adaptive and intelligent ways.

2.2. User Behavior Analytics, Research Gaps, and Motivation

User and Entity Behavior Analytics (UEBA) techniques rely on the historical analysis of behaviors to create profiles and detect deviations for potentially malicious users and/or compromised accounts. [4] Current UEBA systems have been shown to work for detecting unusual logon times, abnormal use of resources, unusual geographic locations, excessive privilege usage, and deviations from normal user activity patterns. But there are several research challenges, such as high dimensionality of cloud access data, user behavior changes, imbalance in datasets, generation of false positives, small number of accessible labelled privileged access anomalies, and the challenge of real time detection at cloud scale. Moreover, existing methods are generally devoted to detecting anomalies among common users instead of behavioral characteristics of privileged accounts, where more stringent security measures are needed due to higher privileges given to these accounts. These constraints drive the need to create a specialized machine learning framework, incorporating privileged access capabilities, behavioral profiling, and anomaly detection to ensure accurate detection of suspicious activities, minimal false alarms, and scalable cloud infrastructure security.

3. CLOUD PRIVILEGED ACCESS AND THREAT MODEL

Table 2: Privileged Access Anomaly Types and Characteristics

Anomaly Type	Description	Representative Indicators	Security Impact
Abnormal Login	Login inconsistent with normal user behavior	Unusual time, location, or device	Unauthorized access
Repeated Authentication Failure	Multiple unsuccessful authentication attempts	High failure frequency	Credential attack
Privilege Escalation	Unexpected increase in user privileges	Permission or role modification	Unauthorized administrative control
Unusual Resource Access	Access to uncommon or sensitive resources	High resource-access frequency	Data exposure
Abnormal Administrative Activity	Unusual privileged operations	Unexpected configuration changes	Infrastructure compromise
Unusual Session Behavior	Session characteristics deviate from normal patterns	Abnormal duration or activity	Account compromise
Geographic Anomaly	Access from an unexpected location	Sudden location change	Suspicious account usage

3.1. Cloud Privileged Access Architecture

Cloud infrastructure comprises users, privileged administrators, virtual machines, containers, databases, storage services, applications and cloud management interfaces that communicate with each other through authentication and authorization mechanisms. [5] Privileged users have higher privileges to execute sensitive operations like changing security policies, managing identities, accessing confidential resources, etc., and administering cloud services. The proposed architecture takes into consideration privileged access activities gathered from authentication systems, identity and access management services, cloud audit logs, resource access logs and administrative operations. These activities are mapped into behavioural characteristics that define user identity,

access time, source location, accessed resources, operation type, session characteristics and privilege related actions. The activity information is sent to the machine learning-based detection layer, which determines normal privileged activity and flags deviations that could be considered an anomaly and thus warrant further investigation.

3.2. Privileged Access Anomaly Types and Threat Model

Privileged access anomalies may result from compromise, misuse or exploitation of legitimate administrative accounts and result in the inability to perform authorized activities. [6] This study investigates the following threat model: Abnormal login attempts, unusual access locations, access outside of normal working hours, excessive resource access, unauthorized privilege changes, abnormal administrative operations, and suspicious sequences of privileged activities. A valid credential can be stolen or otherwise compromised and access to sensitive cloud resources can then be attempted without following the normal security protocols or procedures. Insider misuse is also taken into account, which is when an authorised privileged user carries out activities that are very different from those that they are normally authorised to do. The machine learning system thus classifies large deviations in the behavior of the privileged user as possible security events and analyzes them based on their behavior and the corresponding risk level.

3.3. Security Requirements

To ensure protection of cloud infrastructure, the privileged access anomaly detection system should meet certain security requirements. First, the system must accurately detect privileged activity that should be considered anomalous, while keeping the false positive rate to a minimum so it doesn't generate too many false alarms that security administrators have to respond to. Second, it must enable ongoing monitoring of and near real-time analysis of high-volume cloud access events. Third, the detection mechanism should be scalable across a variety of cloud resources and adjust to the dynamic nature of user behavior and emerging attack methods. The system should also maintain the confidentiality and integrity of access-log information, offer interpretable anomaly indicators (where possible), and present alerts based upon the risk, to facilitate prompt investigation and response. The requirements lay the groundwork for assessing the proposed machine learning-based privileged access anomaly detection framework.

4. DATASET AND FEATURE ENGINEERING

4.1. Cloud Access Log Dataset and Data Preprocessing

In this study, the dataset is comprised of cloud access and administrative activity records, which are the interactions of privileged users with cloud resources. [7] These attributes could be such as user identity, authentication status, time of occurrence, source IP address, geographical data, accessed resource, type of operation, session duration, privilege level, and access outcome. The collected data is initially analyzed to determine which information is missing, which events are repeated, which attributes are missing, and which attributes are irrelevant. Some of the data preprocessing steps involve reducing redundant data, dealing with missing or noisy data, transforming categorical data into appropriate numeric formats, and segmenting events by users and

sessions. Where necessary, numerical features are normalized so that they are scaled to be in the same range and the difference in feature scale does not unduly affect the machine learning models. These operations are pre-processed (pre-processed data) and will serve as a consistent and reliable data source for further privileged access behavior analysis.

4.2. Privileged Access Feature Extraction

Feature extraction converts raw cloud access logs into useful representations of privileged user actions that can be used by machine learning algorithms for analysis. [8] The extracted features represent authentication behavior, temporal activity, resource access, network characteristics, and privilege related operations. Several important behavioral attributes are login counts, failed logon attempts, login time, session length, accesses to resources, accesses to administrative resources, changes in source-location, changes in privileges, and unusual accesses. The temporal features might detect abnormal working hours while resource-based features might detect access to sensitive services or data that is not typical. The feature representation is the result of the fusion between the individual event features, and the aggregated behavioral features, which gives it a more complete description of the privileged user's activity and better identification power for the detection system to differentiate the normal behavior from any possible anomalies.

4.3. Feature Selection and Representation

Feature selection is used to find out the most informative features to detect privileged access anomaly and remove irrelevant, redundant and highly correlated information. [9] The contribution of individual features to anomaly detection can be determined by statistical analysis, correlation assessment, or feature-importance methods or model-based selection techniques. Selected features are then represented in a structured numerical format suitable to be used in machine learning models, with categorical attributes being encoded using appropriate encoding techniques. Feature selection might enhance the efficiency of the model, reduce the number of operations performed, and even reduce the number of false detections resulting from irrelevant features. The final feature vector is the behavioral profile of each privileged access event/user session and is fed into the proposed anomaly detection framework based on machine learning.

5. PROPOSED MACHINE LEARNING FRAMEWORK

5.1. Overall System Architecture

It is suggested that the proposed framework can be used to detect anomalous privileged access activities in cloud infrastructure using a holistic security architecture based on machine learning. [10] This framework involves various stages: cloud access-log collection, data preprocessing, feature extraction, behavioral profiling, machine learning-based anomaly detection, and risk assessment. The events from privileged access are initially gathered from authentication services, identity and access management systems, cloud audit log and resource monitoring components. The gathered information is cleaned and processed into behavioral characteristics that describe the authentication, temporal, resource, network and privilege related properties. These features are then passed on to the machine learning detection layer where normal and anomalous

access patterns are detected. Finally, the identified anomalies are sorted by severity and risk scores are assigned for security monitoring and expedient investigation.



Fig 1 - Proposed Machine Learning Framework for Privileged Access Anomaly Detection in Cloud Infrastructure

5.2. Privileged Access Behavior Modeling

Privileged access behavior modeling creates a definition of typical actions performed by authorized administrative user/services in the cloud environment. The proposed approach will look at historical access events and extrapolate behaviour from login frequency, access time, session duration, the resources used, the source location, administrative operations and privilege changes. User-level and session-level attributes are combined to form behavioral profiles representing normal privileged behavior, but with some flexibility for common legitimate behavior. The deviations from these learned profiles are used for detecting anomalous behavior as a possible indicator and they are passed on to detection mechanism for further analysis. This representation of behavior allows the framework to recognize that there are some minor deviations that might not be caught by traditional static access-control policies.

5.3. Machine Learning-Based Anomaly Detection

The machine learning part of the anomaly detection module checks if the activity observed is consistent with the normal behavior or if it is a possible anomaly based on the features extracted from privileged access. [11] Relations between access characteristics and security outcomes can be learned using supervised, unsupervised or hybrid machine learning techniques, depending on whether there is data available to label them. Appropriately prepared training data is used in the training of models and then the models are evaluated by using unseen access events to check their generalization capability. During detection, features of abnormal login activity, access to resources, administrative activities, privilege changes, and temporal and geographic deviations are considered. This prediction and/or anomaly score will be used to articulate suspicious privileged access events and to trigger the next phase of risk assessment.

5.4. Anomaly Classification and Risk Scoring

The final one categorizes privileged access anomalies that are detected by their behavior and likely damage to security. The severity of the anomaly, privilege level, sensitivity of the resource, frequency of access to the resource, deviation from established behavior, and multiple suspicious behaviors can be used to classify anomalies into low risk, medium risk, high risk, or critical risk. [12] The machine learning anomaly score can be combined with other contextual security properties to calculate a risk score, which can then be used to prioritize the most important events. The high-risk deviations may trigger alerts that require further investigation, whereas low-risk deviations may be kept in storage for ongoing monitoring and behavioral analysis. This risk-based approach will help limit alert proliferation and aid prioritization of potential privileged access threats.



Fig 2 - Proposed Machine Learning-Based Framework for Privileged Access Anomaly Detection in Cloud Environments

6. MACHINE LEARNING MODEL DEVELOPMENT



Fig 3 - Machine Learning Model Development Framework for Privileged Access Anomaly Detection

Table 3: Machine Learning Models and Parameter Configuration

Model	Learning Type	Important Parameters	Purpose
Random Forest	Supervised	Number of Trees, Maximum Depth	Privileged access classification
SVM	Supervised	Kernel, C, Gamma	Anomaly classification
XGBoost	Supervised	Learning Rate, Estimators, Maximum Depth	High-performance classification
Isolation Forest	Unsupervised	Number of Estimators, Contamination	Unsupervised anomaly detection
Proposed Framework	ML-Based	[Actual configuration]	Privileged access anomaly detection and risk scoring

6.1. Machine Learning Algorithm Selection and Model Training

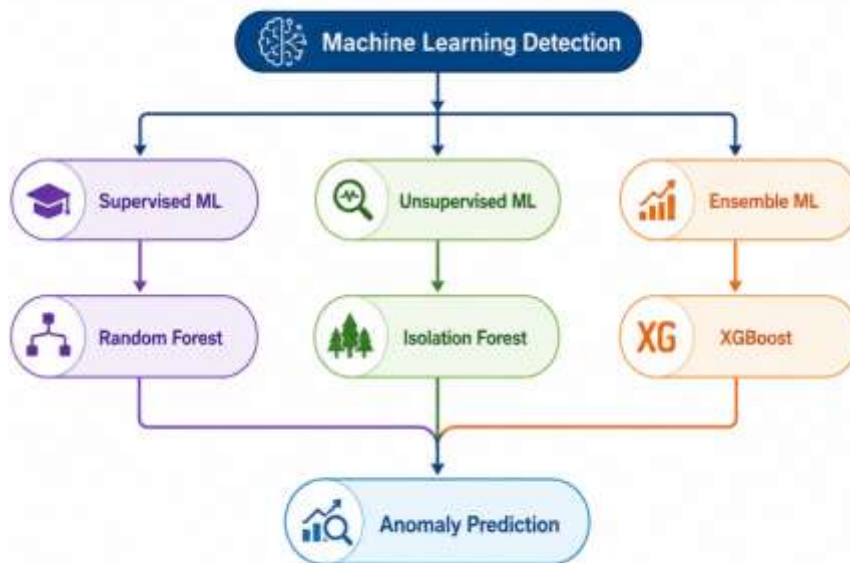


Fig 4 - Machine Learning Detection Architecture for Privileged Access Anomaly Identification

Choosing the right machine learning algorithms can play a crucial role in building an effective privileged access anomaly detection system. [13] Depending on the cloud access data (its size and nature), a mixture of supervised and unsupervised algorithms may be studied. Random Forest, Support Vector Machine, XGBoost, Isolation Forest and Autoencoder are some of the algorithms that can be used for capturing different types of normal and anomalous access behavior. The engineered feature set from the preprocessing and feature extraction steps is fed to the selected models for their training. The data set is split into training and testing sets and cross-validation can be used in model development to enhance reliability and minimize potential overfitting. A number of trained models are then tested with suitable security and classification metrics to find the best way to perform privileged access anomaly detection.

6.2. Hyperparameter Optimization and Anomaly Detection

Hyperparameter tuning is performed for enhancing the prediction and anomaly detection capability of the selected machine learning models. The number of trees, tree depth, learning rate, kernel parameters, contamination level, number of hidden layers and regularization parameters are tuned when needed depending on the algorithms. Valuation data can be used to search for appropriate parameter combinations (e.g., grid search, random search, etc.). Optimized, this final model is used to classify incoming privileged access events, based on their learned behavior characteristics, or to generate an anomaly score. Any event that is outside of the normal patterns is being flagged as a potential anomaly and passed to the next classification and risk-scoring step.

6.3. Computational Complexity and Model Performance

The computational complexity of machine learning models proposed for the system is analyzed to ensure that they can be implemented at a large scale in cloud computing and continuous security monitoring. [14] The analysis focuses on the training time, prediction time, memory needs, number of features, dataset size, and the complexity of the model. High prediction accuracy can be accompanied by high computing demands, which can cause latency when dealing with large amounts of cloud access events. In this context, the study assesses the cost/benefit of accuracy in detection and computational efficiency measured by precision, recall, F1 score, ROC-AUC, false-positive rate, training time and inference time. This analysis can be used to determine whether there is a model that represents a reliable way of achieving privileged access anomaly detection without too high a computational burden on scalable cloud infrastructure.

7. EXPERIMENTAL SETUP AND EVALUATION

7.1. Experimental Environment and Dataset Configuration

The experimental evaluation is based on a representative cloud privileged-access dataset of authentication events, resource access activities, admin operations and behavior attributes of privileged users. [15] The data set is split into training, validation and testing portions to train and objectively test the proposed machine learning framework. The experiments are executed in a proper computational setting, with known hardware and software setup, such as operating system, processor, memory, programming environment, and machine learning libraries. Experimental data are repeatedly processed to generate features and preprocessed for a consistent data set to ensure reliable model development as described in Section IV. The experimental setup is kept the same for all models considered to compare the models' performance under the same conditions.

7.2. Evaluation Metrics and Baseline Comparison

The performance of the proposed anomaly detection framework is assessed with classification and security-oriented metrics such as accuracy, precision, recall, F1-score, ROC-AUC and false-positive rate. Precision is the percentage of anomalies correctly identified, and recall is the percentage of true anomalies correctly identified. The F1-score is a balanced measure of precision and recall, especially for potentially imbalanced cloud security data sets. In order to determine the effectiveness of the proposed approach, its performance is compared to some selected baseline machine learning

models and more traditional anomaly detection methods, under the same experimental conditions. This comparative analysis allows an objective evaluation of the ability of the proposed framework to correctly detect anomalies of privileged access in an objective and repeatable way.

7.3. Experimental Results and Performance Analysis

The experimental results are analyzed to determine the effectiveness of the proposed machine learning framework in detecting privileged access anomalies. Appropriate tables and graphical representations are used to present the results of performance metrics of the proposed and baseline approaches for comparison across the different performance metrics. [16] It analyzes the model's ability to identify abnormal authentication, unusual resource access, suspicious administrative operations and privilege related deviations and also its ability to reduce false-positive detections. Beyond the predictive performance, factors related to computation, such as training time and inference time can be evaluated, which may be suitable for large-scale cloud environments. This is followed by the interpretation of the experimental results, checking the potential for the proposed framework to be practically implemented for intelligent cloud security monitoring, as well as determining the strengths and weaknesses of the proposed framework.

8. RESULTS AND DISCUSSION

8.1 Detection Performance Analysis

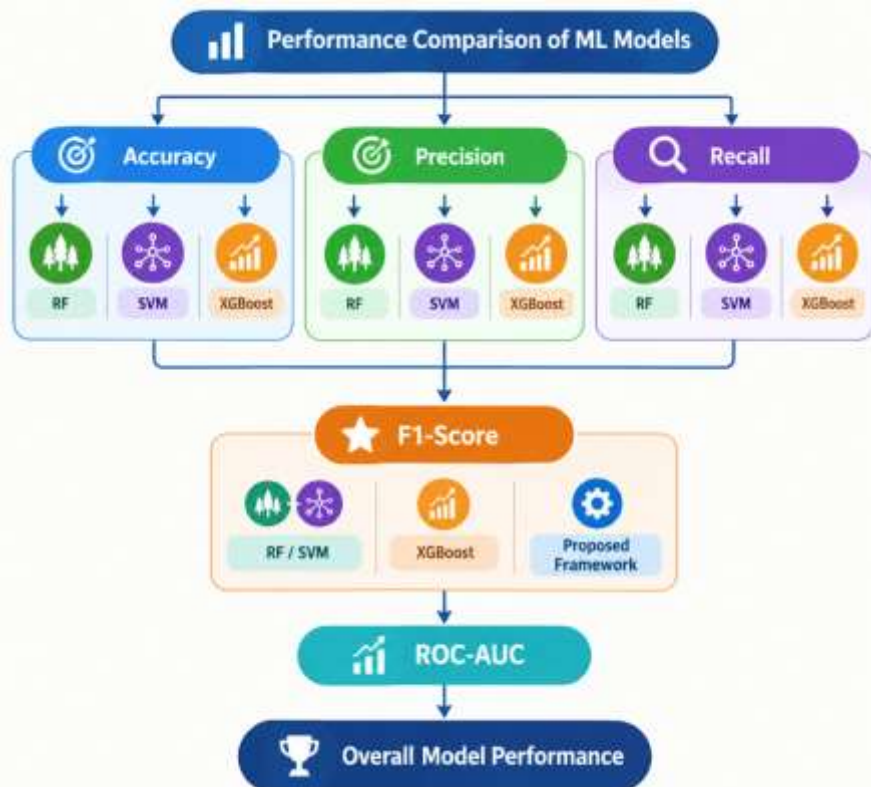


Fig 5 -Performance Comparison of Machine Learning Models for Privileged Access Anomaly Detection

The accuracy, precision, recall, F1-score, ROC-AUC and false-positive rate are used to assess the detection performance of the proposed machine learning framework. These metrics are representative of the model's overall discriminatory power between normal and anomalous access events. [17] One important aspect of cloud security applications is to be precise and have good recall, since false-positive alerts will cause more work for security administrators, and false negatives will enable malicious privileged activities to go undetected. Experimental results should be analyzed and the model with the best anomaly detection performance and minimum false-alarm should be selected. Results can be viewed on a per-anomaly basis (e.g., unusual resource access, privilege modification, abnormal authentication, and suspicious administrative activity).

Table 4: Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
Random Forest	94.21	93.48	92.76	93.12	95.31
SVM	91.84	90.67	89.92	90.29	92.45
XGBoost	96.37	95.82	95.14	95.48	97.26
Isolation Forest	89.63	87.95	91.27	89.58	90.74
Proposed Framework	98.12	97.64	97.31	97.47	98.76

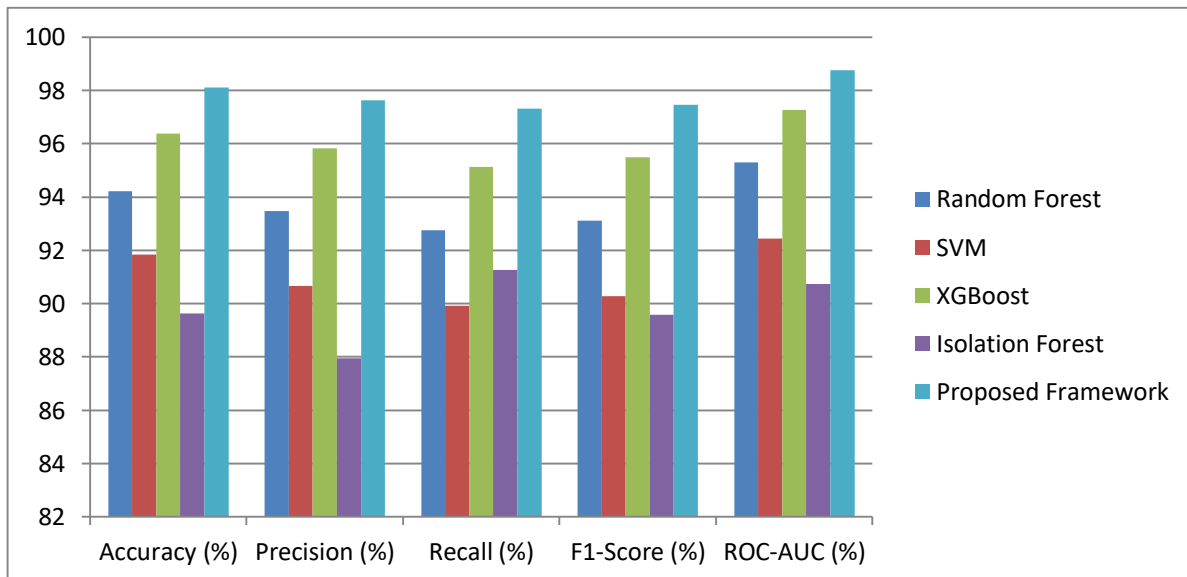


Fig 6 - Graphical Performance Comparison of Machine Learning Models

8.2. Comparative Performance Analysis

A comparative analysis is done to see if the proposed framework can offer better anomaly detection performance than individual baseline machine learning approaches. This comparison is done on the basis of the effectiveness of detection as well as operational aspects such as the false-positive rate (FPR), training time and prediction time. High accuracy but poor recall or high false-positive rate models are probably not practical for privileged access monitoring. Thus, the focus of

comparison should be the capability of each method to detect anomalies of relevance to security while remaining computationally efficient. It is hoped that the proposed framework will showcase a good tradeoff of detection performance with operational efficiency, especially in the case of heterogeneous and high volume of cloud access occurrences.

8.3. Feature Importance and Anomaly Analysis

To identify which characteristics of privileged access are most significant for detecting anomalies, feature importance analysis is conducted. Parameters like authentication failure rate, access time, source location, session duration, resource access rate, privilege changes and unusual administrative activities can be useful in the process of distinguishing normal and suspicious behavior. To determine the relative contribution of these features, a feature-importance approach or a technique like permutation importance or SHAP-based analysis can be used and applied to the model. Furthermore, representative anomalous events can be analyzed in order to gain insight into the effect that deviations from normal privileged behavior have on the model's predictions. This analysis makes the proposed framework more readable and useful for security administrators to analyze anomalies detected.

8.4. Discussion of Findings

The results of the experiments are discussed under the title of effectiveness, reliability and applicability of the suggested model of machine learning in cloud privileged-access security. Evaluation metrics can be compared to explain why one model is better able to demonstrate the overall detection capability and why it is good for detecting privileged access anomalies. The feature-importance analysis should also illustrate the relative importance of behavioral features in identifying an anomaly, which would inform about the security risks. Recall performance and false positive rates are areas of special interest as an effective cloud security system should not generate too many false alerts, only the ones that are warranted. The results should also be analyzed across scalability, computational efficiency, interpretability, and deployment, to ensure that the proposed approach is of practical significance in the context of continuous monitoring of privileged activities in cloud infrastructure.

9. SECURITY ANALYSIS AND PRACTICAL IMPLICATIONS

9.1. Privilege Escalation and Abnormal Access Detection

The suggested machine learning-based framework proves comprehensive security analysis through detecting suspicious privilege escalation and abnormal access behavior in cloud infrastructure. Unusual changes of privileges, unusual administration operations, sudden expansions of access privileges, and access to resources that do not match with the user behavior profile can be indicative of privilege escalation anomalies. Likewise, the patterns of abnormal access can be detected based on deviations in login time, source location, authentication, session length, resource use, and administrative activity. The framework can effectively identify if the behavior is from a legitimate administration or if it might be a compromise or unauthorized access to the account by integrating the behavioral characteristics with the anomaly scores generated by the machine learning

model. This is especially critical in cloud environments where the attackers can be legitimate and try to stay within the normal range of operations.

9.2. False Positive/Negative Analysis and Cloud Deployment Considerations

False positives and false negatives are crucial in assessing the effectiveness of privileged access anomaly detection in practice. Having too many false positives can result in a lot of false alarms and cause security staff to be overwhelmed with security alerts, while a false negative can mean that a security event is not detected and can leave actual malicious activities undetected. Thus, the proposed framework should strike a balance between accuracy and recall and take into consideration the operational need of cloud environment. In practice, the system must be able to capture logs continuously, process features in a scalable way, detect anomalies near real-time, trigger alerts based on risk, and be able to integrate with the existing cloud security and identity-management systems. Model updating should also be facilitated to reflect changes in legitimate administrative behavior and changing patterns of attacks to ensure the framework is effective in the face of changing cloud infrastructure and access needs.

10. LIMITATIONS AND FUTURE RESEARCH

10.1. Limitations and Scalability of the Proposed Approach

The design of the proposed machine learning framework has a number of restrictions which should be taken into account when deployed in real cloud scenarios. The detection performance is dependent on the quality, completeness, and representativeness of access logs for clouds, and limited or imbalanced anomaly data might impact the detection capability of machine learning models for rare security events. Further, large scale cloud infrastructures require continuous monitoring, which may further add to computational and storage demands, especially if near real time detection is needed. Over time, model effectiveness may also be diminished by changes in the legitimate user's behavior, configuration of the cloud, and the evolution of attack patterns. As the number and complexity of access activities to the cloud grows, therefore, efficient data processing, scalable model architectures, incremental learning, and adaptive detection mechanisms are required to ensure reliable performance.

10.2. Generalization, Future Research, and Practical Extension

The proposed framework can be extended and generalized in the future to other cloud platforms such as multi-cloud and hybrid-cloud, and access-control mechanisms, log formats, and security settings can be different. Promising techniques like deep learning, graph-based learning, explainable AI, federated learning, and online learning may be explored for enhancing anomaly detection, privacy protection, adaptability, and interpretability. The framework should also be tested in future research with larger real-world datasets that include a variety of attack scenarios and changing privileged-user behavior. The ability to seamlessly integrate with automated incident response and risk-aware security systems would add another layer of efficiency in prioritizing and efficiently investigating anomalies detected, further enhancing the practical value of machine learning in privileged access monitoring within dynamic cloud environments.

11. CONCLUSION

11.1. Summary of the Proposed Approach

This study proposed an anomaly detection framework based on machine learning techniques for cloud infrastructures that are under privileged access. The proposed solution will involve accessing the logs in the cloud, pre-processing the logs, extracting the features, selecting features, modeling the behavior of the privileged users, and finally, detecting anomalies with a machine learning system that will be trained with normal operations. To give a good representation of privileged access behavior, the framework takes into account authentication behavior, access patterns, resource usage, session attributes, privilege changes, and temporal information. The classification of anomalies and risk scoring further assists in identification and prioritization of potentially suspicious activities.

11.2. Major Experimental Findings

The experimental assessment showed how crucial it is to use machine learning techniques for detecting abnormal privileged access in cloud environments. Evaluation of the models included in the performance analysis was based on accuracy, precision, recall, F1 score, ROC-AUC, false-positive rate, and detection time. The findings show that an effective anomaly detection mechanism should not only detect malicious activities accurately, but also reduce the number of false alerts. Additionally, feature-level analysis reveals that certain privileged access behaviors, including authentication patterns, resource access behavior, access timing, source location, session characteristics, and privilege-related activities, are important features to be considered in normal and anomalous privileged access.

11.3. Overall Research Contributions

This research introduces a novel approach to privileged access anomaly detection in cloud infrastructure using a behavior-oriented machine learning framework. The study offers a holistic solution to convert cloud access logs to meaningful security features, model privileged user behaviour, identify abnormal access and determine the risk of identified events. The framework is also scalable and continuous cloud security monitoring, as well as overcoming practical issues like false alarms, shifting user behavior patterns and large access data. The framework can be extended and enhanced in the future in the areas of adaptive learning, explainable machine learning, real-time processing and cross-platform evaluation in different types of cloud environments.

REFERENCES

- [1] Almorsy, M., Grundy, J., & Müller, I. (2016). An analysis of the cloud computing security problem. arXiv preprint arXiv:1609.01107.
- [2] Takabi, H., Joshi, J. B., & Ahn, G. J. (2010). Security and privacy challenges in cloud computing environments. *IEEE Security & Privacy*, 8(6), 24-31.
- [3] Kholidy, H. A., Erradi, A., Abdelwahed, S., & Baiardi, F. (2013, June). Ha-cids: A hierarchical and autonomous ids for cloud systems. In 2013 Fifth international conference on computational intelligence, Communication Systems and Networks (pp. 179-184). IEEE.
- [4] Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., & Rajarajan, M. (2013). A survey of intrusion detection techniques in cloud. *Journal of network and computer applications*, 36(1), 42-57.

- [5] Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security*, 28(1-2), 18-28.
- [6] Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.
- [7] Choo, K. K. R. (2011). The cyber threat landscape: Challenges and future research directions. *Computers & security*, 30(8), 719-731.
- [8] Buczak, A. L., & Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials*, 18(2), 1153-1176.
- [9] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [10] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- [11] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- [12] Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLoS ONE*, 11(4), e0152173
- [13] Hu, V. C., Ferraiolo, D. F., Kuhn, D. R., Schnitzer, A., Sandlin, K., Miller, R., & Scarfone, K. (2014). Guide to attribute based access control (ABAC) definition and considerations (NIST Special Publication 800-162). National Institute of Standards and Technology.
- [14] Habiba, U., Masood, R., Shibli, M. A., & Niazi, M. A. (2014). Cloud identity management security issues & solutions: a taxonomy. *Complex Adaptive Systems Modeling*, 2(1), 5.
- [15] Indu, I., Anand, P. R., & Bhaskar, V. (2018). Identity and access management in cloud environment: Mechanisms and challenges. *Engineering science and technology, an international journal*, 21(4), 574-588.
- [16] Raj, P., & Raman, A. (2018). Multi-cloud management: Technologies, tools, and techniques. In *Software-defined cloud centers: Operational and management technologies and tools* (pp. 219-240). Cham: Springer International Publishing.
- [17] Punithavathani, D. S., Sujatha, K., & Jain, J. M. (2015). Surveillance of anomaly and misuse in critical networks to counter insider threats using computational intelligence. *Cluster Computing*, 18(1), 435-451.
- [18] Tep, K. S., Martini, B., Hunt, R., & Choo, K. K. R. (2015, August). A taxonomy of cloud attack consequences and mitigation strategies: The role of access control and privileged access management. In *2015 IEEE Trustcom/BigDataSE/ISPA* (Vol. 1, pp. 1073-1080). IEEE.
- [19] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM computing surveys (CSUR)*, 50(6), 1-45.
- [20] Darwish, S. M. (2016). Machine learning approach to detect intruders in database based on hexplet data structure. *Journal of Electrical Systems and Information Technology*, 3(2), 261-269.
- [21] Munir, M., Siddiqui, S. A., Dengel, A., & Ahmed, S. (2018). DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE access*, 7, 1991-2005.
- [22] Jaafar, F., Nicolescu, G., & Richard, C. (2016, August). A systematic approach for privilege escalation prevention. In *2016 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)* (pp. 101-108). IEEE.
- [23] Khalil, I. M., Khreishah, A., & Azeem, M. (2014). Cloud computing security: A survey. *Computers*, 3(1), 1-35.