https://doi.org/10.65025/ICAIC25180f

Cloud AI- Data Breach Detection

M. Fatima
Sagar Institute of Research and Technology, India
mfcollege2050@gmail.com

Kalpana Rai,
Sagar Institute of Research and Technology, India
kalpanarai@gmail.com

Alfiya Jahan
Sagar Institute of Research and Technology, India
alfiya1234@gmail.com

Abstract

In today's digital landscape, data breaches pose a major threat to both organizations and individuals. The exposure of sensitive information can result in financial losses, reputational harm, and legal consequences. To address these risks, Cloud AI- Data Breach Detection offers a powerful solution by harnessing artificial intelligence and cloud computing to identify, mitigate, and prevent breaches effectively. This paper introduces Sky Vault, a cloud AI system designed to simulate a cloud storage platform. Sky Vault replicates core functionalities of popular storage services such as Google Drive and Google Cloud. The proposed approach demonstrates how behavioral analysis can monitor user activities, detect anomalies, and strengthen the security and integrity of cloud storage systems.

Keywords

AI, data breach, data breach detection, Cloud AI, machine learning

1. Introduction

Cloud Guard AI- Data Breach Detection is an advanced solution designed to enhance the security of the cloud storage platform Sky Vault. Sky Vault is a robust cloud storage service similar to Google Drive and Dropbox, enabling users to store, manage, and access their data seamlessly over the Internet anywhere and at any time. Acting as a virtual storage system[1,2], Sky Vault allows users to upload a variety of files, including documents, photos, videos, and other data types, thereby eliminating the need for traditional storage devices. With its reliable accessibility and scalability, Sky Vault offers a convenient, secure, and efficient way to store and share information, making it an ideal solution for personal and professional use in the current digital age. Cloud Guard leveraged advanced machine-learning algorithms and sophisticated behavioral analysis techniques to detect suspicious activities in real time, such as unauthorized logins, abnormal file transfers, and unusual access patterns. Once identified, these activities are promptly reported to the user, enabling swift actions.[3,4] The solution is designed to proactively mitigate security risks, safeguard sensitive cloud data, and ensure user privacy by automating threat detection and response. Through its intelligent monitoring and rapid alerts, Cloud Guard AI provides robust defense against potential breaches, enhancing overall cloud storage security.

2. Cloud Guard AI

A. Cloud guard AI: A proactive approach to cloud security

In light of the growing number of cloud data breaches, traditional security methods often fall short in detecting and mitigating sophisticated threats. Cloud Guard AI offers an innovative, AI-powered solution that enhances cloud storage security through real-time monitoring and anomaly detection.

3. Methodological Approach

The approach adopted in this paper focuses on detecting anomalies through behavioral analysis of user activities. Behavioral analysis involves monitoring, understanding, and analyzing user interactions with the system to establish patterns of normal behaviour. By identifying deviations from these established patterns, the system can detect potentially suspicious or malicious activities that may indicate misuse or security threats. Figure 1 shows Key Components of Anomaly Detection and Threat Response in Cloud Environments. Sky Vault: To implement and demonstrate the concept, we develop an application called Sky Vault, which serves as a simulated cloud storage platform. Sky Vault will be designed to function similarly to well-known storage solutions like Google Drive, Dropbox, or Google Cloud.

The application allows users to perform key operations such as:

- Uploading files: Users can store files securely in the Sky Vault environment.
- Downloading files: Files can be accessed and retrieved by users when needed.
- Deleting files: Users can remove unnecessary or outdated files from the storage.

The Sky Vault application mimics the functionality of real-world cloud storage platforms, providing a controlled environment for monitoring user behaviours. Each user action within the system—such as file uploads, downloads, and deletions—can be logged and analyze to identify patterns of normal activity. Over time, this enables the system to establish a baseline of expected behaviour for each user[5,6]. By analyzing these behaviour patterns, the system can effectively detect anomalous activities that deviate from normal usage.

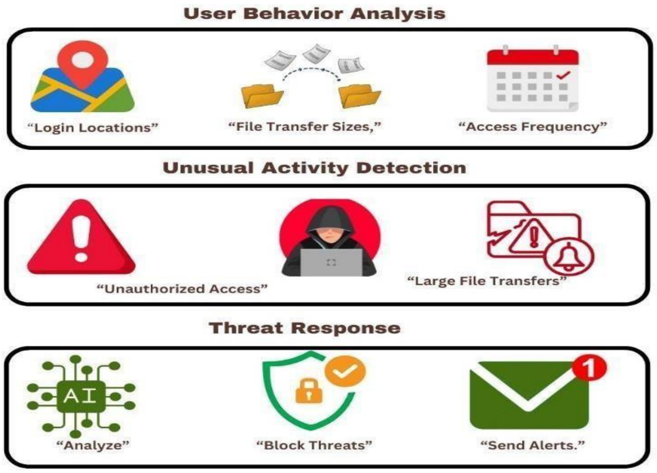

Figure1 Key Components of Anomaly Detection

Examples of such anomalies include:

- Unusual file access patterns: For instance, a user suddenly downloading or deleting a large number of files in a short period of time.
- Suspicious upload activities: Uploading files with unexpected formats, unusually large sizes, or prohibited content.
- Irregular access timing: Accessing the storage system at odd hours or from unusual locations.

The behavioral analysis approach ensures that anomalies are detected proactively and flags activities that might otherwise go unnoticed by traditional rule-based systems. The Sky Vault application acts as the testbed for implementing this approach, providing a realistic yet controlled scenario to evaluate the performance and accuracy of the anomaly detection system. SkyVault application, designed to provide users with a secure and seamless way to access their cloud storage accounts. Users are prompted to enter their email and password in the respective fields, ensuring secure authentication. By clicking the Login button, users can access their accounts and manage stored files. For new users, a convenient "Create an account" link is provided below the login button, enabling quick and easy registration.

The main area allows users to upload files using drag-and-drop or the "Choose File" button, followed by the "Upload" option.

Isolation Forest for Anomaly Detection

An unsupervised machine learning algorithm specifically designed to detect anomalies in high - dimensional datasets by isolating data points that deviate significantly from the norm. Isolation Forest will learn the patterns of normal behaviour of the user by observing the activities logged in Sky Vault. The algorithm functions by recursively and randomly partitioning the data into smaller subsets, making it easier to isolate anomalies that require fewer splits compared to normal data points. For each data point (i.e., a user's action), the algorithm isolates it by randomly choosing features and splitting the data. Normal behaviour, which occurs frequently, will be partitioned in several splits. However, anomalies — rare or abnormal behaviours like an unusual file upload or failed login attempt require fewer splits to be isolated.

Key steps of the Isolation Forest process

- Data Isolation: The algorithm isolates data points by randomly selecting a feature and making binary splits. If a data point (such as a suspicious user activity) is different from the rest of the data, it will be isolated more quickly.
- Anomaly Scoring: Once trained, each data point receives an anomaly score based on how easy it was to isolate. Higher anomaly scores indicate that the behaviour is rare and abnormal. These can be flagged as potentially suspicious or malicious.

Anomaly Detection in Real-Time

Once trained, the Isolation Forest model will analyze the incoming activity logs in real-time. For every user action performed on Sky Vault, the model will:

- Score the activity based on its degree of anomaly.
- Flag any activity that surpasses the threshold as suspicious.

Example of suspicious behaviours that would be flagged:

- Unusual File Uploads: A user uploading an unusually large file that doesn't match their usual activity pattern.
- Rare File Types: Files with rare extensions (e.g., .exe) being uploaded, which are typically outside the scope of the platform's normal file types.
- Multiple Failed Login Attempts: A sudden surge in failed login attempts could indicate a brute force attack or unauthorized access attempt.

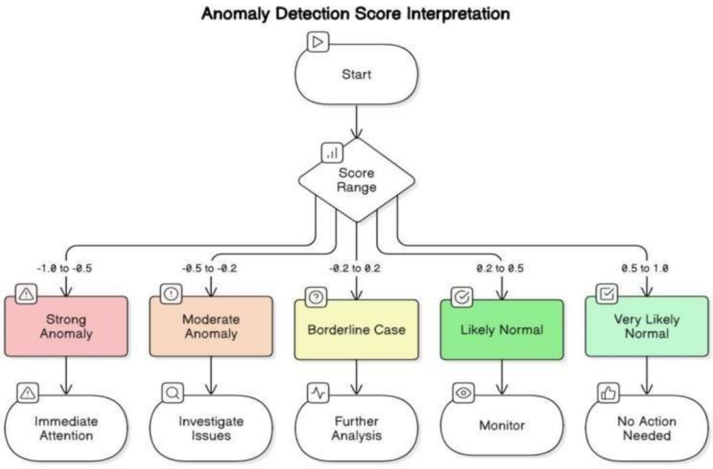

Fig2 Anomaly Detection Score Interpretation

Above figure 2 shows anamoly detection score. By implementing isolation forest, it is possible to detect anamoly and can control it.

Key Parameters Monitored in Isolation Forest Implementation

a. Temporal Features
- File Uploaded At: Tracks the exact timestamp of file uploads, helping identify unusual upload patterns.
- Time-Based Activity Patterns: Analyzes trends in user activities over time, such as frequency, time of day, and irregular gaps, to detect deviations.

b. File-Related Features
- filename: Assesses file types, naming conventions, and extensions to uncover patterns or anomalies (e.g., suspicious file names or uncommon formats).
- File Operations: Monitors activities like uploads, downloads, and deletions to detect abnormal actions.

c. Siordia logic:
- Anomaly Score: The model computes an anomaly score for each activity, quantifying how likely an action isto be anomalous.
- Anomaly Flag: Returns a boolean is anomaly indicator for quick classification, marking activities as normal or suspicious.

4. Conclusion:

Cloud Guard AI has proven its effectiveness in detecting anomalies within the Sky Vault cloud storage platform. The results confirm that the model can accurately identify suspicious activities, playing a crucial role in enhancing the security and integrity of user data. While the system excels at detecting common and highly noticeable anomalies, ongoing optimization and training will focus on improving its sensitivity to rare or more subtle threats.

Continuous learning and real-world data integration will be essential for refining performance, allowing the model to adapt to evolving user behaviors and emerging cybersecurity threats. This paper underscores the importance of machine learning- based anomaly detection in cloud platforms like Sky Vault, where protecting user data is a top priority. With further fine- tuning and regular updates, Cloud Guard AI has the potential to become a highly reliable and efficient tool for safeguarding cloud storage systems against unauthorized activities and emerging threats.

Overall, this study demonstrates the successful application of AI in strengthening cloud security and highlights promising opportunities for future advancements in anomaly detection and cybersecurity.

5. References

[1] R. Karri, J. Rajendran, K. Rosenfeld, M. Tehranipoor, Trustworthy hardware: Identifying and classifying hardware trojans, IEEE Comput., 43 (10) (2010), pp. 39-46

[2] Mouratidis, Secure by design: Developing secure software systems from the group up Intern. J. Secure Software Eng., 2 (3) (2011), pp. 23- 41

[3] Internet Security Threats Report. Symantec, http://www.symantec.com/threatreport/

[4] S.E. Goodman, H.S. Lin (Eds.), Toward a Safer and More Secure Cyberspace, The Nat'l Academics Press (2007)

[5] R.C. Newman, Computer Security: Protecting Digital Resources, (first edition), Jones & Bartlett Publishers (February 20, 2009)

[6] The AI Effect: Amazon Sees Nearly 1 Billion Cyber Threats a Day