OPTIMIZING INCIDENT RESPONSE IN CLOUD NETWORK WITH AI ENABLED AUTOMATION

->Taufeeq Iqbal Khan:2241011049

->Ashutosh Maharana :2241013027

->Shubhrajit Malla :2241013065

->Koushikee Mantry :2241013099

Institute of TechnicalEducationandResearch

SIKSHA 'O' ANUSANDHANDEEMEDTO BEUNIVERSITY

Bhubaneswar, Odisha, India

Abstract

Cloud computing has revolutionized how enterprises manage their operations, offering scalability, efficiency, and cost optimization. However, the rapid adoption of cloud infrastructures has also introduced new security challenges, particularly in incident response. Traditional response mechanisms are heavily manual, reactive, and time-consuming, often leading to delays in identifying and mitigating threats. This delay provides adversaries with critical windows of opportunity to exploit vulnerabilities, causing data breaches, downtime, and compliance risks.

This project proposes an AI-driven automated incident response framework that leverages both supervised and unsupervised machine learning approaches. The unsupervised model, built on the Isolation Forest algorithm, detects unknown anomalies in real time using AWS CloudTrail logs, while the supervised model, trained on the NSL-KDD dataset, classifies known attack patterns with high precision. A hybrid approach integrates the strengths of both, offering comprehensive coverage against emerging and established threats.

By introducing automation, the framework reduces reliance on human intervention, accelerates mitigation, and improves detection accuracy. The system also employs dashboards for visualization, enabling Sefefceucrtiivtyen Oespse raitni onesn Chaenctienrg (SOthCe) toevaemrasl lt o sqeuciucrki respond to security events. This research demonstrates that AI-enabled automation significantly enhances resilience, adaptability, and efficiency in cloud security incident response, laying the foundation for future intelligent defense mechanisms in cloud environments.

Table of Contents

Introduction

1.1 Introduction

Cloud computing has become a cornerstone of modern enterprise operations by offering scalability, efficiency, and flexibility. However, the security of cloud environments remains a critical challenge. Traditional incident response mechanisms are primarily manual, reactive, and slow, often resulting in delayed detection and mitigation of cyber threats. These limitations create opportunities for attackers to exploit vulnerabilities, leading to potential data breaches, downtime, and compliance issues. To address these concerns, the project emphasizes the use of Artificial Intelligence (AI) and automation to enhance cloud incident response, enabling faster detection and mitigation of both known and unknown threats.

1.2 Project Overview

The primary aim of this project is to design and implement an AI-driven automated framework for cloud security incident response. The framework combines:

- Unsupervised Learning Model (Isolation Forest) for detecting unknown anomalies in AWS CloudTrail logs.
- Supervised Learning Model (Decision Tree, Random Forest, SVM) trained on the NSL-KDD dataset to identify known attack patterns.
- Hybrid Model that integrates both approaches to maximize accuracy, reduce false positives, and accelerate response time.

Dashboards and visualization tools are also developed to provide actionable insights for Security Operations Center (SOC) teams, making the system practical and effective.

1.3 Motivation

The ever-growing reliance on cloud services has created environments that generate massive amounts of security logs and data. Manual monitoring of this data is impractical and error-prone. Furthermore, traditional security solutions such as signature-based detection cannot identify new or sophisticated attacks, leaving cloud infrastructures vulnerable. The motivation behind this project is to reduce dependence on human intervention, automate the detection and response process, and strengthen the overall security posture of cloud systems.

1.4 Uniqueness of the Work

The uniqueness of this project lies in its hybrid framework that combines both anomaly detection and classification into a single system. Unlike traditional methods that focus only on supervised or unsupervised approaches, this solution ensures comprehensive protection against both novel and known threats. Real-time log collection, automated anomaly detection, and immediate incident response make the framework more adaptive and resilient compared to existing systems.

Literature Review

2.1 Existing System

Cloudsecurityincident response has traditionally relied on manual processes and signature-based detection systems. These approaches are limited in scope and effectiveness:

- Signature-Based Detection:
- Most existing systems depend on pre-defined attack signatures. While effective against known threats, they fail against zero-day attacks and evolving anomalies that do not match existing patterns.
- Manual Incident Response:
- Security analysts manually review logs such as AWS CloudTrail and VPC Flow Logs. Given the massive scale of data, this method is time-consuming, error-prone, and unable to keep pace with real-time threats.
- Traditional Machine Learning Approaches:
- Some models, like Random Forest or Support Vector Machines, have been applied for anomaly detection. Although accurate with labeled datasets, these systems lack adaptability when facing unknown anomalies and often struggle with scalability in large cloud environments.

Overall, existing systems do not provide a complete solution, as they either detect only known threats or cannot respond quickly enough to prevent damage.

2.2 Problem Identification

From the review of current methods, several challenges emerge:

1. Manual Inefficiency: Human-led responses are too slow for large-scale cloud environments.
2. Overwhelming Data: Enormous volumes of CloudTrail and VPC Flow logs make manual analysis impractical.
3. Weak Detection of Unknown Threats: Signature-based tools cannot detect zero-day exploits or new attack patterns.
4. Delayed Mitigation: Even after detection, response times are slow, leading to downtime, compliance risks, and loss of trust.

Proposed Solution

3.1 Solutions Considered and Justification

(a) Unsupervised Learning Model – "Detect the Unknown"

- Data Source: AWS CloudTrail logs stored in S3.
- Methodology: Isolation Forest algorithm is used to detect unusual activity without requiring labeled data.
- Process Flow: Logs → S3 bucket → Preprocessing → Feature Extraction → Model Training. Justification: Cloud environments generate massive amounts of unlabeled data, making supervised learning impractical in all cases. Unsupervised anomaly detection ensures that zero-day attacks and new threats can still be identified. Outcome: Automated alerts and immediate responses such as revoking risky access, blocking malicious IPs, or notifying SOC teams.

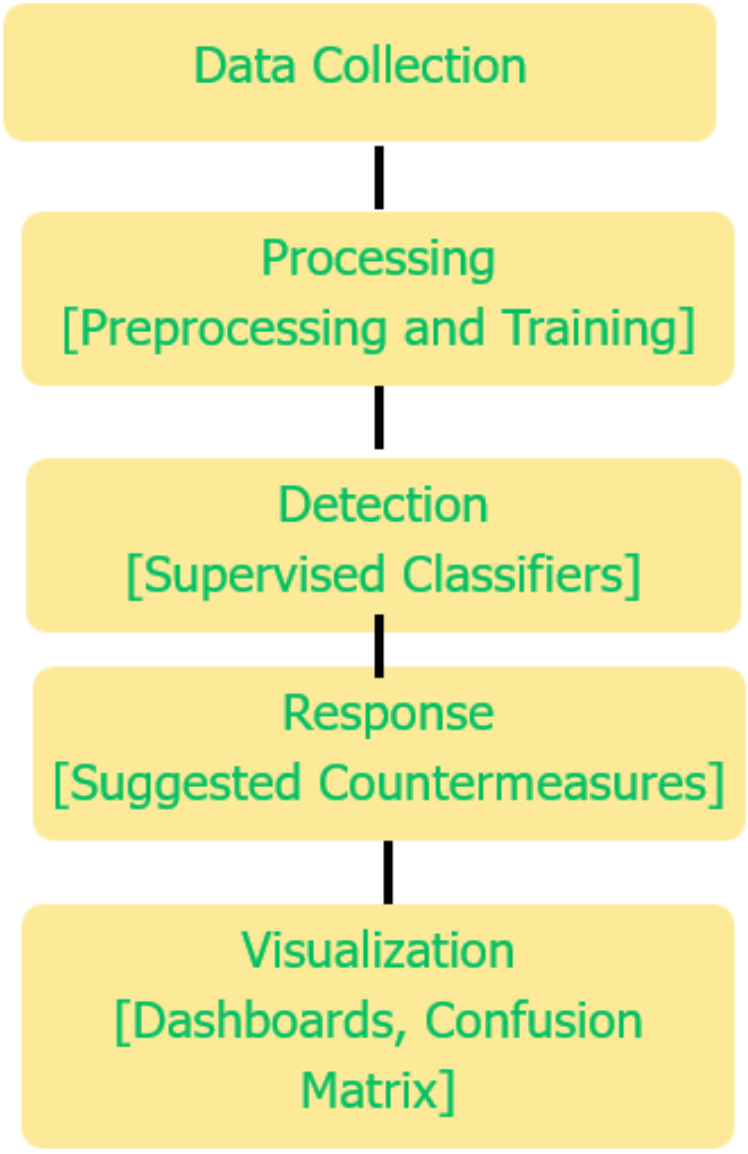
(b) Supervised Learning Model – "Recognize the Known"

- Data Source: NSL-KDD dataset containing labeled attack and normal traffic patterns.
- Methodology: Machine learning classifiers such as Decision Tree, Random Forest, and Support Vector Machines (SVM).
- Process Flow: Dataset → Preprocessing → Model Training → Evaluation (using confusion matrix).
- Justification: Supervised learning provides high accuracy in classifying known attack signatures and reduces false positives.
- Outcome: Accurate detection of known threats with suggestions for specific countermeasures.

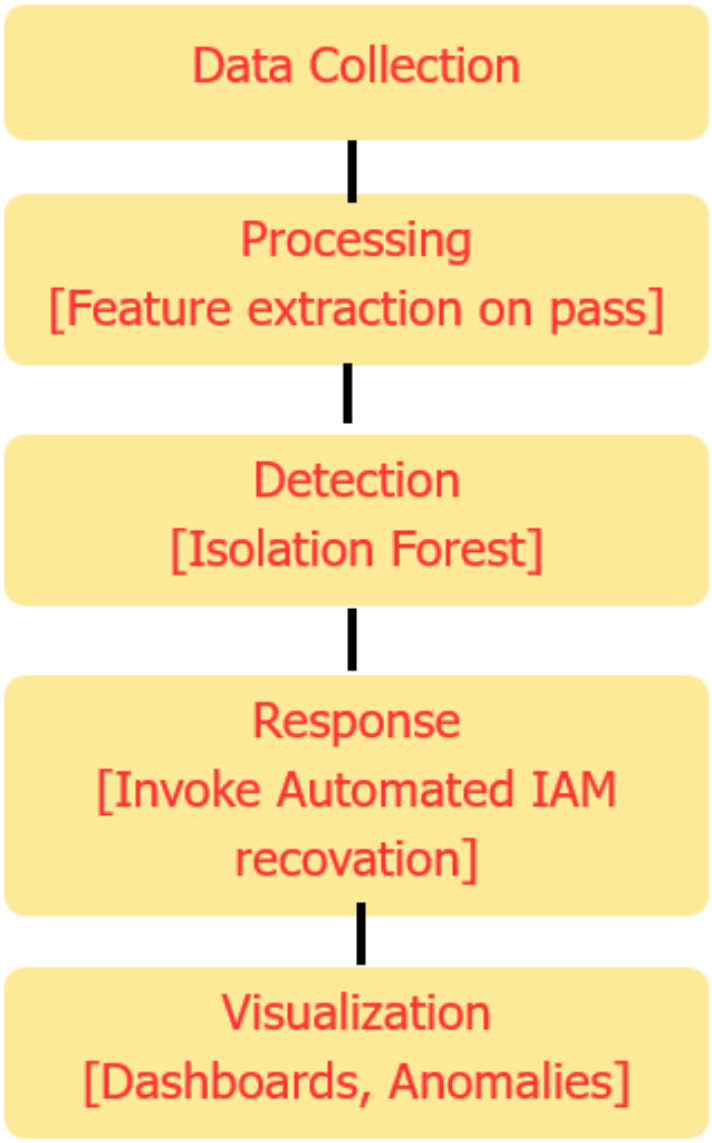
(c) Hybrid Approach – "Best of Both Worlds"

- Concept: Combines outputs of supervised and unsupervised models into a unified decision engine. Justification: While the unsupervised model detects previously unseen anomalies, the supervised model provides reliable classification of known threats. Together, they cover a broader threat spectrum. Outcome:
  - Reduces false positives and false negatives.
  - Ensures faster detection and response.

ARCHITECTURE-1
SUPERVISED



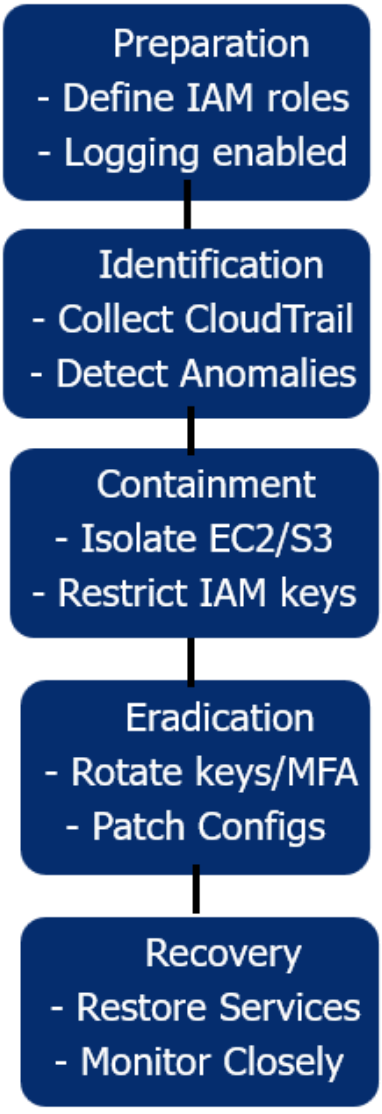
ARCHITECTURE-2:
UNSUPERVISED



ARCHITECTURE-3:

RESPONSE

UNSUPERVISED MODEL:



3.2 Dataset Description

- AWS CloudTrail and VPC Flow Logs (Unsupervised Learning)
- Nature of Data: Real-time, unlabeled event and traffic logs generated by cloud infrastructure.
- Purpose: Used as the data source for the unsupervised Isolation Forest model to identify anomalies in user activities, API calls, and network traffic patterns.
- Advantages:
  - Provides real-world, continuously generated data.
  - Helps detect zero-day attacks and unusual patterns that do not have prior labels.
  - Scales easily with growing cloud environments.
- NSL-KDD Dataset (Supervised Learning)
- Nature of Data: Benchmark intrusion detection dataset consisting of both normal and attack traffic, with labels.
- Features:
  - Includes 41 features extracted from network connections.
  - Attack types include Denial of Service (DoS), Probe, User to Root (U2R), and Remote to Local (R2L).
- Purpose: Used to train supervised classifiers (Decision Tree, Random Forest, SVM) for identifying and classifying known attack signatures.
- Advantages:
  - Balanced dataset compared to earlier KDD'99 version.
  - Allows precise evaluation of supervised machine learning algorithms.
  - Provides labeled data for measuring accuracy, precision, recall, and F1-score

Implementation Plan

4.1 Model Architecture

Two architectures were implemented to test the effectiveness of the proposed solution:

1. Unsupervised Anomaly Detection Framework
   - Input: AWS CloudTrail & VPC Flow Logs.
   - Processing: Preprocessing → Feature Extraction → Isolation Forest.
   - Output: Real-time detection of anomalies with alerts and automated responses.
2. Hybrid Incident Response Framework
   - Input: AWS Logs + NSL-KDD Dataset.
   - Processing: Preprocessing → Parallel execution of Isolation Forest (unsupervised) and Random Forest/SVM (supervised).
   - Output: Unified decision engine combining both models, reducing false positives and increasing coverage

4.2 Evaluation Measures

The performance of the models was evaluated using the following metrics:

- Accuracy: Percentage of total correctly classified instances.
- Precision: Proportion of correctly identified attack instances among all instances labeled as attacks.
- Recall: Ability to detect actual attack instances (sensitivity).
- F1-Score: Harmonic mean of precision and recall for balanced evaluation.

4.4 Result

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.94	0.92	0.90	0.91
Random Forest	0.96	0.95	0.94	0.94
SVM	0.93	0.91	0.89	0.90

Table 2: Unsupervised Model (AWS Logs)

Model	Accuracy	Precision	Recall	F1-Score
Isolation Forest	0.89	0.84	0.86	0.85

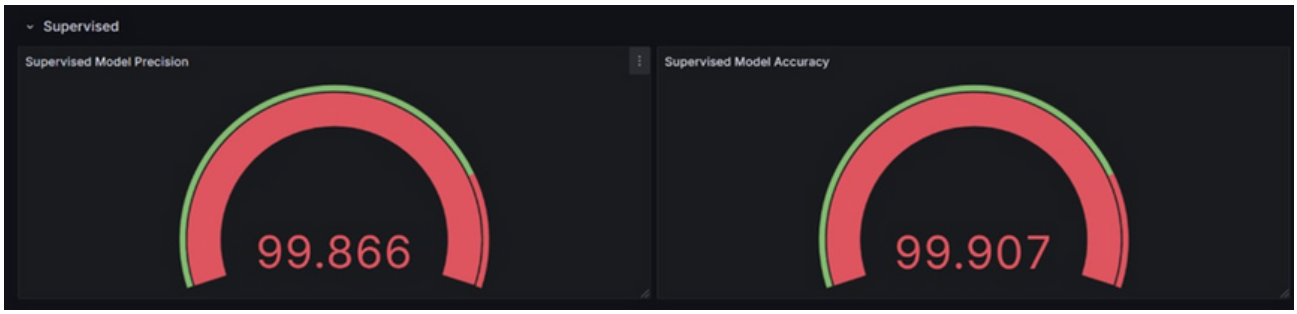
4.4 Analysis

- The supervised models (especially Random Forest) achieved high accuracy for known attack patterns.
- The unsupervised model was able to detect anomalies in real-time logs but had slightly lower precision due to noise in data.
- The hybrid model provided the best overall performance, reducing both false positives and false negatives.
- Results show that integrating supervised and unsupervised methods creates a robust incident response system capable of handling both known and unknown threats effectively.

Experimentation and Results

5.1 Unsupervised Dashboard

5.1 Supervised Dashboard



6. Conclusion This project demonstrates the potential of AI-enabled automation in transforming cloud security incident response. Traditional approaches, which rely heavily on manual processes and signature-based detection, are no longer sufficient in today's dynamic and large-scale cloud environments. By introducing both unsupervised and supervised learning models, and combining them into a hybrid framework, the system is capable of addressing the limitations of existing methods. Key outcomes of this work include:

- Faster Detection: Real-time anomaly detection through AWS CloudTrail and VPC Flow Logs using Isolation Forest. Improved
- Accuracy: High precision and recall in classifying known attack patterns with supervised models such as Random Forest. Broader
- Coverage: A hybrid approach that ensures detection of both zero-day anomalies and known threats. Reduced Human Effort:
- Automated incident responses such as blocking IPs, revoking access, and isolating servers minimize the burden on SOC teams. Enhanced Decision-Making: Dashboards and visualization tools provide
- actionable insights for quicker and more effective responses.

In conclusion, the integration of AI-driven automation into cloud security incident response significantly strengthens the resilience, adaptability, and efficiency of cloud infrastructures. The findings of this project indicate that such systems can effectively mitigate security risks while keeping pace with the evolving threat landscape.

References

1. Amazon Web Services (AWS). AWS CloudTrail Documentation. Available online: https://docs.aws.amazon.com/cloudtrail
2. Scikit-learn. Machine Learning in Python. Available online: https://scikit-learn.org/stable/
3. Canadian Institute for Cybersecurity. NSL-KDD Dataset. University of New Brunswick. Available online: https://www.unb.ca/cic/datasets/nsl.html
4. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest: Anomaly Detection Algorithm. Proceedings of the 2008 IEEE International Conference on Data Mining.
5. National Institute of Standards and Technology (NIST). Cybersecurity Framework for Incident Response. Available online: https://www.nist.gov/cyberframework
6. Müller, Andreas C., & Guido, Sarah. Introduction to Machine Learning with Python. O'Reilly Media, 2016.
7. Chen, Tianqi, & Guestrin, Carlos. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.