National College of Ireland

Adaptive Cloud Access Security Broker

MSc Industry Internship
MSc Cybersecurity

Keerti Ramesha
Student ID: 21144362

School of Computing
National College of Ireland

Supervisor: Prof. Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Keerti Ramesha

Student ID: 21144362

Programme: MSc in Cybersecurity **Year:** 2022-23

Module: Industry Internship

Supervisor:
Submission Due Date: 06/01/2023

Project Title: Adaptive Cloud Access Security Broker (CASB)

Word Count: **6062** **Page Count: 18**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Keerti Ramesha

Date: 06/01/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Adaptive Cloud Access Security Broker (CASB)

Keerti Ramesha
21144362

Abstract

Cloud Security threats are exponentially increasing with individuals and organizations moving their workloads to cloud. It is therefore necessary to monitor all cloud activities from the user end and the application end and enforce security mechanisms to safeguard the data on cloud. Cloud Access Security Brokers (CASB) monitor user activity, inspect the application accesses and warn administrators about potential threats. However, a static CASB solution can have a predefined response state to known threats and may be easily overpowered by various intrusions, malware, and data packets. It may not be well equipped to handle attacks in real time. The solution developed is an adaptive CASB which includes a machine learning model that adapts to real time scenarios as and when new data is added and responds to threats dynamically and regulate data accesses. The research shows that the result of the use of machine learning in CASB systems have the potential to improve the security and usability of cloud-based systems.

1 Introduction

CASBs are becoming increasingly important as more and more organizations adopt cloud services, as the cloud introduces new security challenges that traditional security solutions may not be able to address. By using a CASB, organizations can leverage the benefits of the cloud while still maintaining a high level of security and control over their data. The use of machine learning in Cloud Access Security Brokers (CASBs) is important as organizations today adopt cloud services and face a growing number of security threats. Machine learning can help CASBs improve their ability to detect and prevent security incidents, as well as to analyse and understand the behaviour of users and devices in the cloud.

One of the main benefits of using machine learning in CASBs is its ability to analyse large amounts of data and identify patterns that may indicate a security threat. For example, a CASB with machine learning capabilities can analyse log data from cloud services and detect anomalies that may indicate a cyber-attack, such as a sudden increase in failed login attempts or unusual data access patterns. Another benefit of machine learning in CASBs is its ability to continuously learn and adapt to new security threats as they emerge. By analysing data from multiple sources and using algorithms to identify patterns and trends, a CASB with machine learning capabilities can continuously improve its ability to detect and prevent security incidents. Thus, it is lucrative to research the use of machine learning in CASBs and understand how it can help organizations improve their security posture and better protect their data and users in the cloud.

The aim of this research is to provide organizations with visibility into cloud usage and activity, detecting and preventing data leaks, and providing data loss prevention (DLP) capabilities and ensure overall security and compliance with regulatory requirements. The major contribution of this research is an XDR (Extended Detection and Response) model powered by machine learning that can work in hand with CASB to monitor and detect threats dynamically. Principal component Analysis (PCA) has been used as it is an unsupervised learning method and minimises data loss. PCA can find patterns in the user accesses to the cloud resources and the model is trained dynamically using this data.

This presents us with the following Research questions: How can CASBs improve the security of cloud-based systems and data? How can machine learning be used to improve the capabilities and performance of CASBs? More specifically, how can the Principal component Analysis (PCA) used in CASB solutions to find user access patterns and minimise anomalous access to cloud resources?

This study describes a CASB solution powered with an unsupervised machine learning model for identifying cloud security threats dynamically. Significant research done within this field is detailed in Section 2. The sub-sections in this section cover the different fields of research, such as CASB for cloud security, Machine learning for CASB. Section 3 details the research methodology used. The design components and solution implementation are covered in Sections 4 and 5, respectively. In Section 6, the evaluation of results is covered in more detail. The research is concluded in Section 7, which also addresses future work.

2 Related Work

Cloud computing has become increasingly popular in recent years due to its ability to provide on-demand access to computing resources and data storage. However, the use of cloud computing also introduces new security challenges, as sensitive data and resources are often stored and processed remotely(Kumar and Goyal, 2019). It is crucial for organizations to adopt a SIEM systems which can provide capabilities including real-time monitoring, event correlation, and reporting.(Lee *et al.*, 2017) Access control, which is the process of granting or denying access to resources based on predetermined security policies, is a key component of cloud security. Artificial intelligence (AI) and machine learning techniques have the potential to improve the effectiveness of access control systems.(Rathod, Parekh and Dholariya, 2021) (Kumar *et al.*, 2022) For example, AI can be used to identify and predict security threats, such as malware or unauthorized access attempts, and automate access control decision-making processes.(Mohammad and Pradhan, 2021).

A Cloud Access Security Broker (CASB) is a security solution that helps organizations protect their data and users in the cloud. It acts as a intermediary between the organization's on-premises infrastructure and the cloud, monitoring and enforcing security policies on cloud usage and activity. CASBs offer a range of security features, including monitoring and reporting on cloud usage, enforcing access controls, detecting and preventing data leaks, and providing data loss prevention (DLP) capabilities.(Ahmad, Mehfuz and Beg, 2021)

2.1 Cloud Access Security Broker

A CASB is a security system that acts as a mediator between clients and cloud-based resources, enforcing security policies and providing a secure communication channel

between the two. CASBs can be used to enforce a variety of security measures, including access control, data loss prevention, and threat detection.(Badhwar, 2021b) In the paper (Liu et al., 2017) the authors propose the use of a CASB to facilitate the search and sharing of encrypted data in the cloud. Encrypting data can help to protect it from unauthorized access, but it also makes it more difficult to search and share. The proposed CASB aims to address this challenge by allowing users to search and share encrypted data while still maintaining the security benefits of encryption. A CASB-based approach to encrypted data search and sharing has the potential to improve the security and usability of cloud-based systems. However, the use of a CASB also introduces new challenges and potential risks, such as the potential for errors or bias in the decision-making process and concerns about the transparency and explainability of the system. Therefore, it is important for research on CASBs to consider and address these issues. Further research is needed to fully understand the benefits and drawbacks of using a CASB in this context, as well as to address the challenges and risks associated with its use.

The researchers (Fossum and Andersen, 2021) propose the use of a CASB to improve the security of cloud-based healthcare services. They also propose a discussion framework for evaluating the security of cloud healthcare services that considers various factors such as the security of data in transit, the security of data at rest, the security of access control, and the security of data sharing. The use of a CASB in the healthcare industry has the potential to improve the security of cloud-based healthcare services.

The authors (Ahmad *et al.*, 2022) conduct a systematic review of existing research on CASBs. They use a methodology called RSM (Requirements Specification Method) analysis to identify and evaluate the key features and requirements of CASBs. The authors find that CASBs have the potential to improve the security of cloud-based systems by enforcing security policies, detecting and mitigating threats, and providing a secure communication channel between clients and cloud resources. The systematic literature review suggests that CASBs have the potential to enhance the security of cloud-based systems. However, further research is needed to fully understand the benefits and drawbacks of using a CASB, as well as to address the challenges and risks associated with its use.

The COVID-19 pandemic has led to a significant shift in the way people work, with many organizations implementing remote work policies in order to ensure the continuation of business operations.(Ahmad, Mehfuz and Beg, 2020) One key aspect of securely working from home (WFH) is the use of Cloud Access Security Broker (CASB) policies. CASB policies provide an additional layer of security for cloud-based services, such as those used for remote work. CASB policies can also help to prevent data breaches, as they can detect and block suspicious activity in real-time. CASB policies can help to prevent data breaches and protect sensitive company information. This is especially important given the increased reliance on cloud-based services and the potential for cyber threats to increase during the pandemic. Overall, the use of CASB policies can be an effective way to securely enable WFH during the COVID-19 pandemic. As such, organizations should consider implementing CASB policies as part of their remote work strategies.

There are several benefits to using CASB for secure access to cloud services. It can provide an additional layer of security for users accessing the service. (Twum, B. and K., 2020)This is particularly important for organizations that have a large number of users

accessing the service, as it can help to prevent data breaches and protect sensitive company information. CASB can help to ensure compliance with security protocols and regulations. This can be especially important for regulated industries, such as healthcare and finance, where compliance with data protection laws is critical.(Kemp, 2018) It is an effective pattern for providing secure access to cloud services. It can provide an additional layer of security for users, help to prevent data breaches, and ensure compliance with security protocols and regulations. (Ahmad, Mehfuz and Beg, 2019)As such, organizations should consider implementing CASB as part of their cloud security strategy.

CASBs can provide a powerful and flexible solution for securing access to cloud services. However, there are also challenges to consider, such as the potential for CASBs to create latency or other performance issues, and the need for robust data management and privacy protocols.(Fernandez, Yoshioka and Washizaki, no date)

2.2 Machine learning in CASB Solutions

Machine learning algorithms can be used to identify anomalous behaviour in network traffic or user activity, which can indicate the presence of a potential threat.(Nassif *et al.*, 2021) (Agarwal *et al.*, 2021)These algorithms can also be used to classify different types of cyber threats, such as malware or phishing attacks, and to predict the likelihood of future attacks.(Badhwar, 2021a) By leveraging the vast amounts of data generated in cloud environments, big data analytics can provide more accurate and comprehensive insights into potential security threats. For example, big data analytics can be used to identify patterns and correlations in data that may not be apparent using traditional data analysis techniques.(Mohammad and Pradhan, 2021)

There have also been efforts to integrate machine learning and big data analytics with other security technologies, such as security information and event management (SIEM) systems and intrusion detection and prevention systems (IDPS).(Badhwar, 2021a) These integrations can enhance the ability of these systems to detect and respond to cyber threats in real-time. The researchers (Waskle, Parashar and Singh, 2020), (Meng *et al.*, 2018) propose a method for creating effective IDS that makes use of the random forest classification algorithm and principal component analysis (PCA) is proposed. While the random forest will aid in categorization the PCA will assist in organizing the dataset by lowering its dimensionality. PCA is also proven to be suitable to detect abnormal behaviour in networks. (Al-Fawa'reh *et al.*, 2022)

In the paper by (Bhattacharya *et al.*, 2021), the authors propose the use of a security broker that utilizes AI to dynamically enforce access control policies in the cloud. A security broker acts as a mediator between clients and resources, enforcing security policies and providing a secure communication channel between the two. By using AI to dynamically update and enforce access control policies, the proposed security broker aims to improve the security and usability of cloud-based systems.(Feng, Feng and Dawam, 2020) However, the use of AI in security systems also introduces new challenges and potential risks.(Waguie and Al-Turjman, 2022) For example, there is the potential for bias or errors in the decision-making process, as well as concerns about the transparency and explain ability of AI-based systems. Therefore, it is important for research on the use of AI in security systems to consider and address these issues. Further research is needed to fully understand the benefits

and drawbacks of using AI in access control, as well as to address the challenges and risks associated with its use.

2.3 Research Niche

Table 1: Research Summary

Related Works	Strengths	Limitations
(Bhattacharya *et al.*, 2021)	Dynamic CASB solution discussed	Does not work with real time data
(Rathod, Parekh and Dholariya, 2021)	Observe, Response, Compliance, Awareness approach proposed	Implementation in cloud systems not defined
(Liu *et al.*, 2017)	Framework for encrypted data search using CASB	Performance overhead in huge datasets
(Fernandez, Yoshioka and Washizaki, no date)	Standard CASB framework proposed	No practical applications discussed
(Lee *et al.*, 2017)	SIEM architecture for cloud systems	Proposed model configuration not discussed
(Ahmad, Mehfuz and Beg, 2020)	Proposes CASB policies for remote working	Lacks clarity in terms of implementation
(Ahmad *et al.*, 2022)	RSM analysis on CASB is highlighted	Cloud provider decision-making needs work.
(Kumar *et al.*, 2022)	Security, data protection in cloud computing discussed	Specific ML models could have been discussed
(Kumar and Goyal, 2019)	Outlines the criteria for cloud security, as well as any threats that have been found, known weaknesses, and suggested solutions.	It lacks clarity on tailored solutions to specific cybersecurity issues
(Mohammad and Pradhan, 2021)	Several machine learning algorithms have been discussed based on accuracy, precision and many other factors	Network level security analysis is less accurate
(Nassif *et al.*, 2021)	Provides a systematic review	Work is restricted to small

	of cloud security systems employing machine learning	number of papers
(Kemp, 2018)	Discusses cloud security from a legal perspective	Does not consider compliance with local regulations
(Fossum and Andersen, 2021)	Explores the merits and demerits of using CASB in healthcare services	Poor dataset
(Waskle, Parashar and Singh, 2020)	Results have high accuracy and low error rate (0.21%)	Other methods have not been explored
(Badhwar, 2021b)	Discusses cloud security controls to be employed by organizations	Poor monitoring controls for data visibility
(Ahmad, Mehfuz and Beg, 2019)	Describes a case study solution	Lacks clarity in terms of remediation steps
(Meng *et al.*, 2018)	Deep learning for insider threat detection	False alarm rate is high
(Feng, Feng and Dawam, 2020)	AI strategy for cybersecurity	Compliance issues
(Waguie and Al-Turjman, 2022)	Suggests solutions to edge computing security	Lacks data security over the network
(Agarwal *et al.*, 2021)	Intrusion detection using ML	Complex machine learning approaches need to be considered

3 Research Methodology

A significant element of this research paper is to identity user access patterns in cloud and recognise the threats and diversions from intended behaviour. For this purpose, the technique known as KDD, or Knowledge Discovery in Databases is used to find useful and significant patterns in huge datasets. This process involves many steps as discussed below.

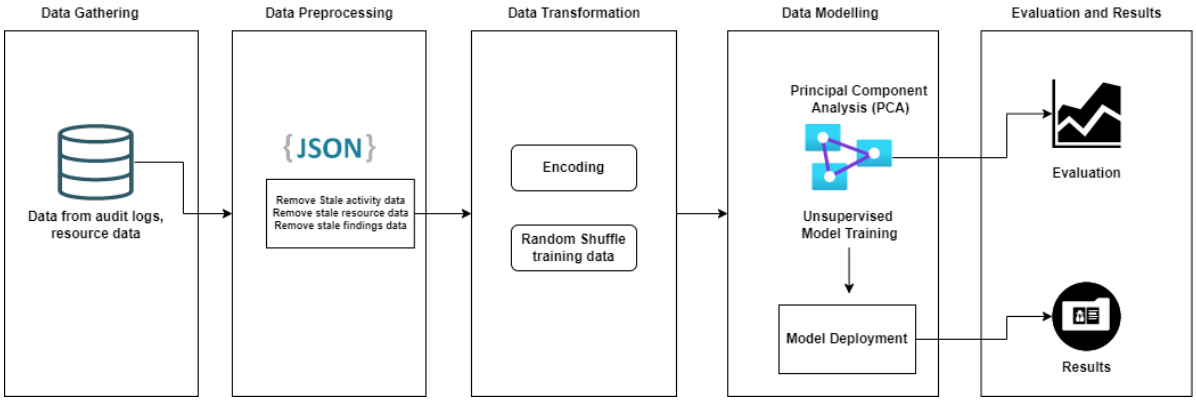

Figure 1 Research Methodology

3.1 Data Gathering

The suggested solution identifies security issues within a tenancy by collecting raw signal data and analysing it according to the detector criteria. Raw signal data is acquired by executing configuration scans on the various services and resources provisioned in each cloud tenancy and ingesting cloud audit data.

All the resources in a cloud tenancy of an organisation were monitored using the proposed CASB solution for incoming threats. Data was gathered from audit logs enabled in the cloud tenancy using APIs. Apart from that, the resource data, activity data was also collected using API adapters.

3.2 Data Preprocessing

This involves cleaning and formatting the data to ensure that it is ready for analysis. The dataset gathered from audit logs is converted to JSON format to group the data for ‘password-guessing’, ‘privilege-escalation’, ‘impossible-travel’, ‘password-spraying’ and ‘persistence’. The data was cleaned to remove stale activity data, stale cloud resource data and stale findings data.

Each of these groups had data in attribute-value pairs. The attributes associated with each resource data include ‘ip address’, ‘id’, ‘tenancyid’, ‘resourcename’, ‘authtype’, ‘eventTime’, ‘source’, ‘userAgent’, and ‘signalDate’.

For the scope of this paper, audit data has been used to identify anomalous accesses to cloud resources. The dataframe consists of ‘ip_address’, ‘total_users’, ‘failed_logins’, ‘successful_logins, ‘failed_logins_users’, ‘successful_login_users.

3.3 Data Transformation

This entails modifying the data in some way to make it more conducive to analysis, as by summarizing or aggregating it. Since, it is an unsupervised model, we do not have to split training and test data as data is identified at real time. The data received after dropping unnecessary columns, are binary encoded. The data is then randomly shuffled to train the PCA model.

3.4 Data Modelling

Principal Component Analysis (PCA) is an unsupervised machine learning approach that reduces the dimensions of a dataset while maintaining as much information as necessary. To

do this, a new set of features termed components that are hybrids of the original features and are uncorrelated with one another are discovered. Additionally, they are limited so that the first component explains the maximum amount of data variability, the second component the next-highest amount, and so on.

3.5 Pattern Evaluation and Presentation

Data patterns or trends that can point to the existence of a threat can be found using PCA. In this case, PCA is used to analyse user access data to spot any irregularities or suspicious activities that can point to a cyberattack.

Each user activity data is analysed to calculate the risk score associated. Each of the problems detected by the CASB solution is categorised as Critical, High, Medium, Low or Minimal based on the problem identified.

At the same time, user accesses are also detected and associated risk score is calculated for each user in the tenancy. Thus, the model identifies the user activity patterns for each user and detects any deviation from the intended behaviour.

4 Design Specification

The proposed Adaptive CASB solution is composed of two components – Data pane and the Control Pane. For this research, the solution is used to monitor resources of a cloud tenancy in an organization.

Monitored resources:

The monitored resources are deployed in the same tenancy across the same Virtual Cloud Network (VCN) in different subnets. These resources include Web server Compute instances (Virtual Machines, Bare Metal instances), Load Balancer, Database System, Storage, Networking, Audit Vault.

Control Pane:

CASB API – This allows users to configure rules for detection.

Administration API – Allows admin users to fetch specific configuration details, data

Data Pane:

Configuration Adapter - Scans all the resources in the tenancy and pushes the resource data to the Signals API

Activity Adapter – Scans the audit logs from the tenancy and pushes the data to the Signals API

Signals API – Validates the resource data and activity data and pushes the valid data to the Detectors

Findings API – Receives the findings (events) data from the detectors and pushes it to the Database for updation

Detector – Applies rules to the tenancy data, creates findings and pushes findings data to Findings API

Correlation Engine – Modifies all the findings for the tenant and creates a problem or an alert in case of a possible security event.

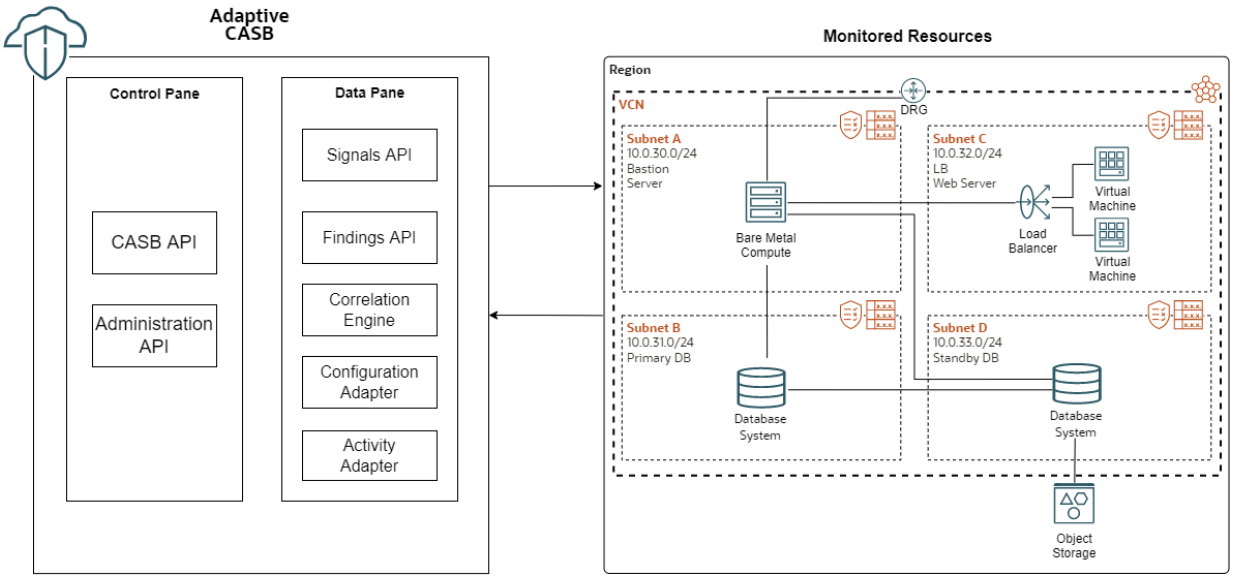

Figure 2 Adaptive CASB Architecture Diagram

The machine learning model used with the Adaptive CASB solution uses a Principal Component analysis (PCA) approach. The CASB collects information from multiple sources, including user activity logs, audit logs and other resources. The database system houses these logs. To eliminate duplications, fix errors, and prepare the data for analysis, the acquired data is pre-processed. Categorical variables may be subjected to processes like normalization, scaling, and encoding. By locating the key patterns in the data and mapping the data onto a lower-dimensional space, PCA is used to reduce the dimensionality of the data. After the data has been reduced, it is examined for anomalies or strange patterns that might point to security threats. In the event of a security event such as password spraying, the CASB solution can identify and notify about the threat.

4.1 Tools used for the project

The programming languages used in this project are Java and Python. Java is used to write the Apache flink job that runs detections on incoming events using the trained model. It is also used for processing incoming events from the auth layer, data aggregation and updation to the database. Python is used for the machine learning code which is a scheduled job responsible for training the model.

The libraries used are pyhocon, numpy, pandas and scikit-learn. Apache flink is also used to process threat detections in real time.

5 Implementation

The Adaptive CASB Solution is deployed in between the cloud users and the applications on cloud. It acts as a gateway which monitors all user accesses and stores data about all the accesses. This same data in real time is used to train the machine learning model to identify anomalous accesses.

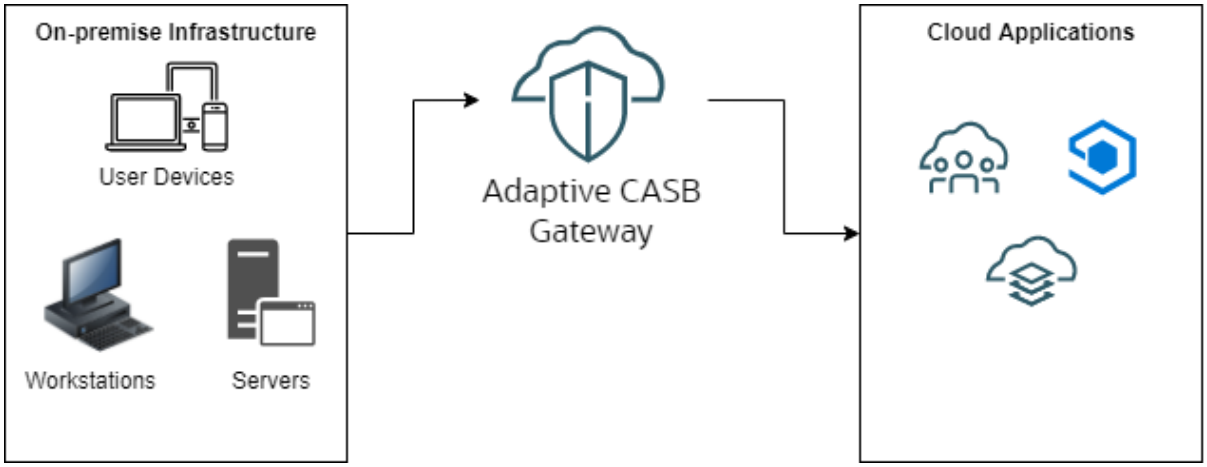

Figure 3 Adaptive CASB implementation in a Cloud tenancy

In the figure above, the internal network and systems of the company, including its servers, workstations, and other hardware, are represented by the on-premises infrastructure. The Adaptive CASB solution's main building block is the CASB gateway. It stands between the organization's on-premises infrastructure and its cloud-based applications. The numerous SaaS, PaaS, and IaaS services that the enterprise utilizes are depicted by the cloud applications. These could be infrastructure services, CRM tools, productivity applications or any other internal applications.

The gateway receives data from the user access such as the location, IP-address, timestamp, resource ID, etc. It also receives data from several API endpoints of the cloud resources.

The data flow diagram is shown below.

CASB Adapters ingests data from the Audit service in order to listen for events that resulted in a change in resource status. These messages are known as Activity Signals. The activity adapter ingests data from the audit logs and the configuration adapter ingests data from the APIs of other cloud resources. To determine the state of the resource, the configuration adapter ingests data utilizing client APIs made available by the services. These signals are referred to as Configuration Signal.

By converting the ingested event and resource state data into the JSON entity, it is shared with the Signals REST API. The validated signal is sent to the Signals Stream for the detector engine to process. The Detector Engine receives incoming signals from the Signals Stream and evaluates rules to determine whether there may be a problem (threat).

If there is a threat or anomalous activity identified, the Detector engine sends the finding to the CASB API which updates the dashboard to tag it as a risk. The reporting dashboard receives the potential issues identified by the Detector Engine. This is called Finding. The Finding entity is forwarded Findings REST API. This is then sent to the Findings Stream and then for the correlation engine to process.

When a new problem is created or an existing problem is linked to a new finding, the correlation engine consumes the incoming finding from the findings stream. The Correlation Engine publishes the problem it has created or changed to the Problems Stream. The issue created or modified by the correlation engine is saved in the database. The Correlation Engine shares the Problem it has created or modified with the Reporting Tenancy.

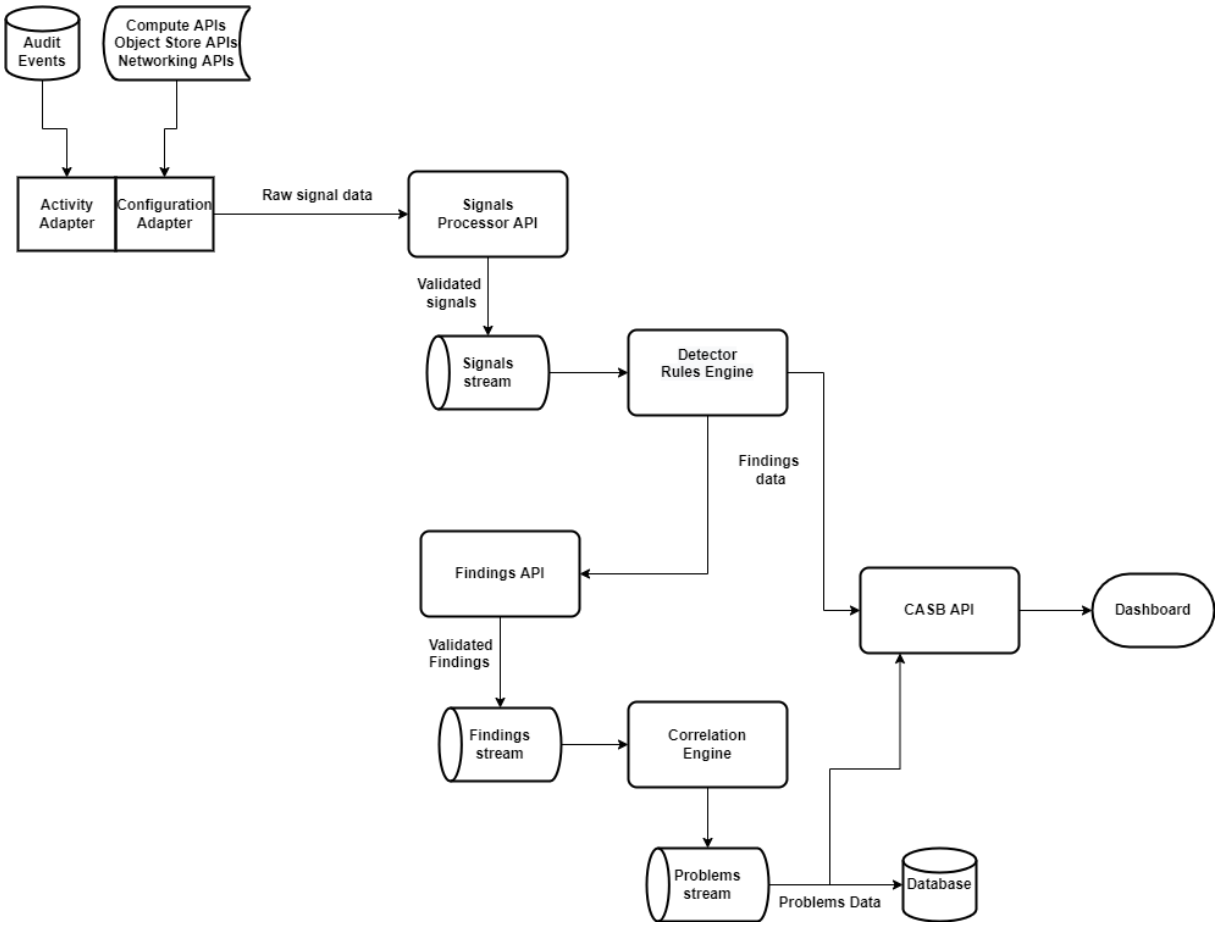

Figure 4 Data flow diagram

In order to configure resources, view issues and associated analytics data, and view problems, End-User interacts via API/CLI/SDK/Console/Terraform. Users interact with the database if they are interested in data about problems, responders, analytics, and associated entities. Thus, when end users interact with the console, they can view the risk score of the users and resources in the cloud tenancy. They can also view and modify the detector rules associated with the cloud tenancy.

This deployment provides visibility to all malicious activity since it can detect all activities dynamically using machine learning model. Principal Component Analysis is an unsupervised learning method, which helps reduce the dimensionality of the data consumed by the CASB Adapters. This makes it suitable for identifying multiple threats including password spraying.

6 Evaluation

This research constitutes of two parts – the CASB solution which detects and notifies the threats and the machine learning model which uses PCA to identify anomalies dynamically and powers the CASB solution to detect the risks dynamically. The results of each of these components are discussed below.

6.1 Principal Component Analysis (PCA)

The data used for this CASB solution is taken in real time and has a huge number of components or attributes. The main objective of this experiment is to use Principal Component analysis to reduce the dimensionality of the data and create a new vector space. Firstly, a dataframe with the suitable column (ip_address) is created which will be appended with all the attributes. Standard scaler is used to reduce the differences between the values. That is, by using Standard Deviation =1 and Mean =0 , the values are scaled down. The number of components in the data is reduced to n=2 to fit the scaled data. As we can see from the figure below, the number of dimensions is reduced and the standardized dataset is scaled down for accurate analysis.

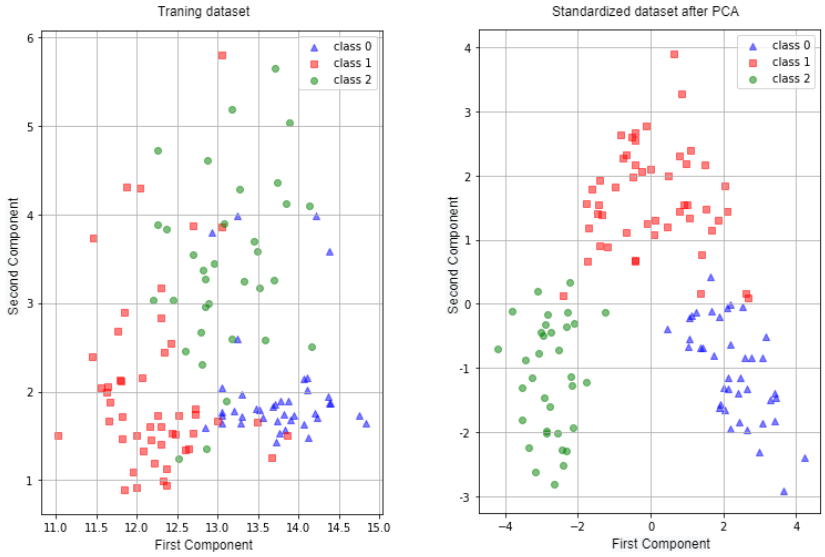

Figure 5 PCA to reduce dimensionality

Data from the audit logs of the cloud resources in a cloud tenancy of an organization is taken in real time, which consists of all the detection rules, IP addresses, Failed and Successful logins. Based on this data, the unsupervised model was trained and the model test accuracy is found to be 98% and the validation accuracy is 85%

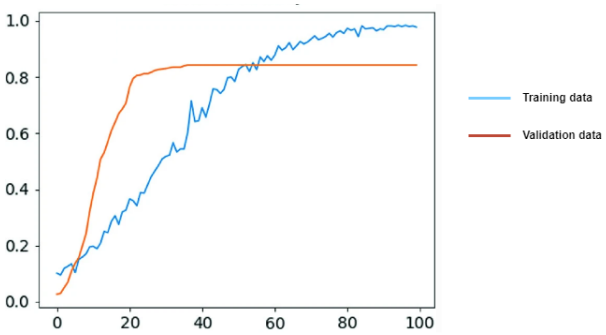

Figure 6 Model Accuracy

6.2 CASB Findings

The objective of this experiment is to calculate the risk score associated with each resource on the cloud tenancy including Users, applications, compute instances, etc. As we can see below, the detector rules and risk levels are configured from the user interface. This is accessed by the CASB solution via the Configuration adapter.

Detector rule	Risk level	Status	Settings configured	Conditional group
VCN deleted	Medium	Enabled	Not allowed	No
VCN created	Low	Enabled	Not allowed	No
VCN Security List ingress rules changed	Medium	Enabled	Not allowed	No
VCN Security List egress rules changed	Medium	Enabled	Not allowed	No
VCN Security List deleted	Medium	Enabled	Not allowed	No
VCN Security List created	Low	Enabled	Not allowed	No
VCN Route Table changed	Medium	Enabled	Not allowed	No
VCN Network Security Group ingress rule changed	Medium	Enabled	Not allowed	No
VCN Network Security Group egress rule changed	Medium	Enabled	Not allowed	No
VCN Network Security Group Deleted	High	Enabled	Not allowed	No
VCN Local Peering Gateway changed	Medium	Enabled	Not allowed	No
VCN Internet Gateway terminated	Low	Enabled	Not allowed	No
VCN Internet Gateway created	Medium	Enabled	Not allowed	No

Figure 7 Detector Rules

Based on these detector rules, the threshold is selected by the CASB solution and it calculates the risk score associated with each resource. The risks are classified as Critical, High, Medium, Low or Minor based on the detector rules

Risk score	Sightings	Tactics	Resource type	First detected	Last detected
55.67	40	4	User	Thu, Jun 2, 2022, 14:14:42 UTC	Wed, Dec 28, 2022, 17:12:25 UTC
47.49	8	3	User	Tue, Aug 30, 2022, 03:03:13 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
42.34	15	2	User	Fri, Sep 9, 2022, 10:14:32 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
40.13	13	4	User	Fri, Aug 26, 2022, 09:13:59 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
34.4	19	5	User	Fri, Jul 1, 2022, 10:22:17 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
30.19	1	1	User	Tue, Dec 6, 2022, 12:15:13 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
23.7	7	2	User	Wed, Aug 17, 2022, 10:24:21 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
23.09	2	1	User	Thu, Dec 1, 2022, 14:12:59 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
22.87	13	2	User	Wed, Jul 20, 2022, 16:27:56 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
21.01	6	2	User	Fri, Sep 2, 2022, 12:14:10 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
18.6	5	2	User	Mon, Sep 26, 2022, 12:25:09 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
12.85	2	1	User	Fri, Sep 2, 2022, 14:13:57 UTC	Tue, Dec 27, 2022, 09:40:49 UTC
12.34	6	4	User	Sat, Aug 27, 2022, 03:03:16 UTC	Tue, Dec 27, 2022, 09:40:49 UTC

Figure 8 Risk score of the resources

The total number of 507 risks are identified by the CASB solution is classified and shown on the dashboard as shown below. The risk score of the users over a 30 day period is also shown.

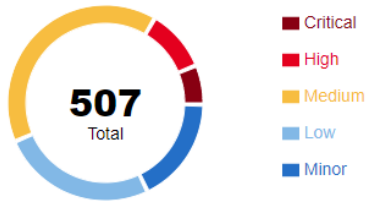

Figure 9 Risk Findings

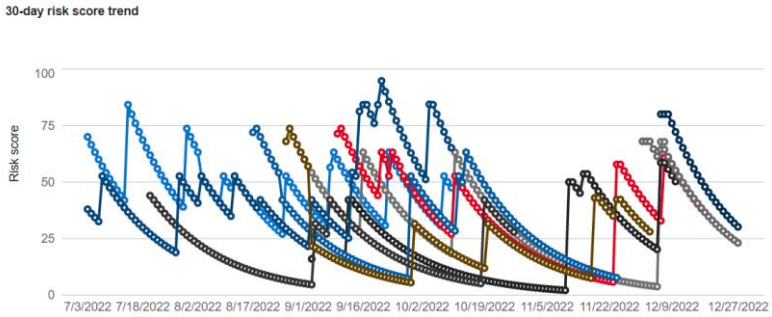

Figure 10 Risk score trend of users in the tenancy

6.3 Discussion

The solution developed is an Adaptive CASB solution, which used machine learning to detect and identify anomalous accesses to the user resources. This paper utilises principal component analysis to minimize the dimensionality of real time data which makes the model more accurate. It can identify the anomalous accesses by evaluating it against a threshold set by the detector rules. The main strength of this solution is that the model can respond to events at a real time and keeps learning as and when more data is collected. This allows the CASB solution to detect and notify threats dynamically and introduce new rules based on the trends observed in the collected audit logs.

7 Conclusion and Future Work

This research focuses on the importance of Cloud access security brokers in cloud deployments and how machine learning can power the standard CASB solutions to respond to threats dynamically at real time. The proposed solution uses principal component analysis to identify user access patterns and detect anomalous accesses. The primary advantage of this technique is that the model can react to events in real time and continues to learn as new data is acquired. This enables the CASB system to dynamically identify and alert risks, as well as propose new rules based on patterns seen in the gathered audit logs.

Some of the limitations of this research includes the lack of access to high volume real time data which can train the model to detect anomalies more accurately. This research was conducted on a small number of resources deployed on a cloud tenancy with restricted access to the audit logs. The model could have been accurately assessed if there was access to more data and the resources monitored were complex and higher in volume. Another limitation of the solution is that it only detects the risks and notifies about the anomalous accesses. There are no measures taken to stop or remediate the findings.

The research can be extended in future by integrating CASB with the remediation solutions such that it completes overall security lifecycle of the cloud resources. This can include responders which can take corrective action based on the problem events generated by the CASB solution. It can also be further enhanced by using various other machine learning models to improve the performance and accuracy such that they are well equipped to handle potential risks.

References

Agarwal, A. *et al.* (2021) 'Classification model for accuracy and intrusion detection using machine learning approach', *PeerJ Computer Science*, 7, p. e437. Available at: https://doi.org/10.7717/peerj-cs.437.

Ahmad, S. *et al.* (2022) 'RSM analysis based cloud access security broker: a systematic literature review', *Cluster Computing*, 25(5), pp. 3733–3763. Available at: https://doi.org/10.1007/s10586-022-03598-z.

Ahmad, S., Mehfuz, S. and Beg, J. (2019) 'Fuzzy Cloud Access Security Broker for Requirements Negotiation and Prioritization', in *2019 International Conference on Power Electronics, Control and Automation (ICPECA)*. *2019 International Conference on Power Electronics, Control and Automation (ICPECA)*, pp. 1–6. Available at: https://doi.org/10.1109/ICPECA47973.2019.8975620.
Ahmad, S., Mehfuz, S. and Beg, J. (2020) 'Securely Work from Home with CASB Policies under COVID-19 Pandemic: A Short Review', in *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*. *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*, pp. 109–114. Available at: https://doi.org/10.1109/SMART50582.2020.9337121.
Ahmad, S., Mehfuz, S. and Beg, J. (2021) 'Enhancing Security of Cloud Platform with Cloud Access Security Broker', in M.S. Kaiser, J. Xie, and V.S. Rathore (eds) *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Singapore: Springer Nature (Lecture Notes in Networks and Systems), pp. 325–335. Available at: https://doi.org/10.1007/978-981-16-0882-7_27.
Al-Fawa'reh, M. *et al.* (2022) 'Cyber threat intelligence using PCA-DNN model to detect abnormal network behavior', *Egyptian Informatics Journal*, 23(2), pp. 173–185. Available at: https://doi.org/10.1016/j.eij.2021.12.001.
Badhwar, R. (2021a) 'Commentary on Insider Threat', in R. Badhwar (ed.) *The CISO's Next Frontier: AI, Post-Quantum Cryptography and Advanced Security Paradigms*. Cham: Springer International Publishing, pp. 345–351. Available at: https://doi.org/10.1007/978-3-030-75354-2_42.
Badhwar, R. (2021b) 'Introduction to Cloud Monitoring Security Controls', in R. Badhwar (ed.) *The CISO's Next Frontier: AI, Post-Quantum Cryptography and Advanced Security Paradigms*. Cham: Springer International Publishing, pp. 289–296. Available at: https://doi.org/10.1007/978-3-030-75354-2_36.
Bhattacharya, D. *et al.* (2021) 'Dynamic Cloud Access Security Broker Using Artificial Intelligence', in A. Joshi, M. Khosravy, and N. Gupta (eds) *Machine Learning for Predictive Analysis*. Singapore: Springer (Lecture Notes in Networks and Systems), pp. 335–342. Available at: https://doi.org/10.1007/978-981-15-7106-0_33.
Feng, X., Feng, Y. and Dawam, E.S. (2020) 'Artificial Intelligence Cyber Security Strategy', in *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, pp. 328–333. Available at: https://doi.org/10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00064.
Fernandez, E.B., Yoshioka, N. and Washizaki, H. (no date) 'Cloud Access Security Broker (CASB): A pattern for secure access to cloud services', p. 8.
Fossum, T. and Andersen, V. (2021) *Investigating Cloud Access Security Broker In A Healthcare Service : Creating A Cloud Access Security Broker (CASB) Discussion Frame-work For Evaluating Security in Cloud Healthcare Services, 107*. Master thesis. University of Agder. Available at: https://uia.brage.unit.no/uia-xmlui/handle/11250/2823634 (Accessed: 29 November 2022).
Kemp, R. (2018) 'Legal aspects of cloud security', *Computer Law & Security Review*, 34(4), pp. 928–932. Available at: https://doi.org/10.1016/j.clsr.2018.06.001.
Kumar, R. and Goyal, R. (2019) 'On cloud security requirements, threats, vulnerabilities and countermeasures: A survey', *Computer Science Review*, 33, pp. 1–48. Available at: https://doi.org/10.1016/j.cosrev.2019.05.002.
Kumar, S. *et al.* (2022) 'Role of Machine Learning in Managing Cloud Computing Security', in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 2366–2369. Available at: https://doi.org/10.1109/ICACITE53722.2022.9823414.
Lee, J.-H. *et al.* (2017) 'Toward the SIEM architecture for cloud-based security services', in *2017 IEEE Conference on Communications and Network Security (CNS)*. *2017 IEEE Conference on Communications and Network Security (CNS)*, pp. 398–399. Available at: https://doi.org/10.1109/CNS.2017.8228696.
Liu, C. *et al.* (2017) 'A Cloud Access Security Broker based approach for encrypted data search and sharing', in *2017 International Conference on Computing, Networking and Communications (ICNC)*. *2017 International Conference on Computing, Networking and Communications (ICNC)*, pp. 422–426. Available at: https://doi.org/10.1109/ICCNC.2017.7876165.
Meng, F. *et al.* (2018) 'Deep Learning Based Attribute Classification Insider Threat Detection for Data Security', in *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*. *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, pp. 576–581. Available at: https://doi.org/10.1109/DSC.2018.00092.
Mohammad, A.S. and Pradhan, M.R. (2021) 'Machine learning with big data analytics for cloud security', *Computers & Electrical Engineering*, 96, p. 107527. Available at: https://doi.org/10.1016/j.compeleceng.2021.107527.

Nassif, A.B. *et al.* (2021) 'Machine Learning for Cloud Security: A Systematic Review', *IEEE Access*, 9, pp. 20717–20735. Available at: https://doi.org/10.1109/ACCESS.2021.3054129.
Rathod, V., Parekh, C. and Dholariya, D. (2021) 'AI & ML Based Anamoly Detection and Response Using Ember Dataset', in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp. 1–5. Available at: https://doi.org/10.1109/ICRITO51393.2021.9596451.
Twum, F., B., J. and K., J. (2020) 'A Comparative Study of Existing Cloud Security System Models as against an Implementation of the CDDI Model Dubbed SecureMyFiles System', *International Journal of Computer Applications*, 177, pp. 17–37. Available at: https://doi.org/10.5120/ijca2020919765.
Waguie, F.T. and Al-Turjman, F. (2022) 'Artificial Intelligence for Edge Computing Security: A Survey', in *2022 International Conference on Artificial Intelligence in Everything (AIE)*. *2022 International Conference on Artificial Intelligence in Everything (AIE)*, pp. 446–450. Available at: https://doi.org/10.1109/AIE57029.2022.00091.
Waskle, S., Parashar, L. and Singh, U. (2020) 'Intrusion Detection System Using PCA with Random Forest Approach', in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 803–808. Available at: https://doi.org/10.1109/ICESC48915.2020.9155656.