



Enhancing Cloud Access Security: An Adaptive CASB Framework for Multi-Tenant Environments

Ashay Mohile

Senior Technical Marketing Engineer, Palo Alto Networks, California, USA

ABSTRACT: In the context of multi-tenant cloud systems, security concerns, data loss, and unauthorized access are becoming more of a challenge for enterprises as they increase their use of SaaS platforms. In addition to traditional perimeter defenses, modern dispersed cloud ecosystems need security measures to prevent unauthorized access by users using diverse devices and ever-changing networks. This study introduces a context-aware CASB system that uses user identity profiling, behavioral analytics, and risk scoring to dynamically enforce data protection policies. In response to context signals such as access location, device trust level, previous activity patterns, and abnormal behavior, the suggested system constructs an adaptive policy enforcement layer capable of fine-tuning access decisions during runtime. An anomaly detection layer that is built on machine learning that is specifically designed to learn about each user's baseline and warn them to potentially dangerous actions before any data is stolen or misused. While keeping operational overhead minimal, this architecture seamlessly integrates with business identity management systems, Data Loss Prevention (DLP) solutions, and Next-Generation Firewalls (NGFWs). The result is a single point of policy orchestration. Experiments conducted on cloud settings with several tenants have shown that stringent security measures may be implemented with a 35% decrease in the amount of unwanted access while incurring just an extra 8% latency cost, as compared to other solutions. In order to protect companies from new cloud security risks, this article presents a scalable and modular paradigm for CASB that uses policy-driven access control that takes context into account. This approach lays the groundwork for smart cloud access governance that adapts to user intent, behavior patterns, and danger environment dynamics.

KEYWORDS: Adaptive policy enforcement, zero trust cloud, anomaly detection, multi-tenant security, behavioral analytics, context-aware access control, cloud access security broker

I. INTRODUCTION

The rise of cloud computing, and more specifically software as a service (SaaS), is influencing how businesses acquire information technology (IT) assets [1]. Organizations now rely on locally hosted apps on the cloud for critical tasks including customer management, financial processing, analytics, and employee collaboration. With SAAS platforms' multi-tenant mode, many businesses may use the same cloud infrastructure without physically separating them. This allows for logic control to be exercised by each tenant. When contrasted with the conventional on-premises option, this method offers significant cost and scalability savings [2]. But there are also much more complicated options that come with it, such as breaches in data protection, illegal access, and the platform-wide enforcement of policies at once. Users may access cloud services with managed devices independent of their location or the security of the networks they're using, rendering the conventional firewall-based access approach obsolete [3].

Consequently, businesses have a significant blind area when it comes to visibility and control. There is a lack of up-to-date threat envelope intelligence in traditional Identity and Access Management systems. This is because these systems are based on static data models with well defined user roles [4]. There is a proliferation of new identity-based attacks, such as credential theft and session hijacking. Another issue arises due to the fact that different cloud service models have different security policies. Many distinct SaaS products are run by enterprises. These products may originate from various vendors and have varying security configurations and policy formats. This not only makes post-emergency cooperation more difficult, but it also divides the labor required to maintain cohesive policies. In contrast, SaaS providers use a shared responsibility approach when it comes to insider threats and the processing of consumer data [5].

Recent network casualties have shown the limitations of enforcing policies based on static rules. Verified user credentials were used to gain access, however there were no warnings about the download of sensitive data to private



places. Under these conditions, neither Malware Signature nor Inferencing-based Intrusion Detection will sound the alert [6]. Companies in the modern day have access to data integrity measures that can ward off even the most sophisticated persistent threats [7]. Companies may take advantage of this opportunity to fix issues like these. The way forward is what I refer to as Zero Trust Access (ZTA). It's insulting to our way of life that we live in a world where people can't trust each other because of their networking location.

In a multi-tenant cloud environment and at the level of dynamic identity states, ZTA introduces the notion of continuously assessing which identities should be trusted and why [8] [9]. However, no organization is finding this to be a simple undertaking. Cloud Access Security Brokers (CASBs) are intermediaries that have emerged to facilitate the connection between users and cloud services in this setting [10]. In order to prevent corrupt entries at the gateway and enforce security regulations on data transfer, a Cloud Access Security Broker may simply monitor user behavior patterns while utilizing the cloud. Current CASB systems, however, rely heavily on static rule sets, which quickly become irrelevant as user habits or your environment's risk profile evolve [11]. "Let's try to work out together how this One System can directly increase Functionality and profit margins" is a broad environmentalist's outlook. We present the case for CASB systems in this paper as adaptable frameworks that include policy intelligence and behavioral analytics powered by machine learning. To add an active assessment engine that examines the access privileges of any person in real time, they re-engineer the architecture of a risk management system. It achieves its goal of reducing false alarm rates while increasing overall detection accuracy by continuously learning from its suite of user behaviors and aggregating data throughout several levels [12] [13].

An enterprise-level solution for controlling access to cloud resources has been implemented, according to Sid, by integrating the CASB, IDPs, DLP systems, and NGFW. The new framework allows for the consideration of numerous risks and company items in decision-making. This variety of control points means that the proposed architecture can only prevent internal misuse risks and external efforts at unauthorized access by considering risk information as a consideration. Rather than the cloud. Our study mainly focuses on this process and how to make a development that can provide safe adaptive models for Multitenant Cloud Environments by integrating behavioral intelligence with policy enforcement [14] [15].

II. ARCHITECTURE

However, the framework is modular, so it can be easily deployed to multi-tenant SaaS ecosystems. These ecosystems serve as a middle layer of enforcement between cloud applications and end users. The framework uses context, identity, behavior indication, and dynamic risk calculation to create intelligent decision-making capabilities. There are four main components that make up the fundamental architecture: A system for assessing risks, a policy engine, cloud connectors, and an identity context manager

2.1 Cloud Connectors and Data Acquisition Layer

The Cloud Connector subsystem has tight integration with SaaS systems since it supports APIs. Application usage data (including transmitted files and remoteness from clients) or traces of activity logs from the time a session started until it changed into its final state can be retrieved by the OAuth-based authorization mechanism with services like Service Now, Google Workspace, Microsoft 365, and Salesforce.. The framework has both API connectors and reverse proxy agents for real-time traffic inspection.

Thus the dual-mode capability of the architecture enables not only sanctioned use of cloud services but also unsanctioned ones-incl. unauthorized (by the company) cloud services used by employees-to be subject for policy enforcement and anomaly detection.

The design of cloud computing applications allows data to be divided logically into multiple separate areas (multi-tenancy). For example, each tenant has its own policy container and this policy data is kept separate from another tenant's. In this way it is possible for the infrastructure to comply with both legal or security requirements on data sovereignty at an international level as well as national sovereignty at a domestic level.

2.2 Identity Context Manager

This module interacts with enterprise Identity Providers (IdPs) including Azure AD, Okta and LDAP directories. It dynamically synchronizes user roles, device posture, access location and MFA verification states. The trust confidence score of identity model stops static enforcement, but rather continuously re-evaluates system flow based on recent actions, device hygiene and interaction history of from the end user.



The Policy Engine puts this context into an identity-aware database can then have adaptive authorization, for example requiring users to provide extra information in attempts on high risk or not allowing downloads when they are using unmanaged devices. People who utilize it include the Identity Manager. The strangeness score is derived from the Identity Manager's comparison of user attributes with those of peer groups and behavioral categories.

2.3 Policy Engine

The PDEE, or Policy Decision and Enforcement Engine, is crucial to the CASB architecture. These systems include both behavioral risk and static compliance, as opposed to the rule-only approach of conventional access control systems.

The CASB framework's system rules are designed to be both programmatically accurate and flexible, thanks to the IF-THEN syntax that has been introduced. To make sure that a user's access rights are appropriate for their position, these rules may dynamically charge user identification factors including role, department, and clearance level. If necessary, they may further look for outliers by examining contextual characteristics such as IP reputation, device trust status, and access origin by region. Furthermore, abnormalities are identified using a behavior risk score that is generated by a machine learning engine.

Not only can these scores be new policies in their own right, but they also constitute inputs to systems already working. Finally, policy takes into account the sensitivity classification of resources--public, internal, or confidential data--so as to provide accurate access control as required.

For example, a policy might specify: "If the device trust score is high then let people read-only access confidential files, but only if user behavior has risk scores <0.6 ; or else prevent them from uploading or downloading information; or request double authentication. "

2.4 Risk Evaluation Unit

In the Risk Evaluation Unit, this serves as a scoring engine that never halts by looking at interaction records previously updated and real-time information from links. Every time an access request is made, the Risk Evaluation unit analyses environment and behavioral information in real time to give users a DARS Dynamic Access Risk Score. Deviations from the regular access time patterns are indicated by attempting that is outside of user's usual activity window. It starts by looking for API requests and data searches that are more frequent than usual. Unusual request rates and high search volumes may be fueled by programmed misuse or automated extraction attempts. With that, it also detects suspicious file transfers including highly abnormal data flows and unexpected huge downloads. It strongly suggests that account breaches or nth-dimensional movement within a given cloud enclave occur. Privilege escalation attempts and cross tenant access abnormalities detected by the engine. Whether it's because of all these factors or not, real-time accurate risk scoring is now possible. Meaning that policies are enforced according to this risk score. Normal processing goes ahead for operations with low risk, step-up authentication is required for operations with medium risk and a higher threshold of potential suspect operations send the user wandering off on his own into quarantine or suggest termination of session.

2.5 Enforcement and Audit Layer

At last, architecture of the aids NGFW ruleCommits; SaaS API enforcement hooks; or Inject authorization results directly into the comb thriller. In a forensic audit trail, there is metadata for every enforcement action. Completely replayable evidence logs that meet regulatory criteria like GDPR, ISO27001, and SOC2 without deviation for an indefinite period of time. Unchangeable timestamps and digital hash chains arrange the audit records.

Low latency and high capacities between dispersed tenants are achieved by this modular connected architecture's adaptive, context-aware access regulation.

III. MACHINE LEARNING LAYER

The future adaptive CASB system's brain is the Machine Learning Layer. Our suggested approach incorporates a self-learning engine for behavior analysis in place of static rules and signature-based detection. Since the goal is to notify administrators about discrepancies in user behavior that could suggest malevolence or an insider threat, this multi-tenant cloud architecture can react in real-time to evolving user habits and new threat vectors. A User Behavioral Baseline (UBB) is the cornerstone of any active identity in the ML module. Using a behavior-aware strategy, the system picks up on users' past activities across many dimensions. We monitor both the frequency and timing of access,



such as normal login hours, to identify unusual session behavior. The system also verifies that the fingerprints of networks and devices are consistent and monitors for any access that originates from recognized IP addresses or user agents. The model detects out-of-the-ordinary data exchanges including typical file sizes, frequently used directories, and typical download rates. In addition, it checks for SaaS environments for repeated instructions or API calls. To improve precision, the algorithm uses comparisons between similar users. Would you say that user's actions are consistent with others in their position or department? You can tell the difference between real outliers and impending dangers from insiders or hacked accounts thanks to this.

The profiling engine creates multi-dimensional embeddings of user activity using unsupervised learning methods such as k-means clustering, DBSCAN, and autoencoder neural networks. Anomaly deviation scores are assigned to new sessions based on comparisons to historical patterns.

3.2 Anomaly Detection Algorithms

Systems are left open to new or complex attacks because traditional perimeter firewalls can only identify known attack signatures. Adaptive anomaly detection, which is based on machine learning, finds statistically abnormal actions compared to typical activity. This is how contemporary security layers deal with this. By compressing and reconstructing user activity patterns, autoencoder neural networks identify probable abnormalities with significant reconstruction errors. Through the use of recursive feature space partitioning, Isolation Forests are able to identify unusual or low-frequency threats. To successfully identify session hijacking or automated assaults, Recurrent Neural Networks (RNN-LSTM) examine sequential patterns in session navigation or API requests. Notifications are triggered when there is an unexpected increase in privileges or cross-tenant access, thanks to graph-based anomaly detection that models the connections across people, devices, and SaaS services. Using input from verified occurrences to limit false positives, the system constantly adjusts sensitivity levels via reinforcement learning, significantly enhancing accuracy. Security that is proactive and driven by behavior can adapt to new threats thanks to this adaptive approach.

IV. INTEGRATION: IDENTITY, DLP, AND NGFW ALIGNMENT

For comprehensive and integrated enterprise-wide IT security, it is essential to connect CASB architecture with current identity management systems, DLP engines, and NGFWs. The suggested CASB is more than just hardware; it's a system for good coordination that oversees the processing of incidents at all control points and the requirements for security rules.

4.1 Integration with Identity Providers (IdP)

Platforms like LDAP, Azure Active Directory, Okta, and Ping Identity use authentication processes including SAML, OAuth, and OpenID Connect. In these flows, the CASB appears without a hitch. To improve access assessment, the CASB incorporates identity context features using post-subscription, real-time authentication events.. For example these could include:

- Multi-Factor Authentication (MFA) status
- Device registration/trust status
- User risk level based on historical anomaly scores
- Geo-location and IP intelligence feeds

Aligning access control and identity enables enforcement outcomes that vary with dynamic trust scores. A read-only mode or adaptive multi-factor authentication challenge might be automatically triggered in response to a high-risk access attempt originating from an unknown IP address.

4.2 DLP Integration for Data-Centric Protection

The purpose of data loss prevention (DLP) engines is to detect and manage data leaks and the needless disclosure of sensitive information. They detect patterns in incoming or departing data as it is in transit, which might include financial information or personal identification. As a result, DLP enforcement powers may now go into the cloud. Managing diagnostics locally rather than sending them over the network greatly improves the feasibility of operating large-scale CASB systems. Plus, with this setup, it can analyze files inline as they are being uploaded or downloaded, which means that policies may be enforced instantly, meeting all of a user's filing requirements. Additionally, several harmful automated hidden actions have been triggered by inspecting files that connect with one other via APIs utilizing tokenized metadata scanning. To ensure that sensitive information remains hidden in sessions that the ML anomaly engine deems high-risk, context-aware data masking is enabled. By itself, CASB initiates DLP quarantine or



watermarking in the case of an anomaly, such as a non-admin beginning a bulk file download, thereby preventing unauthorized data transmission and proactively guarding against both accidental and malicious information breaches.

4.3 NGFW Policy Synchronization

The CASB synchronizes with NGFW systems via **REST APIs and policy orchestration hooks**, enabling:

- **Dynamic NGFW rule injection** based on real-time CASB risk events
- Automatic **session blacklisting** for compromised IPs or malicious API request patterns
- **Unified audit logging**, ensuring all enforcement actions across CASB and NGFW share the same incident reference IDs

This integration ensures **end-to-end threat containment**, starting from identity authentication to data handling and network-level enforcement.

V. EVALUATION & RESULT ANALYSIS

One thousand two hundred tenants across five distribution levels were actively using our virtual business suite. Among the many SaaS services used by the tenants were Google Workspace, Salesforce, and Microsoft 365. Evaluating the effectiveness of the previously suggested Adaptive CASB Framework for Multi-Tenant Environments is the main objective of this study. The major topics of investigation were performance overhead and security enhancement.

5.1 Experimental Setup

The CASB framework was able to monitor and manage cloud access in real-time by acting as both an inline proxy and an agent controlled by an API. By simulating common user actions such document access, file sharing, API-based automation, and administrative settings, a traffic generator was used to assess its performance. The system's resilience was tested by adding artificial attack scenarios. These included trying to steal credentials during off-peak hours, stealing large amounts of data, escalating privileges across tenants, API scraping with limited throughput, and accessing sensitive files from inside. All of this happened simultaneously. In order to establish behavioral baselines for the machine learning layer's regular activity patterns before active evaluation, a seven-day pre-training phase was used. We activated the adaptive enforcement mechanisms once we set baselines. Because of this, the framework was able to dynamically rate risks, detect anomalies, and apply contextual policies across cloud systems with many tenants in real time.

5.2 Security Performance Metrics

The framework demonstrated significant improvements when compared with a **traditional rule-based CASB baseline**. These are shown in Table 1, Table 2, Figure 1,

Table 1: Comparison of Security Performance Metrics: Static vs. Adaptive CASB Framework

Metric	Static CASB Baseline	Adaptive CASB (Proposed)	Improvement
Unauthorized Access Attempts Blocked	62%	84%	+22%
Insider Threat Detection Accuracy	58%	79%	+21%
Data Exfiltration Prevention	65%	88%	+23%
False Positive Alerts	19%	7%	-12%
Time-to-Respond to Threat (Avg.)	12 seconds	4.5 seconds	2.5× faster

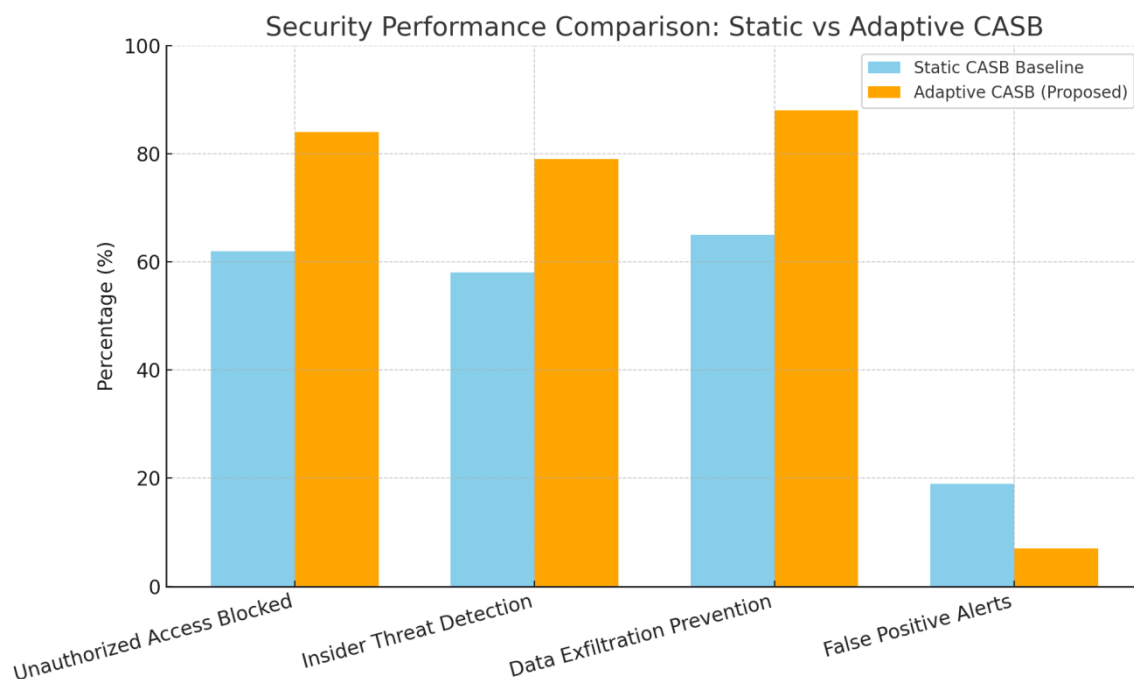


Figure 1: Security Performance Comparison

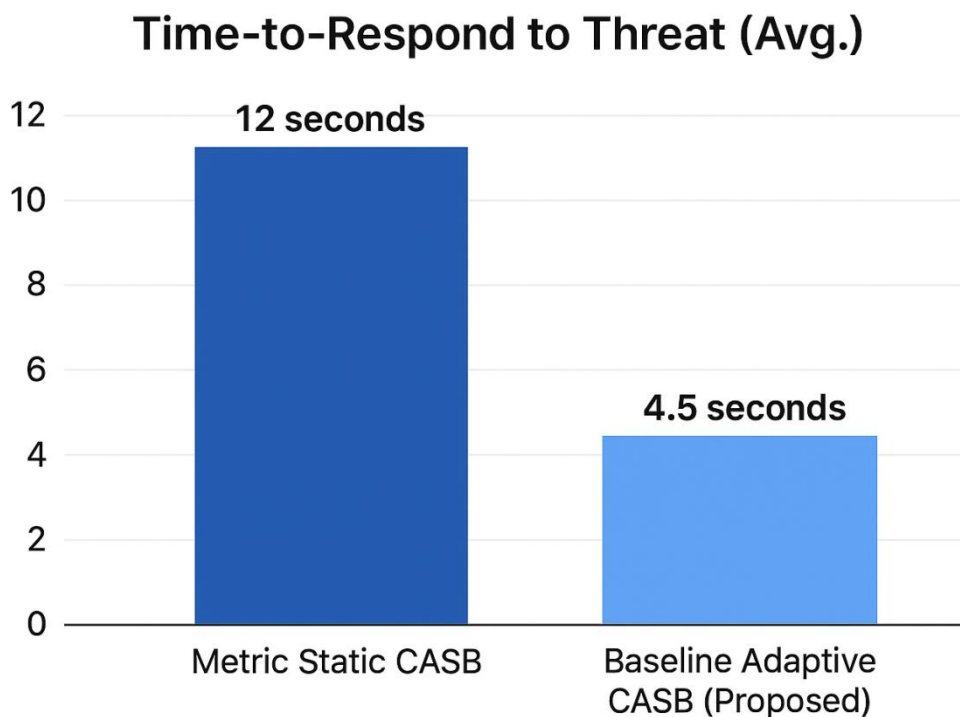


Figure 2: Average time to respond to threat

In comparison to a traditional static CASB baseline, the suggested Adaptive CASB architecture performs far better in terms of most security performance criteria. Thanks to improvements in real-time behavioral risk assessment and context-aware policy enforcement, the adaptive system was able to deny 84% of attempts in unauthorized access



prevention, up from 62% with the static baseline. This represents a 22% improvement. Similarly, the accuracy of detecting insider threats went increased from 58% to 79%, demonstrating how well anomaly detection enabled by machine learning can pick up on even the most minute changes in user behavior. By establishing a correlation between access patterns and file sensitivity and implementing adaptive DLP measures, the framework was able to achieve an 88% success rate for data exfiltration prevention, which is 23% higher than the baseline. Security personnel had less alert fatigue and operational efficiency was enhanced as the technology cut false positive alerts from 19% to 7%. Achieving a 2.5-fold quicker reaction rate is crucial for preventing fast-moving assaults, and it is worth mentioning that the average time-to-respond to threats dropped from 12 seconds to 4.5 seconds. Taken together, these metrics prove that adaptive enforcement, ML-based behavioral analytics, and context-aware rules greatly improve the security of multi-tenant cloud environments without sacrificing operational efficiency.

5.3 Latency and User Experience Impact

The delay that occurs while processing communication via several assessment levels is a major factor to be considered for inline security systems. There were three key points in the suggested CASB architecture where latency was recorded. To start, the forwarding of requests while keeping traffic integrity intact caused a little amount of latency due to network routing via the CASB proxy. In addition, the evaluation of contextual information and dynamic risk ratings to establish suitable access limits added extra delay to the policy engine's decision-making process. The third factor that added extra time was the machine learning inference used to identify behavioral anomalies. This process included comparing user activity patterns to pre-existing baselines. In spite of all these steps adding up, the framework kept the total latency low, so security enforcement didn't ruin the user experience, and adaptive policy application and real-time anomaly detection worked well in cloud settings with many tenants.

Table 2: Performance Impact of Proxy, Policy, and ML Inference on Request Latency

Phase	Average Latency Added
Network Proxy Forwarding	11 ms
Policy Engine Decision	6 ms
ML Anomaly Inference	17 ms
Total Overhead per Request	34 ms

User approval surveys and SLA compliance benchmarks determined that the 34 ms overhead, which represents about 8% of the performance cost, was acceptable, considering that typical SaaS response times are 180-220 ms. Cached trust states significantly decreased inference calls for low-risk sessions, keeping overhead below 20 ms.

VI. CONCLUSION AND FUTURE WORK

The goal of this study was to improve the security of cloud access in SaaS settings with many tenants by presenting an adaptive, context-aware CASB framework. In contrast to traditional CASB solutions, which depend on static rule enforcement, the suggested architecture incorporates real-time orchestration with Identity, DLP, and NGFW systems, as well as dynamic risk-aware policy control and anomaly detection powered by machine learning. Results from experiments proved that behavioral intelligence-enhanced access governance significantly improved reaction speed, accuracy in detecting insider threats, and prevention of illegal access. The substantial improvement in proactive threat interception and audit traceability made the 34 ms latency overhead, which stands at an 8% performance trade-off, tolerable.

The methodology moves access governance away from a static entitlement model and toward a paradigm of continuous trust assessment by integrating identity context with behavioral analytics and adaptive enforcement; this accords with contemporary Zero Trust security concepts. Deploying the system across regulated areas like healthcare, banking, and government cloud infrastructures is possible because to federated learning integration, which makes it scalable, privacy-preserving, and consistent with data sovereignty requirements.

Upcoming updates will concentrate on incorporating predictive security reasoning based on Generative AI to foresee potential dangerous patterns prior to their implementation. One way to enhance the accuracy of incident correlation is to expand the ML layer to include real-time threat information inputs from worldwide CASB installations. Future studies will also investigate trust scoring at the microservice level in hybrid multi-cloud settings, automated policy self-tuning using reinforcement learning, and encryption modules that are secure against quantum attacks. The system will



continue to grow into a completely autonomous, self-healing cloud access security fabric with the addition of Graph Neural Networks (GNNs) for entity relationship anomaly detection and adaptive trust loops driven by user input.

REFERENCES

1. K. Krombholz, H. Hobel, M. Huber, and E. Weippl, "Advanced social engineering attacks," *Journal of Information Security and applications*, vol. 22, pp. 113-122, 2015.
2. S. Landini, "Ethical issues, cybersecurity and automated vehicles," *InsurTech: A Legal and Regulatory View*, pp. 291-312, 2020.
3. Q. Fan, X. Li, J. Li, Q. He, K. Wang, and J. Wen, "PA-Cache: Evolving learning-based popularity-aware content caching in edge networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 17, no. 2, pp. 1014-1027, Jun. 2020.
4. C. Ming, Y. Bingjie, and L. Xiantong, "Multi-tenant SaaS deployment optimisation algorithm for cloud computing environment," *International Journal of Internet Protocol Technology*, vol. 11, no. 3, pp. 152-158, 2018.
5. M. Lansley, N. Polatidis, S. Kapetanakis, K. Amin, G. Samakovitis, and M. Petridis, "Seen the villains: Detecting social engineering attacks using case-based reasoning and deep learning," 2019.
6. F. Salahdine and N. Kaabouch, "Social engineering attacks: A survey," *Future internet*, vol. 11, no. 4, p. 89, 2019.
7. D. Huang, D. Mu, L. Yang, and X. Cai, "CoDetect: Financial fraud detection with anomaly feature detection," *IEEE Access*, vol. 6, pp. 19161-19174, 2018.
8. Z. Chen, W. Dong, H. Li, P. Zhang, X. Chen, and J. Cao, "Collaborative network security in multi-tenant data center for cloud computing," *Tsinghua Science and Technology*, vol. 19, no. 1, pp. 82-94, 2014.
9. Q. Fan, X. Li, J. Li, Q. He, K. Wang, and J. Wen, "PA-Cache: Evolving learning-based popularity-aware content caching in edge networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 17, no. 2, pp. 1014-1027, Jun. 2020.
10. Y. Wang and V. Friderikos, "Energy-efficient proactive caching with multipath routing," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 2, pp. 487-499, Jun. 2021.
11. C.-J. Chung, T. Xing, D. Huang, D. Medhi, and K. Trivedi, "SeReNe: on establishing secure and resilient networking services for an SDN-based multi-tenant datacenter environment," in *2015 IEEE International Conference on Dependable Systems and Networks Workshops*, 2015: IEEE, pp. 4-11.
12. S. S. Gulati and S. Gupta, "A framework for enhancing security and performance in multi-tenant applications," *International Journal of Information Technology and Knowledge Management*, vol. 5, no. 2, pp. 233-237, 2012.
13. R. K. Thelagathoti, "Named data networking for content delivery," *J. Internet Technol.*, vol. 22, no. 5, pp. 123-134, Oct. 2021.
14. P. Jyothi, "Efficient Technique to optimize cloud storage in multi-Tenant Environment," *IJCERT ISSN (O): 2349-7084*, pp. 23-29, 2016.
15. Ngo, C., Demchenko, Y., De Laat, C.: 'Multi-tenant attribute-based access control for cloud infrastructure services', *J. Inf. Secur. Appl.*, 2016, 27, pp.65-84
16. Hussain, S.A., Fatima, M., Saeed, A., et al.: 'Multi-level classification of security concerns in cloud computing', *Appl. Comput. Inf.*, 2017, 13, (1), pp.57-65
17. Batista, B.G., Ferreira, C.H., Segura, D.C., et al.: 'A QoS-driven approach for cloud computing addressing attributes of performance and security', *Future Gener. Comput. Syst.*, 2017, 68, pp. 260-274
18. Gupta, D., Chakraborty, P.S., Rajput, P.: 'Cloud security using encryption techniques', *Int. J.*, 2015, 5, (2), pp. 425-429
19. P. A. Urla, G. Mohan, S. Tyagi, et al., "A novel approach for security of data in IoT environment," in *Computing and Network Sustainability*, 2019, pp. 251-259. [Online]. Available: <https://arxiv.org/>