Data Breaches in Cloud Storage: Risks and Prevention Strategies

Inam Ul Haq
Faculty of Computing and Information Technology, University of the Punjab, Lahore
bitf22m017@pucit.edu.pk
Google Scholar ORCID

Abstract

The shift from on-site infrastructure to cloud computing has fundamentally changed the global digital economy, offering unprecedented scalability and agility. However, this transition has also presented several critical security risks. This review paper provides an exhaustive analysis of data breaches within cloud storage environments, specifically focusing on Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). Through a systematic review of peer-reviewed literature, incident reports and technical documentation from 2015 to 2025, this study discusses cloud breaches, provides comparative analysis of cloud security architectures, prevention strategies and the future of cloud security.

Keywords: Cloud, Data, Security, Storage.

Introduction

Paradox: Scalability vs. Security

Cloud computing has grown from a naive technology into the backbone of the modern internet. According to NIST, it is a system that lets people easily access shared computing power and storage whenever they need it. It allows anyone to use powerful computers and large storage over the internet without owning the physical hardwares.[1] The transition from on-site data centers to cloud models, Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) has accelerated digital transformation across every sector. However, it has also created a paradox. Features that make the cloud economically attractive, centralizing data, sharing resources among many users, and quickly setting things up, also create concentrated risks. Large amounts of data in the cloud creates a high-value target for cybercriminals. In 2020 alone, cybercrime costs were estimated to approach USD 1 trillion globally, a figure made worse by the scale of cloud data breaches.[2]

Persistent Misconfigurations in the Cloud Settings

Despite a decade of maturation in cloud security tooling, the fundamental cause of data breaches remains the same, human error. While providers like AWS and Azure invest billions in securing physical data centers and hypervisors, they cannot prevent a customer from configuring a storage bucket as "public.". A 2025 study highlighted that cloud settings mistakes and incomplete access control systems are primary drivers of data exposure.[3] The complexity of these environments, where a single enterprise

may manage thousands of resources across multiple clouds, has outpaced the cognitive capacity of human administrators. This has led to what is called the "misconfiguration epidemic".

Rise of AI-Driven Threats

The threat landscape is not static. As defenders adopt automated tools, so too do attackers. The emergence of Generative AI and Large Language Models (LLMs) has armed threat actors with new capabilities. "Offensive AI" is now being used to generate sophisticated phishing campaigns to steal cloud credentials and even write code to exploit zero-day vulnerabilities.[4] The intersection of AI and cloud security represents the next frontier of cyber warfare, where AI agents are at the forefront of attacking cloud infrastructures.

Objectives and Scope of the Report

This report aims to provide a detailed report on cloud storage data breaches. The scope includes:

1. **Forensic Analysis:** A deep dive into the technical anatomy of major breaches to identify root causes beyond surface-level explanations.
2. **Architectural Comparison:** An evaluation of how AWS, Azure, and GCP implement security controls and where the gaps lie.
3. **Tooling Efficacy:** A critical comparison of CSPM tools (Prowler, ScoutSuite, Wiz) to guide practitioners in selection.
4. **Future Frameworks:** An analysis of Zero Trust and AI-driven defense mechanisms as the future of cloud security.

Methodology

This review paper uses Systematic Literature Review (SLR) methodology to gather and analyze data from multiple sources. The sources used include high-impact peer-reviewed journals, technical reports from cybersecurity firms (CrowdStrike etc), and official government standards (NIST, CISA).

Background

In traditional IT, security relied on a strong firewall to keep intruders out and users inside were mostly trusted. Cloud computing breaks this model. Now, identity is the key perimeter: anyone with valid credentials can access critical data from anywhere. This makes credential theft as serious as a physical break-in in the old system.

Cloud management relies heavily on APIs, creating a new risk called the ‘Management Plane.’ If an attacker steals management console credentials, they don’t need to hack each server. They can directly issue commands to copy data, create backdoors, or delete entire systems. Such a ‘control plane compromise’ is a worst-case scenario for cloud architects.[1]

A common cause of data breaches is ‘Shadow IT’, when departments or individuals use cloud resources without the central IT team knowing. The Sennheiser breach, discussed later, shows this: a team created a storage bucket for temporary use and then abandoned it. These ‘zombie’ assets stay active, unpatched and

unmonitored, making them easy targets for attackers scanning the internet.

Taxonomy of Cloud Threats

To understand breaches, we must classify the threats. Based on the literature, the following taxonomy defines the modern cloud threat landscape [8]:

Threat Category	Description	Mechanism of Action
Data Breaches	Unauthorized access and exfiltration of sensitive data.	Exploitation of public buckets, unencrypted snapshots, or compromised read-only credentials.
Misconfiguration	Incorrect setup of computing assets, leaving them vulnerable.	Publicly accessible S3 buckets, open security groups (0.0.0.0/0), disabled logging.
Insecure APIs	Vulnerabilities in the interfaces used to manage cloud resources.	Lack of authentication, insufficient input validation, over-permissive tokens.
Account Hijacking	Theft of user credentials to assume legitimate identities.	Phishing, credential stuffing, theft of API keys embedded in code.

Insider Threats	Malicious or negligent actions by authorized users.	Data exfiltration by disgruntled employees, accidental deletion by administrators.
Advanced Persistent Threats (APTs)	Sophisticated, long-term campaigns by nation-states or crime syndicates.	Lateral movement via supply chain attacks, "living off the land" using native tools.

Major Cloud Breaches: Case Studies

The most effective way to understand cloud risk is to analyze failure. Below are some case studies analyzing popular cloud failures.

The Capital One Breach (AWS)

The 2019 breach of Capital One is perhaps the most technically significant cloud breach to date. It did not involve a brute-force attack on a database or a stolen password in the traditional sense. Instead, it exploited the subtle interaction between application code and cloud infrastructure.

Mechanism of Failure:

Entry Point: The attacker found a Server-Side Request Forgery (SSRF) vulnerability in a Web Application Firewall (WAF) on an AWS EC2 instance. Ironically, the security tool (WAF) was misconfigured and could send requests to any URL.[11]

Metadata Exploit: Using the SSRF, the attacker forced the server to access the AWS Instance Metadata Service (IMDSv1) at http://169.254.169.254/latest/meta-data/iam/security-credentials/.[11] IMDSv1 didn't require authentication, so it returned the EC2 instance's temporary credentials (Access Key ID, Secret Key, and Token).

Privilege Escalation & Lateral Movement: The IAM role attached to the WAF had far too many permissions. Instead of just writing logs, it could read sensitive S3 buckets with customer credit card data.[11]

Exfiltration: Using these credentials, the attacker copied the sensitive data to their own storage using the aws s3 sync command.[11]

Implications:

This breach showed two key failures.

1) The SSRF vulnerability lets the attacker reach the cloud control plane.
2) Overly broad IAM permissions violated

the principle of Least Privilege, turning a small flaw into a major breach.

If the WAF role had been limited to writing logs only, the stolen credentials couldn't have been used to steal data.

Microsoft BlueBleed (Azure):

In 2022, the BlueBleed incident exposed data from over 65,000 entities in 111 countries. The breach was significant because it involved a misconfiguration in Microsoft's own infrastructure, showing that even top providers can make human errors.

Mechanism of Failure:

The leak came from a misconfigured Azure Blob Storage container. A threat intelligence firm, SOCRadar, found a public bucket containing SQL Server backups.[12] In Azure, containers can be set to Private, Blob (public read for individual files), or Container (full public access). This bucket was set to allow public access, meaning anyone with the link could view and download its contents.

Data at Risk:

The exposed data totaled 2.4 TB and included sensitive business documents such as Proof-of-Execution (PoE) files, Statements of Work (SoW), product offers, and signed customer contracts.[13] Independent analysis confirmed that personal data and intellectual property were present, despite Microsoft initially suggesting it was mostly metadata.[12]

Implications:

BlueBleed shows the "silent" danger of storage leaks. Unlike ransomware, which disrupts operations, data leaks can remain unnoticed for months or years. This bucket had files dating back to 2017..[13] The incident highlights the importance of continuous external attack surface management (EASM) to monitor what attackers can see.

Sennheiser (AWS):

The Sennheiser breach is a classic example of "data rot", data that is collected, stored, and then forgotten. It occurred in 2021.

Mechanism of Failure:

Researchers discovered an unsecured AWS S3 bucket containing 55 GB of data from 2015–2018. The bucket held customer names, emails, phone numbers, and home addresses.[10]

Root Cause:

The bucket was likely a leftover from an old cloud account or finished project. Because it was dormant, it escaped security audits. This Shadow IT resource remained publicly accessible for three years before being found.[14]

Implications:

The breach highlights the need for proper Cloud Resource Lifecycle Management. Organizations should archive or delete data that is no longer needed. A "clean" cloud is a secure cloud. Old data may go unmonitored but still has high value to attackers for phishing or identity theft.

Pfizer (Google Cloud)

In 2020, a breach at Pfizer showed the dangers of treating "backend" data as low-risk.

Mechanism of Failure:

An unprotected Google Cloud Storage bucket was found containing hundreds of customer support conversations about pharmaceutical prescriptions.[15] The data included mentions of drugs like Lyrica and Viagra, along with patient personal information (PII).

Root Cause:

The data came from an automated system (e.g., chatbot or support tool) and was stored for analysis. Developers assumed these logs were internal and low-risk, so the bucket wasn't secured like the main patient database.[16]

Implications:

This breach shows that logs are sensitive data too. Modern systems often store PII in logs just like in primary databases. Any automated system that saves debug data or conversation history must use the same strong IAM and encryption protections as core applications.

Comparative Analysis of Cloud Security Architectures

The "Big Three" providers; AWS, Azure, and GCP, all offer robust security capabilities, but their architectural philosophies and toolsets differ. Understanding these differences is crucial for multi-cloud security strategies.

Native Control Mapping

The following table synthesizes the primary security controls across the three platforms, derived from the analysis of CSA controls and platform documentation.[17]

Feature Category	Amazon Web Services (AWS)	Microsoft Azure	Google Cloud Platform (GCP)
Identity & Access	AWS IAM (Global), Cognito (App)	Microsoft Entra ID (formerly Azure AD)	Cloud IAM, Cloud Identity
Encryption (Key Mgmt)	AWS KMS, CloudHSM	Azure Key Vault	Cloud KMS, External Key Manager (EKM)

Threat Detection	Amazon GuardDuty	Microsoft Defender for Cloud	Security Command Center (SCC)
Vulnerability Mgmt	Amazon Inspector	Azure Defender for Cloud	Web Security Scanner, Container Analysis
Posture Management	AWS Security Hub, AWS Config	Azure Policy, Secure Score	Policy Intelligence, Recommender
Network Security	AWS Shield, WAF, Network Firewall	Azure DDoS Protection, Azure Firewall	Cloud Armor, VPC Service Controls
Compliance Reporting	AWS Artifact	Service Trust Portal	Compliance Reports Manager

Architectural Differences

AWS: The Modular Builder
AWS uses a modular, service-oriented architecture. Security is made up of separate tools: GuardDuty for threat detection, Macie for data classification, and Inspector for vulnerability scanning, all integrated in Security Hub. This flexibility is powerful but requires careful setup. AWS IAM's "deny-by-default" is strong, but complex JSON policies often cause mistakes.[19]

Azure: The Enterprise Ecosystem
Azure integrates well with existing enterprise systems. Microsoft Entra ID (Azure AD) unifies identities across cloud and Office 365 apps. Its Secure Score gives an easy-to-read view of security for executives.[20] However, complex role inheritance and subscription hierarchies can unintentionally grant permissions.

GCP: Data and Container Specialist
GCP focuses on data analytics and containers. VPC Service Controls create a security perimeter that can block data leaks even if IAM credentials are stolen, addressing attacks like the Capital One breach. Many security features, such as encryption at rest, are on by default, making them harder to disable compared to AWS.[20]

Prevention Strategies: Tools and Frameworks

Given that built-in cloud controls are often misconfigured, the industry relies on frameworks like Zero Trust and tools like CSPM to maintain security hygiene.

Zero Trust Architecture (ZTA):

According to NIST SP 800-207, Zero Trust assumes no implicit trust, every access request must be continuously verified. In cloud storage, it addresses the "soft center" problem of the cloud.[21]

Applying ZTA to Storage:

De-perimeterization: Networks are assumed hostile. Simply placing an S3 bucket in a VPC is not enough. Access must depend on identity, context, and device health, not just network location.[21]

Policy Enforcement Point (PEP): IAM acts as the PEP, checking every request (Read, Write, List) against dynamic policies.

Micro-segmentation: Data is divided into "micro-perimeters." For example, a user with Marketing access cannot see or access HR data, even in the same storage account.[22]

Implementation Challenges:

ZTA is hard to fully implement. While organizations adopt MFA and IAM, they often fail to use real-time context (like a login from a new device or risky country) in access decisions. Automated response to these signals is still limited.[23]

Cloud Security Posture Management (CSPM):

CSPM tools automatically detect cloud misconfigurations by continuously scanning the cloud API and comparing it to a "Golden Standard" (like CIS Benchmarks).

Impact:

Using CSPM can reduce misconfiguration-related cloud security incidents by up to 80%.[24] They provide continuous monitoring, alerting teams to changes, e.g., a bucket that was private yesterday but is public today.

Comparative Analysis of Security Tooling

The market for cloud security tools is divided between community-driven open-source projects and enterprise SaaS platforms. A comparative analysis reveals distinct use cases for each.

Prowler:

Prowler is an open-source tool and is used for hardening and compliance. It is highly effective for engineers who want to run a quick hardening check or integrate security gates into a CI/CD pipeline. Its ability to send findings to Security Hub makes it a viable component of an enterprise dashboard.[25]

ScoutSuite:

ScoutSuite is an open-source tool and is used for audit and assessment. It excels in the consulting use case. A security auditor can run it once, generate an HTML report, and visually walk a client through their risk surface. However, it lacks the "continuous" nature required for day-to-day operations.[29]

Wiz and Orca:

Wiz and Orca are both commercial. They represent the new generation of "CNAPP" (Cloud Native Application Protection Platforms). They solve the "Alert Fatigue" problem inherent in tools like Prowler. Prowler might flag 500 issues; Wiz will tell you which 5 actually matter because they represent a reachable attack path to sensitive data.

The Future of Cloud Security

The Human Element:

Tools like Wiz show why older security approaches failed: alert fatigue. CSPM dashboards can overwhelm teams with thousands of low-severity warnings, causing real threats to be missed. In the Sennheiser breach, the issue wasn't that tools didn't detect the risk, but that the warning was lost in the noise. Modern cloud security focuses on risk prioritization, identifying which misconfigurations could actually lead to a breach.

AI in Cloud Security:

The integration of AI into the threat landscape is accelerating.

Offensive AI: Attackers use AI to scale attacks. AI agents can automatically scan for public buckets, generate SQL injection queries, and craft convincing phishing emails to steal IAM credentials.[30]

Defensive AI: Security teams use AI to detect anomalies that rules-based systems miss. Machine learning can learn "normal" user behavior, like typical API calls or login locations and alert on unusual activity. Tools like AWS GuardDuty and Azure Sentinel use ML to spot threats such as impossible travel or data exfiltration.[31]

Economic and Regulatory Pressure

The cost of data breaches is rising, reaching record highs in 2024.[32] Strict regulations like GDPR, CCPA, and DORA are making organizations treat cloud security as a board-level concern. The BlueBleed incident shows that even exposing metadata can lead to serious reputational and regulatory consequences.

Conclusion

The security of cloud storage is not a solved problem; it is an evolving discipline. Our analysis of major breaches demonstrates that while the underlying cloud infrastructure is secure, the complexity of configuring it, specifically regarding Identity and Access Management remains a treacherous capability gap for many organizations. The "Shared Responsibility Model" often functions as a "Shared Failure Model" when customers lack the visibility and expertise to uphold their end of the bargain.

To mitigate these risks, a transition to a Zero Trust Architecture is mandatory. Trust must be removed from the network and placed in the identity and the data itself. This theoretical framework must be operationalized through robust tooling: CSPM solutions to catch misconfigurations, Graph-based analysis to prioritize risk, and AI-driven detection to catch the sophisticated threats of tomorrow.

Ultimately, the most effective prevention strategy is simplification. By automating the lifecycle of data, decommissioning shadow IT, and enforcing "security-by-design" through Infrastructure as Code, organizations can reduce the surface area for human error. As we move into the AI era, the speed of attack will increase; our defenses must become equally autonomous, predictive, and resilient.

Future Research Directions

The field of cloud storage security is ripe for further investigation. Key areas for future research include:

1. **Automated Self-Healing Systems:** Research into "autonomic" security systems that can not only detect a public bucket but automatically revert it to private status without human intervention, while minimizing business disruption.
2. **AI-Specific Threat Modeling:** Developing standardized frameworks (beyond MITRE ATT&CK) specifically for AI-driven attacks on cloud control planes.
3. **Cryptographic Innovations:** Exploring the commercial viability of Homomorphic Encryption in public clouds, which would allow data to be processed without ever being decrypted, potentially neutralizing the risk of data breaches entirely.
4. **SaaS-to-SaaS Supply Chain Risk:** Investigating the "App-to-App" integration layer, where third-party SaaS tools granted OAuth access to core cloud storage represent a massive, unmonitored shadow access network.

Works cited

1. (PDF) Information Security in Cloud Computing: A ... - ResearchGate, accessed November 19, 2025, https://www.researchgate.net/publication/337007232_Information_Security_in_Cloud_Computing_A_systematic_Literature_Review_and_Analysis
2. Cyber risk and cybersecurity: a systematic review of data availability - PMC - NIH, accessed November 19,

2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8853293/
3. (PDF) DATA BREACHES IN CLOUD ENVIRONMENTS CAUSES ..., accessed November 19, 2025, https://www.researchgate.net/publication/391270943_DATA_BREACHES_IN_CLOUD_ENVIRONMENTS_CAUSES_CONSEQUENCES_AND_PREVENTION
4. CrowdStrike 2025 Threat Hunting Report: AI Becomes a Weapon and a Target, accessed November 19, 2025, https://www.crowdstrike.com/en-us/blog/crowdstrike-2025-threat-hunting-report-ai-weapon-target/
5. Forensic Investigation, Challenges, and Issues of Cloud Data: A ..., accessed November 19, 2025, https://www.mdpi.com/2073-431X/13/8/213
6. nccgroup/ScoutSuite: Multi-Cloud Security Auditing Tool - GitHub, accessed November 19, 2025, https://github.com/nccgroup/ScoutSuite
7. Top 5 Open-Source Cloud Security Tools You Need to Know in 2025, accessed November 19, 2025, https://www.k9security.io/posts/2025/04/top-5-open-source-cloud-security-tools-you-need-to-know-in-2025/
8. (PDF) Taxonomy for Identification of Security Issues in Cloud Computing Environments, accessed November 19, 2025, https://www.researchgate.net/publication/308691801_Taxonomy_for_Identification_of_Security_Issues_in_Cloud_Computing_Environments
9. Top Cloud Security Threats for 2025 - Commvault, accessed November 19, 2025, https://www.commvault.com/explore/top-cloud-security-threats
10. Sennheiser Exposed Personal Data of 28,000 Customers Online | SecureReading, accessed November 19, 2025, https://securereading.com/sennheiser-exposed-personal-data/
11. A Technical Analysis of the Capital One Cloud Misconfiguration | CSA, accessed November 19, 2025, https://cloudsecurityalliance.org/blog/2019/08/09/a-technical-analysis-of-the-capital-one-cloud-misconfiguration-breach
12. Sensitive Data of 65,000+ Entities in 111 Countries Leaked due to a ..., accessed November 19, 2025, https://socradar.io/sensitive-data-of-65000-entities-in-111-countries-leaked-due-to-a-single-misconfigured-data-bucket/
13. Microsoft confirms customer data leak but disputes scope - The Register, accessed November 19, 2025, https://www.theregister.com/2022/10/20/microsoft_data_leak_socradar/
14. Sennheiser exposed personal data of 28,000 customers with leaky S3 bucket | IT Pro - ITPro, accessed November 19, 2025, https://www.itpro.com/cloud/amazon-s3/361864/sennheiser-exposed-data-28000-customers-aws-s3-bucket
15. Pfizer suffers huge data breach on unsecured cloud storage - Pf Media, accessed November 19, 2025, https://pf-media.co.uk/news/pfizer-suffers-huge-data-breach-on-unsecured-cloud-storage/
16. Learning to Prevent Healthcare Data Breaches | Pfizer Cloud Breach - Sonrai Security, accessed November 19, 2025, https://sonraisecurity.com/blog/pfizer-suffers-huge-data-breach-on-unsecured-cloud-storage/
17. Analysis of Cloud Security Controls in AWS, Azure, and Google Cloud, accessed November 19, 2025, https://repository.stcloudstate.edu/cgi/viewcontent.cgi?article=1149&context=msia_etds
18. What's the Difference Between AWS vs. Azure vs. Google Cloud? - Coursera, accessed November 19, 2025, https://www.coursera.org/articles/aws-vs-azure-vs-google-cloud
19. AWS vs Azure vs Google: Cloud Services Comparison - Varonis, accessed November 19, 2025,

https://www.varonis.com/blog/aws-vs-azure-vs-google
20. Comparing AWS, Azure, GCP | DigitalOcean, accessed November 19, 2025, https://www.digitalocean.com/resources/articles/comparing-aws-azure-gcp
21. Zero Trust Architecture - NIST Technical Series Publications, accessed November 19, 2025, https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.SP.800-207.pdf
22. SP 800-207A, A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Cloud Environments - NIST Computer Security Resource Center, accessed November 19, 2025, https://csrc.nist.gov/pubs/sp/800/207/a/final
23. A Systematic Literature Review on the Implementation and Challenges of Zero Trust Architecture Across Domains - MDPI, accessed November 19, 2025, https://www.mdpi.com/1424-8220/25/19/6118
24. What Is CSPM (Cloud Security Posture Management)? | Proofpoint US, accessed November 19, 2025, https://www.proofpoint.com/us/threat-reference/cloud-security-posture-management
25. prowler-cloud/prowler: Prowler is the Open Cloud Security ... - GitHub, accessed November 19, 2025, https://github.com/prowler-cloud/prowler
26. Cloud Security Auditing by Comparing AWS Config, Prowler, and Scout suite, accessed November 19, 2025, https://cloudbridgeusa.com/cloud-security-auditing-by-comparing-aws-config-prowler-and-scout-suite/
27. A Comparative Study of Native vs. Third-Party CSPM Tools: Analysis of AWS S, accessed November 19, 2025, https://journals.indexcopernicus.com/publication/4665795
28. Cloud Vulnerability Assessment Tools with Multi-Cloud Support? : r/cybersecurity - Reddit, accessed November 19, 2025, https://www.reddit.com/r/cybersecurity/comments/1bka4yc/cloud_vulnerability_assessment_tools_with/
29. SECURITY IN CLOUD ENVIRONMENTS - Theseus, accessed November 19, 2025, https://www.theseus.fi/bitstream/10024/884892/2/Kesseli_Tiina.pdf
30. GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools | Google Cloud Blog, accessed November 19, 2025, https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools
31. Cloud Security Posture Management (CSPM) Software - Amazon AWS, accessed November 19, 2025, https://aws.amazon.com/marketplace/solutions/security/cloud-security-posture-management
32. 50+ Cloud Security Statistics in 2025 - SentinelOne, accessed November 19, 2025, https://www.sentinelone.com/cybersecurity-101/cloud-security/cloud-security-statistics/