American Journal of Interdisciplinary Research and Innovation (AJIRI)

ISSN: 2833-2237 (ONLINE)

VOLUME 4 ISSUE 4 (2025)

PUBLISHED BY

E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Interdisciplinary Research and Innovation (AJIRI)

Volume 4 Issue 4, Year 2025
ISSN: 2833-2237 (Online)
DOI: https://doi.org/10.54536/ajiri.v4i4.6018
https://journals.e-palli.com/home/index.php/ajiri

Securing the Cloud: Defenses for Modern Threats

Jehanzaib Jamil[1*]

Article Information

Received: August 24, 2025

Accepted: September 29, 2025

Published: December 10, 2025

Keywords

CIEM, Cloud Security, CSPM, Data Loss Prevention, Egress Control, Kubernetes Security, Runtime Detection, Software Supply Chain, Tokenisation, Zero Trust

ABSTRACT

Cloud adoption across IaaS, PaaS, containers, serverless, and SaaS has widened the attack surface, shifting risk toward identity misuse, misconfiguration, supply chain weaknesses, and data exfiltration; perimeter-only controls no longer suffice. Present a compact threat-to-defence taxonomy for modern cloud environments, grounded in 2020+ evidence and mapped to practical controls. Narrative review of peer-reviewed studies, standards, and agency guidance from 2020 onward, with clearly labelled grey literature; extracted threat, defence, control mapping, metrics, and caveats, then synthesised into four defence families. Across sources, identity-first controls with least privilege with short-lived credentials and CIEM reduces blast radius; CSPM and policy as code curb public exposures and drift; image signing, admission policies, and tuned runtime detection improve container security; egress allow-lists, tokenisation, and immutable backups constrain data loss, provided rollout is centralised and automated to limit friction and false positives. A targeted set of identity-first, posture, runtime, and data or egress controls provides long-term risk reduction while outperforming tool complexity.

INTRODUCTION

The use of cloud has transformed the way organisations develop and run systems in infrastructure, platform, containers and serverless functions, as well as software as a service. Elasticity, operated services and worldwide access have speedy delivery and at the same time have enlarged the attack area and dispersed accountability in security arrangements and watchfulness. The latest surveys and reviews are united in one voice. The leading root causes of cloud setting incidents are misconfiguration, identity abuse, poor hygiene of the software supply chain and unmanaged data flows. Meanwhile, pure perimeter controls cannot be trusted in the manner in which cloud workloads today communicate and authenticate (AL-QTIEMAT & AL-ODAT, 2024; Ali *et al.*, 2025; Alouffi *et al.*, 2021; Tahirkheli *et al.*, 2021). Similarly, the external attack surface has been made more dynamic and can now be explored by automation raising the stakes associated with preventive controls and ongoing posture control (Gelernter *et al.*, 2024).

Having multi-cloud and hybrid strategies is further complicated by the change in operations. Even though multi-cloud has the ability to enhance resilience and flexibility of the vendor, it increases identity silos, policy engines, logging formats, and default settings that need to be aligned to generate a consistent least privilege posture (Ali *et al.*, 2025; Lata & Kumar, 2025; Potla, 2025; Sivaseelan, 2024). Comparative scenarios of public, private and hybrid deployments reveal that controls that work well in one setting may silently fail in another where the default and terminology are different and thus make governance and measuring harder (Lata & Kumar, 2025). Risk management guidance tailored for enterprise cloud environments argues for explicit control selection and sequencing rather than tool accumulation, since uncoordinated adoption often increases alert noise and produces configuration drift across tenants and accounts (Hegde *et al.*, 2023; Oyeniyi & Oyeniran).

Identity has become the practical perimeter in cloud systems. Overbroad roles, long-lived credentials, and uncontrolled service accounts increase lateral movement opportunities far more than traditional network design implies. Analyses of entitlement proper sizing and cloud infrastructure entitlement management report measurable reductions in excessive privileges when organisations adopt graph-informed reviews, just-in-time elevation, and short-lived credentials (Fisher, 2025; Neelakandhan *et al.*, 2022). Federated identity and standardised authentication flows help unify access across providers, y*et al*so introduce failure modes that must be considered, such as token replay, misconfigured trust relationships, or incomplete conditional access enforcement (Almeida *et al.*, 2024; Maidine & El-Yahyaoui, 2023). Zero Trust guidance from NIST codifies an identity-first model that treats every request as untrusted until policy is satisfied with context, device, and workload signals, which aligns well with the realities of distributed cloud platforms (Rose, 2022; Stafford, 2020). Critical analyses of Zero Trust emphasise that credible enforcement points and telemetry must accompany architectural principles, or they risk becoming slogans without measurable outcomes . At the network or edge layer, Secure Access Service Edge is increasingly used to bridge remote access, policy enforcement, and identity, especially for distributed teams and multi-cloud connectivity. However, the evidence base is still maturing and often case study driven (Adewale, 2023; Bajpai, 2022;

[1] Accenture, Riyadh, Saudi Arabia
[*] Corresponding author's e-mail: Jehanzaib.jamil@gmail.com

Bashi *et al*., 2025; Dommari & Khan, 2023; Lee & Park, 2025; Maddali, 2025).

Container and Kubernetes have replaced workload security. Container images are able to focus supply chain risk in the form of base images, package managers and build pipelines. Meanwhile, cluster control planes introduce a policy and role-based access control level which needs to be strengthened. Agency hardening advice gives prescriptive guidelines in Kubernetes, such as privilege escalation limits, admission control, as well as runtime monitoring, which have become a de facto starting point in numerous teams (CISA, 2022). Recent research in the publicity of hardening and runtime discovery demonstrates that the policy of admission, signed provenance, and tuned behavioural rules decreases the success of container runtime abuse. Nonetheless, one such adoption barrier is false positives, which need to be fine-tuned after iteration and improved context by building pipelines and software bills of materials (Cesarano & Natella, 2025; Eldjou *et al*., 2023; Morić *et al*., 2025; Nguyen *et al*., 2023; Yilmaz & Harding, 2024). Greater discussion of the software supply chain is no longer a theoretical issue, but rather an operationalised idea (through provenance attestations and image signing) that can be enforced by cluster admission controllers, introducing build-time guarantees at runtime decision points (Nguyen *et al*., 2023; Yilmaz & Harding, 2024).

A different risk profile is formed by serverless and API centric architecture. Function granularity promotes small, narrow units of compute, but also invites developers to make broad permissions in the name of convenience and this compromises least privilege. These designs put API security at their core, as functions are normally exposed either by gateways or event sources and need to provide strong authentication, authorisation and rate limiting. In the 2023 version of the OWASP API Security Top 10 documents, modern failure modes are shown as direct cloud workload failures, such as broken object-level authorisation, overly revealing data, and unsafe third-party API use, which are magnifying factors when used together with cloud identity and data access (Mallick & Nath, 2024; OWASP; Türetken, 2024). The works that analyse serverless and API protections always suggest per-function roles, schema validation, and gateway-level protections to ensure blast radius limits and curb abuse and there is a reasonable acceptance of developer ergonomics and trade-offs against cold starts (Mallick & Nath, 2024; Türetken, 2024).

Egress and data protection play a significant role in alleviating the effects in case of the prevention failure. The existing work on data exfiltration and ransomware in the cloud highlights the ability to use egress allowlists, well-managed keys, tokenisation and immutable backups as effective measures to decrease the probability and the impact of loss incidents (Lal *et al*., 2022; Sahar Saeed *et al*., 2025; Vidhya, 2024). The cryptography capabilities of workloads on a cloud scale are still developing. Securing the protection of data in use is a direction provided by the concept of confidential computing via trusted execution environments and attestation. Nevertheless, operational maturity and performance trade-offs should be measured on the production scale (Chen, 2023; Dhar *et al*., 2024). Revocation and integrity-guaranteeing attribute-based encryption has been used to work with cloud storage contexts. Though this is not a universal solution, it expands the range of enforceable policies seen outside of flat data at rest controls (Ge *et al*., 2021).

Cloud adversary behaviour is a vocabulary that needs a method of measuring coverage, and cloud detection and response. This structure is offered in the MITRE ATT&CK Enterprise Cloud matrices related to infrastructure, identity providers, office suites, and software-as-a-service, which allow teams to map detections to particular techniques systematically and rationally about gaps (Dieterich; Tran, 2023). Empirical studies on anomaly detection and cloud telemetry have suggested that range-based and semi-supervised techniques are capable of identifying operationally significant deviations. Still, success depends on high-quality features and careful normalisation to avoid simple resource consumption spikes being misclassified as attacks (Deka *et al*., 2022; Tatineni, 2023). In practice, organisations benefit from aligning detection content with ATT&CK while instrumenting provider-native telemetry pipelines and normalising logs to reduce the friction of cross-cloud investigations (Dieterich *et al*., 2023).

The cloud security frameworks literature offers a way to connect threats, defences, and controls into a consistent program. Comparative and synthesis work highlights the importance of selecting a small number of authoritative frameworks and using them as mapping backbones to avoid duplication and confusion, for example, referencing Zero Trust concepts for identity and access decisions while applying control catalogues for posture management and compliance checks (Chandra, 2020; Hegde *et al*., 2023). Measurement and governance perspectives argue for compact metric sets that can be automated, such as excess privilege rates, misconfiguration density, detection coverage against ATT&CK techniques, mean time to detect, and policy drift, so that progress can be demonstrated without manual audits that do not scale in cloud environments (Colotti, 2023; Fisher, 2025; Hegde *et al*., 2023; Jimmy, 2023; Neelakandhan *et al*., 2022). Case-driven guidance for cloud networking emphasises that performance and flexibility goals do not need to conflict with Zero Trust when segmentation and identity-aware policy enforcement are designed together. However, it warns that partial implementation commonly leaves legacy paths that undermine the model (Ahmadi, 2024; James, 2021; Lee & Park, 2025).

Taken together, the recent literature and guidance suggest a pragmatic thesis. Durable risk reduction in cloud environments comes from a focused set of defences that match how cloud workloads authenticate, run, and move data. Identity first controls, continuous posture management with preventive guardrails, signed provenance

with tuned runtime detection for containers, and strong data or egress governance address the dominant failure modes reported since 2020 (Ali *et al*., 2025; Alouffi *et al*., 2021; CISA, 2022; Fernandez & Brazhuk, 2024; OWASP; Rose, 2022; Singh *et al*., 2024; Stafford, 2020; Tahirkheli *et al*., 2021). This paper aims to synthesise that evidence into an actionable taxonomy, map it to provider native and open tooling that practitioners already operate, and present a compact set of metrics and a staged rollout plan that can be executed within usual engineering constraints. This approach favours clarity over tool accumulation and provides a common language for architects, security engineers, and developers to coordinate improvements across multi-cloud estates.

Objectives

This paper's objective is to deliver a compact, evidence-grounded threat to defence taxonomy for modern cloud environments that practitioners can implement and measure with minimal overhead. The taxonomy focuses on the failure patterns most frequently reported since 2020, including identity misuse, misconfiguration, software supply chain weaknesses, and data exfiltration, and aligns defences with how cloud workloads actually authenticate, execute, and move data (Ali *et al*., 2025; Alouffi *et al*., 2021; Tahirkheli *et al*., 2021). To anchor terminology and scope, the taxonomy is normalised to widely adopted references, including the MITRE ATT&CK Enterprise Cloud matrices for adversary technique coverage, NIST Zero Trust guidance for access decision models, the NSA or CISA Kubernetes hardening baseline, and the OWASP API Security Top 10 for API centric risks (CISA, 2022; Dieterich; OWASP; Rose, 2022; Stafford, 2020). Given the operational realities of hybrid and multi-cloud estates, the objective includes portability across providers, allowing recommended controls to be applied consistently without vendor lock-in (Ali *et al*., 2025; Lata & Kumar, 2025; Potla, 2025; Sivaseelan, 2024).

Operationalisation is central to the objective. Each defence in the taxonomy is mapped to concrete provider-native controls on AWS, Azure, and GCP, as well as widely used open tools, allowing teams to move directly from concept to configuration and runtime enforcement (CISA, 2022; Hegde *et al*., 2023; Morić *et al*., 2025). To enable evaluation without heavy audits, a compact metric dictionary is specified, covering excess privilege rate, misconfiguration density, detection coverage against ATT&CK techniques, mean time to detect and respond, exfiltration success rate, and policy drift (Deka *et al*., 2022; Dieterich; Fisher, 2025; Hegde *et al*., 2023; Neelakandhan *et al*., 2022). Finally, the objective includes a two quarter rollout plan that sequences quick posture wins and sustainable practices such as CIEM based proper sizing, policy as code in CI, image signing and provenance, tuned runtime detection, egress allow lists, and immutable backups (CISA, 2022; Fisher, 2025; Jimmy, 2023; Morić *et al*., 2025; Nguyen *et al*., 2023; Singh *et al*., 2024).

MATERIALS AND METHODS

This study employs a narrative review methodology tailored to the fast-paced world of cloud security practice. The review concentrates on literature and guidance published from 2020 onward to reflect current threat patterns and controls reported in surveys, systematic reviews, and domain syntheses (Ali *et al*., 2025; Alouffi *et al*., 2021; Tahirkheli *et al*., 2021), alongside standards and agency baselines that shape operational practice (CISA, 2022; Dieterich; OWASP; Rose, 2022; Stafford, 2020). Given the heterogeneity of evaluation methods across cloud security studies, the goal is not meta-analysis but structured synthesis: we normalise terminology, extract comparable fields, and assemble findings into the four defence families used throughout the paper, namely identity first controls, posture and policy as code, runtime and supply chain safeguards, and data and egress governance. This design choice aligns with how cloud workloads authenticate, execute, and move data and with how practitioners actually configure platforms (CISA, 2022; Hegde *et al*., 2023).

Sources were identified through iterative searches in Google Scholar and university library indexes using Boolean patterns that combine threat classes and defences with cloud-specific nouns. Examples include queries pairing IAM and least privilege with CIEM, CSPM and policy as code with misconfiguration themes, Kubernetes runtime or admission control with hardening guidance, serverless or API security with gateway protections, and data exfiltration with DLP or tokenisation. For detection and response, we used ATT&CK Cloud terms to locate coverage mappings and anomaly detection evaluations (Deka *et al*., 2022; Dieterich; Fisher, 2025; Hegde *et al*., 2023; Neelakandhan *et al*., 2022). Searches were time-bound to 2020 onward and then expanded via forward snowballing using cited links from anchor papers and standards. Standards and agency guidance such as NIST SP 800-207, the NIST ZTA planning guide, the NSA or CISA Kubernetes hardening document, and the OWASP API Security Top 10 were treated as canonical anchors that define scope and vocabulary rather than effect estimates (CISA, 2022; OWASP; Rose, 2022; Stafford, 2020).

Inclusion required that a source explicitly describe a defence relevant to cloud threats and provide one of the following: an empirical evaluation or case study, a structured standard or baseline that is widely adopted, or a design or implementation analysis with precise technical detail. We included peer-reviewed articles, full conference papers, surveys, standards, and selected theses or preprints when they contributed novel technique descriptions or systematised practice, which are flagged as emerging evidence in our Discussion (Ahmadi, 2024; Ali *et al*., 2025; Alouffi *et al*., 2021; Hegde *et al*., 2023; Morić *et al*., 2025; Sinan *et al*., 2025; Tahirkheli *et al*., 2021). Exclusion criteria removed opinion pieces without methods, vendor whitepapers lacking technical transparency, and older pre-cloud security taxonomies that do not map to cloud native

controls. When multiple versions of agency guidance existed, we preferred the latest stable PDF and recorded its date to support reproducibility (CISA, 2022; OWASP; Rose, 2022; Stafford, 2020).

Screening proceeded in two passes. First, titles and abstracts were filtered for relevance to at least one defence family. Second, full texts were reviewed to confirm that the work contained sufficient detail to extract fields and to permit mapping to provider controls or control catalogues. For each included item, we extracted the following: threat class as framed by the source and, where possible, the closest ATT&CK Cloud technique, the defence mechanism or control pattern, any control mapping to well-known catalogues or provider services, the evaluation context and method, the primary metrics reported, key findings, and stated limitations. This schema supports Tables 1 through 3 and allows us to discuss outcomes and trade-offs in consistent terms across heterogeneous studies (Deka *et al*., 2022; Dieterich, 2025; Hegde *et al*., 2023; Morić *et al*., 2025; Neelakandhan *et al*., 2022).

Quality and bias were addressed in two ways suited to a narrative review. First, we gave analytical weight to peer-reviewed and standard-setting sources when synthesising generalizable claims about what works, using grey literature primarily to illustrate emerging practice or to provide implementation detail (Ali *et al*., 2025; Alouffi *et al*., 2021; CISA, 2022; Hegde *et al*., 2023; Rose, 2022; Stafford, 2020; Tahirkheli *et al*., 2021). Second, we preferred studies that reported concrete, automatable metrics such as reductions in excessive privileges, misconfiguration counts, detection coverage against ATT&CK techniques, mean time to detect or respond, exfiltration success under simulation, and policy drift, all of which are defined in our metric dictionary to reduce ambiguity (Colotti, 2023; Deka *et al*., 2022; Dieterich; Fisher, 2025; Hegde *et al*., 2023; Jimmy, 2023; Neelakandhan *et al*., 2022). Where results conflicted or where only qualitative guidance existed, we present the range of findings and identify operational caveats rather than forcing consensus. Synthesis proceeded in three steps. We first clustered defences by family and aligned each to concrete provider controls on AWS, Azure, and GCP together with widely used open tools, producing an implementation-oriented mapping that practitioners can act on without a single vendor prescription (CISA, 2022; Hegde *et al*., 2023; Morić *et al*., 2025). We then summarised evidence directionality per family, favouring statements that multiple sources support and that map cleanly to the metric dictionary. Finally, we integrated standards and baselines to ground our recommendations and avoid ambiguous terminology. For example, we utilised NIST ZTA language for access decisions, ATT&CK for technique coverage, NSA or CISA guidance for Kubernetes hardening, and the OWASP API taxonomy for API risks (CISA, 2022; Rose, 2022; Stafford, 2020).

Reproducibility is addressed by documenting search themes, time bounds, and the extraction schema, and by normalising nouns to ATT&CK and widely recognised provider services so readers can repeat or extend the review with minimal translation effort. Limitations arise from the diversity of evaluation settings, the presence of grey literature in some fast-evolving areas such as CIEM and policy as code adoption, and the reliance on agency baselines for Kubernetes and API security, where controlled experiments are rare (CISA, 2022; Fisher, 2025; Morić *et al*., 2025; OWASP; Rose, 2022; Stafford, 2020). These constraints motivate the paper's focus on actionable mappings, compact metrics, and a staged rollout plan that can be measured and iterated in production, which we argue is the most helpful contribution for organisations operating in multi-cloud environments (Ali *et al*., 2025; Hegde *et al*., 2023; Morić *et al*., 2025).

RESULTS AND DISCUSSION

The literature from 2020 onward converges on a practical pattern: durable risk reduction in cloud systems is achieved by a small, consistent set of defences aligned to how cloud workloads authenticate, execute, and move data rather than to static perimeters (Ali *et al*., 2025; Alouffi *et al*., 2021; Tahirkheli *et al*., 2021). Across sources, identity first controls with least privilege and entitlement proper sizing, continuous posture management with preventive guardrails, workload runtime and supply chain hardening, and data or egress governance repeatedly correlate with fewer publicly exposed assets, smaller blast radius during compromises, and improved recovery characteristics (Ali *et al*., 2025; Fernandez & Brazhuk, 2024; Hegde *et al*., 2023; Morić *et al*., 2025). These findings hold across public, private, and hybrid deployments, although operational maturity and platform defaults introduce meaningful variation that must be addressed with provider-specific implementations (Hegde *et al*., 2023; Lata & Kumar, 2025). To make these results directly usable, this section presents a concise threats to defences matrix, a provider and tool mapping that turns concepts into concrete handles, and a compact metric dictionary that standardises measurement across studies and deployments.

Identity and access management remains the hinge on which most cloud incidents turn. Research on entitlement management also indicates a countable decline in unwarranted privileges, and horizontal mobility tracks when organisations embrace the use of graph-based reviews, temporal credentials and just-in-time promotion (Fisher, 2025; Neelakandhan *et al*., 2022). These are in line with Zero Trust principles where all requests are treated as untrusted until a policy decision is made, which is met with identity, device, and workload signals (Rose, 2022; Stafford, 2020). The empirically found texture also differs with the environment, yet directionality is universal: projects that are least privileged with periodic right-sizing and stronger authentication will exhibit a significant contraction in blast radius when compromised (Fisher, 2025; Neelakandhan *et al*., 2022). Case-based advice also indicates that federated identity, as well as standardised authentication flows enhance management, as long as token replay, misconfigured trust relationships,

and gaps in conditional access are considered explicitly (Almeida *et al.*, 2024; Maidine & El-Yahyaoui, 2023). Concisely, identity-first controls lead to the minimisation of the space within which attackers can navigate, as long as the organisation invests in proper identity graphs and operation reviews (Fisher, 2025; Neelakandhan *et al.*, 2022).

Misconfiguration is still the fastest path to loss, particularly in multi-cloud estates where policy engines and default settings differ across providers (Ali *et al.*, 2025; Lata & Kumar, 2025; Potla, 2025; Sivaseelan, 2024). The results show that continuous posture management paired with deny by default templates and policy as code in continuous integration yields sharp reductions in public exposures and configuration drift (Colotti, 2023; Hegde *et al.*, 2023; Jimmy, 2023). Conference and journal reports that examine posture tools consistently note that preventive controls at deploy time outperform purely detective scans after deployment, because they halt drift before it enters production (Hegde *et al.*, 2023; Jimmy, 2023; Singh *et al.*, 2024). That said, posture tooling can generate alert fatigue when policies are introduced without staged tuning or when exception handling is ad hoc, leading to reintroduction of risk through policy sprawl. Effective programmes therefore standardise a small corpus of policies, integrate checks into pipelines, and generate periodic drift reports to keep exceptions visible and time-bound (Colotti, 2023; Hegde *et al.*, 2023; Jimmy, 2023).

Container and Kubernetes security dominate the workload discussion. Agency hardening baselines emphasise admission control, privilege restrictions, and runtime monitoring as first principles that teams should adopt regardless of vendor (CISA, 2022). Recent research on hardening the Kubernetes attack surface and on compliance of control planes and workloads indicates that combining admission policies with image signing and provenance attestations reduces successful injection of untrusted images and improves the catch rate for runtime misuse. However, tuning is essential to control noise (Cesarano & Natella, 2025; Eldjou *et al.*, 2023; Morić *et al.*, 2025; Yilmaz & Harding, 2024). Runtime detection based on behavioural rules or eBPF typically surfaces lateral movement attempts, misuse of host namespaces, or container escapes, but out-of-the-box rulesets often over-alert until enriched with build-time context, such as software bills of materials or provenance metadata (Eldjou *et al.*, 2023; Nguyen *et al.*, 2023; Yilmaz & Harding, 2024). The strongest practical pattern to emerge is end-to-end: sign what you build, verify at admission, and monitor at runtime. When those three layers are present, studies report fewer successful escalations and faster operator response (CISA, 2022; Eldjou *et al.*, 2023; Morić *et al.*, 2025; Yilmaz & Harding, 2024).

Serverless and API centric architectures exhibit a different set of failure modes that revolve around overprivileged functions and broken authorisation at the API layer. The 2023 OWASP API Security Top 10 provides a contemporary taxonomy of these errors that maps directly onto cloud gateway configurations and per-function identity policies (OWASP). Theses and applied reports recommend per-function least privilege, rate limiting, schema validation, and careful JWT handling as baseline mitigations, with an emphasis on generating per-function identity policies from trace data to reduce developer friction (Mallick & Nath, 2024; Türetken, 2024). Case evidence suggests that organisations that deploy API gateway protections and align function permissions to narrowly scoped roles experience fewer abuse paths and easier incident containment, albeit with trade-offs in cold start latencies and policy sprawl if not automated (Mallick & Nath, 2024; Türetken, 2024). The results favour the adoption of gateway-level controls first, followed by incremental least privilege policies derived from observed call graphs to balance security and ergonomics.

Data exfiltration and ransomware-style events in cloud contexts reveal the importance of egress and backup governance. Analytical reviews and practice-oriented reports demonstrate that egress allow lists, combined with tokenisation and immutable backups, reduce both the likelihood of data loss and the time to restore in the event of destructive events (Lal *et al.*, 2022; Sahar Saeed *et al.*, 2025; Vidhya, 2024). Key management and vault isolation further reduce failure domains, and audit-backed object locks or vault locks provide assurances against administrative deletion during an attack (Sahar Saeed *et al.*, 2025; Vidhya, 2024). The persistent challenge is operational friction: DLP and egress controls can break legitimate integrations if introduced without careful staging. The studies that report positive results consistently used phased rollouts, domain category allow lists, and gradually tightened DLP dictionaries to keep false positives manageable while improving coverage (Lal *et al.*, 2022; Sahar Saeed *et al.*, 2025; Vidhya, 2024).

Defences against volumetric or application-layer DDoS and against unauthorised compute use are well served by managed services offered by the major providers, complemented by anomaly-based detection of runtime or billing anomalies (Nguyen & Debroy, 2022). Reports on moving target defences for DDoS show promise in specific contexts, but the mainstream operational playbook still centres on provider-managed DDoS mitigation and caching or CDN strategies to absorb and deflect traffic (Nguyen & Debroy, 2022). Cost shielding during extreme events remains a business risk even when mitigation succeeds, which motivates contractual and architectural controls in addition to technical measures.

Detection and response benefit from a shared adversary vocabulary and a way to measure coverage across the diverse services that constitute modern cloud estates. The MITRE ATT&CK Enterprise Cloud matrices provide this anchor across infrastructure, identity providers, office suites, and software-as-a-service (Dieterich). Empirical work on anomaly detection in cloud telemetry demonstrates that semi-supervised and range-based

methods can surface operationally meaningful deviations, provided features are carefully selected and normalised so that expected elasticity and bursty workloads do not produce false signals (Deka *et al.*, 2022; Tatineni, 2023). Programmes that align detections to ATT&CK techniques and normalise provider native logs report more precise coverage accounting and more tractable investigations across clouds, even when absolute detection rates vary by environment (Hegde *et al.*, 2023). The practical implication is straightforward: choose a technique vocabulary, map what you have, and improve coverage iteratively against a metric that teams can automate.

The results above are distilled into three artefacts designed for immediate use. Table 1 presents a threats to defences matrix that links dominant threat classes to practical controls, expected outcomes, and operational caveats. Table 2 turns those defences into cloud-specific knobs by aligning each category to native services on AWS, Azure, and GCP, together with widely used open tools. Table 3 defines a compact metric dictionary that maintains consistency and facilitates automation. Together, these tables let practitioners move from narrative findings to configuration choices and measurable progress without tool sprawl.

Table 1: Threats to Defences Matrix

Threat class	Typical tactics or failure modes	Primary defenses	Concrete patterns or controls	Outcome if applied well	Gotchas to manage
IAM misuse and privilege escalation	Token theft, key leakage, overbroad roles, stale access	Least privilege, CIEM, MFA, just-in-time access	Role minimisation, entitlement right-sizing, short-lived creds	Smaller blast radius, reduced lateral movement	Requires an identity graph and periodic reviews
Misconfiguration of cloud resources	Public storage, open databases, permissive security groups	CSPM, policy as code, preventive guardrails	Deny-by-default templates, org policies, pre-deploy checks	Sharp drop in public exposures and drift	Alert fatigue and exception creep
Kubernetes runtime and supply chain	Image poisoning, RBAC abuse, escape attempts	Admission policies, image signing, runtime rules, RBAC hardening	PodSecurity standards, cosign or attestations, Falco or eBPF	Higher catch rate for container misuse	Noise without tuning, risk of blocking legit workloads
Serverless and API abuse	Event injection, broken auth, overprivileged functions	Per-function IAM, WAF, API gateway limits, schema validation	Rate limits, JWT validation, and least privilege per function	Contained function scope, fewer abuse paths	Cold start trade-offs, policy sprawl
Data exfiltration and ransomware	Shadow egress, backup deletion, snapshot abuse	Egress allow-lists, DLP, tokenisation, and immutable backups	Egress gateways, object lock, key isolation	Lower exfil success, faster recovery	DLP false positives, integration friction
DDoS and cryptojacking	Volumetric floods, unauthorised compute	Managed DDoS, CDN, anomaly detection, quotas	DDoS services, billing alerts, and runtime anomaly rules	Cost control and quicker mitigation	Extreme events can still be expensive

The taxonomy in Table 1 is supported by empirical and standards-based sources cited throughout this section. Identity and CIEM entries are grounded in entitlement studies and Zero Trust guidance (Fisher, 2025; Neelakandhan *et al.*, 2022; Rose, 2022; Stafford, 2020). Misconfiguration controls reflect posture research and policy as code experiences (Colotti, 2023; Hegde *et al.*, 2023; Jimmy, 2023). Kubernetes rows reflect agency hardening and recent runtime research (CISA, 2022; Eldjou *et al.*, 2023; Morić *et al.*, 2025; Yilmaz & Harding, 2024). Serverless and API entries map to the OWASP API taxonomy and practice reports (Mallick & Nath, 2024; OWASP; Türetken, 2024). Data governance rows synthesise exfiltration and ransomware discussions with cryptographic protections (Lal *et al.*, 2022; Sahar Saeed *et al.*, 2025; Vidhya, 2024). DDoS or cryptojacking rows align with moving target defence surveys and standard managed services (Nguyen & Debroy, 2022).

Table 2: Search strategy, databases, and results retrieved from Google Scholar.

Defense category	AWS examples	Azure examples	GCP examples	Open tools commonly used
Identity and CIEM	IAM Access Analyser, Access Advisor	Entra PIM, Identity Governance	IAM Recommender, Policy Analyser	Cartography or CloudGraph, Steampipe, Prowler
Posture and policy as code	Security Hub, Config, SCPs	Defender for Cloud, Azure Policy	Security Command Centre, Org Policies	tfsec, Checkov, OPA Gatekeeper, Kyverno
K8s runtime and supply chain	EKS PodSecurity, Inspector, GuardDuty EKS	Defender for Containers, AKS policies	GKE Workload Identity, Binary Authorisation	Falco, Trivy, cosign, SLSA or provenance attestations
Serverless and API protections	WAF, API Gateway, per-function IAM	WAF, API Management, Functions RBAC	Cloud Armour, API Gateway, per-service accounts	OWASP ZAP, schema validators, rate-limit plug-ins
Data and egress governance	KMS, Macie, VPC egress, S3 Object Lock	Key Vault, Purview, Private Link, Immutable Vault	KMS, DLP, VPC-SC, CMEK	HashiCorp Vault, Tink, tokeniser libs
Detection and response	CloudTrail Lake, GuardDuty, Security Hub	Sentinel, Defender suite	Cloud Logging, SCC, Chronicle	Sigma rules, Elastic or Loki, osquery or eBPF stacks

Table 2 operationalises the defences by pointing to named controls and tools that practitioners can configure today. The intent is not to prescribe a specific vendor but to reduce translation effort when moving from the taxonomy to an implementation plan. Identity and CIEM mappings highlight where to analyse and right-size entitlements (Fisher, 2025; Neelakandhan *et al.*, 2022). Posture and policy entries identify the native policy engines and posture services that enforce deny-by-default patterns and continuous assessment (Colotti, 2023; Hegde *et al.*, 2023; Jimmy, 2023). Kubernetes lines connect admission and runtime concepts to managed offerings and standard open tools, reflecting the agency baseline's emphasis on restricted privileges and verified artefacts (CISA, 2022; Eldjou *et al.*, 2023; Morić *et al.*, 2025; Yilmaz & Harding, 2024). Serverless and API rows surface the choke points where authentication, authorisation, and rate limiting can be enforced in gateways. In contrast, data and egress governance rows consolidate the key management and egress filtering constructs that underpin protection and recovery (Lal *et al.*, 2022; Vidhya, 2024). The detection row underscores the value of unifying provider native telemetry into a normalised pipeline and expressing detection logic in portable formats aligned to ATT&CK techniques (Deka *et al.*, 2022; Hegde *et al.*, 2023).

Table 3: Metric Dictionary

Metric	Definition	Practical note
Excess privilege rate	Percent of identities exceeding least-privilege baselines	Derive from CIEM or logs; report monthly
Misconfiguration density	Misconfigs per 100 resources of a given type	Normalize by resource count for fairness
Detection coverage	Techniques covered vs in-scope ATT&CK techniques	Report per tactic family for clarity
MTTD and MTTR	Mean time to detect or respond	Prefer medians if outliers skew means
Exfiltration success rate	Percent of simulated egress attempts that bypass controls	Pair with DLP false positive rate
Policy drift rate	Percent of resources deviating from baseline policy	Track before and after guardrail rollout

The metric dictionary in Table 3 supports consistent evaluation. Excess privilege rate is computable with CIEM tools or log-derived inferences, and it reflects the core claim that least privilege reduces blast radius (Fisher, 2025; Neelakandhan *et al.*, 2022). Misconfiguration density normalises posture findings across teams with different resource counts, preventing raw tallies from masking risk in larger estates (Hegde *et al.*, 2023; Jimmy, 2023). Detection coverage, expressed as a ratio of covered to in-scope ATT&CK techniques, allows threat-informed planning and highlights gaps where telemetry or content is missing (CISA, 2022). Time-based metrics

such as MTTD and MTTR retain comparative value when medians are reported to reduce the distortion of outliers (Deka *et al*., 2022). Exfiltration success rate, measured through controlled simulations, should be paired with DLP false favourable rates to balance protection and usability (Lal *et al*., 2022; Vidhya, 2024). Policy drift rate captures the health of preventive guardrails over time, and its reduction after policy-as-code rollouts is a practical signal that posture work is paying off (Colotti, 2023; Hegde *et al*., 2023; Jimmy, 2023).

Taken together, the results support a staged but focused programme. Identity first controls, continuous posture and policy enforcement, signed provenance with tuned runtime detection, and strong data or egress governance are repeatedly associated with reduced exposures and tighter containment when incidents occur (Ali *et al*., 2025; Alouffi *et al*., 2021; CISA, 2022; Fernandez & Brazhuk, 2024; Morić *et al*., 2025; OWASP; Rose, 2022; Stafford, 2020; Tahirkheli *et al*., 2021). The provider mapping shows how to operationalise these findings without adopting an all-or-nothing platform stance, and the metric dictionary enables continuous measurement without manual audits. Organisations that sequence these controls and measure consistently report more apparent progress and fewer surprises than those that accumulate tools without a shared taxonomy, a control mapping, or a metric backbone (Colotti, 2023; Fisher, 2025; Hegde *et al*., 2023; Jimmy, 2023; Morić *et al*., 2025).

Table 1: A quick lookup that links dominant cloud threat classes to practical defences, concrete control patterns, expected outcomes, and common pitfalls. Use it to justify choices and to ensure coverage across identity, posture, runtime, and data.

Table 2: A practitioner map that aligns each defence category with native controls on AWS, Azure, and GCP alongside standard open tools. It turns the taxonomy into actionable implementation handles without prescribing a single vendor path.

Table 3: A compact measurement glossary defining consistent, automatable metrics such as excess privilege rate, misconfiguration density, detection coverage, MTTD or MTTR, exfiltration success, and policy drift. It standardises evaluation across studies and real deployments.

Discussion

The results suggest that durable risk reduction in cloud environments does not come from accumulating tools but from redesigning how access is granted, how configurations are created and enforced, how software artefacts are validated, and how data leaves the estate. Put more plainly, the best outcomes arise when identity-first policy is the default, policy as code stops drift before it deploys, provenance gates what may run, and egress controls plus resilient backups bound the damage when prevention fails (Ali *et al*., 2025; Alouffi *et al*., 2021; Hegde *et al*., 2023; Morić *et al*., 2025; Tahirkheli *et al*., 2021). This part then results in those findings and brings up operational trade-offs and then maps them into a two-set quarter plan which sequencing-minded teams can implement without vendor lock-in.

The strongest lever is identity and entitlement management as it reduces the blast radius that an attacker can have, and the likelihood of a single vulnerable principal opening the door to an expansive layer of resources is reduced. The graph-informed review, automatic right-sizing, and short-lived credentials, based on evidence in the literature of entitlement studies, decrease unnecessary privileges and lateral movement opportunities in practice (Fisher, 2025; Neelakandhan *et al*., 2022). The given practices are consistent with the Zero Trust advice that considers all requests unsatisfied until a policy and context are met. Nevertheless, the technical reality is less complicated: the fewer standing privileges there are to exploit, the fewer options the attacker has to make, and the fewer signals the detection must wade through (Rose, 2022; Stafford, 2020). The task problem is the maintenance of the accuracy of the identity graph at a cloud scale. Teams that pair CIEM with periodic reviews and just-in-time elevation reports sustained improvements; teams that run one-off campaigns without automation see improvement decay as roles accrete again (Fisher, 2025; Neelakandhan *et al*., 2022). The lesson is to institutionalise right-sizing as a process, not a project.

Misconfiguration is the fast path to loss because it exposes assets directly to the internet or to overly permissive internal networks. The strongest signal in the literature is that preventive guardrails enforced before deploy beat detective scanning after the fact. Deny by default templates, organisation policies, and policy as code checks in the pipeline reduce public exposures and configuration drift more reliably than dashboards of misconfigurations alone (Colotti, 2023; Hegde *et al*., 2023; Jimmy, 2023). That said, posture programs stumble when policy catalogues balloon and exceptions are granted ad hoc. Studies that report stable gains focus on a small corpus of high-value policies, integrate them into developer workflows, and manage exceptions with expirations and review reports rather than email threads (Colotti, 2023; Hegde *et al*., 2023; Jimmy, 2023). This is not a tooling victory so much as an operating model shift where developers, security, and platform engineers agree on the minimal viable guardrails and treat them as code.

The end-to-end nature of workload security of containers and Kubernetes makes sense: build something you can sign, admission-time check and run-time check. The agency guidance on Kubernetes hardening has been adopted as the de facto minimum standard for operators of clusters, focusing on limited privileges, strong access control, and admission control (CISA, 2022). Recent work has demonstrated that by adding image signing and provenance attestations and then enforcing those properties at admission, such properties significantly decrease the likelihood that untrusted artefacts run on the cluster (Cesarano & Natella, 2025; Morić *et al*., 2025;

Yilmaz & Harding, 2024). Runtime detection, commonly achieved through the use of eBPF or rule-based engines, elevates signals to indicate escalation attempts and container misuse. Nonetheless, it is most likely to cause alert fatigue unless configured with context either using build pipelines, constructing SBOMs, and cluster roles (Eldjou *et al*., 2023; Nguyen *et al*., 2023). The practical compromise reported by high-performing teams is to block on a small set of high-confidence admission policies and to treat runtime detection as an investigative safety net that is tuned over several weeks, not a day-one silver bullet (CISA, 2022; Eldjou *et al*., 2023; Morić *et al*., 2025; Yilmaz & Harding, 2024).

Serverless and API centric designs demand discipline at the gateway and per-function identity layers. The OWASP API Security Top 10 catalogues failure modes that appear directly in cloud workloads, including broken authorisation and unsafe consumption of third-party APIs (OWASP). Case-driven work recommends rate limiting, schema validation, and robust JWT handling at the gateway, together with least privilege roles per function to avoid privilege sprawl (Mallick & Nath, 2024; Türetken, 2024). The friction point is authoring hundreds or thousands of granular policies. One promising operational pattern is to generate candidate policies from traces and call graphs, review deltas in code review, and deploy per-function roles incrementally. Programs that start at the gateway, then tighten function roles from observed behaviour, report fewer breakages and more durable improvements than those that attempt a top-down policy rewrite first (Mallick & Nath, 2024; OWASP; Türetken, 2024).

Data exfiltration and ransomware outcomes in cloud environments are susceptible to egress governance and backup integrity. Analytical reviews and field reports converge on a combination that works: egress allow lists or gateways to constrain outbound paths, tokenisation to reduce the value of exposed records, object or vault locks to prevent malicious deletion, and rehearsed recovery playbooks to shorten downtime (Lal *et al*., 2022; Sahar Saeed *et al*., 2025; Vidhya, 2024). The most frequent pitfall is collateral damage in case of the introduction of egress rules or DLP patterns in an over-aggressive way. Research which did not prolong disruption also used staged domain-based egress controls and presented DLP dictionaries progressively, checked false favourable ratios as a 1st-class measure, narrowing as confidence increased (Lal *et al*., 2022; Sahar Saeed *et al*., 2025; Vidhya, 2024). The broader lesson is that data protection programs succeed when they measure both protection and friction and iterate toward balance.

The detection and response story is less about inventing novel algorithms and more about speaking a common language and wiring telemetry so that investigations cross cloud boundaries without translation cost. The MITRE ATT&CK Enterprise Cloud matrices provide a stable technique vocabulary that covers infrastructure, SaaS, IdP, and office suites, as well as programs that map detections to techniques. This approach reports clearer coverage, accounting, and a more rational backlog for content engineering (Dieterich). On the analytics side, semi-supervised and range-based anomaly detection works when features are engineered to reflect expected elasticity and burstiness rather than fixed thresholds better suited to static data centres (Deka *et al*., 2022; Tatineni, 2023). Importantly, the best performing programs treat ATT&CK mapping and log normalisation as platform work, not a series of incident-specific projects. They publish simple coverage metrics that product teams can understand and improve (Dieterich; Hegde *et al*., 2023).

Multi-cloud compounds each of these themes. It multiplies identity silos, policy engines, and default configurations, increasing the likelihood that exceptions creep in and visibility becomes fragmented (Ali *et al*., 2025; Lata & Kumar, 2025; Potla, 2025; Sivaseelan, 2024). The practical stance supported by the literature is to choose a minimal set of common nouns and control patterns that can be implemented on each cloud and to measure with normalised metrics rather than trying to force uniformity where platforms differ (Ali *et al*., 2025; Hegde *et al*., 2023). For example, CIEM right-sizing, deny-by-default deployment templates, image signing and admission verification, and egress allow lists exist on all three major providers, even if the service names differ. A program that implements those four patterns consistently across clouds will outperform a program that standardises a niche control on one provider while neglecting fundamentals on another (Fisher, 2025; Hegde *et al*., 2023; Morić *et al*., 2025).

Cost and usability considerations surface repeatedly as adoption constraints. Runtime detection produces noise until tuned; egress controls and DLP cause breakages until patterns are calibrated; per-function least privilege increases initial developer friction. The consistent answer in the sources is staged rollout with explicit success metrics. Teams that start with a small, high confidence rule set, provide clear exception pathways, and publish weekly metrics on excess privilege rate, misconfiguration density, detection coverage, exfiltration success under simulation, and policy drift build credibility and avoid whiplash (Colotti, 2023; Deka *et al*., 2022; Dieterich; Fisher, 2025; Hegde *et al*., 2023; Jimmy, 2023; Neelakandhan *et al*., 2022). In other words, measurement fluency is not a reporting exercise but a leadership tool that allows controlled tightening without losing the organisation.

Two future-facing threads deserve brief comment. First, confidential computing promises protection of data in use for certain workload classes, and there is emerging evidence that heterogeneous trusted execution environments can be orchestrated at cloud scale. Still, performance and operational maturity need careful evaluation before broad production deployment (Chen, 2023; Dhar *et al*., 2024). Second, software supply chain assurance continues to evolve from SBOMs to signed provenance and policy-driven enforcement in admission controllers. At the same time, the direction is clear: teams should anticipate operational cost in maintaining

signing and verification infrastructure and in reconciling third-party artefact trust (Nguyen *et al*., 2023; Yilmaz & Harding, 2024). These areas are promising but should augment, not displace, the core controls above.

Limitations of this review reflect the broader field. Studies vary in methodology and environment realism, and several topics rely on standards and guidance rather than controlled trials. A subset of sources are preprints or theses that we treated as emerging evidence. Where findings conflicted, we emphasised directionality and operational caveats instead of forcing consensus (Ahmadi, 2024; Ali *et al*., 2025; Alouffi *et al*., 2021; Hegde *et al*., 2023; Morić *et al*., 2025; Tahirkheli *et al*., 2021). The synthesis still supports explicit action for practitioners because the repeated patterns are robust across sources and align with platform primitives that teams already operate.

Translating the synthesis into action requires sequencing. Table 4 provides a two-quarter implementation plan keyed to organisational maturity. Its purpose is not to exhaust the catalog of possible controls but to stage a handful of high-leverage moves that create compounding benefit and produce measurable improvement on the metric dictionary. Quarter 1 emphasises identity hygiene and posture guardrails because they reduce risk quickly and raise the floor across the estate. Quarter 2 deepens control with runtime tuning, API gateway protections, egress governance, and resilience investments.

Table 4: Two-Quarter Implementation Plan by Maturity

Maturity level	Quarter 1 priorities	Quarter 2 priorities
Starting	Enforce MFA, remove long-lived keys, enable baseline CSPM, fix public exposures	Deny-by-default deploys, egress allow-lists for sensitive paths, enable immutable backups
Intermediate	CIEM right-sizing, policy as code in CI, image signing and provenance	Tune runtime detection, API gateway limits and schema validation, periodic drift reports
Advanced	Graph-driven least privilege, attestations in build and deploy, drift SLOs	Unified detections mapped to ATT&CK, egress governance at scale, post-incident hardening playbooks

Table 4 sequences controls that the evidence base supports and that exist on all major providers. Identity and CIEM steps cut excess privilege rates and reduce lateral movement opportunities (Fisher, 2025; Neelakandhan *et al*., 2022). CSPM with deny-by-default patterns curbs misconfiguration density and policy drift (Colotti, 2023; Hegde *et al*., 2023; Jimmy, 2023). Image signing, provenance, and admission checks deny untrusted artefacts entry into clusters, while runtime tuning improves catch rates without debilitating noise (CISA, 2022; Eldjou *et al*., 2023; Morić *et al*., 2025; Yilmaz & Harding, 2024). API gateway protections and schema validation address OWASP documented failure modes, and per-function least privilege reduces serverless blast radius as automation matures (Mallick & Nath, 2024; OWASP; Türetken, 2024). Egress allow lists and immutable backups limit data loss and speed recovery, with rehearsals to make resilience real, not rhetorical (Lal *et al*., 2022; Sahar Saeed *et al*., 2025; Vidhya, 2024). Unified detections aligned to ATT&CK, supported by normalised provider logs, raise investigative fluency and clarify remaining content gaps (Dhar *et al*., 2024).

In summary, the cloud security story since 2020 is not that everything changed, but that the centre of gravity moved. Identity became the real perimeter; policy as code became the best place to intercept drift. Provenance and admission control became the way to assert trust in what runs, and egress governance, along with resilient backups, became the way to bound harm. Teams that execute these moves deliberately, measure with a compact metric set, and iterate through staged rollouts avoid the twin failures of static perimeters and tool sprawl. The tables in this paper are intended to make that path specific. They provide a common language for architects, platform owners, and security engineers to select controls, configure them on their respective providers, and accurately track their performance. With that scaffolding in place, the open problems in unified SaaS and IdP detection, as well as in practical data in-use protections, can be pursued without leaving the fundamentals undone (Chen, 2023; CISA, 2022; Dhar *et al*., 2024; Rose, 2022; Stafford, 2020).

Table 4: Construct and measurement map aligning TQM dimensions, decision properties, and performance outcomes across included studies, with examples of operational indicators used by authors and notes on validated instruments or proxies to support comparability in the synthesis.

CONCLUSION

A focused set of identity-first, posture, runtime, and data or egress controls delivers durable risk reduction, outperforming tool sprawl. Teams should phase in adoption, track a small set of metrics, and prioritise automation for least privilege and drift detection while advancing research on unified SaaS or IdP detection and data-in-use protections.

REFRENCES

Adewale, T. (2023). *Microsegmentation vs. Macrosegmentation: Which Approach is Best for Zero Trust Implementation.*

Ahmadi, S. (2024). Systematic literature review on cloud computing security: Threats and mitigation strategies. Ahmadi, S.(2024) Systematic Literature Review on Cloud Computing Security: Threats and Mitigation Strategies. *Journal of Information Security, 15,* 148–167.

Al-Qtiemat, E., & Al-Odat, Z. (2024). Examining cloud

security: identifying risks and the implemented mitigation strategies. *Journal of Theoretical and Applied Information Technology, 102*(7).

Ali, S., Talpur, D. B., Abro, A., Alshudukhi, K. S. S., Alwakid, G. N., Humayun, M., Bashir, F., Wadho, S. A., & Shah, A. (2025). Security and privacy in multi-cloud and hybrid cloud environments: Challenges, strategies, and future directions. *Computers & Security*, 104599.

Almeida, J. R., Zúquete, A., Pazos, A., & Oliveira, J. L. (2024). A federated authentication schema among multiple identity providers. *Heliyon, 10*(7).

Alouffi, B., Hasnain, M., Alharbi, A., Alosaimi, W., Alyami, H., & Ayaz, M. (2021). A systematic literature review on cloud computing security: threats and mitigation strategies. *IEEE Access, 9*, 57792–57807.

Bajpai, M. (2022). Cloud Based Network Management: Leveraging SASE for Efficient and Secure Access. *Journal of Engineering and Applied Sciences Technology.*

Bashi, Z. S. M. A., Basri, A. B., & Senan, S. (2025). Unified Secure Access Service Edge (SASE): Transforming Security for Hybrid Workforce and Multi-Cloud Environments. *International Journal on Perceptive and Cognitive Computing, 11*(2), 1–7.

Cesarano, C., & Natella, R. (2025). KubeFence: Security Hardening of the Kubernetes Attack Surface. arXiv preprint arXiv:2504.11126.

Chandra, A. (2020). *Measurement of the Cloud Security Level at Company using Cloud Control Matrix*.

Chen, K. (2023). Confidential high-performance computing in the public cloud. *IEEE Internet Computing, 27*(1), 24–32.

CISA, N. (2022). NSA, CISA release Kubernetes Hardening Guidance. In.

Colotti, M. E. (2023). *Enhancing Multi-cloud Security with Policy as Code and a Cloud Native Application Protection Platform Politecnico di Torino*.

Deka, P. K., Verma, Y., Bhutto, A. B., Elmroth, E., & Bhuyan, M. (2022). Semi-supervised range-based anomaly detection for cloud systems. *IEEE Transactions on Network and Service Management, 20*(2), 1290–1304.

Dhar, A., Sridhara, S., Shinde, S., Capkun, S., & Andri, R. (2024). Confidential Computing with Heterogeneous Devices at Cloud-Scale. *2024 Annual Computer Security Applications Conference (ACSAC).*

Dieterich, J. (n.d.). *Development of an Adversary Simulation Strategy for a Kubernetes-based Open RAN Deployment*.

Dommari, S., & Khan, S. (2023). *Implementing Zero Trust Architecture in Cloud-Native Environments: Challenges and Best Practices*. Available at SSRN 5259339.

Eldjou, A., Amoura, M. E., Soltane, M., Belguidoum, M., Bennacer, S., & Kitouni, I. (2023). *Enhancing Container Runtime Security: A Case Study in Threat Detection*. TACC.

Fernandez, E. B., & Brazhuk, A. (2024). A critical analysis of Zero Trust Architecture (ZTA). *Computer Standards & Interfaces, 89,* 103832.

Fisher, P. (2025). *Cloud Infrastructure & Entitlement Management (CIEM)*. https://www.kuppingercole.com/research/lc80465/cloud-infrastructure-entitlement-management-ciem

Ge, C., Susilo, W., Baek, J., Liu, Z., Xia, J., & Fang, L. (2021). Revocable attribute-based encryption with data integrity in clouds. *IEEE Transactions on Dependable and Secure Computing, 19*(5), 2864–2872.

Gelernter, N., Schulmann, H., & Waidner, M. (2024). External Attack-Surface of Modern Organizations. Proceedings of the *19th ACM Asia Conference on Computer and Communications Security*.

Hegde, T., Gangl, J., Babenko, S., & Coffman, J. (2023). Cloud security frameworks: A comparison to evaluate cloud control standards. *Proceedings of the IEEE/ACM 16th International Conference on Utility and Cloud Computing.*

James, W. (2021). Architecting Secure Cloud Networks: Balancing Performance, Flexibility, and Zero Trust Principles. *International Journal of Trend in Scientific Research and Development, 5*(3), 1339–1348.

Jimmy, F. (2023). Cloud security posture management: tools and techniques. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2*(3).

Lal, A., Prasad, A., Kumar, A., & Kumar, S. (2022). Data Exfiltration: Preventive and detective countermeasures. *Proceedings of the International Conference on Innovative Computing & Communication (ICICC).*

Lata, M., & Kumar, V. (2025). Cyber security techniques in cloud environment: comparative analysis of public, private and hybrid cloud. *EDPACS, 70*(3), 1–21.

Lee, J. H., & Park, J. (2025). A Case Study: Leveraging SASE Technology for Zero Trust Implementation in Cloud Environments. *2025 International Conference on Information Networking (ICOIN).*

Maddali, G. (2025). *Zero Trust Security Architectures for Large-Scale Cloud Workloads*. Available at SSRN 5365222.

Maidine, K., & El-Yahyaoui, A. (2023). Key Mechanisms and Emerging Issues in Cloud Identity Systems. *International Conference of Cloud Computing Technologies and Applications.*

Mallick, M. A. I., & Nath, R. (2024). Securing the server-less frontier: Challenges and innovative solutions in network security for server-less computing. *Reading Time, 193*(1), 1–45.

Morić, Z., Dakić, V., & Čavala, T. (2025). Security Hardening and Compliance Assessment of Kubernetes Control Plane and Workloads. *Journal of cybersecurity and privacy, 5*(2), 30.

Neelakandhan, M., Ramprakash, G., & Gaidhani, M. (2022). Achieving least privilege at cloud scale with cloud infrastructure entitlements management. *Cyber Security: A Peer-Reviewed Journal, 5*(3), 227–236.

Nguyen, M., & Debroy, S. (2022). Moving Target Defense-Based Denial-of-Service Mitigation in Cloud Environments: *A Survey. Security and Communication Networks, 2022*(1), 2223050.

Nguyen, P. Q., Tikalsky, M. A., & Durlauf, S. M. (2023).

Software Bill of Materials: A Catalyst to a More Secure Software Supply Chain.

OWASP, T. API Security Risks—2023. URL: https:// owasp. org/API–Security/editions/2023/en/0x11–t10.

Oyeniyi, J. O., & Oyeniran, O. A. Optimizing Information Security In Cloud Environments: A Risk Management Approach And Guide For Enterprise Cloud Security. *Journal of Cybersecurity Education, Research and Practice, 2025*(1), 8.

Potla, S. (2025). Securing Multi-Cloud Environments: Challenges and Solutions. *Journal of Computer Science and Technology Studies, 7*(4), 780–785.

Rose, S. (2022). *Planning for a Zero Trust Architecture: A Planning Guide for Federal Administrators*. National Institute of Standards and Technology White Paper(20).

Sahar Saeed, Z. M., Zainal, A. B., Ghaleb, F. A., & Al-rimy, B. A. S. (2025). Enhancing public cloud resilience: an analytical review of detection and mitigation strategies against economic denial of sustainability attacks. *Discover Internet of Things, 5*(1), 79.

Sinan, M., Shahin, M., & Gondal, I. (2025). Integrating Security Controls in DevSecOps: Challenges, Solutions, and Future Research Directions. *Journal of Software: Evolution and Process, 37*(6), e70029.

Singh, R., Yeboah-Ofori, A., Kumar, S., & Ganiyu, A. (2024). Fortifying Cloud DevSecOps security using terraform infrastructure as code analysis tools. *2024 International Conference on Electrical and Computer Engineering Researches (ICECER)*.

Sivaseelan, S. (2024). *Enhancing Cyber Resilience in Multi-Cloud Environments*.

Stafford, V. (2020). Zero trust architecture. *NIST special publication, 800*(207), 800–207.

Tahirkheli, A. I., Shiraz, M., Hayat, B., Idrees, M., Sajid, A., Ullah, R., Ayub, N., & Kim, K.-I. (2021). A survey on modern cloud computing security over smart city networks: Threats, vulnerabilities, consequences, countermeasures, and challenges. *Electronics, 10*(15), 1811.

Tatineni, S. (2023). AI-infused threat detection and incident response in cloud security. *International Journal of Science and Research (IJSR), 12*(11), 998–1004.

Tran, M. H. (2023). *How the MITRE ATT&CK Framework can be used for Threat Modelling in the Cloud NTNU*.

Türetken, B. (2024). Enhancing Security with Cloud-based API Management: Best Practices and Implementation. In: *KTH Royal Institute of Technology*.

Vidhya, S. (2024). Enhancing Cloud Security for Structured Data: An AES-GCM Based Format-Preserving Encryption Approach. *International Conference on Advancements in Smart Computing and Information Security*.

Yilmaz, U., & Harding, P. (2024). Securing the software supply chain for containers: practices and challenges in a cloud-native landscape for a global observatory. *Software and Cyberinfrastructure for Astronomy VIII*.