Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 2, Issue 4, 2024

AN INTELLIGENT CYBERSECURITY FRAMEWORK FOR DATA PROTECTION IN CLOUD COMPUTING ENVIRONMENTS

Partha Chakraborty*, Hemayet Uddin Himel, Monjira Bashir, Mohammad Somon Sikder,Harleen Kaur, Md Talha Bin Ansar, Jobanpreet kaur, Md Kazi Tuhin
International American University, Los Angeles, California, USA, Email: parthachk64@gmail.com
Pacific States University, Los Angeles, California, USA, Email: P25506@psuca.edu
International American University, Los Angeles, California, USA, Email: monjiratrisha@gmail.com
Pacific States University, Los Angeles, California, USA, Email: p25681@psuca.edu
Westcliff University, Irvine, California, USA, Email: H.kaur.4088@westcliff.edu
Yeshiva University, Manhattan, New York, USA, Email: mtalhabi@mail.yu.edu
Westcliff University, Irvine, California, USA, Email: j.kaur.244@westcliff.edu
Yeshiva University, Manhattan, New York, USA, Email: muhammadkazituhin@gmail.com

DOI:- https://doi.org/10.5281/zenodo.21538159

Keywords
Cloud Security; Data Protection; Artificial Intelligence; Machine Learning; Intrusion Detection; Zero Trust; Automated Incident ResponseArticle History
Received: 11 Sep, 2024
Accepted: 20 Oct, 2024
Published: 26 Oct, 2024

Corresponding Author: *
Partha Chakraborty*

Abstract
Cloud computing offers flexible and scalable infrastructure, platforms, and applications; but, its shared, programmable, and distributed framework introduces interconnected threats to sensitive information. This research formulates ICF-CloudSecure, a sophisticated, data-oriented cybersecurity architecture that incorporates multifactor authentication, identity and access management, cryptographic safeguards, AI-driven threat analysis, machine learning-based intrusion detection, ongoing risk assessment, and automated incident management. The study employs a design-science approach that includes threat modeling, security requirements engineering, modular architecture creation, feature pipeline specification, risk assessment, and analytical validation via requirements-to-control traceability and failure scenario evaluation. Multisource cloud telemetry is standardized, and features related to identity, network, workload, data access, and time are derived. Supervised and anomaly-detection methods subsequently produce calibrated evidence instead of independent enforcement determinations. A constrained risk engine integrates threat intelligence, vulnerability exposure, asset significance, data sensitivity, contextual irregularities, and validated mitigating measures to trigger policy-driven, least-privilege actions. The design-level assessment illustrates traceability across objectives of confidentiality, integrity, availability, scalability, dependability, threat detection, and privacy via preventative, detective, responsive, and evidentiary measures. Architecture delineates inference from enforcement, facilitates reversible response protocols, and maintains human oversight, auditability, and multi-cloud adaptability. A validation methodology that may be replicated includes datasets that are segregated by tenant and time, cross-dataset evaluations, ablation experiments, calibration assessments, robustness evaluations, and operational metrics are also delineated. ICF-CloudSecure thus offers a verifiable basis for intelligent and responsible cloud data safeguarding.

1. Introducion

From hardware virtualization, cloud computing has developed into a distributed application substrate that incorporates serverless services, software-defined networking, managed databases, elastic compute, and globally replicated storage. Rapid provisioning and resource pooling are made possible by this transition, but it also distributes security responsibility among tenants, providers, apps, identities, and geographical areas. Tenants are still in charge of identity governance, configuration, application security, encryption selections, and legitimate data use; cloud providers safeguard the managed services and underlying infrastructure. Therefore, rather of occurring as separate failures, vulnerabilities in multi-tenancy, virtualization, management interfaces, access control, configuration, and operational visibility may interact (Alouffi et al., 2021; Subramanian & Jeyaraj, 2018; Tabrizchi & Kuchaki Rafsanjani, 2020). Security based just on a single network boundary becomes inadequate as lawful sessions travel across APIs, containers, geographies, and service providers.

The tenant's control surface is determined by the cloud service model. Tenants of Infrastructure as a Service are mostly in charge of workloads, virtual networks, operating systems, and data controls. Platform as a Service adds provider-specific runtimes, APIs, and identity systems while reducing infrastructure administration. Software as a Service prioritizes account security, configuration, contractual assurance, and data loss protection because it provides less direct technical control. Deployments that are hybrid or multi-cloud further split trust and policy enforcement between separate environments. These variations necessitate a portable security strategy that can assess each request based on the identity involved, workload and device posture, resource sensitivity, session behavior, and requested activity. Therefore, rather than being optional architectural features, continuous verification and least-privilege enforcement are operational requirements (Buck et al., 2021; Syed et al., 2022).

Data safeguarding in these contexts must encompass information that is stored, transmitted, actively utilized, and preserved in backups or duplicates. Confidentiality mandates that only sanctioned individuals may access comprehensible data; integrity necessitates trustworthy and tamper-proof modification logs; and availability demands robust services and retrievable information. Privacy incorporates purpose limitation, data minimization, tenant segregation, and responsible processing. Cryptographic measures safeguard information, although their efficacy relies on key creation, renewal, invalidation, separation, and audit capability. Similarly, identity and access management should regulate human users, workloads, devices, and APIs across their entire lifecycles (Acar et al., 2018; Indu et al., 2018). A breached credential or undue privilege can circumvent robust encryption, whereas a breakdown in key management can reveal data accessed via a genuine identity. Consequently, efficient safeguarding must integrate identity, cryptography, workload, network, and data-access regulations.

Cloud vulnerabilities take advantage of these interconnections. Theft of credentials, insecure application programming interfaces, misconfigurations, unprotected storage, susceptible programs, and cross-tenant vulnerabilities can lead to data breaches. Insider activity necessitates both behavioral and environmental examination with signatures (Homoliak et al., 2019). Malware and ransomware jeopardize core systems, management interfaces, and backup solutions, while distributed denial-of-service assaults deplete network or application resources (Al-rimy et al., 2018; Gibert et al., 2020; Ucci et al., 2019). Phishing, token pilfering, and zero-day vulnerabilities pose further difficulties for deterministic defenses (Cil et al., 2021; Doriguzzi-Corin et al., 2020; Sahingoz et al., 2019). In numerous implementations, authentication, encryption, intrusion detection, monitoring, and reaction are executed as distinct solutions. This disintegration results in uncoordinated alarms, erratic policy choices, and postponed containment, since data from one control is not methodically integrated into the decisions of another.

Artificial intelligence and machine learning can mitigate some aspects of this coordination issue by scrutinizing extensive amounts of identity, network, host, workload, and data-access telemetry. Support vector machines and random forests are capable of modeling structured flow and event characteristics, whereas convolutional, recurrent, and autoencoder-based models can discern local, sequential, or latent patterns linked to both recognized and novel attacks (Shaukat et al., 2020; Vinayakumar et al., 2019; Xin et al., 2018). Nonetheless, an exceptionally effective classifier does not inherently constitute a secure decision-making system. Class imbalance, idea drift, unrepresentative data, avoidance, poisoning, inconsistent explanations, and inaccurately calibrated confidence might render direct automation hazardous (Biggio & Roli, 2018; Rosenberg et al., 2021). Model outputs ought to serve as calibrated evidence inside a regulated policy framework, with uncertainty, human supervision, rollback criteria, and proportionality of action clearly delineated.

The primary research deficiency is in the lack of a cohesive and verifiable framework that connects evidence gathering, contextual risk evaluation, minimal access privileges, cryptographic safeguards, and validated incident management. Current methodologies typically enhance a singular element, like intrusion categorization, authentication, encryption, or zero-trust access, without delineating the process of integrating diverse signals into a constrained risk determination or the manner in which that determination should initiate a secure and verifiable reaction. Dataset-specific accuracy offers scant insight into cross-environment generalization, privacy, resilience, calibration, response safety, and operational dependability. A cloud data-protection system must consequently include an integrated control architecture alongside a replicable assessment method that can assess security and operational results across many tenants, timeframes, attack scenarios, and cloud platforms.

This research tackles the deficiency by creating ICF-CloudSecure, a sophisticated, data-oriented cybersecurity framework for cloud computing settings. The architecture

integrates federated authentication, adaptive multifactor authentication, role- and attribute-centric access control, envelope encryption, multisource telemetry analysis, supervised and anomalous detection of threats, ongoing risk assessment, and policy-driven incident management. It converts identity, network, workload, data access, and temporal evidence into a defined risk score that facilitates one of many measured outcomes: permit, limit, contest, segregate, or intensify. Three research inquiries direct the study: (1) In what manner may preventative, detective, responsive, and evidentiary controls be amalgamated within a closed and verifiable cloud-security decision-making cycle? How can diverse threat evidence be transformed into contextual risk without permitting model outputs to evolve into unregulated enforcement measures? What assessment methods can evaluate the framework's efficacy, generalizability, resilience, confidentiality, and operational dependability?

A design-science research technique is employed to identify the issue, extract security specifications, develop the modular framework, formalize the feature and risk management pipeline, and assess the design via requirements-to-control traceability and failure-scenario evaluation. The research presents four primary contributions. Initially, it suggests a modular control plane that integrates identification, cryptography, detection, risk assessment, and reaction capabilities without necessitating the uncontrolled centralization of tenant information. Secondly, it delineates multisource feature-engineering and evidence-fusion methodology wherein supervised and anomaly models yield calibrated evidence instead of direct enforcement directives. Third, it presents a constrained risk framework and a reversible response model that maintains human supervision, auditability, and minimal privilege. Fourth, it delineates a replicable validation process employing tenant- and time-segregated data, cross-dataset evaluations, ablation experiments, calibration assessments, adversarial resilience examinations, and operational metrics. The subsequent sections of the document delineate the technical underpinnings, introduce the ICF-CloudSecure framework, evaluate its design outcomes, and outline the empirical validation approach along with prospective research avenues.

Literature Review

Cloud Computing Security

Cloud Security Surveys detail the many layers of the attack surface, from the physical environment to the hypervisors, virtual networks, applications, management APIs and tenant configuration. Subramanian & Jeyaraj (2018) highlighted the issues of virtualisation, location, access, recovery and threat; Tabrizchi & Kuchaki Rafsanjani (2020) categorised the threats and countermeasures based on the cloud components. The systematic review conducted by Alouffi et al. (2021) revealed that none of the mechanisms can solve the problem of shared responsibility. These are observations that favour defence-in-depth but don't offer significant guidance on combining identity, workload, network, and

data signals to make a single operational decision.

Data Protection Techniques

Encryption – both in transit and at rest – is necessary but insufficient. Envelope encryption, tenant-specific encryption keys, keys protected with master keys, key rotation and key revocation minimise the risk of a key compromise. While homomorphic encryption can enable computation without plain-text disclosure, it comes at a cost and faces implementation challenges that prevent widespread use (Acar et al., 2018). According to the IAM research, federation, role- and attribute-based authorisation, single sign-on (SSO), and lifecycle governance are key areas highlighted, while interoperability and privilege management are among the areas in which every organisation struggles. (Indu et al., 2018) Although blockchain research implies audit and tamper-evident, not only does scalability, privacy and governance need to be taken into account prior to implementing blockchain, it also needs to be demonstrated to be effective.

Artificial Intelligence in Cybersecurity

Security based on AI, from telemetry to prediction or anomaly scores. According to the reviews by Xin et al. (2018) and Shaukat et al. (2020), intrusion detection, malware detection, spam detection, and fraud detection have been widely addressed using supervised, unsupervised, and deep learning methods. When a score results in a change in access or containment, it is important to consider explainability, which may be unstable with post hoc explanations, but should be designed in connection with model choice, an interpretable set of features, and audience-specific evidence (Barredo Arrieta et al., 2020). That has to be where AI plays a role as complementary to, and alongside, policy and analyst judgment, not in place of.

Intrusion Detection Systems

Signature-based intrusion detection is accurate at detecting known signatures but weak at detecting new or obfuscated attacks. Anomaly detection provides good coverage of novelty but usually produces more false alarms. Poor data quality, feature selection, class imbalance, and realistic validation continue to pose challenges, as indicated by surveys by Ahmad et al. [2], Khraisat et al. [3], and Salo et al. [4]. Although public datasets such as CSE-CIC-IDS2018 and Bot-IoT can help increase attack variety, the recorded environments and label construction still do not resemble production-like clouds (Koroniotis et al., 2019; Ring et al., 2019; Sharafaldin et al., 2018).

Machine-Learning-Based Threat Detection

There are also some examples of how representation learning can be utilised for classification, or online anomaly detection (Al-Qatf et al., 2018; Diro & Chilamkurti, 2018; Meidan et al., 2018; Mirsky et al., 2018), based on sparse autoencoder–SVM models, deep autoencoders, and distributed deep learning, respectively. Extensive comparative analyses of within-dataset performance of deep IDS are presented, and the experimental results indicate that architecture, preprocessing, and class distribution significantly impact performance (Aldweesh et al., 2020; Ferrag et al., 2020; Shone et al., 2018; Vinayakumar et al., 2019). To reduce the computation without

compromising the predictive ability, feature selection and multi-stage or Bayesian optimisation can be used (Injadat et al., 2018, 2021). NetFlow feature sets are appealing for their cloud portability, as they don't require payload inspection; however, the level of detail across the various classes can diminish (Sarhan et al., 2021).

But the point of inference varies with place and time of operation, as well. A study on lightweight models demonstrates that they can detect abnormal device behaviour from structured telemetry data, while real-time deep IDS studies focus on bounded latency and continuous data processing (Hasan et al., 2019; Thirimanne et al., 2022). These results can apply to cloud environments, where streams of events are generated by managed services rather than a table of features. Web deployment introduces some issues, though, that benchmark classification will not address: labels can take time to reach the deployment, alerts can occur multiple times even with the same code, workloads can spike, regional failures can occur, and adversaries can learn the threshold for defences and adapt accordingly. As such, the salient features of a deployable system are model health monitoring, evidence collection and correlation, graceful degradation, and a secure avenue for detection-to-enforcement.

Zero Trust Security

Zero trust eliminates trust based on location in the network. Evaluation of each request is based on identity, device, workload, observed behaviour and sensitivity of resources. While there is strong conceptual consensus on continuous verification and least privilege, there is limited consensus on how to measure them, as well as on migration and interoperability (Buck et al., 2021; Syed et al., 2022). Taking this approach, ICF-CloudSecure is not a zero trust product per se; it is a principle used as an authorisation mechanism.

Research Gap

As Table 1 indicates, while various studies exist that offer valuable components, few offer some. Only a few describe the integration of cryptographic data protection, contextual IAM, heterogeneous detection, risk calibration and automated response. Edge and federated learning research also reveals the need to address remaining issues related to distributed privacy, device heterogeneity, and model security (Alwarafy et al., 2021; Li et al., 2020; Mothukuri et al., 2021; Nguyen et al., 2021; Xiao et al., 2019). As such, the central gap is architectural and evaluative: cloud architectures ought to coordinate controls and exhibit generalisation, privacy, and answer security, rather than delivering a stand-alone accuracy score.

Table 1: Comparative Analysis of Previous Studies

Authors Year	Technique	Key findings	Limitations
Subramanian & Je-2018 yaraj	Cloud security review	Consolidates virtualisation, access, storage, and recovery challenges.	Review-level guidance; no adaptive control loop.

Indu et al. 2018	Cloud IAM review	Explains federation, authorisation, identity lifecycle, and interoperability.	Detection and response are outside scope.
Al-Qatf et al. (2018).	Sparseautoenc	Learns compact features before supervised intrusion classification.	Dataset dependence and limited operational context.
Mirsky et al. 2018	Autoencoder ensemble	Supports efficient online, unsupervised anomaly detection.	Anomaly scores require threshold calibration and investigation.
Salo et al. 2018	Systematic IDS review	Identifies data-mining and feature-selection tradeoffs.	Literature heterogeneity prevents simple ranking.
Sharafaldin et al. 2018	CSE-CIC-IDS2018 dataset	Provides contemporary benign and attack traffic scenarios.	Laboratory traffic cannot reproduce every cloud workload.
Khraisat et al. 2019	IDS taxonomy	Compares signatures, anomalies, datasets, and evaluation issues.	Does not integrate IAM or encryption.
Ring et al. 2019	Dataset survey	Defines properties for judging NIDS dataset suitability.	Finds persistent representativeness and comparability gaps.
Vinayakumar et al. 2019	Deep IDS	Evaluates deep architectures across intrusion datasets.	Computational cost and cross-domain drift remain concerns.
Koroniotis et al. 2019	Bot-IoT dataset	Adds realistic IoT botnet and normal traffic for forensic analytics.	IoT testbed differs from enterprise cloud behaviour.
Aldweesh et al. 2020	Deep-IDS survey	Taxonomises deep anomaly detection and open problems.	Reports inconsistent preprocessing and validation practices.
Doriguzzi-Corin et al. (2020).	Lightweight CNN	Demonstrates resource-conscious DDoS detection.	Focused on DDoS rather than full data protection.
Injadat et al. 2021	Optimised ML pipeline	Combines sampling, feature selection, and hyperparameter tuning.	Benchmark gains need production and drift validation.
Buck et al. (2021).	Zero-trust review	Finds continuous verification and least privilege central to zero trust.	Operational metrics and evidence of migration are limited.
Syed et al. (2022).	ZTA survey	Maps zero-trust components, applications, and challenges.	Broad architecture; no cloud-specific ML response design.

Proposed Intelligent Cybersecurity Framework

ICF-Cloud Secure is a modular security control plane that connects the preventive, detective, and corrective security functions without requiring all telemetry/canonical data to be stored in a single repository. Novelty is that the physical access request will go through a closed decision loop: authentication, authorisation, cryptographic constraints, complemented by physical detectors which monitor access, classification in the context of risk, and one of four processes: give "go," restrict access, challenge access, isolate access control or escalate to higher authority. Although the logical ordering of the modules is maintained in Figure 1, implementations can distribute the modules across various cloud regions and tenant boundaries as needed.

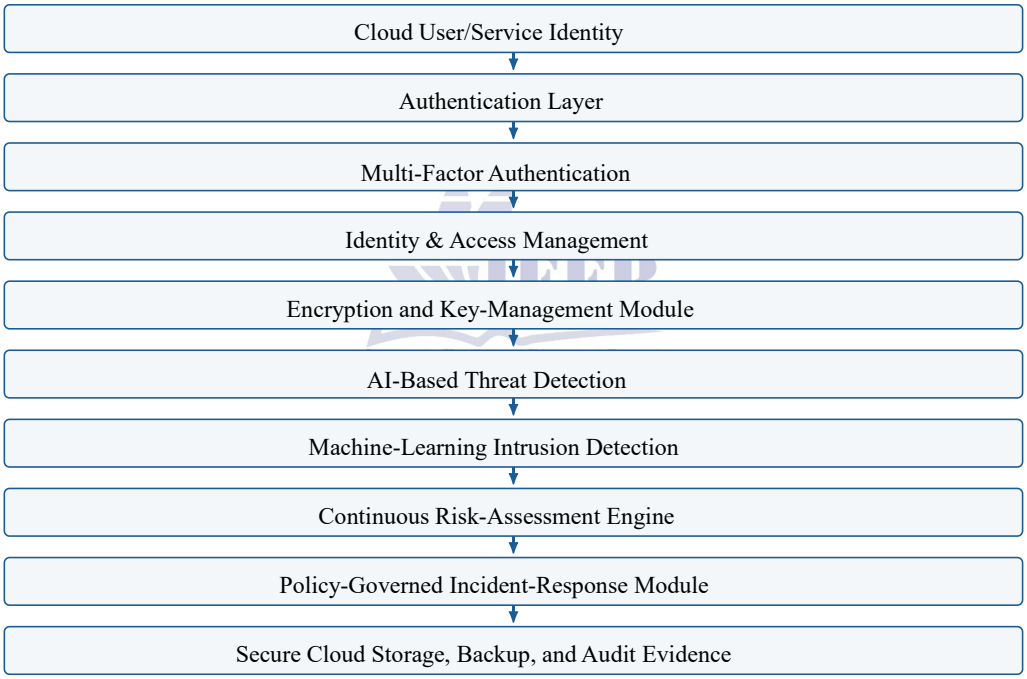

Figure 1: Proposed Intelligent Cybersecurity Framework (ICF-CloudSecure).

Authentication, MFA, and Access Control

The authentication layer verifies the identities of the humans, workloads, devices and APIs via a federated identity provider. Where possible, passwordless public-key methods are selected; where that is not an option, then

passwords are rate-limited, screened for breaches and have a secure recovery capability. MFA adds independent factors and boosts assurance for privileged or out-of-the-norm sessions. Unlike embedded secrets, the workload identities will have short-lived, auto-rotated credentials.

IAM uses two types of access control: role-based access control for an unchanging aspect of a job and attribute-based conditions that can change based on the circumstance. A policy decision takes into account the identity of the person, the posture of the device, the condition of the workload, the sensitivity of the resource, the location, the session behaviour, and the requested operation. Limited standing access with default denial, separation of duties, just-in-time access and rotating entitlements. In addition to allowing or denying, every decision includes a set of obligations that apply in various modes (read-only mode, step-up MFA, masking, recording the session, or reducing the lifetime of the token). This is in line with the zero-trust research (Buck et al., 2021; Syed et al., 2022), which implements continuous verification.

Encryption and Privacy Protection

The encryption module provides protection for the traffic using authenticated transport security and protection for the stored objects using envelope encryption. For each tenant, or for each dataset (or sensitivity class), a unique data-encryption key is created and encrypted with the key-encryption key in the hardware-backed key-management service. The ability to create new key versions with rotation, but not to silently revoke access to ciphertext from past rotations, along with dual control and immutable key-use logs, aids in investigation. If high-sensitivity field, they can be tokenised prior to Analytics. Mr.kürzlich minimiert, pseudonymisiert, regionalisiert und markiert die Verzichtbarkeit und spart Zeit. For chosen computations where there is a performance-critical justification, homomorphic computing/confidential computing techniques can be added.

Data Preprocessing and Feature Extraction

Cloud audit events, IAM decisions, API calls, NetFlow field records, DNS summaries, endpoint alerts, workload runtime events, storage access, and key-management logs are just a few examples of events that sensors can capture. The pipeline performs the following functions: validates schema, timestamps, provenance and signatures, removes duplicates, encodes categorical, scales continuous, and groups records into entity-based time windows. Before adding scalers or selectors to the training and test partitions, make sure they are also separated chronologically or by tenant to prevent leakage.

For event e_i, the feature vector is

$$\mathbf{x}_i = [\mathbf{x}id_i, \mathbf{x}net_i, \mathbf{x}host_i, \mathbf{x}data_i, \mathbf{x}time_i], (1)$$

The features include: Processes: outcome related to authentication failures and privilege changes; network: direction of move, volume of data (bytes), duration of time, fan-out of other processes involved, and protocol involved; host: summary of process and integrity information; and temporal: sequence,

periodicity, deviation from baseline information for an entity. However, such an approach of collecting payload is avoided unless authorised and portable flow features and metadata are preferred (Sarhan et al., 2021).

AI Threat Detection and ML Intrusion Detection

A combination of rules, threat intelligence, graph relations and model scores is fed into the AI layer. If a known indicator violates a policy, it is still considered a deterministic indicator because it is explainable and fast. Attacks in labelled data are classified using a supervised learner: an SVM provides a strong margin-based baseline, a random forest can handle nonlinear, structured features, a CNN can learn local interactions between features, and an LSTM can incorporate sequences of attacks. Novelty detection on rare or zero-day behaviour as provided by an autoencoder or isolation model. Thus, Malware, Phishing, DDoS, and Insider evidence can be merged into a common evidence model, which suggests that one is not necessarily better than another (Gibert et al., 2020; Sahingoz et al., 2019; Tounsi & Rais, 2018).

The models return calibrated probabilities, or anomaly percentiles, not actions to take for enforcement. Class weighting and stratified sampling help address imbalance; feature selection is used to minimise latency; drift is estimated using time-based validation; and evasion and poisoning are investigated with adversarial tests. Explanations reveal features that influenced outcomes, like causes, as well as benign or comparable baselines and supporting events. A model registry keeps track of the data lineage, the model version, the threshold, the model owner, whether it has undergone fairness/privacy review, and the rollback criteria.

The threat detection workflow is completed in 4 gates. A deterministic gate first validates the integrity of telemetry. It implements “deny rules” to protect against the use of already known malicious events and malformed telemetry events to consume valuable inference. Then, a classification gate rates the labelled attack hypotheses, and an anomaly gate compares the entity with itself and a group of peers. Some observations are very weak, such as the discovery of a new device, the sudden and unexpected granting of access privileges, the quick appearance of enumerated objects and outbound transfers within a defined period of time, etc., which are correlated by the correlation gate. Evidence is deduplicated, i.e., multiple alerts are created about the same event aren't added to the list to increase the risk. The system starts to increase observation and requires analyst review where the model confidence is low, or the detectors agree. Continuous monitors track the distributions of features, analyst dispositions, prediction confidence and calibration, and alert yield against approved baselines. A newly trained model can go to enforcement only once drift has caused a threshold adjustment, shadow retraining, or rollback. However, only if drift has caused threshold adjustment, shadow retraining and/or rollback can a newly trained model enter enforcement.

Risk Assessment and Incident Response

The risk engine converts evidence into a bounded score:

$$R = 100\sigma(w_tT + w_vV + w_aA + w_dD + w_cC - w_mM),$$

Threat evidence (T), Vulnerability and Exposure (V), Asset Criticality (A), Data Sensitivity (D), Contextual Anomaly (C), Verified Mitigating Control Strength (M) and logistic mapping (σ). Weights are not learned from the results of incidents but are hand-governed and tested. “Low” is monitored access; “Medium” is time-limited sessions or step-up logon; “High” is restricted from sensitive operations and an analyst case is created; and “Critical” results in pre-approved containment.

The incident-response module provides playbooks for each version. Automated reversible countermeasures can include temporary network rules, token revocation, key suspension, restoring the victim system's state to an earlier time, or quarantining the victim workload when evidence thresholds are crossed and validated. It's destructive – can't be done without human authorisation. All actions have their own idempotency key, timeout, rollback path, evidence link, and a notified owner. All business data is encrypted, as well as audit data -- which is immutably stored -- and backup data that has been tested for the immutability of that data, and stored in separate trust domains, leaving no way for ransomware to be able to delete your primary data, as well as your backup of it.

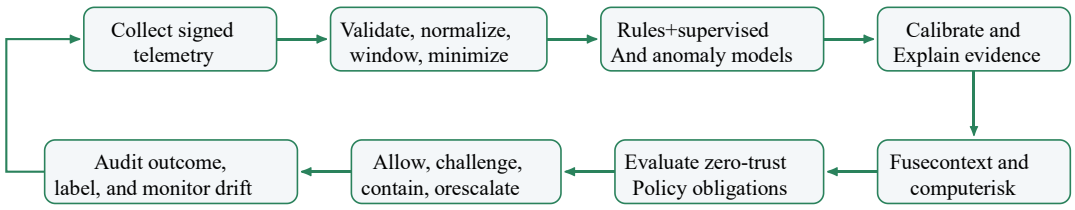

Figure 2: Framework workflow from telemetry collection to governed response and feedback.

Results and Discussion

This article documents at the design level one outcome: that all of the stated goals are mapped to at least one "preventive," "detective," "responsive", and "evidence" mechanism by the ICF-CloudSecure. The architecture eliminates some of the "disconnects" that normally occur. The two (ID assurance/behavioral evidence) affect the SAME access decision. Second, it's security telemetry, not an infrastructure in itself that one would need to invest in encryption and

key events. Thirdly, detection generates a risk output for a playbook that is managed, not a freewheeling automated response. These properties enable defence-in-depth and provide visibility into failures. These, on their own, can't prove any sort of empirical superiority.

Table 2: Security Features of the Proposed Framework

Property	Primary mechanisms	Intended effect	Verification evidence
Confidentiality	MFA, least privilege, envelope encryption, tokenisation	Restrict intelligible data to authorised purposes	Access tests, key logs, leakage tests
Integrity	Authenticated encryption, signed telemetry, immutable audit, snapshots	Detect unauthorised change and preserve provenance	Tamper tests, hash and audit verification
Availability	DDoS analytics, rate limits, isolation, separate backups	Maintain service and recoverable data under attack	Load, failover, and restore exercises
Scalability	Streaming feature: modular models, regional inference	Scale analysis without a centralised bottleneck	Throughput, latency, and queue depth tests
Reliability	Calibrated risk, rollback, health checks, model registry	Bound automation errors and support recovery	Chaos tests, rollback success, calibration error
Threat detection	Rules, supervised ensemble, anomaly detection	Cover known, variant, and previously unseen behaviour	Accuracy, precision, recall, F1, DR, FPR, ROC
Privacy	Minimisation, pseudonymization, federated option, retention	Reduce unnecessary personal-data exposure	Data inventory, retention and privacy audits

Table 3: Performance Comparison (Reproducible Benchmark Template)

Model	Accuracy	Precision	Recall	F1	DR	FPR
SVM	TBE	TBE	TBE	TBE	TBE	TBE
Random forest	TBE	TBE	TBE	TBE	TBE	TBE
CNN	TBE	TBE	TBE	TBE	TBE	TBE
LSTM	TBE	TBE	TBE	TBE	TBE	TBE
ICF-CloudSecure	TBE	TBE	TBE	TBE	TBE	TBE

The suggested evaluation is performed with at least two data sources in different time periods and from different sources (e.g. CSE-CIC-IDS2018 + a Bot-IoT / NetFlow-generated data source + cloud pilot for the tenant/time, respectively). Although public data have the benefit of providing repeatability, it is also necessary to carry out cross-dataset testing to appreciate the differences between the datasets, as documented in different surveys of the datasets (Koroniotis et al., 2019; Ring et al., 2019; Sharafaldin et al., 2018). Only the training folds are fitted with the pre-processing. Once the hyperparameters are set (using nested validation), the untouched test set is evaluated once only. The MFA context, anomaly detection/risk fusion and response feedback are all removed with the ablations to identify the benefit of each module.

For binary detection, the metrics are

$$\text{Accuracy} \quad \text{2(Precision)(Recall)} \quad \frac{TP}{TP+FN}, \qquad F1 = \qquad = \qquad ,$$

$$\text{Recall} = \text{DR} \qquad \text{(4) Precision + Recall}$$

$$\text{FPR} = \frac{FP}{FP+TN}. \tag{5}$$

For multi-class attacks, macro and per-class accuracy are a must, as they might be hiding failures in the rare classes. Bootstrap resampling at the session or entity level should be used to calculate confidence intervals, rather than assuming independent events in the bootstrap resampling. The operational evaluation should also include information on p95 inference latency, events per second (EPS), analyst alerts per day, mean time to containment, action rollback rate, recovery point achievement, and key service availability. Calibration curves are used to assess the expected calibration error and determine whether the stated risk of 0.8 occurs with about the same frequency.

The comparison(s) should represent the various actions of the baselines. An SVM test whether it is possible to recognize a small amount of classes with a compact decision surface, a random forest test whether it is possible to recognize a lot of classes with a robust nonlinear learning on a structured feature set, a CNN test if the local interactions between a set of ordered features can improve the recognition task, and finally an LSTM test if the sequence context can enhance the recognition task. In addition to working as a detector ensemble, ICF-CloudSecure can be viewed as an end-to-end policy system. Returning to norms that attained high areas under the ROC curve may be detrimental due to an excessive false-positive rate, and high

AUROC with poor calibration at the tightening threshold may also be unhelpful. The results of comparison should then be the paired bootstrap difference and the practical effect size; this means that the difference in cost paid for each incident that has been successfully contained should be reported. The degradation of the data felt is especially elucidative; a lower difference in degradation is better than the top score on the source dataset.

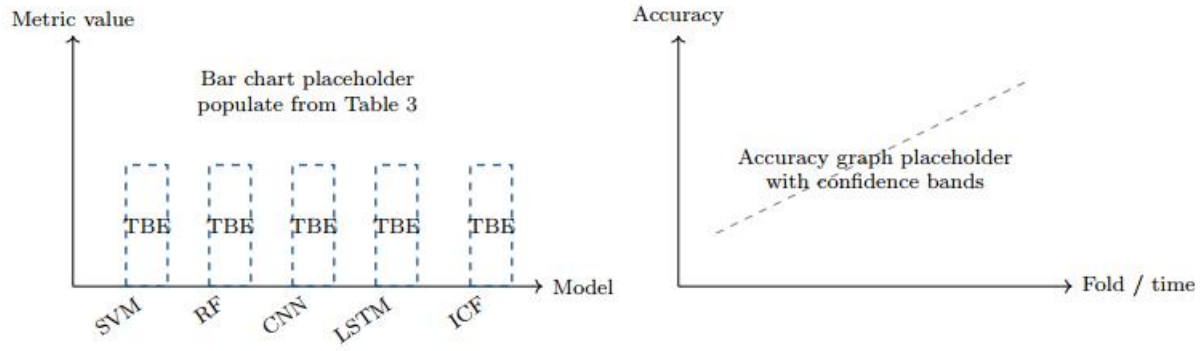

Figure 3: Performance comparison and accuracy graph placeholders; no empirical values are asserted.

Some of the main features of the framework are: Modularity, Data-centric Protection, Explicit Uncertainty, and Safe Automation. Classical and deep models can be swapped out without redesigning the IAM or encryption protocol, and a regional inference tier can scale on its own. The possibility of favouring metadata enhances privacy, supports pseudonymous identifiers for entities (for feature extraction), and enables local or federated training. Reliability is enhanced if response actions can be carried out in reverse and verified by third parties. These design considerations are not for the sake of making a single claim of accuracy, but include elements drawn from IDS reviews as well as reviews of so-called "zero-trust" and cloud security (Alouffi et al., 2021; Buck et al., 2021; Khraisat et al., 2019).

In practice, that means a gradual transition toward using the new protocol, instead of an upheaval in the transition to a "replacement project. The initial access to observe audit, identity, network, and key-management logs is followed by risk recommendations that can be added to any existing security operations, and later the tenant can authorise select reversible playbooks. Providers can offer a view of the same interfaces as the managed controls, and regulated tenants get the same policy as well.

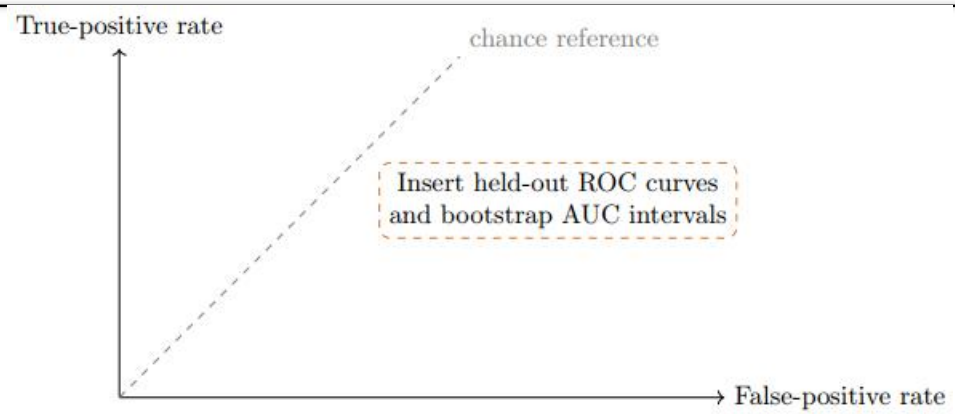

Figure 4: ROC curve placeholder for the four baselines and ICF-CloudSecure.

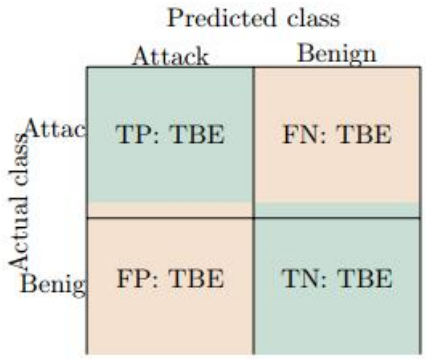

Figure 5: Confusion matrix placeholder with counts to be obtained from the untouched test set.

Ownership, evidence (keyness) and approval boundaries associated with tenants. A shared case record with the access decision, detector explanation, impacted data, response actions, and rollback results benefits security teams. Access to it is now traceable for access review, retention, incident handling and recovery testing by compliance teams. Scaling a high-volume service doesn't need to be about providing a central model with unrestricted access to the raw tenant content – it's about being able to infer it at the region level and having the models controlled by the registry.

There are still several challenges for practical deployment, such as telemetry completeness, provider-specific APIs, visibility of encrypted traffic, concept drift and adversarial manipulation, and class imbalance. Models with high depth could be impacted by latency, explanation requirements, and extreme sensitivity of responses, which could disrupt legitimate work. Even though Federated learning is a technique that can reduce the sharing of raw data, it still faces poisoning, inference, non-identically distributed data, and aggregation-security problems (Li et al., 2020; Mothukuri et al., 2021; Nguyen et al., 2021). The whole framework must be phased in, which means: watch, advise, authorise reversible responses, and then, once

significance safety has been achieved, increase powers. It is, however, currently limited by the lack of a prototype and landmark empirical implementation, and the provided protocol spells out that limitation and makes it potentially testable.

Conclusion and Future Work

This study aimed to create an intelligent cybersecurity framework to enhance cloud Data confidentiality, integrity, availability, Scalability, reliability, threat detection and privacy. ICF-CloudSecure adds federated authentication, adaptive MFA, heterogeneous AI and machine-learning detection, calibrated and applied risk assessment, policy-governed incident response, and protected storage. Not another stand-alone classifier is proposed as a contribution. It constitutes a closed, auditable, and observable control loop in which evidence of identity, workload, network, cryptographic, and/or data access is collectively used to make least-privilege decisions and/or proportionate responses.

Results of the literature synthesis revealed that building blocks such as cloud security, IDS, cryptography, and zero trust are mature areas of research and have yet to be integrated and evaluated. The design-level analysis identified all objectives with their implementation mechanisms, and the evidence to verify them. It has also provided accuracy, precision, recall, F1-score, detection rate, false-positive rate, ROC analysis, and confusion matrices as benchmark scores for all the models discussed, such as SVM, random forest, CNN, LSTM, and the proposed model. Notably, no numerical results were created or implemented without data. This helps maintain the gap between expected benefits and measured performance and provides a solid basis for future experiments by providing strong evidence of acceptance.

The next step in the work is to deploy the architecture in a multi-tenant testbed, and validate it under various drift/evasion/poisoning/partial telemetry/response-failure scenarios, in a time- and tenant-separated manner. Explainable AI should generate evidence that is counterfactual and analyst-oriented, rather than ranking features in general (Barredo Arrieta et al., 2020). Secure aggregation, poisoning defence, and privacy leakage can be studied to enable federated learning to train across organisations without any local records being shared with others. Transparency mechanisms such as blockchain or append-only systems may help strengthen the cross-party audit, but are only worthwhile when needed and require careful consideration of the governance and performance costs associated with their implementation (Casino et al., 2019; Taylor et al., 2020). The applications that can be secured under the same concept should not only be user sessions, but also workload and data identities. There must be a combination of minimisation, differential privacy in AI, confidential computation, and cryptographic processing, depending on the sensitivity of the workload, for privacy-preserving AI. Finally, although response latency and cloud bandwidth consumption can be reduced using edge inference, the integrity of the model deployed at the edge, device attestation, and security with constrained resources need to be tested

(Alwarafy et al., 2021; Xiao et al., 2019). These add-ons would make the current blueprint into a quantifiable, common, and robust cloud data protection platform.

References

Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. *ACM Computing Surveys*, *51*(4), 1–35. https://doi.org/10.1145/ 3214303

Ahmad, Z., Khan, A. S., Shiang, C. W., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, *32*(1), e4150. https://doi.org/10.1002/ett.4150

Aldweesh, A., Derhab, A., & Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowledge-Based Systems*, *189*, 105124. https://doi.org/10.1016/j.knosys.2019.105124

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. https://doi.org/10.1109/ACCESS.2021.3073203

Al-Qatf, M., Lasheng, Y., Al-Habib, M., & Al-Sabahi, K. (2018). Deep learning approach combining sparse autoencoder with SVM for network intrusion detection. *IEEE Access*, *6*, 52843–52856. https://doi.org/10.1109/ACCESS.2018.2869577

Al-rimy, B. A. S., Maarof, M. A., & Shaid, S. Z. M. (2018). Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. *Computers & Security*, *74*, 144–166. https://doi.org/10.1016/j.cose.2018.01.001

Alwarafy, A., Al-Thelaya, K. A., Abdallah, M., Schneider, J., & Hamdi, M. (2021). A survey on security and privacy issues in edge-computing-assisted internet of things. *IEEE Internet of Things Journal*, *8*(6), 4004–4022. https://doi.org/10.1109/JIOT.2020.3015432

Barredo Arrieta, A., D´ıaz-Rodr´ıguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garc´ıa, S., Gil-Lo´pez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, *84*, 317–331.

https://doi.org/10.1016/j.patcog.2018.07.023

Buck, C., Olenberger, C., Schweizer, A., Vo¨lter, F., & Eymann, T. (2021). Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust. *Computers & Security*, *110*, 102436. https://doi.org/10.1016/j.cose.2021.102436

Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and Informatics*, *36*, 55–81. https://doi.org/10.1016/j.tele.2018.11.006

Cil, A. E., Yildiz, K., & Buldu, A. (2021). Detection of DDoS attacks with a feedforward-based deep neural network model. *Expert Systems with Applications*, *169*, 114520. https://doi.org/10.1016/j.eswa.2020.114520

Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using a deep learning approach for the Internet of Things. *Future Generation Computer Systems*, *82*, 761–768. https://doi.org/10.1016/j.future.2017.08.043

Doriguzzi-Corin, R., Millar, S., Scott-Hayward, S., Martínez-del-Rincón, J., & Siracusa, D. (2020). LUCID: A practical, lightweight deep learning solution for DDoS attack detection. *IEEE Transactions on Network and Service Management*, *17*(2), 876–889. https://doi.org/10.1109/TNSM.2020.2971776

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419. https://doi.org/10.1016/j.jisa.2019.102419

Gibert, D., Mateu, C., & Planes, J. (2020). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. *Journal of Network and Computer Applications*, *153*, 102526. https://doi.org/10.1016/j.jnca.2019.102526

Hasan, M., Islam, M. M., Zarif, M. I. I., & Hashem, M. M. A. (2019). Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Things*, *7*, 100059. https://doi.org/10.1016/j.iot.2019.100059

Homoliak, I., Toffalini, F., Guarnizo, J., Elovici, Y., & Ochoa, M. (2019). Insight into insiders and IT: A survey of insider threat taxonomies, analysis, modelling, and countermeasures. *ACM Computing Surveys*, *52*(2), 1–40. https://doi.org/10.1145/3303771

Indu, I., Anand, P. M. R., & Bhaskar, V. (2018). Identity and access management in cloud environment: Mechanisms and challenges. *Engineering Science and Technology, an International Journal*, *21*(4), 574–588. https://doi.org/10.1016/j.jestch.2018.05.010

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Multi-stage optimised machine learning framework for network intrusion detection. *IEEE Transactions on Network and Service Management*, *18*(2), 1803–1816. https://doi.org/10.1109/TNSM.2020.3014929

Injadat, M., Salo, F., Nassif, A. B., Essex, A., & Shami, A. (2018). Bayesian optimisation with machine learning algorithms towards anomaly detection. *2018 IEEE Global Communications Conference (GLOBECOM)*, 1–6. https://doi.org/10.1109/GLOCOM.2018.8647714

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, *2*(1), 20. https://doi.org/10.1186/s42400019-0038-7

Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of a realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. *Future Generation Computer Systems*, *100*, 779–796. https://doi.org/10.1016/j.future.2019.05.041

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60. https://doi.org/10.1109/MSP.2020.2975749

Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., & Elovici, Y. (2018). N-BaIoT network-based detection of IoT botnet attacks using deep autoencoders. *IEEE Pervasive Computing*, *17*(3), 12–22. https://doi.org/10.1109/MPRV.2018.03367731

Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. *Proceedings of the 25th Network and Distributed System Security Symposium*. https://doi.org/10.14722/ndss.2018.23204

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, *115*, 619–640. https://doi.org/10.1016/j.future.2020.10.007

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for

internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *23*(3), 1622–1658. https://doi.org/10.1109/COMST.2021.3075439

Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. *Computers & Security*, *86*, 147–167. https://doi.org/10.1016/j.cose.2019.06.005

Rosenberg, I., Shabtai, A., Elovici, Y., & Rokach, L. (2021). Adversarial machine learning attacks and defence methods in the cyber security domain. *ACM Computing Surveys*, *54*(5), 1–36. https://doi.org/10.1145/3453158

Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning-based phishing detection from URLs. *Expert Systems with Applications*, *117*, 345–357. https://doi.org/10.1016/j.eswa.2018.09.029

Salo, F., Injadat, M., Nassif, A. B., Shami, A., & Essex, A. (2018). Data mining techniques in intrusion detection systems: A systematic literature review. *IEEE Access*, *6*, 56046–56058. https://doi.org/10.1109/ACCESS.2018.2872784

Sarhan, M., Layeghy, S., Moustafa, N., & Portmann, M. (2021). NetFlow datasets for machine learning-based network intrusion detection systems. *Big Data Technologies and Applications*, 117–135. https://doi.org/10.1007/978-3-030-72802-1 9

Sharafaldin, I., Habibi Lashkari, A., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterisation. *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, 108–116. https://doi.org/10.5220/0006639801080116

Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., & Xu, M. (2020). A survey on machine learning techniques for cyber security in the last decade. *IEEE Access*, *8*, 222310–222354. https://doi.org/10.1109/ACCESS.2020.3041951

Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792

Subramanian, N., & Jeyaraj, A. (2018). Recent security challenges in cloud computing. *Computers & Electrical Engineering*, *71*, 28–42. https://doi.org/10.1016/j.compeleceng.2018.06.006

Syed, N. F., Shah, S. W., Shaghaghi, A., Anwar, A., Baig, Z., & Doss, R. (2022). Zero trust architecture (ZTA): A comprehensive survey. *IEEE Access*, *10*, 57143–57179.

https://doi.org/10.1109/ACCESS.2022.3174679

Tabrizchi, H., & Kuchaki Rafsanjani, M. (2020). A survey on security challenges in cloud computing: Issues, threats, and solutions. *The Journal of Supercomputing*, *76*, 9493–9532. https://doi.org/10.1007/s11227-020-03213-1

Taylor, P. J., Dargahi, T., Dehghantanha, A., Parizi, R. M., & Choo, K.-K. R. (2020). A systematic literature review of blockchain cyber security. *Digital Communications and Networks*, *6*(2), 147–156. https://doi.org/10.1016/j.dcan.2019.01.005

Thirimanne, S. P., Jayawardana, L., Yasakethu, L., Liyanaarachchi, P., & Hewage, C. (2022). Deep neural network-based real-time intrusion detection system. *SN Computer Science*, *3*, 145. https://doi.org/10.1007/s42979-022-01031-1

Tounsi, W., & Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security*, *72*, 212–233. https://doi.org/10.1016/j.cose.2017.09.001

Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers & Security*, *81*, 123–147. https://doi.org/10.1016/j.cose.2018.11.001

Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access*, *7*, 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334

Xiao, Y., Jia, Y., Liu, C., Cheng, X., Yu, J., & Lv, W. (2019). Edge computing security: State of the art and challenges. *Proceedings of the IEEE*, *107*(8), 1608–1631. https://doi.org/10.1109/JPROC.2019.2918437

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., & Wang, C. (2018). Machine learning and deep learning methods for cybersecurity. *IEEE Access*, *6*, 35365–35381. https://doi.org/10.1109/ACCESS.2018.2836950