

Review

Security Datasets for Intrusion Detection and Prevention: A Structured Review and Dataset-Selection Framework

Hakan Güler ^{1,*} , Aytug Boyacı ²  and Mustafa Ulaş ³ ¹ Department of Software Engineering, Faculty of Technology, Fırat University, Elazığ 23119, Türkiye² Department of Computer Engineering, Air Force Academy, National Defence University, Istanbul 34149, Türkiye; aytug.boyaci@msu.edu.tr³ Department of Artificial Intelligence and Data Engineering, Faculty of Engineering, Fırat University, Elazığ 23119, Türkiye; mustafaulas@firat.edu.tr

* Correspondence: hakanguler@firat.edu.tr

Abstract

Security datasets are central to the evaluation of intrusion detection and prevention systems, but their suitability differs substantially across domains, data sources, attack scenarios, labeling practices, and reproducibility conditions. This study presents a structured evidence-mapping review of security datasets reported in intrusion detection and prevention research between 2018 and 2025. The final evidence map includes 42 primary dataset-use publication records, 82 dataset or evidence-source mentions, and 69 unique datasets, corpora, or evidence sources after duplicate, retracted, and irrelevant records were removed. The review organizes datasets across network-based IDS, IPS, firewall, VPN, WAF, endpoint, email filtering, IAM, DDoS, Windows, Linux-Apache, NetFlow, cloud, and IoT/IloT security settings. In addition to adoption-frequency analysis, the study assesses representative dataset families in terms of documentation, labeling information, feature representation, class balance, public availability, reproducibility, and practical usability limitations. The findings show that established benchmarks remain widely reused, but recent studies increasingly rely on domain-specific datasets for IoT/IloT, DDoS, cloud, edge, host, and cyber-physical environments. The review also proposes and illustrates a dataset-selection framework that links dataset choice to security objectives, operational context, telemetry requirements, attack coverage, and evaluation protocol. The results support more transparent, context-aware, and reproducible dataset selection for intrusion detection and prevention research.

Keywords: security datasets; intrusion detection; intrusion prevention; dataset selection; cybersecurity benchmarking



Academic Editors: George Drosatos
and Christos Bouras

Received: 23 June 2026

Revised: 16 July 2026

Accepted: 22 July 2026

Published: 27 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The performance, reliability, and external validity of intrusion detection and prevention systems depend substantially on the quality and suitability of the datasets used for their development and evaluation. Security data may originate from network traffic, host and system logs, cloud services, web applications, IoT devices, and other operational environments. Such heterogeneous data sources provide essential evidence for detecting malicious activity, interpreting security events, and evaluating machine-learning-based security mechanisms [1,2]. However, the increasing diversity of security environments has made dataset selection more difficult, because available datasets differ in their data-

generation processes, attack coverage, labeling practices, feature representations, and levels of real-world representativeness.

1.1. Research Gap and Study Objectives

Although intrusion detection and prevention research has produced a large body of studies on system design, detection methods, and deployment environments [3–5], the evidence related to dataset selection remains fragmented. Available security datasets differ considerably in their collection environments, traffic-generation methods, attack categories, labeling procedures, feature representations, class distributions, accessibility, and documentation quality. Consequently, a dataset that is appropriate for one security objective may be unsuitable for another. This study addresses this gap through a structured review of security datasets reported in intrusion detection and prevention research between 2018 and 2025. The study aims to identify the datasets used across major security domains, compare them using common evaluation criteria, clarify dataset-adoption patterns, and provide a practical framework for selecting datasets according to security objectives and machine-learning tasks.

1.2. Scope of Security Data Sources

Security datasets used for intrusion detection and prevention are generated from heterogeneous operational environments rather than from a single type of network record. Relevant sources include network-security appliances and traffic-monitoring platforms, such as firewalls, proxy servers, virtual private networks, and web application firewalls [6–12]. Additional evidence may be obtained from endpoint-protection systems, data-loss prevention platforms, email-filtering systems, identity and access-management infrastructures, DDoS-protection mechanisms, and anti-malware solutions [13–21]. The security analysis may also rely on host and server logs, including Windows and Linux event records, web-server logs, database logs, NetFlow records, cloud-service logs, and IoT-generated data [22–31]. These sources differ substantially in their semantics, collection mechanisms, granularity, and attack visibility. Therefore, dataset selection should not be based only on dataset popularity or classification accuracy as reported in prior studies; it should also account for the target security environment, the available telemetry, the intended analytical task, and the operational relevance of the represented attacks.

1.3. Research Questions and Contributions

The review is guided by the following research questions:

- RQ1. Which security datasets have been reported in intrusion detection and prevention research published between 2018 and 2025?
- RQ2. How do these datasets differ with respect to their operational environment, traffic source, attack coverage, labeling procedure, feature representation, class balance, accessibility, and reproducibility?
- RQ3. How has dataset adoption evolved across network-based IDS, IoT/IloT, DDoS, cloud, web-application, and host-oriented security settings?
- RQ4. What recurring limitations affect the suitability of current datasets for machine-learning- and deep-learning-based security analysis?
- RQ5. Which dataset characteristics should guide dataset selection for different security objectives and analytical tasks?

This study makes four contributions. First, it consolidates datasets reported across multiple security domains within a common analytical structure. Second, it distinguishes between reviewed publications, unique datasets, and dataset-adoption frequency to avoid ambiguity in usage statistics. Third, it evaluates datasets using common criteria related

to realism, attack diversity, labeling, class balance, availability, documentation, and reproducibility. Finally, it proposes a dataset-selection framework that links dataset characteristics to security objectives, deployment environments, and machine-learning tasks.

1.4. Dataset-Oriented Perspective

The diversity of security data sources creates a fundamental dataset-selection problem. Datasets derived from network traffic, security appliances, operating-system logs, web applications, cloud platforms, and IoT devices do not represent the same operational conditions or analytical objectives. Their suitability depends on whether the intended task concerns network intrusion detection, application-layer attacks, host-based monitoring, distributed denial-of-service detection, cloud-service analysis, or IoT security. Accordingly, this review does not treat security datasets as interchangeable benchmarks. Instead, it evaluates them according to the environment from which the data originate, the type of observable evidence they provide, the attacks they represent, and the practical constraints that affect their use. This perspective provides the basis for the taxonomy, comparative analysis, and dataset-selection framework presented in the following sections.

2. Review Design and Study Selection

This study adopted a structured review design to identify and analyze publicly documented security datasets used in intrusion detection and prevention research. The review focused on studies published between 2018 and 2025 that explicitly introduced, used, or evaluated datasets relevant to intrusion detection and prevention. The identification process combined predefined keyword searches with the screening of reference lists and citation links to capture dataset-related studies across journal and conference venues.

2.1. Search Strategy

The literature identification process was conducted as a structured evidence-mapping search rather than as a database-specific systematic review. The initial discovery of candidate studies was performed primarily through Google Scholar, because the review aimed to capture dataset-use evidence distributed across journals, conference proceedings, preprint servers, dataset-provider pages, and publisher platforms. After candidate records were identified, bibliographic metadata, publication details, full-text availability, and dataset information were verified through the corresponding publisher or repository pages, including IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, Wiley Online Library, MDPI, arXiv, and official dataset-provider websites where applicable. The search focused on publications issued between 2018 and 2025. The final search update and reference verification were completed during the revision stage in July 2026. Earlier studies were retained only when they represented the original description, release, technical documentation, or foundational source of a dataset that continued to be used in studies published during the review period. The search used combinations of dataset- and security-oriented terms. Representative search expressions included the following: “intrusion detection dataset”, “intrusion prevention dataset”, “cybersecurity dataset”, “network traffic dataset”, “IoT security dataset”, “IIoT intrusion detection dataset”, “DDoS dataset”, “cloud security dataset”, “web attack dataset”, “web application firewall dataset”, “host-based intrusion dataset”, “security log dataset”, “NetFlow intrusion detection dataset”, and “edge IoT security dataset”. These expressions were combined with dataset names and domain-specific terms such as NSL-KDD, UNSW-NB15, CICIDS2017, CSE-CIC-IDS2018, CIC-DDoS2019, BoT-IoT, ToN-IoT, Edge-IIoTset, SWaT, WADI, InSDN, HTTP CSIC 2010, ISCXVPN, and Loghub.

Backward reference screening and citation-link exploration were then used to identify original dataset papers, dataset documentation, and additional dataset-use studies. Only

English-language publications or dataset documents with sufficient English bibliographic information were retained. Duplicate records were removed by comparing titles, author lists, publication years, DOIs, and dataset names. When multiple records referred to the same dataset, the dataset was treated as a single unique dataset, while each eligible publication-level use was recorded separately for adoption-frequency analysis.

2.2. Eligibility Criteria and Screening Process

The screening process was designed to identify publications that provided verifiable dataset-related evidence for intrusion detection, intrusion prevention, or closely related security-analysis tasks. Publications were considered eligible when they introduced a new security dataset or corpus, used one or more public or clearly identified datasets in an intrusion detection or prevention experiment, provided technical documentation for a dataset reused during the review period, or supplied domain-specific context necessary to interpret dataset use in a particular security environment. Records were excluded when they did not clearly identify the dataset or corpus used, did not provide dataset-based evaluation, described only general cybersecurity concepts without dataset evidence, duplicated another publication already included in the evidence map, or were unrelated to intrusion detection, intrusion prevention, security logs, traffic analysis, or cybersecurity dataset selection. During the revision stage, the corpus was re-screened to remove retracted, duplicated, and clearly irrelevant references. This additional verification step was performed to ensure that the cited studies were relevant to the corresponding security domain and that each dataset-use record could be traced to a verifiable publication or dataset source.

The retained records were classified as original dataset sources, dataset-use studies, or supporting context studies. Original dataset sources describe the release, construction, or technical documentation of a dataset. Dataset-use studies apply one or more datasets in experimental security analysis. Supporting context studies provide methodological or domain-specific background but were not counted as primary dataset-adoption evidence unless they reported a clearly identifiable dataset use. This distinction was used to avoid conflating general background literature with publication-level dataset adoption. Although the review focused on publications from 2018 to 2025, a limited number of pre-2018 references were retained when they served as original dataset descriptions, foundational benchmark sources, protocol- or flow-representation references, or technical documentation necessary to interpret datasets reused during the review period. These older references were not used to expand the temporal scope of the review. Instead, they were retained only to document the origin, structure, or technical assumptions of datasets that continued to appear in eligible 2018–2025 studies. Pre-2018 references that did not provide such dataset-origin, benchmark, or technical-context value were removed during the revision screening.

2.3. Dataset Documentation and Quality Assessment

To support a consistent dataset-level comparison, each retained corpus was examined using a documentation and usability checklist. The checklist focused on practical characteristics that directly affect the interpretation, reproducibility, and operational relevance of intrusion detection and prevention experiments. These characteristics included the collection or operational environment, traffic or telemetry source, attack coverage, labeling information, feature representation, class-balance information, public availability, documentation quality, and reproducibility support.

Each criterion was coded qualitatively as reported, partially reported, or not reported. A criterion was marked as reported when the publication or dataset documentation provided sufficient information for independent interpretation or reuse. It was marked as partially reported when the information was mentioned but incomplete, ambiguous, or de-

pendent on secondary sources. It was marked as not reported when the information could not be verified from the reviewed publication or dataset documentation. This assessment was not intended to rank datasets using a single numerical score. Instead, it was used to identify which datasets provide sufficient documentation for specific security objectives and which datasets require caution because of incomplete labeling, limited attack coverage, unclear collection assumptions, or restricted reproducibility. The corpus-level checklist results are reported later in the manuscript as a dataset documentation and usability assessment table. This table complements the study-to-dataset evidence map by showing how representative datasets differ in terms of availability, labeling, feature representation, class balance, and reproducibility. The assessment also supports the dataset-selection framework proposed in the Discussion Section.

2.4. Data Extraction and Evidence Synthesis

For each retained publication, information was extracted on the security domain, dataset or corpus name, dataset role, operational environment, data or telemetry source, traffic-generation approach, represented attack categories, labeling information, feature format, class-balance information, public availability, and dataset documentation. When the same dataset appeared in multiple publications, it was treated as one unique corpus for dataset-level analysis, while each eligible publication-level use was recorded separately for adoption-frequency analysis.

The extracted evidence was synthesized at three complementary levels. The first level was the publication level, which recorded how individual studies used or reported datasets. The second level was the corpus level, which grouped repeated mentions of the same dataset and summarized its documentation, usability, and reproducibility characteristics. The third level was the domain level, which organized datasets according to security settings such as network-based IDS, IoT/IIoT, DDoS, web-application security, host and endpoint security, NetFlow analysis, cloud systems, and log-based security.

Publication-level performance results were not directly aggregated because the reviewed studies differed substantially in preprocessing, feature engineering, train-test partitioning, class definitions, attack taxonomies, evaluation metrics, and experimental objectives. Instead, the synthesis focused on dataset characteristics, reported adoption patterns, domain suitability, and limitations that affect external validity. This approach supports the objective of the review, which is to guide dataset selection rather than to rank machine-learning models according to reported accuracy.

Figure 1 summarizes the evidence-map construction and synthesis workflow used in the review. The figure reports the final evidence-map structure rather than database-specific retrieval totals, because the review was designed as a structured evidence-mapping review focused on verifiable dataset-use evidence. Detailed corpus statistics are reported in Section 2.5, including the number of publication-level dataset-use records, dataset or evidence-source mentions, and unique datasets or corpora retained after duplicate, retracted, and irrelevant records were removed.

2.5. Review Corpus and Evidence-Map Statistics

The final evidence map was constructed from 42 primary dataset-use publication records. These records reported 82 dataset, corpus, or evaluation-evidence mentions across the reviewed security domains. After repeated mentions of the same dataset or evidence source were consolidated, the evidence map contained 69 unique datasets, corpora, or evaluation-evidence sources. These counts refer to the publication-level evidence map used for dataset adoption and domain-level analysis, rather than to the total number of bibliographic search hits returned by Google Scholar or publisher websites. Because the

review was designed as a structured evidence-mapping review rather than as a database-specific systematic review, the study does not report database-by-database retrieval totals. Instead, the reported counts represent the final cleaned and verifiable evidence corpus after duplicate, retracted, and irrelevant references were removed during the revision process. This distinction is important because the objective of the review is to compare documented dataset use and dataset suitability across security domains, rather than to estimate the total volume of cybersecurity dataset literature.

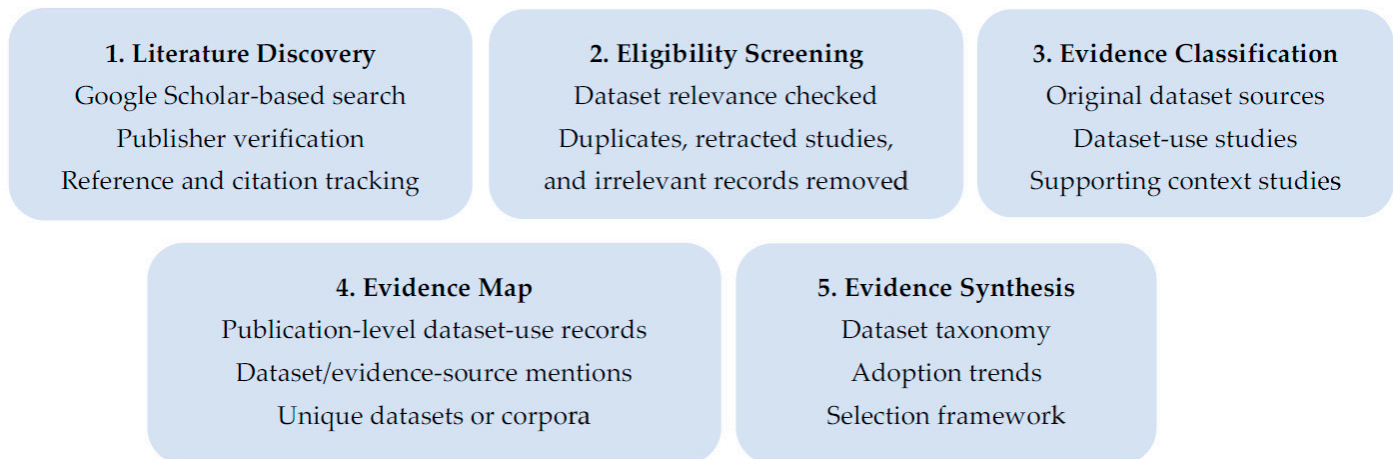


Figure 1. Structured evidence-map construction and synthesis workflow.

2.6. Review Scope and Contribution

Previous review studies have examined intrusion detection systems, automated log analysis, IoT security, and DDoS detection from domain-specific or method-specific perspectives [4,23,32–37]. These studies provide valuable insights into their respective research areas. However, their primary scopes differ from the objective of the present study, which focuses on comparing security datasets across multiple operational environments and supporting dataset selection for intrusion detection and prevention research. The present review organizes security datasets across network-based IDS, IoT/IIoT, DDoS, cloud, web-application, and host-oriented security settings. Rather than comparing machine-learning models based solely on reported accuracy values, the study evaluates datasets according to their operational environment, traffic source, attack coverage, labeling, class balance, availability, documentation, and reproducibility. These dimensions are subsequently translated into a dataset-selection framework.

3. Dataset Taxonomy and Comparative Analysis

Security datasets were organized according to the operational environments in which they are generated and the security objectives they support. The taxonomy includes network-based IDS, IoT/IIoT, DDoS, cloud and distributed systems, web-application security, and host- or endpoint-oriented analysis. This structure enables a cross-domain comparison of datasets that differ in collection mechanisms, traffic semantics, attack visibility, feature representations, and labeling practices. Within each domain, the analysis emphasizes the datasets used, their practical strengths and limitations, and the contexts in which they are appropriate. Accordingly, the following sections synthesize dataset-related evidence rather than providing a study-by-study narrative.

3.1. Network-Based IDS Datasets

Network-based IDS research continues to rely on a relatively small group of recurring benchmark datasets, particularly, NSL-KDD, UNSW-NB15, CICIDS2017, and the CIC 2018

family. These datasets have been reused in studies addressing conventional network intrusion detection, IoT-connected environments, vehicular networks, and privacy-preserving or hybrid IDS architectures [38–42]. Their continued use supports experimental comparability; however, dataset selection should also consider the represented environment, attack taxonomy, labeling process, class distribution, and temporal relevance rather than reported detection accuracy alone.

Representative studies in this group frequently evaluate IDS models using established benchmark datasets rather than introducing new datasets. For example, NSL-KDD and UNSW-NB15 were used to evaluate deep-learning-based intrusion detection models [38,40], whereas CICIDS2017, CSE-CIC-IDS2018, and related CIC datasets were adopted in host-network hybrid, vehicular, and 5G-V2X security settings [39,41,42]. These datasets remain useful because they support a comparison across studies; however, their suitability depends on the extent to which their traffic characteristics and attack categories reflect the intended deployment environment. Other studies demonstrate that dataset selection is also shaped by implementation context, including lightweight in-vehicle IDS architectures, adaptive network IDS designs, mobile-network security, and defenses against adversarial attacks [43–46]. Such studies are valuable for understanding deployment constraints and methodological choices. Nevertheless, their reported performance values are not directly aggregated in this review because differences in preprocessing, feature selection, class definitions, and evaluation protocols limit cross-study comparability. Additional evaluations using UNSW-NB15 and WSN-DS illustrate that feature-selection and classifier choices can materially affect reported detection performance. This further reinforces the need to interpret dataset results together with preprocessing choices, class definitions, and task design rather than accuracy values alone [47].

3.2. IoT and Cyber-Physical Security Datasets

IoT and cyber-physical security datasets extend conventional network benchmarks by incorporating device telemetry, industrial communication patterns, embedded platforms, and operational technology environments. In this group, studies employ or discuss datasets such as CICIDS2017, BoT-IoT, Edge-IIoTset, ToN-IoT, SWaT, WADI, NF-UNSW-NB15, and SDN-oriented dataset collections. These datasets support research on botnet detection, industrial-control-system security, federated intrusion detection, embedded-device monitoring, and cyber-physical attack analysis. Compared with conventional network IDS benchmarks, IoT and IIoT datasets may better represent heterogeneous devices, resource-constrained deployments, and domain-specific attack behavior. However, their suitability depends on the represented environment, traffic-generation process, label quality, attack diversity, and availability of contextual information. Therefore, results obtained from one IoT or industrial dataset should not be generalized automatically to other IoT, enterprise-network, or cyber-physical environments [48–56]. Recent IoT-oriented studies also extend the evaluation landscape toward DDoS detection, metaverse-connected environments, and comparative assessments of machine learning algorithms [57–59]. Such settings increase the need to distinguish between datasets designed for general network intrusion detection and those representing constrained IoT devices, high-volume DDoS traffic, or emerging application environments. Dataset selection in these cases should consider whether the available traffic, attack labels, and feature representations adequately reflect the target deployment context.

3.3. Emerging and Specialized Security Environments

Security datasets are increasingly being evaluated in specialized environments that extend beyond conventional enterprise-network intrusion detection. These include DDoS-

focused scenarios, metaverse-connected IoT environments, cyber-physical systems, healthcare infrastructures, and studies that incorporate synthetic data to address limitations in available training samples [60–66]. These specialized settings may provide more application-specific evidence than general-purpose benchmarks; however, they also introduce new challenges related to representativeness, dataset scale, label reliability, and transferability. For example, a dataset designed for healthcare, cyber-physical, or metaverse-connected systems may not adequately reflect the traffic patterns, attack behavior, or operational constraints found in conventional enterprise networks. Therefore, such datasets should be selected primarily when their deployment context aligns with the intended security objective.

3.4. Intrusion-Prevention and Deployment-Oriented Datasets

Research on intrusion prevention and deployment-oriented security systems spans adaptive recognition, cloud and IoT environments, honeypot-assisted architectures, rule-based monitoring, and programmable data-plane implementations [67–73]. These studies show that dataset suitability is influenced not only by attack labels but also by deployment constraints, including traffic visibility, response latency, hardware resources, and the point at which prevention decisions are executed. A distinct data contribution in this group is the live enterprise VoIP attack dataset introduced for machine-learning-based intrusion detection and prevention research [74]. Such operational datasets may provide stronger contextual relevance than generic benchmarks; however, their scope, accessibility, documentation, and reproducibility should be assessed carefully before they are used as general evaluation benchmarks. Existing reviews of IoT IDS/IPS systems are considered supporting context rather than primary dataset-use evidence [34,75].

3.5. Security-Device Testbeds and Web-Application Security Evidence

Firewall- and web-application-security research commonly relies on device-specific testbeds, policy configurations, traffic traces, HTTP requests, and attack simulations rather than on a small set of universally adopted benchmark datasets [76–87]. These studies address firewall performance, policy consistency, secure reconfiguration, next-generation firewall operation, application-layer DDoS detection, honeypot-assisted tuning, and threat-prevention test environments. For dataset-selection purposes, this evidence is useful primarily for identifying the telemetry that an evaluation environment should expose, including packet or flow statistics, HTTP request patterns, policy events, attack labels, and response outcomes. Web-application security additionally requires datasets that preserve request syntax, payload structure, benign application behavior, and attack context. Existing WAF research shows the relevance of both signature-based and machine-learning-supported inspection, while also highlighting the importance of attack diversity, label reliability, and reproducible evaluation procedures [78,83,84,86–88].

3.6. Proxy, VPN, and Encrypted-Traffic Security Evidence

Proxy and VPN environments create security-analysis challenges because traffic may be encrypted, mediated, cached, or transformed before it is observed by a monitoring system. In these settings, dataset suitability depends on the visibility of usable traffic features, the availability of attack labels, and the extent to which the collected data represent realistic deployment conditions. Reverse-proxy security research illustrates the relevance of request- and transaction-level evidence for detecting web-based threats, whereas VPN traffic-identification studies demonstrate the importance of flow-level and encrypted-traffic features [89,90]. Accordingly, datasets intended for proxy or VPN-related security analysis should document the observation point, encryption assumptions, traffic granularity, represented applications, and labeling method. Without this information, the results may not

transfer reliably to enterprise gateways, remote-access infrastructures, or encrypted-traffic monitoring scenarios.

3.7. Web-Application Security Datasets

Web-application security requires datasets that preserve the structure and semantics of HTTP requests, user inputs, application responses, and malicious payloads. In contrast to flow-based network datasets, web-security datasets should capture application-layer attack characteristics, including SQL injection, cross-site scripting, phishing-related requests, malicious automation, and evasion attempts [91–104]. Studies on machine-learning-supported web application firewalls show that dataset quality is strongly affected by the diversity of benign requests, payload obfuscation techniques, attack-label granularity, and the realism of the underlying web-application environment [92–104]. Therefore, datasets intended for WAF evaluation should not be selected solely according to classification accuracy. They should also document request sources, preprocessing procedures, attack-generation methods, application context, and the availability of reproducible train–test splits.

3.8. Host- and Endpoint-Oriented Security Evidence

Host- and endpoint-oriented security analysis relies on telemetry generated by user devices, operating systems, endpoint detection and response platforms, and malware-protection mechanisms. Relevant evidence may include process activity, file and registry events, authentication records, command execution traces, endpoint alerts, and remediation outcomes. Studies addressing endpoint-security systems, ransomware pre-attack detection, persistent threats, device-mediated data theft, attack experiments against endpoint protection, and insider-threat mitigation demonstrate the diversity of such evidence sources [105–110]. Compared with widely reused network benchmarks, endpoint-oriented datasets are often more dependent on the monitored operating system, deployed protection stack, user behavior, and attack scenario. Their suitability should therefore be assessed according to the available telemetry, label provenance, attack lifecycle coverage, privacy constraints, and reproducibility of the collection environment. Dataset selection is particularly important when the intended task involves malware detection, ransomware analysis, insider-threat monitoring, or host-based intrusion detection.

3.9. Data-Loss Prevention and Information-Flow Evidence

Data-loss prevention research relies on evidence related to sensitive-data identification, information flows, user actions, access policies, and data-transfer events. Datasets and evaluation environments in this area should capture both legitimate and policy-violating behavior while preserving sufficient context to distinguish accidental exposure from malicious exfiltration [111–115]. For dataset selection, the main considerations include the type of protected information, the monitored communication or storage channel, the granularity of user and system activity, label provenance, and privacy constraints. Because real DLP data may contain sensitive organizational information, reproducible public datasets are comparatively limited. Consequently, researchers should report the collection context, anonymization procedure, class definitions, and policy rules used to generate labels.

3.10. Email, Phishing, and Message-Based Security Evidence

Email and phishing security research relies on message-level evidence, including email headers, sender and domain attributes, URLs, attachments, body text, thread context, and user-interaction signals. Studies addressing spam, phishing, spoofing, disposable-email use, subscription-email bombing, and malicious-email detection indicate that data representation and label definitions materially influence evaluation outcomes [116–133]. For dataset selection, researchers should consider whether the data preserve realistic benign communi-

cation patterns alongside malicious messages, whether labels distinguish spam, phishing, spoofing, and other email-borne threats, and whether privacy-preserving preprocessing has altered important contextual information. Dataset documentation should also report the collection source, language coverage, class distribution, feature-extraction process, and train–test split strategy. These details are essential because models trained on isolated message content may not generalize reliably to operational email-security environments.

3.11. Identity and Access Management Security Evidence

Identity and access management research relies on security evidence related to authentication events, authorization decisions, access-control policies, user and device identities, credential use, and privilege changes. Such evidence may originate from cloud platforms, enterprise systems, IoT environments, healthcare infrastructures, and decentralized identity frameworks [134–147]. For dataset selection, the key considerations include whether identity-related records preserve temporal sequences of access events, the granularity of authentication and authorization labels, the representation of normal and malicious behavior, and the privacy constraints associated with personal or biometric information. Datasets intended for IAM security analysis should also document the access-control model, monitored assets, identity lifecycle stage, and policy context. These details are necessary to distinguish anomalous access behavior from legitimate operational variation.

3.12. DDoS Security Datasets and Evaluation Evidence

DDoS-oriented security research covers heterogeneous deployment settings, including IoT networks, SDN infrastructures, cloud services, 5G environments, vehicular systems, electric-vehicle charging stations, and critical multi-energy systems [36,37,148–163]. These settings differ substantially in traffic volume, protocol composition, attack intensity, topology, and mitigation requirements. Consequently, a DDoS dataset that is suitable for one environment may not provide a valid basis for evaluating a model intended for another. DDoS dataset selection should therefore consider the represented attack families, traffic-generation process, observation level, feature format, class distribution, and temporal characteristics of the attack traces. Publicly documented datasets, such as the smart-home DDoS dataset reported for Fibaro Home Center 3 environments, are particularly valuable when they provide clear collection procedures, labels, and contextual information [164]. Prior reviews also emphasize the diversity of DDoS detection and mitigation environments, reinforcing the need for deployment-aware dataset selection rather than the reliance on a single generic benchmark [35–37].

3.13. Malware and Anti-Malware Evaluation Evidence

Malware and anti-malware evaluation requires evidence that captures malicious behavior, evasion techniques, fileless execution, host activity, and the defensive response of security controls. The relevant studies examine Android malware evasion, fileless malware behavior, anti-malware effectiveness, malware-analysis approaches, and verification-oriented protection mechanisms [165–171]. For dataset selection, the main considerations include whether the data represent static artifacts, dynamic execution traces, behavioral indicators, evasion techniques, and realistic benign activity. Malware-related datasets should also document the collection source, labeling procedure, malware family or attack category, execution environment, and potential data leakage between training and test samples. These requirements are particularly important because overly controlled malware datasets may not reflect the operational behavior encountered by endpoint-protection systems.

3.14. System, Host, and Log-Based Security Datasets

System- and host-based security analysis relies on event records generated by operating systems, applications, authentication services, audit mechanisms, and security-monitoring tools. Relevant evidence may include Windows event logs, Linux system and audit logs, command histories, process activity, file-access events, application traces, and multi-source operational logs [172–184]. These datasets support alert detection, anomaly detection, malware analysis, forensic investigation, and advanced persistent threat monitoring. Public resources such as Loghub and the recently reported Linux APT evaluation datasets are particularly valuable because they provide reusable log collections and documented security scenarios [183,185]. However, system-log datasets may vary substantially in log format, event semantics, collection period, label quality, and privacy treatment. For dataset selection, researchers should assess the monitored operating system, log source, event granularity, attack lifecycle coverage, labeling approach, anonymization procedure, and availability of parsing or preprocessing guidance. Secure log-collection mechanisms and integrity protections may also influence whether a dataset is suitable for forensic or incident-response research [186].

3.15. Virtualized and Edge-Cloud Security Environments

Virtualized, edge, and cloud-native environments introduce additional constraints for security-data collection and intrusion-detection evaluation. In these environments, traffic may traverse virtual switches, containers, network functions, edge nodes, and software-defined infrastructure components. Consequently, datasets should document the virtualization layer, deployment topology, monitoring location, resource constraints, and the relationship between observed events and the underlying workload [187–192]. Security evaluations in such environments should also consider whether the collected data reflect containerized, virtual-machine-based, or edge-based deployments. Differences in isolation mechanisms, network visibility, workload mobility, and resource allocation may affect both attack observability and model transferability. Therefore, datasets obtained from virtualized environments should be selected only when their infrastructure assumptions are compatible with the intended deployment scenario.

3.16. Web-Server Log Security Datasets

Web-server security analysis relies on access logs, error logs, request paths, client identifiers, timestamps, response codes, and server-side event traces. Such evidence can support the detection of attack-oriented scanning, suspicious request behavior, application-layer anomalies, and vulnerability-exploitation attempts [193–196]. For dataset selection, web-server log collections should document the server configuration, logging format, observation period, benign workload characteristics, attack-generation process, and label definitions. These details are essential because access logs collected from a controlled environment may not represent the diversity of normal users, automated clients, web crawlers, or adversarial behaviors encountered in operational deployments. Dataset usability also depends on whether raw logs, parsing guidance, and reproducible train–test splits are available.

3.17. Database and Transaction-Log Security Evidence

Log-based and transaction-oriented security evidence can support intrusion-detection and forensic analysis by preserving event sequences, access behavior, system states, audit records, and recovery-related actions. Unlike packet- or flow-level traffic, these records are usually tied to a specific operating system, application, database platform, or cloud service. For dataset selection, such collections should document the platform, logging configuration,

event granularity, query or event representation, label definitions, and the relationship between normal and suspicious activity. Because log and transaction records may contain sensitive organizational or personal information, anonymization and access restrictions can limit public availability and reproducibility. Therefore, studies using log-based evidence should clearly report how records were collected, sanitized, labeled, and partitioned for evaluation [174,177,181–183,197–203].

3.18. NetFlow and Flow-Based Security Datasets

Flow-based security datasets aggregate network communications into records that describe properties such as source and destination information, protocol, duration, packet counts, byte volumes, and timing characteristics. These datasets are widely used for traffic analysis, user identification, brute-force detection, DDoS detection, malware-related traffic analysis, and large-scale IoT network monitoring [204–215]. The suitability of a flow-based dataset depends strongly on how flows are generated and exported. Dataset documentation should therefore report the flow definition, exporter or protocol format, sampling strategy, aggregation settings, timeout values, monitored topology, feature set, labeling procedure, and observation period. These details are important because sampling, aggregation, NAT effects, and deployment choices can substantially alter traffic characteristics and reduce direct comparability across datasets [205,209,216,217]. For intrusion detection and prevention research, flow-based datasets are particularly useful when packet payloads are unavailable, encrypted, or impractical to process at scale. However, researchers should verify that the represented attacks, traffic diversity, and flow-generation assumptions are compatible with the intended operational environment.

3.19. Cloud and Distributed-System Security Evidence

Cloud and distributed-system security analysis relies on logs generated by virtual machines, containers, cloud services, orchestration platforms, network functions, and distributed application components. Relevant evidence may include exception logs, audit records, workflow traces, monitoring events, metrics, and multi-source operational logs [202,203,218–227]. These data sources support cloud forensics, anomaly detection, execution-failure diagnosis, workflow reconstruction, and incident investigation. However, cloud-log datasets may differ substantially in deployment topology, service model, logging format, tenant isolation, event granularity, and access restrictions. For this reason, datasets collected from cloud environments should document the cloud platform, monitored components, collection period, data-sanitization method, event labels, and relationship between normal operations and security-relevant incidents. For intrusion detection and prevention research, cloud datasets are most useful when they preserve correlations between events generated by multiple components and provide sufficient contextual information to reconstruct attack sequences or abnormal operational behavior. Researchers should therefore assess the availability of raw records, parsing guidance, labeling procedures, and reproducible evaluation splits before using cloud logs as benchmark data.

3.20. IoT Log and Edge-Security Evidence

IoT and edge-security datasets may combine network traffic, device telemetry, distributed logs, protocol events, and context generated by constrained devices. Such evidence is relevant to intrusion detection, attack identification, IoT forensics, adaptive edge-security policies, and blockchain-assisted IDS configurations [228–233]. Compared with conventional network datasets, IoT log collections should document the device types, communication protocols, deployment topology, observation point, data-collection period, attack-generation process, and labeling method. These details are essential because device heterogeneity, limited resources, encrypted communication, and distributed data collection

can substantially affect both attack visibility and model transferability. Accordingly, IoT and edge-security datasets should be selected only when their operational assumptions align with the intended deployment environment.

3.21. Study-to-Dataset Evidence Map

Table 1 presents the cleaned study-to-dataset evidence map used in this review. To improve readability and to support domain-level interpretation, the evidence map is organized into domain-specific blocks rather than as a single undifferentiated list. These blocks cover network-based IDS and IoT/IIoT datasets, intrusion prevention and perimeter-security datasets, web and application-security datasets, host-, endpoint-, email-, and identity-related datasets, DDoS datasets, log and NetFlow datasets, and cloud, edge, and cyber-physical security datasets.

Table 1. Domain-grouped study-to-dataset evidence map derived from the reviewed publications.

Security Domain	Dataset(s) Reported	Representative Study	Reference	Year	Dataset-Related Evidence
IDS	UNSW-NB15	An automated intrusion detection system in IoT system using attention-based deep bidirectional sparse auto encoder model	Swathi et al. [38]	2024	Superior to traditional techniques; attack detection in IoT environments.
	CH, SA-KIA Soul, SA-Chevrolet Spark (Car Hacking Dataset)	BIDS: An efficient intrusion detection system for in-vehicle networks using a two-stage binarised neural network on low-cost FPGA	A. Rangsikunpum et al. [43]	2024	Known: Fuzzy, RPM Spoofing, Gear Spoofing; Unknown: DoS.
	BoT-IoT; CSE-CIC-IDS2018; ToN-IoT; UNSW-NB15; NSL-KDD	A privacy-preserving self-supervised-learning-based intrusion detection system for 5G-V2X networks. Ad hoc networks.	S. Hossain et al. [39]	2025	FL/SSL; IoT attack detection.
	CSE-CIC-IDS2018; NDSec-1	Machine-learning-enabled hybrid intrusion detection system with host data transformation and an advanced two-stage classifier	Chen et al. [41]	2024	Stream-event-message combined features; binary and multi-class.
	CICIDS2017; CSE-CIC-IDS2018; CIC-DDoS2019	Apollon: a robust defense system against adversarial machine-learning attacks in intrusion detection systems	Paya et al. [46]	2024	Comparison with various ML classifiers.
	CICIDS2017; BoT-IoT-2018	CyberAIBot: Artificial Intelligence in an intrusion detection system for CyberSecurity in the IoT	Serrano et al. [48]	2025	DL Management + Technical Cluster; attack detection.
	Edge-IIoTset	Deep-learning-enabled intrusion detection system for industrial IOT environment	Nandanwar et al. [49]	2024	Conv1D + GRU; for IIoT.

Table 1. Cont.

Security Domain	Dataset(s) Reported	Representative Study	Reference	Year	Dataset-Related Evidence
IDS	SWaT; WADI	A feature selection based on genetic algorithm for intrusion detection of industrial control systems	Y. Fang, Y. Yao, X. Lin et al. [51]	2024	GA-based feature selection; ICS.
	CIC-ToN-IoT; NF-CSE-CIC-IDS2018; NF-UNSW-NB15; InSDN; InSecLab-IDS2021	Fed-Evolver: An automated evolving approach for federated intrusion detection system using adversarial autoencoder in SDN-enabled networks	Duy et al. [53]	2024	FL + AAE; stable performance with few labels.
	Edge-IIoTset	A convolutional autoencoder architecture for robust network intrusion detection in embedded systems.	Borgioli et al. [55]	2022	Embedded environment; single-class training; poisoning resistance.
	Edge-IIoTset; ToN-IoT	Advancing IoT security: A novel intrusion detection system for evolving threats in Industry 4.0 using optimized convolutional sparse Ficks law graph point trans-Net.	Mathina et al. [56]	2025	Real-time packet-based detection; poisoning resistance.
	BoT-IoT; ToN-IoT	Metaverse-IDS: Deep-learning-based intrusion detection system for Metaverse-IoT networks	Gaber et al. [32]	2023	Metaverse-IoT traffic; 12 attack classes.
	UNSW-NB15; NSL-KDD; BoT-IoT	SYN-GAN: A robust intrusion detection system using GAN-based synthetic data for IoT security	Rahman et al. [64]	2024	GAN-based training with synthetic data; low latency.
IPS	NSL-KDD; UNSW-NB15	AIRP: A system-based selection of adaptive intrusion recognition and prevention	Alqahtani et al. [234]	2021	Probe/DoS/U2R/R2L included; selection + detection.
	Aposemat IoT-23	PS-IPS: Deploying intrusion prevention system with machine learning on programmable switch.	Lee et al. [76]	2024	ML-based IPS on programmable switches; benign/malicious traffic.
Firewall	OWASP 2022; GetSimple CMS; Piwigo; OpenCATS	Dynamic web application firewall detection supported by cyber mimic defense approach.	Mariusz Sepczuk [79]	2023	XSS, SQLi, intelligence gathering; ModSecurity + OWASP CRS 3.3.2.
	Artificial topologies and policies; realistic scenarios	Optimizing distributed firewall reconfiguration transients.	Daniele Brighenti, Fulvio Valenza [81]	2022	University, data center, GÉANT2, Internet2; FATO/MaxSMT.

Table 1. Cont.

Security Domain	Dataset(s) Reported	Representative Study	Reference	Year	Dataset-Related Evidence
Firewall	Traffic generated with fit.py	Toward a modern secure network based on next-generation firewalls: recommendations and best practices.	Oussama Lamdakkara et al. [82]	2024	Architectural resilience; detection/blocking reports.
	SDSC; CLARKNET; own attack traces	HTTP-request-pattern-based signatures for early application layer DDoS detection: A firewall agnostic approach.	Amit Praseed, P. Santhi Thilagam [84]	2022	A1–A7 application layer DDoS; SampEn/PTA.
	Wireshark captured packets	DDoS attack detection model using machine-learning algorithm in next generation firewall.	Manthan Patel et al. [85]	2024	ML-based DDoS detection with header features.
	Real FW configurations; synthetic policies	Can my firewall system enforce this policy?	Lorenzo Ceragioli, Pierpaolo Degano, Letterio Galletta [86]	2022	IFCL; iptables/ipfw/pf expressiveness comparison.
	Ransomware dataset; honeynet outputs	H-DOCTOR: Honeypot based firewall tuning for attack prevention. Measurement: sensors.	M.R. Amal, P. Venkadesh [87]	2023	Ransomware behavior; tuning and accuracy.
	Self-multicast traffic and attacks	A suggested testbed to evaluate multicast network and threat prevention performance of next-generation firewalls.	Göksel Uçtu et al. [88]	2021	0–10 Gbps multicast; ZAPROXY; custom malware/exploits.
VPN	ISCVPN	AAE-DSVDD: A one-class classification model for VPN traffic identification.	Sicai Lv et al. [235]	2023	7 traditional encrypted traffic + 7 VPNs; sequential features derived from 23 features.
WAF	HTTP CSIC 2010	Web application firewall based on anomaly detection using deep learning	Sezer Toprak, Ali Gökhan Yavuz [10]	2022	Normally labeled data is used in training/testing.
Endpoint Protection	Malware Images: Visualization and Automatic Classification	Digital forensics as advanced ransomware pre-attack detection algorithm for endpoint data protection.	Jian Du et al. [107]	2022	25 ransom family clusters; 56 columns, 65,535 rows.
Email Filtering	Enron	A novel and secured email classification and emotion detection using hybrid deep neural network.	A. Poobalan et al. [124]	2024	Six datasets from Enron; spam/raw classification.
	SpamText; LingSpam; SpamAssassin; ENTR; GS	Email spam detection using hierarchical attention hybrid deep-learning method.	Sultan Zavrak et al. [125]	2023	Same- and cross-dataset experiments; spam detection.
IAM	IITDv1; MMUv1	Privacy-preserving cancelable multi-biometrics for identity information management.	Zhou et al. [145]	2025	NIR 320 × 240; 224 subjects, 2240 images.

Table 1. Cont.

Security Domain	Dataset(s) Reported	Representative Study	Reference	Year	Dataset-Related Evidence
DDoS	CICEV2023	DDoS detection in electric vehicle charging stations: A deep-learning perspective via CICEV2023 dataset.	Anlı et al. [152]	2024	Normal/attack data at EV charging stations.
	CIC-DDoS2019; KDD-CUP	Machine-learning-based detection of DDoS attacks on IoT devices in multi-energy systems.	H.A. Sakr et al. [155]	2024	Training splits 80%/60%/40%; Acc/Prec/Rec/F1.
	InSDN; Kaggle DDoS	DDoS attack detection and mitigation using deep neural network in SDN environment.	Hnamte et al. [157]	2024	Detection/mitigation with DNN on SDN.
Windows	BG/L; Hadoop	AutoLog: Anomaly detection by deep autoencoding of system logs	Marta Catillo et al. [174]	2022	215 transaction logs containing hardware/software errors.
	Windows Audit Logs	Malware detection on windows audit logs using LSTMs.	Markus Ring et al. [177]	2021	Normal and malicious activities; public dataset.
Linux-Apache	Successful CI/CD Deployment Logs	Anomaly detection for Linux system log	Rongjun Ma [181]	2020	Linux logs collected during CI/CD; training successful deployments.
	Linux System Logs	Deep learning for anomaly detection in Linux system log	Minttu Susanna Mäkinen [182]	2019	73 log files; truncated to 10k lines; 3145 key types.
	Various Operating System Logs	Anomaly detection in operating system logs with deep-learning-based sentiment analysis	Hudan Studiawan et al. [183]	2020	Logs from multiple OSES (Casper, Hadoop, Spark, Windows, etc.).
NetFlow	BoT-IoT; ToN-IoT; CSE-CIC-IDS2018	Intrusion detection system for large-scale IoT NetFlow networks using machine learning with modified arithmetic optimization algorithm.	S. Fraihat et al. [213]	2023	NF-BoT-IoT-v2; NetFlow features from packet captures.
	UPC-I	Analysis of the impact of sampling on NetFlow traffic classification.	V. Carela-Español et al. [210]	2011	Effect of sampling on classification.
Cloud	EXi; Ci; COR(Ri); D(Ri)	An approach to cloud execution failure diagnosis based on exception logs in OpenStack	Pinjia He, Jieming Zhu, Shilin He, Jian Li, Michael R. Lyu [202]	2018	Exception logs and node information associated with request IDs.
	logfiles	Cloud forensics: Evidence collection and preliminary analysis	N. Dalezios, S. Shiales, N. Kolokotronis [203]	2020	73-sample relational dataset analyzed with furthest-first clustering.
IoT	ToN-IoT; CICIDS2017	Sandpiper optimization with hybrid deep-learning model for blockchain-assisted intrusion detection in IoT environment	Alkhonaini et al. [236]	2024	For Industry 4.0/IoT/IIoT; OS telemetry + network traffic.

Table 1 should be interpreted as a domain-grouped publication-level evidence map rather than as a direct count of unique datasets. A single dataset may be reported in multiple studies, while one publication may use more than one dataset. Therefore, the evidence map distinguishes publication-level dataset use from a unique dataset identity and provides the basis for the adoption-frequency, temporal, and domain-level analyses presented in the following section. The domain-grouped structure is intended to improve readability and to avoid treating datasets from substantially different operational settings as directly interchangeable.

Table 2 shows that widely reused network and IoT datasets generally provide a stronger documentation of the collection environment, attack coverage, feature representation, and public availability. However, several dataset families still require caution because of outdated traffic profiles, class imbalance, domain-specific collection assumptions, limited transferability, or incomplete reproducibility information. This assessment supports the later dataset-selection framework by linking dataset choice to documentation quality, operational context, and intended analytical use.

Table 2. Dataset documentation and usability assessment for representative corpus families.

Dataset/Corpus	Main Domain	Environment/Source	Attack Coverage and Labeling	Features/Class Balance	Availability and Reproducibility	Main Usability Caution
NSL-KDD	Network IDS	Reported; benchmark network traffic	Reported; labeled attack categories	Partially reported; known class-distribution limitations	Reported; widely available	Outdated traffic and limited realism
UNSW-NB15	Network IDS	Reported; synthetic and real network traffic	Reported; multiple attack families	Reported; feature set documented	Reported; publicly available	Controlled generation assumptions should be considered
CICIDS2017	Network IDS	Reported; enterprise-like network environment	Reported; multiple attack scenarios	Reported; flow features available, imbalance present	Reported; publicly available	Preprocessing choices affect comparability
CSE-CIC-IDS2018	Network IDS/host-network	Reported; large-scale network and host-related setting	Reported; multiple attack categories	Reported; high-volume feature data	Reported; publicly available	Large-scale preprocessing and class imbalance require caution
CIC-DDoS2019	DDoS	Reported; DDoS-focused traffic generation	Reported; DDoS attack families	Reported; flow-based features	Reported; publicly available	Mainly suitable for DDoS-specific evaluation
BoT-IoT	IoT/IIoT	Reported; IoT testbed environment	Reported; botnet and IoT attack categories	Reported; strong class imbalance	Reported; publicly available	Extreme imbalance may inflate performance if not handled
ToN-IoT	IoT/IIoT	Reported; network, telemetry, and OS data	Reported; heterogeneous attack classes	Reported; multimodal features	Reported; publicly available	Cross-source integration and preprocessing should be documented

Table 2. Cont.

Dataset/Corpus	Main Domain	Environment/Source	Attack Coverage and Labeling	Features/Class Balance	Availability and Reproducibility	Main Usability Caution
Edge-IIoTset	IIoT/edge	Reported; edge and IIoT environment	Reported; multiple IIoT/IoT attacks	Reported; feature representation documented	Reported; publicly available	Environment-specific assumptions limit generalization
SWaT	Industrial control systems	Reported; water-treatment testbed	Reported; process-oriented attacks	Reported; sensor and actuator data	Partially reported; access and use conditions may vary	ICS-specific process behavior limits transferability
WADI	Industrial control systems	Reported; water-distribution testbed	Reported; physical-process attacks	Reported; sensor and process data	Partially reported; access and use conditions may vary	Domain-specific process dynamics should be considered
InSDN	SDN security	Reported; software-defined network environment	Reported; SDN-related attacks	Reported; flow-level features	Reported; publicly available	SDN topology assumptions affect transferability
N-BaIoT	IoT botnet	Reported; IoT device traffic	Reported; botnet attack behavior	Reported; device-level feature sets	Reported; publicly available	Device-specific behavior may not generalize to all IoT systems
HTTP CSIC 2010	Web/WAF	Reported; HTTP request dataset	Reported; normal and anomalous web requests	Partially reported; request-level features depend on preprocessing	Reported; publicly available	Older web-application behavior and payloads may limit realism
ISCXVPN	VPN/encrypted traffic	Reported; VPN and non-VPN traffic	Partially reported; focused on traffic identification rather than attacks	Reported; flow-level features	Reported; publicly available	More suitable for VPN traffic classification than intrusion detection
Loghub	Log-based analysis	Reported; multi-system log collections	Partially reported; labels vary by log source	Partially reported; heterogeneous log formats	Reported; publicly available	Not all included logs are security-specific
Windows Audit Logs	Host security	Reported; Windows event/audit records	Reported; benign and malicious activity labels	Reported; event-sequence features	Reported; publicly available	Operating-system and scenario dependence should be considered
CICEV2023	DDoS/EV charging	Reported; electric-vehicle charging context	Reported; normal and attack traffic	Reported; deep-learning-oriented evaluation data	Reported; publicly available	Highly relevant to EV charging but not general network IDS

Table 2. *Cont.*

Dataset/ Corpus	Main Domain	Environment/ Source	Attack Coverage and Labeling	Features/Class Balance	Availability and Reproducibility	Main Usability Caution
CICIoT2023	IoT security	Reported; large-scale IoT environment	Reported; multiple IoT attack scenarios	Reported; large-scale feature data	Reported; publicly available	Large-scale data handling and preprocessing must be documented
ROSPaCe	CPS/IoT	Reported; ROS2-based cyber-physical and IoT environment	Reported; intrusion scenarios	Reported; multi-layer CPS/IoT data	Reported; publicly available	Robotics/CPS- specific assumptions limit direct transfer

4. Dataset Adoption Trends

This section analyzes dataset adoption using the study-to-dataset evidence map presented in Table 1. The analysis distinguishes three concepts: the number of reviewed publications, the number of unique datasets, and the frequency with which individual datasets were reported across the reviewed studies. This distinction is necessary because repeated use of a dataset does not represent the introduction of a new dataset.

4.1. Dataset Adoption Frequency

Dataset adoption frequency was calculated as the number of retained publication records in Table 1 that reported a given dataset. A publication using multiple datasets contributed one count to each reported dataset, whereas repeated mentions of the same dataset within a single publication were counted once. Context-only reviews, retracted publications, and records lacking verifiable dataset information were not considered primary dataset-use evidence. The resulting frequencies should be interpreted as adoption counts within the reviewed corpus rather than as global measures of dataset use across all cybersecurity literature. Dataset name variants or shorthand forms referring to the same official corpus were standardized and merged. Derived or flow-based versions were considered separately only when they represented distinct data representations, preprocessing pipelines, or collection processes. Table 3 reports the most frequently adopted datasets in the primary study-to-dataset evidence map.

Table 3. Dataset adoption frequency in the primary study-to-dataset evidence map.

Dataset	Number of Included Studies	Main Security Settings
BoT-IoT	5	IoT/IIoT, network IDS
ToN-IoT	5	IoT/IIoT, network IDS
UNSW-NB15	5	Network IDS, IPS, flow-based IDS
CSE-CIC-IDS2018/CICIDS2018	4	Network IDS, host-network analysis
NSL-KDD	3	Network IDS, IPS
CICIDS2017	3	Network IDS, IoT-related IDS
Edge-IIoTset	3	IIoT, embedded-device IDS
CIC-DDoS2019	2	Network IDS, DDoS detection
InSDN	2	SDN-based DDoS and federated IDS

To address the temporal component of RQ3, the publication-level dataset-use records were also grouped by publication year and security domain. Table 4 reports the year-by-year distribution of the primary dataset-use records included in the evidence map.

The table should be interpreted as a temporal distribution within the reviewed corpus rather than as a complete estimate of all cybersecurity dataset publications in each year. One pre-2018 foundational NetFlow reference retained for methodological context is not included in the annual totals.

Table 4. Temporal distribution of primary dataset-use records by security domain, 2018–2025.

Year	Network IDS/IoT-IIoT	IPS/Firewall/ WAF/VPN	DDoS	Host/Endpoint/ Email/IAM	Log/NetFlow/ Cloud/Edge	Total Records
2018	0	0	0	0	1	1
2019	0	0	0	0	1	1
2020	0	0	0	0	3	3
2021	0	2	0	1	0	3
2022	1	4	0	1	1	7
2023	1	3	0	1	1	6
2024	8	3	3	1	1	16
2025	3	0	0	1	0	4
Total	13	12	3	5	8	41

The temporal distribution shows that the reviewed evidence shifted from log-, cloud-, and system-oriented studies in the earlier part of the review period toward IoT/IIoT, DDoS, edge, and domain-specific intrusion-detection datasets in more recent years. The strongest concentration appears in 2024, where the evidence map includes a larger number of IoT/IIoT, network IDS, DDoS, and deployment-oriented studies. This supports the interpretation that dataset adoption has broadened beyond conventional network IDS benchmarks toward specialized and operationally contextualized security datasets.

4.2. Dataset Adoption by Security Domain

Dataset adoption varies across security domains because different environments require different types of telemetry, attack coverage, and labeling detail. Network-based IDS studies predominantly reuse established benchmark datasets, whereas IoT/IIoT and DDoS research increasingly relies on domain-specific datasets. In contrast, firewall, proxy, VPN, WAF, endpoint, email, IAM, and cloud-security studies often use scenario-specific traces, operational logs, or custom testbeds rather than a common benchmark corpus. The domain-level analysis should therefore be interpreted as an overview of dataset suitability and reported adoption patterns, rather than as a universal ranking of dataset quality. Table 5 summarizes the main dataset types and evidence sources associated with each security domain.

Table 5. Dataset characteristics and selection considerations by security domain.

Security Domain	Representative Dataset or Evidence Sources	Key Dataset-Selection Considerations
Network-Based IDS	NSL-KDD; UNSW-NB15; CICIDS2017; CICIDS2018; CSE-CIC-IDS2018	Attack taxonomy, traffic realism, feature representation, class balance, temporal relevance
Intrusion Prevention Systems	NSL-KDD; UNSW-NB15; operational prevention traces; programmable-switch testbeds	Response latency, prevention point, traffic visibility, attack labels, deployment constraints

Table 5. *Cont.*

Security Domain	Representative Dataset or Evidence Sources	Key Dataset-Selection Considerations
IoT/IIoT/Cyber-Physical Security	BoT-IoT; ToN-IoT; Edge-IIoTset; SWaT; WADI; N-BaIoT	Device diversity, protocol coverage, operational context, attack-generation process, telemetry availability
DDoS Detection	CIC-DDoS2019; CICIDS2017; CICEV2023; smart-home and SDN traces	Attack intensity, traffic volume, topology, class distribution, observation level
Web and WAF Security	CSIC 2010; HTTP request traces; web-application logs; attack payload collections	Request semantics, benign traffic diversity, payload realism, label granularity, reproducibility
Proxy and VPN Security	Reverse-proxy traces; encrypted-traffic flows; ISCXVPN; VPN traffic records	Encryption assumptions, observation point, flow features, application diversity, labeling procedure
Host and Endpoint Security	Windows event logs; Linux audit logs; malware traces; endpoint telemetry	Event granularity, operating-system context, privacy treatment, attack lifecycle coverage
Email, DLP, and IAM Security	Email headers and message corpora; information-flow records; authentication and access logs	Sensitive-data protection, privacy constraints, label provenance, normal-behavior diversity
Cloud and Distributed Systems	Cloud audit logs; workflow traces; orchestration logs; multi-source operational records	Service model, topology, tenant isolation, event correlation, access restrictions
NetFlow and Flow-Based Analysis	NetFlow/IPFIX records; flow traces; aggregated traffic datasets	Flow definition, sampling settings, aggregation policy, timeout values, topology, feature availability

4.3. Dataset Characteristics and Methodological Suitability

Dataset characteristics influence the types of analytical methods for which a dataset is practically suitable. Structured benchmark datasets with established feature sets and clearly defined classes are frequently used to evaluate conventional machine-learning pipelines, including feature selection, classification, and baseline comparison. In contrast, datasets containing high-volume traffic, temporal dependencies, heterogeneous IoT telemetry, or complex attack patterns may be more suitable for deep-learning and distributed-learning experiments. This relationship should not be interpreted as a direct comparison of model superiority. Reported performance values are affected by preprocessing, feature engineering, train–test splits, class definitions, and evaluation metrics. Accordingly, Table 6 summarizes the methodological settings for which representative datasets are commonly suitable, rather than ranking machine-learning methods according to the reported accuracy.

Table 6. Representative dataset characteristics and methodological suitability.

Analytical Setting	Suitable Dataset Characteristics	Representative Examples	Main Caution
Conventional Machine Learning	Structured tabular or flow-based features; clearly defined classes; manageable dataset size	NSL-KDD; UNSW-NB15; CICIDS2017	Results depend on preprocessing, feature selection, and class definitions

Table 6. *Cont.*

Analytical Setting	Suitable Dataset Characteristics	Representative Examples	Main Caution
Deep Learning	High-volume traffic; temporal dependencies; heterogeneous features; complex attack patterns	BoT-IoT; ToN-IoT; CICIDS2018; Edge-IIoTset	High accuracy may not transfer to unseen attacks or other environments
Federated and Distributed Learning	Partitionable data; multiple sources or nodes; decentralized or SDN-related settings	InSDN; CIC-ToN-IoT; NF-UNSW-NB15	Partition strategy, non-IID data, and communication assumptions must be reported
Host and Log-Based Analysis	Event sequences; timestamps; process, audit, or authentication records	Windows event logs; Linux audit logs; Loghub collections	Log semantics, anonymization, and label quality may limit comparability

5. Discussion

5.1. Interpretation of Dataset Adoption Patterns

The evidence map indicates that dataset adoption is concentrated around a limited number of established benchmarks, while the range of domain-specific datasets has broadened in recent studies. Established datasets continue to be reused because they provide familiar feature structures, accessible benchmark settings, and a basis for comparison with previous work. At the same time, the increasing presence of IoT/IIoT, DDoS, cloud, web-application, and host-oriented datasets reflects a shift toward deployment-specific evaluation settings. This pattern should not be interpreted as evidence that frequently reused datasets are universally more appropriate or more realistic. Adoption frequency may also reflect public availability, ease of preprocessing, legacy comparability, and the existence of established experimental pipelines. Therefore, Tables 2, 3 and 5 should be interpreted as indicators of documentation quality, reported use, and domain fit rather than as rankings of dataset quality or model performance.

5.2. Cross-Domain Dataset Limitations

Across the reviewed security domains, several limitations recur despite differences in the data source and deployment context. Many datasets are collected in controlled testbeds and may not fully reflect the variability, background noise, user behavior, and operational constraints of real environments. Other recurring issues include the limited attack diversity, outdated traffic profiles, class imbalance, incomplete labeling procedures, restricted protocol coverage, and insufficient documentation of collection and preprocessing steps. These limitations affect the external validity of the reported results. A model evaluated on a well-labeled benchmark may achieve strong performance while remaining poorly suited to unseen attacks, changing traffic conditions, encrypted communications, or different organizational environments. This concern is particularly important when the results are transferred across domains, such as from enterprise networks to IoT deployments, from flow-based data to host logs, or from synthetic web requests to operational application traffic. The availability of public datasets also creates a trade-off between accessibility and realism. Public benchmarks support a reproducible evaluation, whereas operational datasets may offer stronger contextual relevance but are often constrained by privacy, confidentiality, licensing, and limited documentation. Consequently, dataset selection should consider both scientific reproducibility and operational relevance.

5.3. Implications for Dataset Selection

Dataset selection should begin with the intended security objective rather than with the popularity of a benchmark. Researchers should first define the target environment, such as an enterprise network, IoT deployment, cloud platform, web application, endpoint, or encrypted-traffic setting. They should then select datasets whose data source, observation point, attack types, feature representation, and operational assumptions are compatible with that environment. A practical selection process should also assess the label quality, class distribution, temporal relevance, documentation quality, public availability, licensing conditions, and reproducibility support. When possible, the evaluation should include more than one dataset or complementary data source to reduce the risk that the conclusions depend on the characteristics of a single benchmark. For example, a model intended for IoT DDoS detection should be evaluated using data that reflect IoT protocols, device behavior, and realistic attack intensity, rather than relying only on a conventional enterprise-network benchmark. Similarly, host-based detection models require event-level telemetry and attack-lifecycle information that cannot be inferred reliably from flow-based datasets alone. The dataset-selection framework proposed in this study is intended to support these environment-aware decisions.

Based on the cross-domain evidence, the proposed dataset-selection framework consists of five sequential decisions: (i) defining the security objective and target deployment environment; (ii) specifying the required telemetry, attack coverage, and analytical task; (iii) assessing dataset realism, labeling quality, class balance, documentation, availability, and reproducibility; (iv) selecting one or complementary datasets whose collection assumptions match the intended environment; and (v) reporting the preprocessing, data splits, and evaluation procedures so that the findings can be interpreted in their operational context. This framework is intended as a practical decision guide rather than a universal scoring or ranking mechanism.

5.4. Illustrative Application of the Dataset-Selection Framework

To demonstrate how the proposed dataset-selection framework can be operationalized, this subsection applies it to an illustrative IoT/IloT intrusion-detection scenario. The target use case is the evaluation of a model intended for Industry 4.0 environments in which both network traffic and device- or process-level telemetry may be relevant. In the first step, the operational setting is defined as IoT/IloT rather than as a generic enterprise network. This immediately narrows the candidate datasets to those that reflect embedded, device-level, industrial, or cyber-physical assumptions. In the second step, candidate datasets are compared according to the data representation and telemetry source. For example, BoT-IoT provides botnet-oriented IoT traffic, ToN-IoT includes heterogeneous network, operating-system, and telemetry sources, Edge-IloTset is more closely aligned with edge and embedded IloT environments, CICIoT2023 provides recent large-scale IoT attack evidence, and ROSPaCe is more suitable when ROS2-based cyber-physical or robotic assumptions are central to the evaluation. Therefore, the framework does not select a dataset only because it is frequently used; it links dataset choice to the intended operational context. In the third step, the attack coverage and labeling are checked against the intended threat model. If the evaluation focuses on botnet, denial-of-service, reconnaissance, or IoT-oriented attack behavior, BoT-IoT, ToN-IoT, Edge-IloTset, and CICIoT2023 may be more appropriate than older general-purpose benchmarks. If the evaluation requires cyber-physical or robotics-specific behavior, ROSPaCe or industrial-control datasets such as SWaT and WADI may provide better contextual alignment. However, these datasets should not be treated as directly interchangeable because their collection environments, feature representations, and attack assumptions differ substantially. In the fourth step,

documentation and reproducibility are assessed using the checklist reported in Table 2. Candidate datasets are examined in terms of availability, labeling information, feature representation, class balance, preprocessing requirements, and reproducibility support. A dataset with strong adoption frequency but weak documentation or unclear preprocessing assumptions may be less suitable than a more specialized dataset that better matches the operational environment. In the final step, the selected dataset or dataset combination is aligned with the evaluation protocol. For an IoT/IIoT intrusion-detection study, a defensible selection could combine a general IoT benchmark such as BoT-IoT or ToN-IoT with a more recent or context-specific corpus such as Edge-IIoTset, CICIoT2023, or ROSPaCe. This combination would allow the study to test both the benchmark comparability and operational relevance. The example shows that the proposed framework functions as a decision-support process for matching datasets to security objectives, rather than as a fixed ranking of datasets.

5.5. Limitations of This Review

This review has several limitations. First, the included evidence is limited to publicly documented studies and datasets identified within the defined review scope and publication period. Relevant datasets or technical documentation that were not publicly accessible, insufficiently described, or difficult to identify may not have been captured. Second, the reviewed studies differ substantially in their experimental protocols, preprocessing methods, feature representations, class definitions, train–test splits, and evaluation metrics. For this reason, the review does not perform a meta-analysis or directly rank detection models according to the reported accuracy values. Instead, it focuses on the dataset characteristics, reported adoption patterns, and practical suitability for different security settings. Finally, dataset availability, licensing conditions, documentation quality, and public access may change over time. The findings should therefore be interpreted as a structured snapshot of the reviewed evidence rather than as a permanent or exhaustive catalogue of all cybersecurity datasets.

6. Conclusions

This structured review examined security datasets reported in intrusion detection and prevention research published between 2018 and 2025. The analysis showed that established benchmarks remain widely adopted, while datasets designed for IoT/IIoT, DDoS, cloud, web-application, host, and other specialized environments have expanded the available evaluation landscape. The findings confirm that dataset selection should not rely solely on popularity or reported classification accuracy. Dataset suitability depends on the alignment between the target security environment and the dataset’s operational context, traffic source, attack coverage, labeling quality, feature representation, class distribution, documentation, availability, and reproducibility. The study contributes a cross-domain dataset taxonomy, a study-to-dataset evidence map, and a dataset-selection perspective for intrusion detection and prevention research. These contributions are intended to help researchers select more appropriate evaluation data and interpret experimental results with greater attention to operational relevance and external validity. Future work should prioritize realistic, well-documented, and continuously updated datasets that capture evolving attacks and heterogeneous security environments. Public datasets that combine network traffic, host telemetry, application-layer evidence, cloud logs, and IoT data would be particularly valuable for evaluating next-generation security systems.

Author Contributions: Conceptualization, H.G. and A.B.; methodology, H.G. and M.U.; validation, H.G. and A.B.; investigation, H.G., A.B. and M.U.; resources, H.G.; writing—review and editing,

H.G., M.U. and A.B.; visualization, M.U. and A.B.; project administration, M.U.; funding acquisition, M.U. and A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Firat University Scientific Research Projects Coordination Unit under the Comprehensive Research Project program, with project number MF.26.91 titled “Systematic Analysis of Data Sets Used in Cybersecurity and Artificial Intelligence Systems”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were generated in this study. The datasets and sources discussed in this review are available from the original publications, repositories, or dataset-provider pages cited in the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Karlsen, E.; Luo, X.; Zincir-Heywood, N.; Heywood, M. Benchmarking Large Language Models for Log Analysis, Security, and Interpretation. *J. Netw. Syst. Manag.* **2024**, *32*, 59. [\[CrossRef\]](#)
2. Gupta, U. Secure management of logs in internet of things. *arXiv* **2015**, arXiv:1507.05085.
3. Ishaq, K.; Javed, H.A. Implementing Snort Intrusion Prevention System (IPS) for Network Forensic Analysis. *arXiv* **2023**, arXiv:2308.13589.
4. Liao, H.J.; Lin, C.H.R.; Lin, Y.C.; Tung, K.Y. Intrusion detection system: A comprehensive review. *J. Netw. Comput. Appl.* **2013**, *36*, 16–24. [\[CrossRef\]](#)
5. Scarfone, K.; Mell, P. Intrusion detection and prevention systems. In *Handbook of Information and Communication Security*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 177–192.
6. Abie, H. An overview of firewall technologies. *Teletronikk* **2000**, *96*, 47–52.
7. Mehanna, N.; Rudametkin, W.; Laperdrix, P.; Vastel, A. Free Proxies Unmasked: A Vulnerability and Longitudinal Analysis of Free Proxy Services. *arXiv* **2024**, arXiv:2403.02445.
8. Azwee, K.; Alkhattali, M.; Dow, M. Exploring the Effectiveness of VPN Architecture in Enhancing Network Security for Mobile Networks: An Investigation Study. *Int. J. Netw. Secur. Its Appl.* **2023**, *15*, 33–40. [\[CrossRef\]](#)
9. Ghanbari, Z.; Rahmani, Y.; Ghaffarian, H.; Ahmadzadegan, M.H. Comparative approach to web application firewalls. In *Proceedings of the 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*; IEEE: Piscataway, NJ, USA, 2015; pp. 808–812.
10. Toprak, S.; Yavuz, A.G. Web application firewall based on anomaly detection using deep learning. *Acta Infologica* **2022**, *6*, 219–244. [\[CrossRef\]](#)
11. Scano, C.; Floris, G.; Montaruli, B.; Demetrio, L.; Valenza, A.; Compagna, L.; Ariu, D.; Piras, L.; Balzarotti, D.; Biggio, B. Modsec-learn: Boosting modsecurity with machine learning. In *Proceedings of the International Symposium on Distributed Computing and Artificial Intelligence*; Springer Nature: Cham, Switzerland, 2024; pp. 23–33.
12. Floris, G.; Scano, C.; Montaruli, B.; Demetrio, L.; Valenza, A.; Compagna, L.; Ariu, D.; Piras, L.; Balzarotti, D.; Biggio, B. ModSec-AdvLearn: Countering adversarial SQL injections with robust machine learning. *IEEE Trans. Inf. Forensics Secur.* **2025**, *20*, 6693–6705. [\[CrossRef\]](#)
13. Kumar, N.; Kasbekar, G.S.; Manjunath, D. Application of data collected by endpoint detection and response systems for implementation of a network security system based on zero trust principles and the eigentrust algorithm. *ACM SIGMETRICS Perform. Eval. Rev.* **2023**, *50*, 5–7. [\[CrossRef\]](#)
14. Xi, N.; Chen, C.; Zhang, J.; Sun, C.; Liu, S.; Feng, P.; Ma, J. Information flow based defensive chain for data leakage detection and prevention: A survey. *arXiv* **2021**, arXiv:2106.04951.
15. Braghin, S.; Simioni, M.; Sinn, M. DLPFS: The Data Leakage Prevention FileSystem. In *Proceedings of the International Conference on Applied Cryptography and Network Security*; Springer: Cham, Switzerland, 2022; pp. 380–397.
16. Androutsopoulos, I.; Paliouras, G.; Karkaletsis, V.; Sakkis, G.; Spyropoulos, C.D.; Stamatopoulos, P. Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach. *arXiv* **2000**, arXiv:cs/0009009.
17. Glöckler, J.; Sedlmeir, J.; Frank, M.; Fridgen, G. A systematic review of identity and access management requirements in enterprises and potential contributions of self-sovereign identity. *Bus. Inf. Syst. Eng.* **2024**, *66*, 421–440.
18. Olabanji, S.O.; Olaniyi, O.O.; Adigwe, C.S.; Okunleye, O.J.; Oladoyinbo, T.O. AI for identity and access management (IAM) in the cloud: Exploring the potential of artificial intelligence to improve user authentication, authorization, and access control within cloud-based systems. *Auth. Access Control Cloud-Based Syst.* **2024**, *17*, 38–56. [\[CrossRef\]](#)

19. Golduzian, A. Predict And Prevent DDOS Attacks Using Machine Learning and Statistical Algorithms. *arXiv* **2023**, arXiv:2308.15674.
20. Li, J.; Sisodia, D.; Feng, Y.; Shi, L.; Zhang, M.; Early, C.; Reiher, P. Adaptive Distributed Filtering of DDoS Traffic on the Internet. *arXiv* **2023**, arXiv:2305.13296.
21. Erdoğan, F.; Alıcı, M.C. Derin öğrenme tabanlı antivirüs modellerinin açık kaynak kodlu çeşitli atak kütüphaneleri kullanılarak atlatılması. *İstanbul Sabahattin Zaim Üniversitesi Fen. Bilim. Enstitüsü Derg.* **2021**, *3*, 66–71. [\[CrossRef\]](#)
22. Koppolu, N.R. Utilizing Event Logs of Windows Operating System in Digital Crime Investigations. *Int. J. Eng. Res. Technol.* **2021**, *10*, 654–661.
23. Korzeniowski, Ł.; Goczyła, K. Landscape of automated log analysis: A systematic literature review and mapping study. *IEEE Access* **2022**, *10*, 21892–21913. [\[CrossRef\]](#)
24. Nadeem, S.F.; Huang, C.Y. Data Visualization in Cybersecurity. In *Proceedings of the 2018 International Conference on Computational Science and Computational Intelligence (CSCI)*; IEEE: Piscataway, NJ, USA, 2018; pp. 48–52.
25. Adams, C.N.; Snider, D.H. Effective data visualization in cybersecurity. In *SoutheastCon*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8.
26. Garae, J.; Ko, R.K.; Apperley, M. A full-scale security visualization effectiveness measurement and presentation approach. In *Proceedings of the 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)*; IEEE: Piscataway, NJ, USA, 2018; pp. 639–650.
27. Velinov, A.; Zdravev, Z. Analysis of Apache Logs Using Hadoop and Hive. *TEM J.* **2018**, *7*, 645–650. [\[CrossRef\]](#)
28. Fang, R.; Hsiao, H.I.; He, B.; Mohan, C.; Wang, Y. High performance database logging using storage class memory. In *Proceedings of the 2011 IEEE 27th International Conference on Data Engineering*; IEEE: Piscataway, NJ, USA, 2011; pp. 1221–1231.
29. Shih, W.C.; Yang, C.T.; Jiang, C.T.; Kristiani, E. Implementation and visualization of a NetFlow log data lake system for cyberattack detection using distributed deep learning. *J. Supercomput.* **2023**, *79*, 4983–5012.
30. Rajalakshmi, J.R.; Rathinraj, M.; Braveen, M. Anonymizing log management process for secure logging in the cloud. In *Proceedings of the 2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]*; IEEE: Piscataway, NJ, USA, 2014; pp. 1559–1564.
31. Diro, A.A.; Chilamkurti, N. Distributed attack detection scheme using deep learning approach for Internet of Things. *Future Gener. Comput. Syst.* **2018**, *82*, 761–768. [\[CrossRef\]](#)
32. Gaber, T.; Awotunde, J.B.; Torky, M.; Ajagbe, S.A.; Hammoudeh, M.; Li, W. Metaverse-IDS: Deep learning-based intrusion detection system for Metaverse-IoT networks. *Internet Things* **2023**, *24*, 100977. [\[CrossRef\]](#)
33. Abed, R.A.; Hamza, E.K.; Humaidi, A.J. A modified CNN-IDS model for enhancing the efficacy of intrusion detection system. *Meas. Sens.* **2024**, *35*, 101299. [\[CrossRef\]](#)
34. Rizvi, S.; Labrador, G.; Guyan, M.; Savan, J. Advocating for hybrid intrusion detection prevention system and framework improvement. *Procedia Comput. Sci.* **2016**, *95*, 369–374. [\[CrossRef\]](#)
35. Huraj, L.; Šimon, M.; Lietava, J. Dataset of DDoS attacks on Fibaro home center 3 for smart home security. *Data Brief.* **2024**, *57*, 110991. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Verma, A.; Saha, R.; Kumar, G.; Conti, M. PETRAK: A solution against DDoS attacks in vehicular networks. *Comput. Commun.* **2024**, *221*, 142–154. [\[CrossRef\]](#)
37. Chahal, J.K.; Bhandari, A.; Behal, S. DDoS attacks & defense mechanisms in SDN-enabled cloud: Taxonomy, review and research challenges. *Comput. Sci. Rev.* **2024**, *53*, 100644. [\[CrossRef\]](#)
38. Swathi, K.; Bindu, G.H. An automated intrusion detection system in IoT system using Attention based Deep Bidirectional Sparse Auto Encoder model. *Knowl.-Based Syst.* **2024**, *305*, 112633. [\[CrossRef\]](#)
39. Hossain, S.; Senouci, S.M.; Brik, B.; Boualouache, A. A privacy-preserving Self-Supervised Learning-based intrusion detection system for 5G-V2X networks. *Ad. Hoc Netw.* **2025**, *166*, 103674. [\[CrossRef\]](#)
40. Bamber, S.S.; Katkuri, A.V.R.; Sharma, S.; Angurala, M. A hybrid CNN-LSTM approach for intelligent cyber intrusion detection system. *Comput. Secur.* **2025**, *148*, 104146. [\[CrossRef\]](#)
41. Chen, Z.; Simsek, M.; Kantarci, B.; Bagheri, M.; Djukic, P. Machine learning-enabled hybrid intrusion detection system with host data transformation and an advanced two-stage classifier. *Comput. Netw.* **2024**, *250*, 110576. [\[CrossRef\]](#)
42. Korium, M.S.; Saber, M.; Beattie, A.; Narayanan, A.; Sahoo, S.; Nardelli, P.H. Intrusion detection system for cyberattacks in the Internet of Vehicles environment. *Ad. Hoc Netw.* **2024**, *153*, 103330. [\[CrossRef\]](#)
43. Rangasikunpam, A.; Amiri, S.; Ost, L. BIDS: An efficient Intrusion Detection System for in-vehicle networks using a two-stage Binarised Neural Network on low-cost FPGA. *J. Syst. Archit.* **2024**, *156*, 103285. [\[CrossRef\]](#)
44. Farrukh, Y.A.; Wali, S.; Khan, I.; Bastian, N.D. Ais-nids: An intelligent and self-sustaining network intrusion detection system. *Comput. Secur.* **2024**, *144*, 103982. [\[CrossRef\]](#)
45. Xu, B. Design of intrusion detection system for intelligent mobile network teaching. *Comput. Electr. Eng.* **2023**, *112*, 109013. [\[CrossRef\]](#)

46. Paya, A.; Arroni, S.; García-Díaz, V.; Gómez, A. Apollon: A robust defense system against adversarial machine learning attacks in intrusion detection systems. *Comput. Secur.* **2024**, *136*, 103546. [\[CrossRef\]](#)
47. Zhang, Y.; Zhu, D.; Wang, M.; Li, J.; Zhang, J. A comparative study of cyber security intrusion detection in healthcare systems. *Int. J. Crit. Infrastruct. Prot.* **2024**, *44*, 100658. [\[CrossRef\]](#)
48. Serrano, W. CyberAIBot: Artificial Intelligence in an Intrusion Detection System for CyberSecurity in the IoT. *Future Gener. Comput. Syst.* **2024**, *166*, 107543.
49. Nandanwar, H.; Katarya, R. Deep learning enabled intrusion detection system for Industrial IOT environment. *Expert Syst. Appl.* **2024**, *249*, 123808. [\[CrossRef\]](#)
50. Shahin, M.; Maghanaki, M.; Hosseinzadeh, A.; Chen, F.F. Advancing network security in industrial IoT: A deep dive into AI-enabled intrusion detection systems. *Adv. Eng. Inform.* **2024**, *62*, 102685. [\[CrossRef\]](#)
51. Fang, Y.; Yao, Y.; Lin, X.; Wang, J.; Zhai, H. A feature selection based on genetic algorithm for intrusion detection of industrial control systems. *Comput. Secur.* **2024**, *139*, 103675. [\[CrossRef\]](#)
52. Shao, J.M.; Zeng, G.Q.; Lu, K.D.; Geng, G.G.; Weng, J. Automated federated learning for intrusion detection of industrial control systems based on evolutionary neural architecture search. *Comput. Secur.* **2024**, *143*, 103910. [\[CrossRef\]](#)
53. Duy, P.T.; Luong, T.D.; Quyen, N.H.; Pham, V.H. Fed-Evolver: An automated evolving approach for federated Intrusion Detection System using adversarial autoencoder in SDN-enabled networks. *Internet Things* **2024**, *28*, 101397. [\[CrossRef\]](#)
54. Lu, Q.; Gao, Q.; Li, J.; Xie, X.; Guo, W.; Wang, J. Distributed cyber-physical intrusion detection using stacking learning for wide-area protection system. *Comput. Commun.* **2024**, *215*, 91–102. [\[CrossRef\]](#)
55. Borgioli, N.; Aromolo, F.; Phan, L.T.X.; Buttazzo, G. A convolutional autoencoder architecture for robust network intrusion detection in embedded systems. *J. Syst. Archit.* **2024**, *156*, 103283. [\[CrossRef\]](#)
56. Mathina, P.A.; Valarmathi, K. Advancing IoT security: A novel intrusion detection system for evolving threats in industry 4.0 using optimized convolutional sparse Ficks law graph point trans-Net. *Comput. Secur.* **2025**, *148*, 104169. [\[CrossRef\]](#)
57. Hizal, S.; Cavusoglu, U.; Akgun, D. A novel deep learning-based intrusion detection system for IoT DDoS security. *Internet Things* **2024**, *28*, 101336.
58. Samantaray, M.; Barik, R.C.; Biswal, A.K. A comparative assessment of machine learning algorithms in the IoT-based network intrusion detection systems. *Decis. Anal. J.* **2024**, *11*, 100478. [\[CrossRef\]](#)
59. Sayed, G.I.; Hassanien, A.E. An optimized and intelligent metaverse intrusion detection system based on rough sets. *Internet Things* **2024**, *28*, 101360. [\[CrossRef\]](#)
60. Najar, A.A. A Robust DDoS Intrusion Detection System Using Convolutional Neural Network. *Comput. Electr. Eng.* **2024**, *117*, 109277. [\[CrossRef\]](#)
61. Puccetti, T.; Nardi, S.; Cinquilli, C.; Zoppi, T.; Ceccarelli, A. ROSPaCe: Intrusion detection dataset for a ROS2-based cyber-physical system and IoT networks. *Sci. Data* **2024**, *11*, 481. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Manivannan, D. Recent endeavors in machine learning-powered intrusion detection systems for the Internet of Things. *J. Netw. Comput. Appl.* **2024**, *229*, 103925. [\[CrossRef\]](#)
63. Chintapalli, S.S.N.; Singh, S.P.; Frnda, J.; Divakarachari, P.B.; Sarraju, V.L.; Falkowski-Gilski, P. OOA-modified Bi-LSTM network: An effective intrusion detection framework for IoT systems. *Heliyon* **2024**, *10*, e29410. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Saheed, Y.K.; Abdulganiyu, O.H.; Majikumna, K.U.; Mustapha, M.; Workneh, A.D. ResNet50-1D-CNN: A new lightweight resNet50-One-dimensional convolution neural network transfer learning-based approach for improved intrusion detection in cyber-physical systems. *Int. J. Crit. Infrastruct. Prot.* **2024**, *45*, 100674. [\[CrossRef\]](#)
65. Rahman, S.; Pal, S.; Mittal, S.; Chawla, T.; Karmakar, C. SYN-GAN: A robust intrusion detection system using GAN-based synthetic data for IoT security. *Internet Things* **2024**, *26*, 101212. [\[CrossRef\]](#)
66. Hore, S.; Ghadermazi, J.; Shah, A.; Bastian, N.D. A sequential deep learning framework for a robust and resilient network intrusion detection system. *Comput. Secur.* **2024**, *144*, 103928. [\[CrossRef\]](#)
67. Zhong, M.; Lin, M.; Zhang, C.; Xu, Z. A Survey on Graph Neural Networks for Intrusion Detection Systems: Methods, Trends and Challenges. *Comput. Secur.* **2024**, *141*, 103821. [\[CrossRef\]](#)
68. Habiba, M.; Islam, M.R.; Mueen, S.M.; Ali, A.S. Edge intelligence for network intrusion prevention in IoT ecosystem. *Comput. Electr. Eng.* **2023**, *108*, 108727. [\[CrossRef\]](#)
69. Ajay, P.; Nagaraj, B.; Kumar, R.A.; Suthana, V.; Keziah, M.R. DBN-protected material Enhanced intrusion prevention sensor system defends against cyber attacks in the IoT devices. *Meas. Sens.* **2024**, *34*, 101263. [\[CrossRef\]](#)
70. Baykara, M.; Das, R. A novel honeypot based security approach for real-time intrusion detection and prevention systems. *J. Inf. Secur. Appl.* **2018**, *41*, 103–116. [\[CrossRef\]](#)
71. Shalabi, K.; Al-Haija, Q.A.; Al-Fayoumi, M. A Blockchain-based Intrusion Detection/Prevention Systems in IoT Network: A Systematic Review. *Procedia Comput. Sci.* **2024**, *236*, 410–419. [\[CrossRef\]](#)
72. Kumar, A.; Abhishek, K.; Ghalib, M.R.; Shankar, A.; Cheng, X. Intrusion detection and prevention system for an IoT environment. *Digit. Commun. Netw.* **2022**, *8*, 540–551. [\[CrossRef\]](#)

73. Chiba, Z.; Abghour, N.; Moussaid, K.; Lifandali, O.; Kinta, R. A deep study of novel intrusion detection systems and intrusion prevention systems for Internet of Things networks. *Procedia Comput. Sci.* **2022**, *210*, 94–103. [\[CrossRef\]](#)
74. Turner, C.; Jeremiah, R.; Richards, D.; Joseph, A. A rule status monitoring algorithm for rule-based intrusion detection and prevention systems. *Procedia Comput. Sci.* **2016**, *95*, 361–368. [\[CrossRef\]](#)
75. Christabelle, A.; Dristi, D.; Syed, A.; Tannish, G.; Maheen, H.; Ali, R. Dataset of attacks on a live enterprise VoIP network for machine learning based intrusion detection and prevention systems. *Comput. Netw.* **2021**, *97*, 108283. [\[CrossRef\]](#)
76. Lee, A.Y.P.; Wang, M.I.C.; Hung, C.H.; Wen, C.H.P. PS-IPS: Deploying Intrusion Prevention System with machine learning on programmable switch. *Future Gener. Comput. Syst.* **2024**, *152*, 333–342. [\[CrossRef\]](#)
77. Sheth, C.; Thakker, R. Performance evaluation and comparative analysis of network firewalls. In *Proceedings of the 2011 International Conference on Devices and Communications (ICDeCom)*; IEEE: Piscataway, NJ, USA, 2011; pp. 1–5.
78. Salah, K.; Elbadawi, K.; Boutaba, R. Performance modeling and analysis of network firewalls. *IEEE Trans. Netw. Serv. Manag.* **2011**, *9*, 12–21.
79. Sepczuk, M. Dynamic web application firewall detection supported by cyber mimic defense approach. *J. Netw. Comput. Appl.* **2023**, *213*, 103596. [\[CrossRef\]](#)
80. Al-Shaer, E.S.; Hamed, H.H. Discovery of policy anomalies in distributed firewalls. In *IEEE Infocom*; IEEE: Piscataway, NJ, USA, 2004; Volume 4, pp. 2605–2616.
81. Bringhenti, D.; Valenza, F. Optimizing distributed firewall reconfiguration transients. *Comput. Netw.* **2022**, *215*, 109183. [\[CrossRef\]](#)
82. Lamdakar, O.; Ameer, I.; Eleyatt, M.M.; Carlier, F.; Ait Ibourek, L. Toward a modern secure network based on next-generation firewalls: Recommendations and best practices. *Procedia Comput. Sci.* **2024**, *238*, 1029–1035. [\[CrossRef\]](#)
83. Gouda, M.G.; Liu, A.X. A model of stateful firewalls and its properties. In *Proceedings of the 2005 International Conference on Dependable Systems and Networks (DSN'05)*; IEEE: Piscataway, NJ, USA, 2005; pp. 128–137.
84. Praseed, A.; Thilagam, P.S. HTTP request pattern based signatures for early application layer DDoS detection: A firewall agnostic approach. *J. Inf. Secur. Appl.* **2022**, *65*, 103090. [\[CrossRef\]](#)
85. Patel, M.; Amritha, P.P.; Sudheer, V.B.; Sethumadhavan, M. DDoS attack detection model using machine learning algorithm in next generation firewall. *Procedia Comput. Sci.* **2024**, *233*, 175–183. [\[CrossRef\]](#)
86. Ceragioli, L.; Degano, P.; Galletta, L. Can my firewall system enforce this policy? *Comput. Secur.* **2022**, *117*, 102683. [\[CrossRef\]](#)
87. Amal, M.R.; Venkadesh, P. H-DOCTOR: Honeypot based firewall tuning for attack prevention. *Meas. Sens.* **2023**, *25*, 100664. [\[CrossRef\]](#)
88. Uçtu, G.; Alkan, M.; Doğru, İ.A.; Dörterler, M. A suggested testbed to evaluate multicast network and threat prevention performance of Next Generation Firewalls. *Future Gener. Comput. Syst.* **2021**, *124*, 56–67. [\[CrossRef\]](#)
89. Ryu, J.; Noh, D.K.; Kang, K. FlashPage: A read cache for low-latency SSDs in web proxy servers. *Eng. Sci. Technol. An. Int. J.* **2024**, *51*, 101639. [\[CrossRef\]](#)
90. Gautam, N. Performance analysis and optimization of web proxy servers and mirror sites. *Eur. J. Oper. Res.* **2002**, *142*, 396–418. [\[CrossRef\]](#)
91. Fitzgerald, W.M.; Foley, S.N. Aligning Semantic Web applications with network access controls. *Comput. Stand. Interfaces* **2011**, *33*, 24–34. [\[CrossRef\]](#)
92. Epp, N.; Funk, R.; Cappel, C.; Lorenzo-Paraguay, S. Anomaly-based web application firewall using http-specific features and one-class svm. In *Workshop Regional de Segurança Da Informação e de Sistemas Computacionais*; UFSM: Sta Maria, Brazil, 2017.
93. Iliadis, L. Cognitive web application firewall to critical infrastructures protection from phishing attacks. *J. Comput. Model.* **2019**, *9*, 1–26. [\[CrossRef\]](#)
94. Razzaq, A.; Hur, A.; Shahbaz, S.; Masood, M.; Ahmad, H.F. Critical analysis on web application firewall solutions. In *Proceedings of the 2013 IEEE Eleventh International Symposium on Autonomous Decentralized Systems (ISADS)*; IEEE: Piscataway, NJ, USA, 2013; pp. 1–6.
95. Dawadi, B.R.; Adhikari, B.; Srivastava, D.K. Deep learning technique-enabled web application firewall for the detection of web attacks. *Sensors* **2023**, *23*, 2073. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Tekerek, A.; Gemci, C.; Bay, O.F. Development of a hybrid web application firewall to prevent web based attacks. In *Proceedings of the 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*; IEEE: Piscataway, NJ, USA, 2014; pp. 1–4.
97. Liang, J.; Zhao, W.; Ye, W. Anomaly-based web attack detection: A deep learning approach. In *Proceedings of the 2017 VI International Conference on Network, Communication and Computing*; ACM: New York, NY, USA, 2017; pp. 80–85.
98. Zhou, L.; Yau, W.C.; Gan, Y.S.; Liong, S.T. E-WebGuard: Enhanced neural architectures for precision web attack detection. *Comput. Secur.* **2025**, *148*, 104127. [\[CrossRef\]](#)
99. Mukhtar, B.I.; Azer, M.A. Evaluating the modsecurity web application firewall against SQL injection attacks. In *Proceedings of the 2020 15th International Conference on Computer Engineering and Systems (ICCES)*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.

100. Işiker, B.; Soğukpınar, İ. Machine learning based web application firewall. In *Proceedings of the 2021 2nd International Informatics and Software Engineering Conference (IISEC)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
101. Asghar, H.; Anwar, Z.; Latif, K. A deliberately insecure RDF-based Semantic Web application framework for teaching SPARQL/SPARUL injection attacks and defense mechanisms. *Comput. Secur.* **2016**, *58*, 63–82. [\[CrossRef\]](#)
102. Jemal, I.; Haddar, M.A.; Cheikhrouhou, O.; Mahfoudhi, A. SWAF: A smart web application firewall based on convolutional neural network. In *Proceedings of the 2022 15th International Conference on Security of Information and Networks (SIN)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
103. Shaheed, A.; Kurdy, M.B. Web application firewall using machine learning and features engineering. *Secur. Commun. Netw.* **2022**, *2022*, 5280158. [\[CrossRef\]](#)
104. Pillai, S.; Sharma, A. Hybrid unsupervised web-attack detection and classification—A deep learning approach. *Comput. Stand. Interfaces* **2023**, *86*, 103738. [\[CrossRef\]](#)
105. Clincy, V.; Shahriar, H. Web application firewall: Network security models and configuration. In *Proceedings of the 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*; IEEE: Piscataway, NJ, USA, 2018; Volume 1, pp. 835–836.
106. Solanki, P.S.; Singh, A.; Sao, S.; Atkekar, N.D. Protection of Research Data and Devices from Malware Attacks Using Endpoint Security System in Network. *Int. J. Sci. Res. Netw. Secur. Commun.* **2024**, *12*, 3.
107. Du, J.; Raza, S.H.; Ahmad, M.; Alam, I.; Dar, S.H.; Habib, M.A. Digital Forensics as Advanced Ransomware Pre-Attack Detection Algorithm for Endpoint Data Protection. *Secur. Commun. Netw.* **2022**, *2022*, 1424638. [\[CrossRef\]](#)
108. Hudson, B. *Advanced Persistent Threats: Detection, Protection and Prevention*; Sophos Ltd.: Burlington, MA, USA, 2014.
109. Verma, S.; Singh, A.; Lamba, R.; Verma, A. Data theft prevention & end point protection from PnP Devices. In *Proceedings of the National Conference on Communication Technologies & Its Impact on Next Generation Computing*; IJCA: New York, NY, USA, 2012.
110. Lewis, T.M.; Rimal, B.P. Effects of removing user-land hooks in endpoint protection during attack experiments. *IEEE Access* **2024**, *12*, 15820–15844. [\[CrossRef\]](#)
111. Chandel, S.; Yu, S.; Yitian, T.; Zhili, Z.; Yusheng, H. Endpoint protection: Measuring the effectiveness of remediation technologies and methodologies for insider threat. In *Proceedings of the 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CYBERC)*; IEEE: Piscataway, NJ, USA, 2019; pp. 81–89.
112. Hauer, B. Data and information leakage prevention within the scope of information security. *IEEE Access* **2015**, *3*, 2554–2565. [\[CrossRef\]](#)
113. Wüchner, T.; Pretschner, A. Data loss prevention based on data-driven usage control. In *Proceedings of the 2012 IEEE 23rd International Symposium on Software Reliability Engineering*; IEEE: Piscataway, NJ, USA, 2012; pp. 151–160.
114. Abelson, M. Data loss prevention and useful utilities. *Am. J. Orthod. Dentofac. Orthop.* **2000**, *117*, 367–369. [\[CrossRef\]](#)
115. Rogowski, W. The right approach to data loss prevention. *Comput. Fraud Secur.* **2013**, *2013*, 5–7. [\[CrossRef\]](#)
116. Charbonneau, S. The role of user-driven security in data loss prevention. *Comput. Fraud Secur.* **2011**, *2011*, 5–8. [\[CrossRef\]](#)
117. Ayo, F.E.; Ogundele, L.A.; Olakunle, S.; Awotunde, J.B.; Kasali, F.A. A hybrid correlation-based deep learning model for email spam classification using fuzzy inference system. *Decis. Anal. J.* **2024**, *10*, 100390. [\[CrossRef\]](#)
118. Bera, D.; Ogbanufe, O.; Kim, D.J. Towards a thematic dimensional framework of online fraud: An exploration of fraudulent email attack tactics and intentions. *Decis. Support Syst.* **2023**, *171*, 113977. [\[CrossRef\]](#)
119. Poobalan, A.; Ganapriya, K.; Kalaivani, K.; Parthiban, K. A novel and secured email classification using deep neural network with bidirectional long short-term memory. *Comput. Speech Lang.* **2025**, *89*, 101667. [\[CrossRef\]](#)
120. Yaseen, Q. Spam email detection using deep learning techniques. *Procedia Comput. Sci.* **2021**, *184*, 853–858. [\[CrossRef\]](#)
121. Magdy, S.; Abouelseoud, Y.; Mikhail, M. Efficient spam and phishing emails filtering based on deep learning. *Comput. Netw.* **2022**, *206*, 108826. [\[CrossRef\]](#)
122. Sethuraman, S.C.; Vs, D.P.; Reddi, T.; Reddy, M.S.T.; Khan, M.K. A comprehensive examination of email spoofing: Issues and prospects for email security. *Comput. Secur.* **2024**, *137*, 103600. [\[CrossRef\]](#)
123. Letmathe, P.; Noll, E. Analysis of email management strategies and their effects on email management performance. *Omega* **2024**, *124*, 103002. [\[CrossRef\]](#)
124. Krishnamoorthy, P.; Sathiyarayanan, M.; Proença, H.P. A novel and secured email classification and emotion detection using hybrid deep neural network. *Int. J. Cogn. Comput. Eng.* **2024**, *5*, 44–57. [\[CrossRef\]](#)
125. Zavrak, S.; Yilmaz, S. Email spam detection using hierarchical attention hybrid deep learning method. *Expert Syst. Appl.* **2023**, *233*, 120977. [\[CrossRef\]](#)
126. Al-Subaiey, A.; Al-Thani, M.; Alam, N.A.; Antora, K.F.; Khandakar, A.; Zaman, S.A.U. Novel interpretable and robust web-based AI platform for phishing email detection. *Comput. Electr. Eng.* **2024**, *120*, 109625. [\[CrossRef\]](#)
127. Doshi, J.; Parmar, K.; Sanghavi, R.; Shekokar, N. A comprehensive dual-layer architecture for phishing and spam email detection. *Comput. Secur.* **2023**, *133*, 103378. [\[CrossRef\]](#)
128. Srinivasarao, U.; Sharaff, A. Email thread sentiment sequence identification using PLSA clustering algorithm. *Expert Syst. Appl.* **2022**, *193*, 116475. [\[CrossRef\]](#)

129. Diale, M.; Celik, T.; Van Der Walt, C. Unsupervised feature learning for spam email filtering. *Comput. Electr. Eng.* **2019**, *74*, 89–104. [\[CrossRef\]](#)
130. Lopes, C.; Cortez, P.; Sousa, P.; Rocha, M.; Rio, M. Symbiotic filtering for spam email detection. *Expert Syst. Appl.* **2011**, *38*, 9365–9372. [\[CrossRef\]](#)
131. Laha, A.; Yasar, M.T.; Cheng, Y. Substop: An analysis on subscription email bombing attack and machine learning based mitigation. *High-Confid. Comput.* **2022**, *2*, 100086. [\[CrossRef\]](#)
132. Alanazi, R.; Alanazi, S. A hybrid NLP and domain validation technique for disposable email detection. *Alex. Eng. J.* **2024**, *102*, 200–210. [\[CrossRef\]](#)
133. Muralidharan, T.; Nissim, N. Improving malicious email detection through novel designated deep-learning architectures utilizing entire email. *Neural Netw.* **2023**, *157*, 257–279. [\[CrossRef\]](#) [\[PubMed\]](#)
134. Bountakas, P.; Xenakis, C. Helped: Hybrid ensemble learning phishing email detection. *J. Netw. Comput. Appl.* **2023**, *210*, 103545. [\[CrossRef\]](#)
135. Roy, U.; Ghosh, N. BloAC: A blockchain-based secure access control management for the Internet of Things. *J. Inf. Secur. Appl.* **2024**, *87*, 103897. [\[CrossRef\]](#)
136. Liao, C.H.; Guan, X.Q.; Cheng, J.H.; Yuan, S.M. Blockchain-based identity management and access control framework for open banking ecosystem. *Future Gener. Comput. Syst.* **2022**, *135*, 450–466. [\[CrossRef\]](#)
137. Indu, I.; Anand, P.R.; Bhaskar, V. Identity and access management in cloud environment: Mechanisms and challenges. *Eng. Sci. Technol. An. Int. J.* **2018**, *21*, 574–588. [\[CrossRef\]](#)
138. Sharma, D.H.; Dhote, C.A.; Potey, M.M. Identity and access management as security-as-a-service from clouds. *Procedia Comput. Sci.* **2016**, *79*, 170–174. [\[CrossRef\]](#)
139. Golightly, L.; Modesti, P.; Garcia, R.; Chang, V. Securing distributed systems: A survey on access control techniques for cloud, blockchain, IoT and SDN. *Cyber Secur. Appl.* **2023**, *1*, 100015. [\[CrossRef\]](#)
140. Kunz, M.; Puchta, A.; Groll, S.; Fuchs, L.; Pernul, G. Attribute quality management for dynamic identity and access management. *J. Inf. Secur. Appl.* **2019**, *44*, 64–79. [\[CrossRef\]](#)
141. Singh, I.; Singh, B. Access management of IoT devices using access control mechanism and decentralized authentication: A review. *Meas. Sens.* **2023**, *25*, 100591. [\[CrossRef\]](#)
142. Partanen, J.; Kylänpää, M.; Loukusa, S.; Korkiakoski, M.; Salonen, J. Device life cycle management requirements for identity and access management in the factory of future environment. *Procedia Comput. Sci.* **2022**, *204*, 235–244. [\[CrossRef\]](#)
143. Sutradhar, S.; Karforma, S.; Bose, R.; Roy, S.; Djebali, S.; Bhattacharyya, D. Enhancing identity and access management using hyperledger fabric and oauth 2.0: A block-chain-based approach for security and scalability for healthcare industry. *Internet Things Cyber-Phys. Syst.* **2024**, *4*, 49–67. [\[CrossRef\]](#)
144. Deng, W.; Li, J.; Yan, H.; Koe, A.S.V.; Wang, J.; Peng, C. Self-sovereign identity management in ciphertext policy attribute based encryption for IoT protocols. *J. Inf. Secur. Appl.* **2024**, *86*, 103885. [\[CrossRef\]](#)
145. Zhou, Z.; Liu, Y.; Zhu, X.; Zhang, S.; Liu, Z. Privacy-preserving cancelable multi-biometrics for identity information management. *Inf. Process. Manag.* **2025**, *62*, 103869. [\[CrossRef\]](#)
146. Bradford, M.; Earp, J.B.; Grabski, S. Centralized end-to-end identity and access management and ERP systems: A multi-case analysis using the Technology Organization Environment framework. *Int. J. Account. Inf. Syst.* **2014**, *15*, 149–165. [\[CrossRef\]](#)
147. Cremonezi, B.; Vieira, A.B.; Nacif, J.; Silva, E.F.; Nogueira, M. Identity management for Internet of Things: Concepts, challenges and opportunities. *Comput. Commun.* **2024**, *224*, 72–94. [\[CrossRef\]](#)
148. Murcia, J.M.B.; Cánovas, E.; García-Rodríguez, J.; Zarca, A.M.; Skarmeta, A. Decentralised Identity Management solution for zero-trust multi-domain Computing Continuum frameworks. *Future Gener. Comput. Syst.* **2025**, *162*, 107479. [\[CrossRef\]](#)
149. Farzaneh, B.; Shahriar, N.; Al Muktedir, A.H.; Towhid, M.S.; Khosravani, M.S. DTL-5G: Deep transfer learning-based DDoS attack detection in 5G and beyond networks. *Comput. Commun.* **2024**, *228*, 107927. [\[CrossRef\]](#)
150. Rao, G.S.; Patra, P.S.K.; Narayana, V.A.; Reddy, A.R.; Reddy, G.V.; Eshwar, D. DDoSNet: Detection and prediction of DDoS attacks from realistic multidimensional dataset in IoT network environment. *Egypt. Inform. J.* **2024**, *27*, 100526. [\[CrossRef\]](#)
151. Aslam, N.; Srivastava, S.; Gore, M.M. DDoS SourceTracer: An Intelligent Application for DDoS Attack Mitigation in SDN. *Comput. Electr. Eng.* **2024**, *117*, 109282. [\[CrossRef\]](#)
152. Anli, Y.A.; Ciplak, Z.; Sakaliuzun, M.; Izgu, S.Z.; Yildiz, K. DDoS detection in electric vehicle charging stations: A deep learning perspective via CICEV2023 dataset. *Internet Things* **2024**, *28*, 101343. [\[CrossRef\]](#)
153. Ouhssini, M.; Afdel, K.; Agherrabi, E.; Akouhar, M.; Abarda, A. DeepDefend: A comprehensive framework for DDoS attack detection and prevention in cloud computing. *J. King Saud. Univ.-Comput. Inf. Sci.* **2024**, *36*, 101938. [\[CrossRef\]](#)
154. Wang, L.; Chen, S.; Zhang, X.; Liu, J. Runtime reliability fractional distribution change analytics against cloud-based systems DDoS attacks. *J. Syst. Softw.* **2025**, *220*, 112265. [\[CrossRef\]](#)
155. Sakr, H.A.; Fouda, M.M.; Ashour, A.F.; Abdelhafeez, A.; El-Afifi, M.I.; Abdellah, M.R. Machine learning-based detection of DDoS attacks on IoT devices in multi-energy systems. *Egypt. Inform. J.* **2024**, *28*, 100540. [\[CrossRef\]](#)

156. Alasali, T.; Dakkak, O. A novel DDoS detection method using multi-layer stacking in SDN environment. *Comput. Electr. Eng.* **2024**, *120*, 109769. [[CrossRef](#)]
157. Hnamte, V.; Najar, A.A.; Nhung-Nguyen, H.; Hussain, J.; Sugali, M.N. DDoS attack detection and mitigation using deep neural network in SDN environment. *Comput. Secur.* **2024**, *138*, 103661. [[CrossRef](#)]
158. Bakar, R.A.; De Marinis, L.; Cugini, F.; Paolucci, F. FTG-Net-E: A hierarchical ensemble graph neural network for DDoS attack detection. *Comput. Netw.* **2024**, *250*, 110508. [[CrossRef](#)]
159. Feng, H.; Zhang, W.; Liu, Y.; Zhang, C.; Ying, C.; Jin, J.; Jiao, Z. Multi-domain collaborative two-level DDoS detection via hybrid deep learning. *Comput. Netw.* **2024**, *242*, 110251. [[CrossRef](#)]
160. Bala, B.; Behal, S. AI techniques for IoT-based DDoS attack detection: Taxonomies, comprehensive review and research challenges. *Comput. Sci. Rev.* **2024**, *52*, 100631. [[CrossRef](#)]
161. Setitra, M.A.; Fan, M. Detection of DDoS attacks in SDN-based VANET using optimized TabNet. *Comput. Stand. Interfaces* **2024**, *90*, 103845. [[CrossRef](#)]
162. Dash, S.K.; Dash, S.; Mahapatra, S.; Mohanty, S.N.; Khan, M.I.; Medani, M.; Abdullaev, S.; Gupta, M. Enhancing DDoS attack detection in IoT using PCA. *Egypt. Inform. J.* **2024**, *25*, 100450. [[CrossRef](#)]
163. Ouhssini, M.; Afdel, K.; Akouhar, M.; Agherrabi, E.; Abarda, A. Advancements in detecting, preventing, and mitigating DDoS attacks in cloud environments: A comprehensive systematic review of state-of-the-art approaches. *Egypt. Inform. J.* **2024**, *27*, 100517. [[CrossRef](#)]
164. Nalayini, C.M.; Katiravan, J.; Geetha, S.; Ji, C.E. A novel dual optimized IDS to detect DDoS attack in SDN using hyper tuned RFE and deep grid network. *Cyber Secur. Appl.* **2024**, *2*, 100042. [[CrossRef](#)]
165. Najar, A.A.; Naik, S.M. Cyber-secure SDN: A CNN-based approach for efficient detection and mitigation of DDoS attacks. *Comput. Secur.* **2024**, *139*, 103716. [[CrossRef](#)]
166. Mirza, S.; Abbas, H.; Shahid, W.B.; Shafqat, N.; Fugini, M.; Iqbal, Z.; Muhammad, Z. A malware evasion technique for auditing android anti-malware solutions. In *Proceedings of the 2021 IEEE 30th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*; IEEE: Piscataway, NJ, USA, 2021; pp. 125–130.
167. Patnaik, N. The Best Antivirus, Antimalware Solution for Home Owners And Corporates. *Int. J. Adv. Res. Comput. Sci.* **2015**, *6*, 26.
168. Garuba, M.; Liu, C.; Washington, N. A comparative analysis of anti-malware software, patch management, and host-based firewalls in preventing malware infections on client computers. In *Proceedings of the Fifth International Conference on Information Technology: New Generations (ITNG 2008)*; IEEE: Piscataway, NJ, USA, 2008; pp. 628–632.
169. Alharbi, B.; Alzahrani, H.; Asseri, A.; Taramisi, K. Anti-malware efficiency evaluation framework. In *Proceedings of the 2020 2nd International Conference on Computer and Information Sciences (ICCIS)*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–4.
170. Chakkaravarthy, S.S.; Sangeetha, D.; Vaidehi, V. A survey on malware analysis and mitigation techniques. *Comput. Sci. Rev.* **2019**, *32*, 1–23. [[CrossRef](#)]
171. Safarkhanlou, A.; Souri, A.; Norouzi, M.; Sardroud, S.E.H. Formalizing and verification of an antivirus protection service using model checking. *Procedia Comput. Sci.* **2015**, *57*, 1324–1331. [[CrossRef](#)]
172. Zelinka, I.; Das, S.; Sikora, L.; Šenkeřík, R. Swarm virus-next-generation virus and antivirus paradigm? *Swarm Evol. Comput.* **2018**, *43*, 207–224. [[CrossRef](#)]
173. Oliner, A.J.; Aiken, A.; Stearley, J. Alert detection in system logs. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*; IEEE: Piscataway, NJ, USA, 2008; pp. 959–964.
174. Catillo, M.; Pecchia, A.; Villano, U. AutoLog: Anomaly detection by deep autoencoding of system logs. *Expert Syst. Appl.* **2022**, *191*, 116263. [[CrossRef](#)]
175. He, S.; Zhu, J.; He, P.; Lyu, M.R. Experience report: System log analysis for anomaly detection. In *Proceedings of the 2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE)*; IEEE: Piscataway, NJ, USA, 2016; pp. 207–218.
176. Chen, Z.; Liu, J.; Gu, W.; Su, Y.; Lyu, M.R. Experience report: Deep learning-based system log analysis for anomaly detection. *arXiv* **2021**, arXiv:2107.05908.
177. Ring, M.; Schlör, D.; Wunderlich, S.; Landes, D.; Hotho, A. Malware detection on windows audit logs using LSTMs. *Comput. Secur.* **2021**, *109*, 102389. [[CrossRef](#)]
178. Xu, J.; Kalbarczyk, Z.; Iyer, R.K. Networked Windows NT system field failure data analysis. In *Proceedings 1999 Pacific Rim International Symposium on Dependable Computing*; IEEE: Piscataway, NJ, USA, 1999; pp. 178–185.
179. Dolak, R.; Janakova, M.; Botlik, J. Process Mining of Events Log from Windows. In *Proceedings of the SIMPDA*; IFIP: Mödling, Austria, 2018; pp. 73–77.
180. Jia, Z.; Shen, C.; Yi, X.; Chen, Y.; Yu, T.; Guan, X. Big-data analysis of multi-source logs for anomaly detection on network-based system. In *Proceedings of the 2017 13th IEEE Conference on Automation Science and Engineering (CASE)*; IEEE: Piscataway, NJ, USA, 2017; pp. 1136–1141.
181. Ma, R. Anomaly Detection for Linux System Log. Master's Thesis, University of Twente, Enschede, The Netherlands, 2020.
182. Mäkinen, M. Deep Learning for Anomaly Detection in Linux System Log. Master's Thesis, Aalto University, Espoo, Finland, 2019.

183. Studiawan, H.; Sohel, F.; Payne, C. Anomaly detection in operating system logs with deep learning-based sentiment analysis. *IEEE Trans. Dependable Secur. Comput.* **2020**, *18*, 2136–2148. [\[CrossRef\]](#)
184. Karim, S.S.; Afzal, M.; Iqbal, W.; Al Abri, D. Advanced Persistent Threat (APT) and intrusion detection evaluation dataset for linux systems 2024. *Data Brief.* **2024**, *54*, 110290. [\[CrossRef\]](#) [\[PubMed\]](#)
185. Dusane, P.; Sujatha, G. Logea: Log extraction and analysis tool to support forensic investigation of linux-based system. In *Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*; IEEE: Piscataway, NJ, USA, 2021; pp. 909–916.
186. Zhu, J.; He, S.; He, P.; Liu, J.; Lyu, M.R. Loghub: A large collection of system log datasets for ai-driven log analytics. In *Proceedings of the 2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*; IEEE: Piscataway, NJ, USA, 2023; pp. 355–366.
187. Karande, V.; Bauman, E.; Lin, Z.; Khan, L. SGX-Log: Securing system logs with SGX. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*; Association for Computing Machinery: New York, NY, USA, 2017; pp. 19–30.
188. White, J.; Pilbeam, A. A survey of virtualization technologies with performance testing. *arXiv* **2010**, arXiv:1010.3233.
189. Bonafiglia, R.; Cerrato, I.; Ciaccia, F.; Nemirovsky, M.; Risso, F. Assessing the performance of virtualization technologies for NFV: A preliminary benchmarking. In *Proceedings of the 2015 Fourth European Workshop on Software Defined Networks*; IEEE: Piscataway, NJ, USA, 2015; pp. 67–72.
190. Liu, L.; Masfary, O.; Li, J. Evaluation of server virtualization technologies for Green IT. In *Proceedings of 2011 IEEE 6th International Symposium on Service Oriented System (SOSE)*; IEEE: Piscataway, NJ, USA, 2011; pp. 79–84.
191. Padala, P.; Zhu, X.; Wang, Z.; Singhal, S.; Shin, K.G. Performance evaluation of virtualization technologies for server consolidation. *HP Labs Tec. Rep.* **2007**, *137*, 19.
192. Gaspar, A.; Langevin, S.; Armitage, W.D. Virtualization technologies in the undergraduate IT curriculum. *IT Prof.* **2007**, *9*, 10–17. [\[CrossRef\]](#)
193. Sandström, K.; Vulgarakis, A.; Lindgren, M.; Nolte, T. Virtualization technologies in embedded real-time systems. In *Proceedings of the 2013 IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA)*; IEEE: Piscataway, NJ, USA, 2013; pp. 1–8.
194. Li, Z.; Cai, G.; Yu, Q.; Liang, P.; Mo, R.; Liu, H. Bug priority change: An empirical study on Apache projects. *J. Syst. Softw.* **2024**, *212*, 112019. [\[CrossRef\]](#)
195. Denysenko, M. Detecting Anomalies in Log Data from Apache Web Server by Means of Cluster Analysis. Master's Thesis, Merseburg University of Applied Sciences, Merseburg, Germany, 2017.
196. Lakka, S.; Michalakelis, C.; Varoutas, D.; Martakos, D. Exploring the determinants of the OSS market potential: The case of the Apache web server. *Telecommun. Policy* **2012**, *36*, 51–68. [\[CrossRef\]](#)
197. Choi, H.; Lee, S. Forensic analysis of SQL server transaction log in unallocated area of file system. *Forensic Sci. Int. Digit. Investig.* **2023**, *46*, 301605. [\[CrossRef\]](#)
198. Wagner, J.; Nissan, M.I.; Rasin, A. Database memory forensics: Identifying cache patterns for log verification. *Forensic Sci. Int. Digit. Investig.* **2023**, *45*, 301567. [\[CrossRef\]](#)
199. Adedayo, O.M.; Olivier, M.S. Ideal log setting for database forensics reconstruction. *Digit. Investig.* **2015**, *12*, 27–40. [\[CrossRef\]](#)
200. Pavlou, K.E.; Snodgrass, R.T. Forensic analysis of database tampering. *ACM Trans. Database Syst. (TODS)* **2008**, *33*, 1–47. [\[CrossRef\]](#)
201. Frühwirth, P.; Kieseberg, P.; Schrittwieser, S.; Huber, M.; Weippl, E. InnoDB database forensics: Enhanced reconstruction of data manipulation queries from redo logs. *Inf. Secur. Tech. Rep.* **2013**, *17*, 227–238. [\[CrossRef\]](#)
202. Yuan, Y.; Shi, W.; Liang, B.; Qin, B. An approach to cloud execution failure diagnosis based on exception logs in openstack. In *Proceedings of the 2019 IEEE 12th International Conference on Cloud Computing (CLOUD)*; IEEE: Piscataway, NJ, USA, 2019; pp. 124–131.
203. Saibharath, S.; Geethakumari, G. Cloud forensics: Evidence collection and preliminary analysis. In *Proceedings of the 2015 IEEE International Advance Computing Conference (IACC)*; IEEE: Piscataway, NJ, USA, 2015; pp. 464–467.
204. Sarhan, M.; Layeghy, S.; Moustafa, N.; Portmann, M. Netflow datasets for machine learning-based network intrusion detection systems. In *Proceedings of the International Conference on Big Data Technologies and Applications*; Springer International Publishing: Cham, Switzerland, 2020; pp. 117–135.
205. Bakhshandeh, A.; Eskandari, Z. An efficient user identification approach based on NetFlow analysis. In *Proceedings of the 2018 15th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC)*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.
206. Hofstede, R.; Čeleda, P.; Trammell, B.; Drago, I.; Sadre, R.; Sperotto, A.; Pras, A. Flow monitoring explained: From packet capture to data analysis with NetFlow and ipfix. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 2037–2064. [\[CrossRef\]](#)
207. Maghsoudlou, A.; Gasser, O.; Poese, I.; Feldmann, A. FlowDNS: Correlating NetFlow and DNS streams at scale. In *Proceedings of the 18th International Conference on Emerging Networking EXperiments and Technologies*; ACM: New York, NY, USA, 2022; pp. 187–195.

208. Najafabadi, M.M.; Khoshgoftaar, T.M.; Calvert, C.; Kemp, C. Detection of ssh brute force attacks using aggregated NetFlow data. In *Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*; IEEE: Piscataway, NJ, USA, 2015; pp. 283–288.
209. Bin, L.; Chuang, L.; Jian, Q.; Jianping, H.; Ungsunan, P. A NetFlow based flow analysis and monitoring system in enterprise networks. *Comput. Netw.* **2008**, *52*, 1074–1092. [\[CrossRef\]](#)
210. Carela-Español, V.; Barlet-Ros, P.; Cabellos-Aparicio, A.; Solé-Pareta, J. Analysis of the impact of sampling on NetFlow traffic classification. *Comput. Netw.* **2011**, *55*, 1083–1099. [\[CrossRef\]](#)
211. Grill, M.; Nikolaev, I.; Valeros, V.; Rehak, M. Detecting DGA malware using NetFlow. In *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*; IEEE: Piscataway, NJ, USA, 2015; pp. 1304–1309.
212. Hou, J.; Fu, P.; Cao, Z.; Xu, A. Machine learning based DDoS detection through NetFlow analysis. In *Proceedings of the MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
213. Fraihat, S.; Makhadmeh, S.; Awad, M.; Al-Betar, M.A.; Al-Redhaei, A. Intrusion detection system for large-scale IoT NetFlow networks using machine learning with modified Arithmetic Optimization Algorithm. *Internet Things* **2023**, *22*, 100819. [\[CrossRef\]](#)
214. Lee, M.; Duffield, N.; Kompella, R. Two samples are enough: Opportunistic flow-level latency estimation using NetFlow. In *Proceedings of the 2010 Proceedings IEEE INFOCOM*; IEEE: Piscataway, NJ, USA, 2010; pp. 1–9.
215. Verde, N.V.; Ateniese, G.; Gabrielli, E.; Mancini, L.V.; Spognardi, A. No NAT'd user left behind: Fingerprinting users behind NAT from NetFlow records alone. In *Proceedings of the 2014 IEEE 34th International Conference on Distributed Computing Systems*; IEEE: Piscataway, NJ, USA, 2014; pp. 218–227.
216. Kemp, C.; Calvert, C.; Khoshgoftaar, T. Utilizing NetFlow data to detect slow read attacks. In *Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI)*; IEEE: Piscataway, NJ, USA, 2018; pp. 108–116.
217. Choi, B.Y.; Bhattacharyya, S. On the accuracy and overhead of cisco sampled NetFlow. In *Proceedings of ACM SIGMETRICS Workshop on Large Scale Network Inference (LSNI)*; ACM: New York, NY, USA, 2005; pp. 1–6.
218. Zang, H.; Nucci, A. Optimal NetFlow deployment in IP networks. In *Proceedings of the 19th International Teletraffic Congress (ITC), Beijing, China, 29 August–2 September 2005*; Beijing University of Posts and Telecommunications Press: Beijing, China, 2005; Volume 14, pp. 621–630.
219. Lemoudden, M.; Amar, M.; El Ouahidi, B. A binary-based mapreduce analysis for cloud logs. *Procedia Comput. Sci.* **2016**, *83*, 1213–1218. [\[CrossRef\]](#)
220. Vrabie, M.; Savage, S.; Voelker, G.M. BlueSky: A cloud-backed file system for the enterprise. In *FAST: Proceedings of the 10th USENIX conference on File and Storage Technologies*; USENIX Association: San Francisco, CA, USA, 2012; Volume 12, pp. 1–14.
221. Dalezios, N.; Shiaeles, S.; Kolokotronis, N.; Ghita, B. Digital forensics cloud log unification: Implementing CADF in Apache CloudStack. *J. Inf. Secur. Appl.* **2020**, *54*, 102555. [\[CrossRef\]](#)
222. Tak, B.C.; Tao, S.; Yang, L.; Zhu, C.; Ruan, Y. Logan: Problem diagnosis in the cloud using log-based reference models. In *Proceedings of the 2016 IEEE International Conference on Cloud Engineering (IC2E)*; IEEE: Piscataway, NJ, USA, 2016; pp. 62–67.
223. Yuan, Y.; Anu, H.; Shi, W.; Liang, B.; Qin, B. Learning-based anomaly cause tracing with synthetic analysis of logs from multiple cloud service components. In *Proceedings of the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*; IEEE: Piscataway, NJ, USA, 2019; Volume 1, pp. 66–71.
224. Agrawal, V.; Kotia, D.; Moshirian, K.; Kim, M. Log-based cloud monitoring system for OpenStack. In *Proceedings of the 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService)*; IEEE: Piscataway, NJ, USA, 2018; pp. 276–281.
225. Xu, J.; Chen, P.; Yang, L.; Meng, F.; Wang, P. Logdc: Problem diagnosis for declaratively-deployed cloud applications with log. In *Proceedings of the 2017 IEEE 14th International Conference on E-Business Engineering (ICEBE)*; IEEE: Piscataway, NJ, USA, 2017; pp. 282–287.
226. Zhou, P.; Wang, Y.; Li, Z.; Tyson, G.; Guan, H.; Xie, G. Logchain: Cloud workflow reconstruction & troubleshooting with unstructured logs. *Comput. Netw.* **2020**, *175*, 107279. [\[CrossRef\]](#)
227. Farshchi, M.; Schneider, J.G.; Weber, I.; Grundy, J. Metric selection and anomaly detection for cloud operations using log and metric correlation analysis. *J. Syst. Softw.* **2018**, *137*, 531–549. [\[CrossRef\]](#)
228. Batmaz, B.; Doğan, A. Accelerating DTLS on SoC FPGA for secure IoT applications. *Internet Things* **2024**, *28*, 101423. [\[CrossRef\]](#)
229. Halgamuge, M.N.; Niyato, D. Adaptive edge security framework for dynamic IoT security policies in diverse environments. *Comput. Secur.* **2025**, *148*, 104128. [\[CrossRef\]](#)
230. Hu, W.; Yu, Y.; Hao, X.; Yeoh, P.L.; Guo, L.; Li, Y. Cross-chain mapping blockchain: Scalable data management in massive IoT networks. *Digit. Commun. Netw.* **2024**, *11*, 1125–1140. [\[CrossRef\]](#)
231. Neto, E.C.P.; Dadkhah, S.; Ferreira, R.; Zohourian, A.; Lu, R.; Ghorbani, A.A. CICIOT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment. *Sensors* **2023**, *23*, 5941. [\[CrossRef\]](#) [\[PubMed\]](#)
232. Alsaedi, A.; Moustafa, N.; Tari, Z.; Mahmood, A.; Anwar, A. TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems. *IEEE Access* **2020**, *8*, 165130–165150. [\[CrossRef\]](#)

233. Sejan, M.A.S.; Rahman, M.H.; Aziz, M.A.; Tabassum, R.; Baik, J.I.; Song, H.K. Powerful graph neural network for node classification of the IoT network. *Internet Things* **2024**, *28*, 101410. [[CrossRef](#)]
234. Alqahtani, A.S.; Abuhasel, K.A.; Alquraish, M. AIRP: A system-based selection of adaptive intrusion recognition and prevention. *Environ. Technol. Innov.* **2021**, *23*, 101640. [[CrossRef](#)]
235. Lv, S.; Wang, C.; Wang, Z.; Wang, S.; Wang, B.; Zhang, Y. AAE-DSVDD: A one-class classification model for VPN traffic identification. *Comput. Netw.* **2023**, *236*, 109990. [[CrossRef](#)]
236. Alkhonaini, M.A.; Alohal, M.A.; Aljebreen, M.; Eltahir, M.M.; Alanazi, M.H.; Yafoz, A.; Alsini, R.; Khadidos, A.O. Sandpiper optimization with hybrid deep learning model for blockchain-assisted intrusion detection in iot environment. *Alex. Eng. J.* **2025**, *112*, 49–62. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.