International Journal of Emerging Engineering and Technology (IJEET)
ISSN (e): 2958-3764
Volume: 4, Number: 2, Pages: 77- 87, Year: 2025

IoT-Based Zero-Trust Security Architecture: A Comprehensive Analysis

Erej Azeem[1,*], Irshad Ahmed Sumra[1], Mehrooj Shahid[2], and Abdul Sattar[2]

[1] Department of Informatics and Systems, School of Science and Technology, University of Management and Technology Lahore, 54000, Pakistan
[2] Department of Computer Science, Lahore Garrison University, Lahore. Pakistan
[*] Corresponding author: Erej Azeem (Email: F2025375013@umt.edu.pk[1])

Received: 02/01/2025, Revised: 23/04/2025, Accepted: 29/05/2025

***Abstract*— Traditional perimeter-based security models have become ineffective in the face of the explosive growth in the number of Internet of Things (IoT) endpoints deployed at critical infrastructures. To cope with this weakness, network engineering is moving toward Zero-Trust Architecture (ZTA) frameworks, based on the principle of "never trust, always verify". But enterprise zero-trust models deployed directly in deeply embedded, resource-constrained nodes create a critical operational conflict: the Energy-Security Paradox. The survey paper offers an in-depth and multidimensional analysis of scientific publications from 2024 to 2026 to chart the convergence of cutting-edge IoT Zero-Trust deployments under stringent hardware and computational constraints. Using a Systematic Literature Review (SLR) methodology, a technical taxonomy was developed to sort eligible studies into a structured classification of studies in the three operational layers: physical hardware layers, trust anchors, lightweight cryptography (LWC), and decentralized behavioural intelligence. Silicon-derived fingerprints are still environmentally sensitive, standalone LWC ciphers do not have extensive cross-layer protocol integration and decentralized threat detection models are still highly susceptible to adversarial manipulation and poisoning attacks. This survey paper outlines some of the most important open challenges relevant to the next decade of research, including Byzantine-resilient aggregation frameworks.**

***Index Terms*—Anomaly Detection, Cross-Layer Co-Design, Edge-Fog Computing, Energy-Security Paradox, Internet of Things, Lightweight Cryptography, Zero-Trust Architecture.**

I. INTRODUCTION

From industrial to domestic environments and critical infrastructure, Internet of Things (IoT) endpoints are multiplying and scaling rapidly, far outstripping legacy digital security strategies [1]. The security perimeter that was previously relied upon is no longer holding up to the modern network environments, which are characterized by their high heterogeneity, multi-domains and open structures. They can be exploited by an insecure application programming interface (API) or compromised edge node as an unchecked vector for adversarial lateral movement [2]. In recent years, wamidgrowing challenges in perimeter security, modern network engineering has shifted towards the Zero-Trust Architecture (ZTA) framework. In contrast to the concept of implicit operational confidence associated with network topology or spatial location [3], ZTA is based on the fundamental principle of "never trust, always verify" (Fig. 1). It, on the other hand, requires ongoing identity verification, context-dependent access, micro-segmentation and strict enforcement of least-privilege access rules across all currently running digital assets, irrespective of their source location or operational life [4].

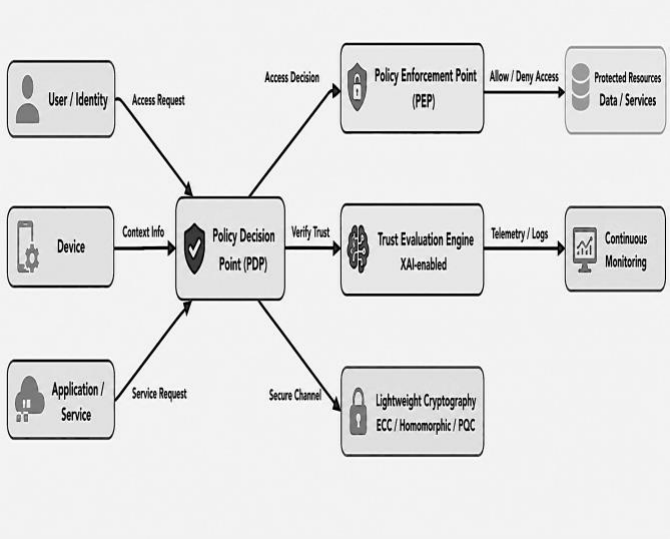

Fig. 1. Conceptual Concept-based logical core components of ZTA and orchestration with distinct functional view of Control Plane and Data Plane. The diagram represents interaction scenario of Policy Decision Point (PDP) dynamically receiving and interpreting telemetry and context derived from the Trust Evaluation Engine to make informed decision for the control access rules implemented by Policy Enforcement Point (PEP) on various heterogeneous IoT device. (Redrawn based on the structural models proposed in NIST [3].

While the concept of Zero-Trust is effective, implementing powerful Zero-Trust models directly at the edge creates a significant engineering challenge – the "Energy-Security Paradox" [5]. Dense, frequent, cryptographic handshakes; multi-factor public-key infrastructure (PKI); and centralized, high-throughput log analysis are heavily relied upon in typical enterprise ZTA deployments. But the other extreme represents a large number of IoT deployments, those with deeply embedded, resource constrained nodes, typically 8-bit or 16-bit microcontroller-based, with only a limited amount of RAM, and long-term autonomous battery constraints [6]. Classical cryptographic suites such as Rivest-Shamir-Adleman (RSA) or typical Advanced Encryption Standard (AES) implementations cause impractical computational delays, consume local storage resources, and soon deplete physical battery cells, especially in resource-constrained devices. This operational mismatch prevents the immediate implementation of enterprise-grade ZTNA without changes [7].

In the last few years (2024-2026), scientific studies have emergedthat take ah convergent approach to the zero-trust concept and resource-constrained IoT security. In one research paradigm, emphasis is placed on verifying identity and optimizing cryptography at the silicon level. This method aims to optimize the device's security perimeter at the local level through innovative hardware primitives and light cryptographic algorithms with minimal memory and energy consumption [8]. At the same time, a complementary research track targets at the network level orchestration and decentralized behavioral analytics, investigating the same area from a different perspective. This approach is based on decentralized threat detection systems, ledger verification and decentralized machine learning, which dynamically calculate trust metrics based on context, and do not overload individual low-power endpoints [9].

This survey paper seeks to identify the intersection between these security paradigms by offering an exhaustive, multifaceted analysis of the state of the art in the implementation of IoT Zero-Trust frameworks from 2024 to 2026. In contrast to a proposed design of such an architecture, this survey paper offers an evaluation of current trends in the literature based on two essential perspectives that intersect each other:

The advancement of techniques to generate unforgeable identification codes at the hardware level together with cryptographic protocols that would allow the implementation of such protocols within hardware constraints.

The creation of distributed intelligence systems that can monitor and evaluate the endpoint activity without violating its privacy and sustainability.

The rest of the paper has been strategically structured into various sections that will break down the relationship between ZTA and IoT systems with constrained resources. Section II presents the Systematic Literature Review (SLR) research focus and design, providing an explanation of the explicit search approach, databases, and the inclusion and exclusion criteria applied to filter contemporary literature ranging from 2024 to 2026. Section III presents categorization of Zero-Trust Components, which is based on three pillars of analysis, including physical layer trust anchors, LWC-based dynamic access control, and decentralized behavioral intelligence systems. Section IV focuses on the discussion of the architectural trade-offs, security margin, and friction introduced by the Energy-Security Paradox. Section V offers some key open challenges and future directions, such as Byzantine fault tolerant aggregation and automatic near zero time revocation, for the next decade of edge security. Section VI summarizes and concludes the paper.

While several excellent review papers have examined the intersection of Internet of Things (IoT) security and Zero-Trust Architecture (ZTA) between 2024 and 2026, existing literature predominantly approaches the domain from high-level enterprise orchestrations, cloud-to-edge gateways, or general architectural principles. These surveys frequently assume resource capabilities that are absent in deeply embedded systems, thereby failing to address the acute trade-offs imposed by the Energy-Security Paradox.

This survey distinguishes itself from contemporary reviews by focusing strictly on the unique constraints of 8-bit and 16-bit low-power edge endpoints. Rather than treating zero-trust as a monolithic software policy, this paper provides a systematic, cross-layer analysis that bridges the gaps between underlying silicon primitives, hardware roots of trust, lightweight protocol designs, and edge-native distributed machine learning. Furthermore, unlike ad-hoc literature roundups, this work employs a mathematically rigorous Systematic Literature Review (SLR) methodology governed by the PRISMA framework to categorize and audit the literature from 2024 to 2026.

To clearly delineate the boundaries of this work, Table I provides a comparative analysis of recent prominent surveys in the ZTA-IoT domain against our survey across core coverage domains, selection methodology, and identified research gaps.

II. Scope and Methodology

In order to produce a systematic review of existing literature and assess the current state of affairs, this research utilizes a Systematic Literature Review (SLR) methodological approach. SLRs involve specific criteria for selecting a relevant database and defining a search string that directly targets the intersection of ZTA and constrained IoT environments [10].

TABLE I: COMPARATIVE ANALYSIS OF RECENT ZERO-TRUST IOT SURVEYS VS. THIS WORK

Survey Reference	Review Period	Methodology	Core Coverage Areas	Addressed Research Gaps / Limitations	This Survey's Unique Focus & Contribution
Al-Tamimi et al. [1]	2024	General Narrative Review	High-level enterprise ZTA, cloud-edge gateway orchestration, macro-segmentation.	Pre-supposes strong main-line power and enterprise computing overheads; lacks analysis of hardware anchoring techniques.	Considers hardware resource constraints only, and software-hardware primitives across all layers.
Weinberg & Cohen [2]	2024	Ad-hoc Literature Survey	Widely used new technologies, identity management, generic network perimeter.	Reproducibility criterion is missing; macro-level perspective but no low power crypto metrics.	Analyzes exact execution trade-offs due to hardware limitations (RAM/latency) in the context of the Energy-Security Paradox.
Mushtaq et al. [3]	2025	Systematic Literature Review (SLR)	Multi-domain ZTA deployments, cloud migration in enterprises, generic asset management.	High-level taxonomy mapping across several industries; no focus on the vulnerabilities of physical layer hardware.	Identifies architectural weaknesses (PUF environmental dependence, machine learning poisoning) of edge devices explicitly.
Neagu et al. [11]	2026	Comprehensive Taxonomy Review	Privacy-preserving technologies, widely used anomaly detection, generic trust at the edge.	Lots of emphasis on the mathematical aspects of privacy preserving algorithms; little hardware co-design.	Ensures an exclusive emphasis on 2024-2026 cryptographic protocol implementations and ASICs.
This Survey	2024–2026	Quantitative SLR (PRISMA Framework)	Cross-Layer Co-Design: RoT in physical hardware, LWC, and DBI.	Multi-domain deployment issues, noise margin in PUF, adversarial poisoning in edge AI covered.	Establishes a hardware-to-intelligence taxonomy specifically for highly constrained 8/16-bit IoT devices.

A. Search Strategy and Data Sources

As part of the initial search process, the following data sources were considered to be the primary venues for acquiring scholarly articles published from 2024 through 2026:

1) *IEEE Xplore Digital Library*
2) *ACM Digital Library*
3) *ScienceDirect (Elsevier)*
4) *SpringerLink*
5) *MDPI Digital Publishing*

A search string was built using specific keywords to filter out publications that address the operational and hardware restrictions in zero-trust environments. The main search string used in the database search was composed as follows:

("Zero Trust" OR "ZTA" OR "Zero Trust Network Access") AND ("IoT" OR "Internet of Things" OR "Edge" OR "Resource-Constrained") AND ("Cryptographic" OR "Anomaly Detection" OR "Federated Learning")

B. Inclusion and Exclusion Criteria

To filter out results that do not meet the highest standards of technical relevance, a set of inclusion and exclusion criteria was carefully defined when filtering the database results.

Inclusion Criteria (IC):

Temporal Relevance: Studies published or accepted for publication from 2024 to 2026 were prioritized, reflecting the latest advancements in the field of Zero-Trust Architecture (ZTA) within resource-limited IoT settings. For background, context and popular security mechanisms, foundational and seminal publications prior to 2024 have been added where necessary [10].

Architectural Focus: Zero Trust principles ("never trust, always verify") were studied, evaluated or implemented within an Internet of Things (IoT), edge or resource constrained computing environment.

Articles were eligible for inclusion only if they were published in peer-reviewed journals, conference papers or reputable scholarly publications.

Exclusion Criteria (EC):

Works concentrating purely on the cloud infrastructure of companies, desktops' network, or data centers without any discussion of edge/Internet of Things (IoT) devices.

Use of traditional cryptographic systems like pure RSA-4096 or multiple iterations of cryptography that do not resolve the energy-security trade-off [5].

Unpublished works like, white papers, book reviews, and extended abstracts without experimental verification.

C. Literature Screening and Classification

In order to provide a clear and systematic selection procedure, the present literature review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) as shown in Fig. 2. The selection procedure was systemically divided into four separate chronological stages, which helped to reduce the number of identified papers from the total of n = 160 to n = 55.

Phase 1: Identification (n = 160)

The selection lifecycle started when the major search term was run against the metadata of five digital repositories: IEEE Xplore, Scopus, Web of Science, ACM Digital Library and Google Scholar (n = 160). The first automated query retrieved n = 160 papers as candidate papers.

Phase 2: Preliminary Screening (n = 85)

The n = 160 works identified in Phase 1 were then subjected to preliminary screening filters in terms of publication type, language, and duplication metrics as part of Phase 2. A total of 75 papers were removed at this stage, as follows:

Duplicate Removals: 31 papers from the indexes identified as duplicates were removed.

Document Type Failures: n = 28 informal items (non-peer-reviewed extended abstracts, editorial introductions, industry white papers and book reviews) were not included.

Language/Completeness Filters: 16 publications or short posters that were in a language other than English and did not have full documentation were excluded at the end of this phase there were still 85 unique papers left, namely the resulting pool.

Phase 3: Eligibility Evaluation (n = 72 down to n = 55)

The remaining 85 papers underwent two successive filters: Abstract and Title Review (n = 85 to n = 72). The titles, keywords, and abstracts of these papers were scrutinized for the presence of generic network security architectures that were not explicitly discussed with regards to resource constrained parameters or Zero-Trust components. In this process, 13 papers were rejected, leaving 72 papers for full text review.

Full-Text Analysis (n = 72 → n = 55): A more thorough reading of the remaining 72 papers was performed on the remaining papers in relation to the following core exclusion criteria. During this hard pass, 17 papers were rejected for the following reasons: No edge context, or lack thereof, in the hardware footprint; No device hardware footprint, or lack thereof, in the context; Cloud/data-center orchestration only. The implementation of legacy, high overhead cryptographic setups (typically standard RSA-4096 or multi-loop ciphers) that did not consider or address the Energy-Security Paradox. Too few performance evaluations that map empirical overhead metrics (run time execution latency, storage footprint or energy drain, etc.).

Phase 4: Primary study inclusion (n = 55)

After 17 failed technical requirement checks out of the full text in the set of 72 articles evaluated, a final set of exactly n = 55 studies were deemed suitable for technical assessment and to build a initial technical taxonomy (Physical layer Trust anchor n = 15 ; lightweight cryptographic access controls n = 22 ; and decentralized behavioral threat intelligence n = 18).

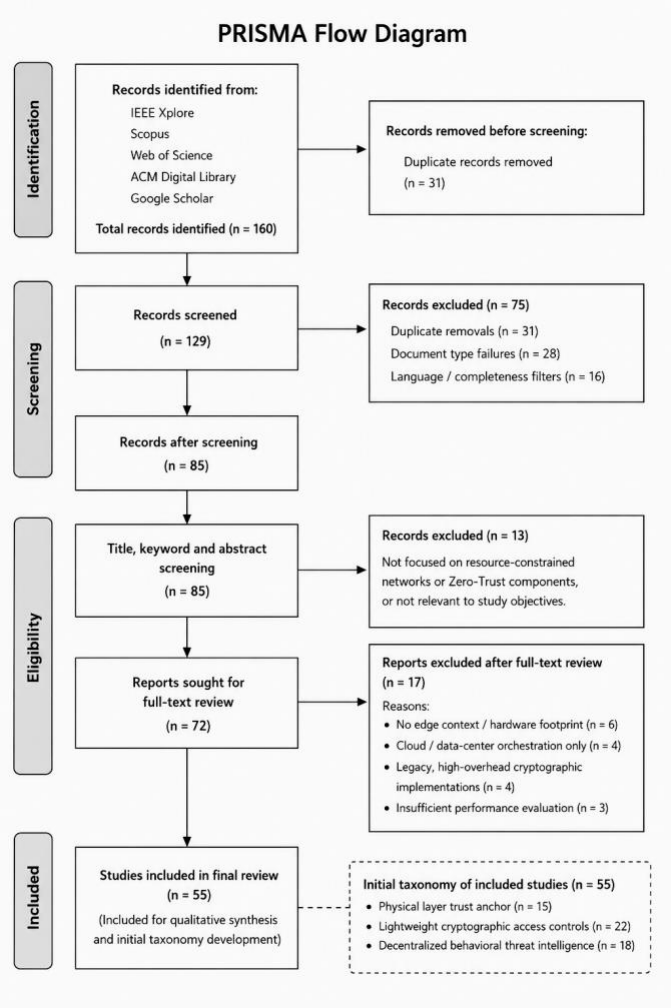

Fig. 2. PRISMA flow diagram of our systematically implemented four-step literature screening and selection procedure (adapted from [11]).

Specifically, in Phase 1, entries of the raw metadata were generated by the combined search engine query. In Phase 2, duplicates were removed, and title-abstract screening was done to sieve away general cybersecurity papers without ZTA-specific features.

Phase 3 entailed full-text screening for confirming whether any of the selected works included performance measurement criteria such as latency, overheads, or power consumption costs.

Lastly, Phase 4 involved systematic classification of the collected works based on their focus under a technical taxonomy of four operational pillars, which comprised physical hardware-based identity assurance, lightweight cryptography support, and edge-fog-based behavioral intelligence.

III. Categorization of Zero-Trust Components

A. Physical Hardware Layer: Trust Anchors and Device Identity

One of the most critical aspects in building an effective ZTA is establishing absolute verification of identity of the device. Cryptographic keys implemented via software are very susceptible to attacks based on side-channel analysis, memory dumps, and even physical access to the chip. For that reason, the most recent research is concentrated on moving identity verification to hardware primitives [12].

Physical Unclonable Functions (PUFs): When examining the 2024-2026 time-frame, one can note a major trend for adoption of SRAM-based and ring-oscillator PUFs to build a unique uncloneable device fingerprint. In order to prove the effectiveness of physical primitives in establishing reliable security metrics, Lightweight cryptographic executions were tested on physical IoT microcontroller platforms demonstrating how hardware-optimized primitives help to avoid long cryptographic executions on resource-constrained hardware [12]. In order to tackle the problem of intrinsic noise associated with silicon fingerprints, Error correction techniques were tested and overhead involved into implementing soft-decision fuzzy extractors for constructing cryptographic keys based on noisy SRAM startup states [13]. The latter technique is implementing mutual authentication for distributed IoT nodes via edge gateways by using dual PUF [14].

Hardware Roots of Trust (RoT): In addition to identifying unique devices, runtime execution contexts need to be protected against malicious/vulnerable code running at the application layer. The design of ARM TrustZone-M was explored, highlighting how the microsegmented hardware architecture enables isolation by defining secure/non-secure worlds in processing. The result is that the device will be assured that even when its main application runtime running in a host operating system is fully compromised, its ZTA policy enforcement and compliance mechanisms will continue to be completely uncompromised [15].

To achieve similar assurance during boot cycles, various RA frameworks were examined, explaining how the hardware-based RoT mechanism ensures isolation of code to validate the true state of a remote IoT device [16]. To protect against peripheral tampering and untrusted memory-mapped I/O accesses, an efficient security architecture called SANCUS was designed that offers hardware-based authentication and protection of software modules in low-end microcontrollers. As proven by their design, simple cryptographic hardware support is sufficient to guarantee isolation of peripheral and memory states, thwarting any attempts to exploit the hardware via physical insertion or memory probing attacks [17].

B. Lightweight Cryptography and Dynamic Access Control

Once a hardened identity is in place, the data plane requires a method to securely encrypt all communication through a means that does not significantly drain a device's power reserves. Traditional corporate cipher suites such as AES-256 or RSA-4096 were infeasible given the computation power limitations of battery-powered IoT devices. Accordingly, considerable research effort was directed toward specialized, lightweight cryptographic (LWC) primitives that meet ZTA's continuous validation demands.

Lightweight Permutations and Ciphers: The Ascon cipher suite, adhering to the rigorous formalization pathways of NIST, is the prevailing standard of modern ZTA architectures. Ascon-128 was adapted in a micro-segmentation context, which demonstrated energy-efficiency improvement over standard AES, while retaining identical levels of mathematical security [18]. To minimize memory requirements, CLEFIA block cipher was adapted into a streamlined form for use in smart grid sensor data communication and drastically reduce the SRAM required per microtransaction for data intensive periodic exchanges [19]. Such lightweight cryptographic implementations were even utilized in a dynamic scenario and presented a hybrid approach to multi-domain verification by combining an elliptic curve based handshake localized to a network segment and attribute based security controls, showcasing the benefit of opportunistically utilizing lightweight cryptographic solutions to mitigate performance impacts on dispersed edge devices [20].

Optimized Authentication and Signcryption: One key direction has been to combine encryption & digital signatures into a single low overhead cryptographic operation, which is called signcryption, to enforce constant authentication without doubling transport payloads. An identity-based signcryption scheme was developed that completely eliminates the verification loops that are needed to process standard PKI certificates, greatly reducing transport packet sizes and processing delays, and enabling a critical cryptographic performance bottleneck. This has enabled resource constrained devices to transmit zero-trust tokens at will without the associated transmission delays and computational overhead. They also formulated an identity-based signcryption scheme that completely circumvents standard Public Key Infrastructure (PKI) certificate verification loops, clearing the bottleneck of costly handshake processing delays at the edge device [21].

Dynamic Key Management and Micro-Segmentation: If a system requires constant authentication, it also needs to perform frequent session re-keying to prevent reply attacks. A decentralized key-graph system has been described to enable edge gateways to locally re-key devices without contacting the central cloud, a local attribute-based encryption system was developed to maintain finely grained perimeters by ensuring that only a device with the desired environmental parameters

may receive and process the desired command based on the current dynamic policy established by the Policy Decision Point [22].

C. Behavioral Intelligence and Decentralized Architecture

The fourth element of this taxonomy involves dynamic monitoring of enterprise threat postures. Given that the highly mutable resource constrained edge environment is a breeding ground for zero day attacks, tracking credentials statically is not viable. The ability to continuously analyze the actual behavior of devices for insider threat, data injection payloads and lateral movement needs to be part of the zero-trust network access architecture. To fulfill the behavioral tracking requirements as clearly outlined in the search strategy methodology section above modern ZTA architectures leverage a distributed anomaly detection model extensively. Shifting the computation burden from central cloud gateways, these suites of decentralized intelligence provide real time assessment while retaining low-power end-point operational durability and data privacy [23].

Federated Learning for Edge Anomaly Detection:

Forcing raw device log telemetry into a central cloud server causes critical privacy issues and causes high congestion within the network due to resource limitations on the local side. In order to avoid these problems, a zero-trust model for edge devices using federated learning (FL) was proposed on top of a block-chain framework. The model prevents the sending of raw data by means of localized IoT nodes to train anomaly detection models by using their own data, and sharing only the learned model updates to the edge-fog coordinators [24].

Still, the usual federated learning aggregation algorithms require a large computation capacity on low-power end-points, this issue was tackled by developing a quantized FL that compressed the weights by up to 80%, allowing low-power gateways to participate in a cooperative threat intelligence network [25]. When considering an active containment, such distributed FL models can provide instantaneous mitigation, e.g., how a decentralized FL scheme on the edge could classify a compromised robot endpoint in just 1.2 seconds after an unusual command injection on the actuators was shown [26].

Edge-Fog Orchestration and Offloading:

To protect devices from excessive battery consumption that may arise during resource intensive Policy Enforcement Points (PEP), intensive behavioral verification's are offloaded to edge-fog infrastructure. MEC framework was designed where local PEP agents stream heavily compressed behavior digests to proximal fog devices running custom, high throughput Policy Decision Points (PDPs) [27]. Where device capabilities are insufficient for basic cryptographic challenge computation, intervening proxies are employed. A proxy-assisted authentication framework designed to conserve battery on extremely resource constrained sensors through leveraging edge servers for computationally intensive cryptographic handshake validation was also constructed [28].

Spatiotemporal Graph Neural Networks for Contextual Threat Verification:

Representing device interactions as dynamic, spatiotemporal graphs has yielded intelligent context aware verification techniques for use across disparate multi-domain IoT environments. The use of GNNs was introduced in fog layer PDP controllers in order to learn device interactions across multiple IoT verticals [29]. Topological relationships of network participants allow for real-time examination of interactions between devices; thus enabling automatic detection of attempted lateral movement and network segmentation before threats propagate through the data plane. The learning derived from topological analysis can further be extended by correlating real-world locations to digital data movement for additional verification; a spatiotemporal GNN was designed that takes location data and correlates physical asset location with network traffic. This contextual layer supports zero-trust's premise that physical location should not provide implicit security, such that any alteration in location should revoke access to the network [30].

IV. Discussion And Critical Analysis

An analysis of the thematic classification of literature from 2024–2026 demonstrates the evident movement in how academics try to solve the Energy-Security Paradox. However a structural analysis has shown the substantial structural trade-offs and implicit risk exposure and operational friction throughout the three pillars of the taxonomy (Table II).

TABLE II
ARCHITECTURAL TRADE-OFFS ACROSS ZERO-TRUST LAYERS

Layer	Advantages	Limitations
Physical Hardware RoT	High stability, no stored keys	High cost, noisy enrollment
Lightweight Cryptography	Low power, high performance	Weak cross-layer integration
Behavioral Intelligence	Privacy preserving	Vulnerable to poisoning attacks

A. Physical Hardware RoT vs. Deployment Scalability

Although, integration of silicon-level trust anchors (e.g., SRAM-PUF) provides a strong protection against physical/side-channel extraction of the keys as there is no need to keep static private key in NVM. However, thorough examination of these integrated systems shows a trade-off between the reliability and environmental stability of cryptographic schemes. A silicon derived "fingerprint" will inherently depend on the environment (temperature variation and voltage drops) that affects the cell start up behavior and introduces noise into the produced cryptographic keys (Fig. 3).

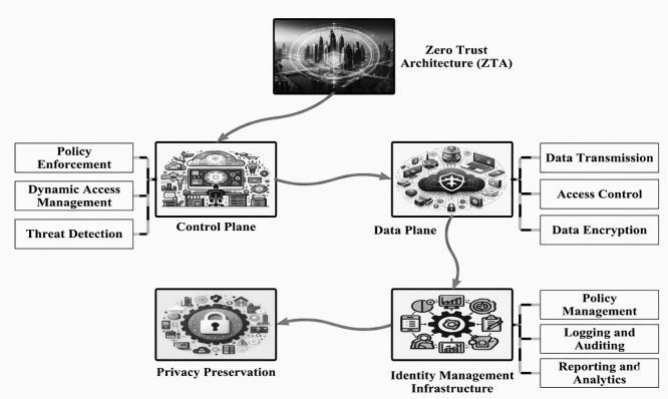

Fig. 3. Decentralized functional orchestration between control and data plane of ZTA with the adoption of blockchain based identity management infrastructure [6].

Residual security to power side-channel attacks is known to exist in the finalists for the NIST lightweight cryptography competition where leakage in physical traces can break the operation of these ciphers on a constrained endpoint, without implementing masking countermeasures. In modern implementations, post-processing fuzzy extractors and helper-data extraction algorithms are widely used to correct for unstable identities. Unfortunately, these error-correction schemes, although important, also implement significant computational power and memory requirement and the complexity is shifted from the cryptographic core to the error correction part [31].

A low overhead soft-decision helper data algorithm implementation for SRAM PUFs was provided, demonstrating that computation efficiency can be reduced sufficiently to enable operation within a heavily constrained embedded architecture by using mathematical optimization on the fuzzy extraction pathways [32]. However, these mathematical optimizations are no match for the fractured, non-uniform manufacturing process that governs the consumer IoT sector, thus relegating the widespread implementation of a standard secure enclave like ARM TrustZone-M to high cost, vendor-specific chip-sets, limiting application scaling across diverse IoT verticals.

B. Lightweight Cryptography Security Margins

The introduction of NIST standardized algorithms (Ascon), has greatly enhanced the efficiency of data plane encryption [33]. The literature, however, reveals that emphasis is placed on optimizing isolated primitives instead of achieving the full protocol life cycle. However, in resource constrained verticals, if the local session initialization vector (IV) or nonce generation is compromised then an identity-based signcryption protocol will be vulnerable to key leakage and lose its optimization advantage of the channel. This results in unpredictable processing latency profile on edge nodes when reacting to the dynamic environmental conditions [34].

C. Behavioral Privacy and Artificial Intelligence Limitations

One of the reasons for decentralizing threat detection is Federated Learning (FL), By keeping raw telemetry data local, FL satisfies the privacy requirements of today's edge networks [35]. This, however, presents a new attack vector – adversarial manipulation. Since it is impossible to verify the local training data by the central Policy Decision Point (PDP), the updates of the local model weights should be blindly trusted. Standard edge-fog FL loops are particularly prone to model-poisoning and gradient manipulation attacks, due to their structural weakness.

Moreover, using Graph Neural Networks (GNNs) to perform topology-aware lateral threat tracking results in a high computational complexity in the graph-matching phase. In the case of thousands of assets that are connected in a fog layer, the time required to pass the information from one asset to another can exceed the time the malicious payload reaches each individual asset, making it impossible to isolate assets at late stages. To reduce these topological processing overheads, other approaches involve representing the explicit hardware and network configuration as a rigid classification boundary, such as Neagu et al. [36] multi-dimensional structural taxonomy to standardize anomaly detection profiles, which was designed for highly constrained endpoints (Fig. 4).

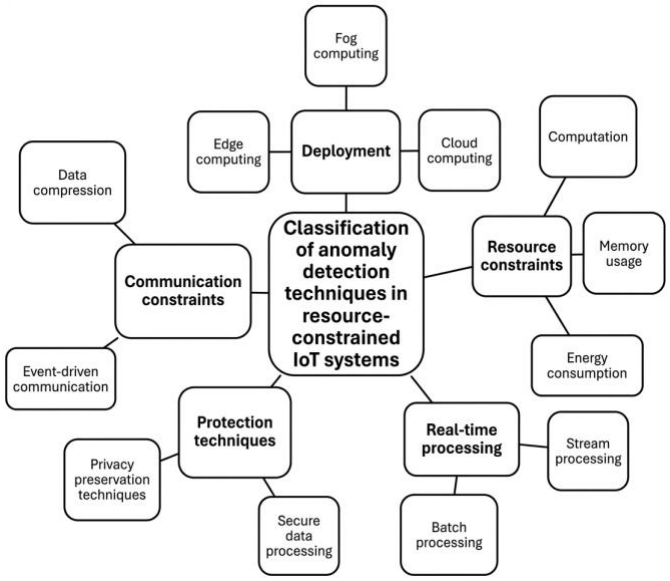

Fig. 4. Concept-based taxation of ZTA with Multi-dimensional Vectors for deployment, Communication, Real-time Streaming and hardware resource limits for anomaly detection models in deep embedded IoT systems [34].

D. Practical Real-World Deployment Bottlenecks

These models and frameworks, when adopted in production environments, reveal deep systemic weaknesses. An analysis of these practical implementation vectors reveals four main engineering hurdles:

Interoperability Among Heterogeneous IoT Devices:

The topologies of real-world IoTs are made up of a multi-generational, multi-vendor pool of devices that use widely different communication protocols (Zigbee, BLE, LoRaWAN, and MQTT). Such environments is asymmetric with respect to cryptographic capabilities between 32-bit devices that are hardware-isolated from the network [17] and legacy devices based on 8-bit or 16-bit microprocessors with limited battery life that do not support cross-layer protocol interoperability [19].

Standardization Challenges and Fragmentation of Ecosystems:

One of the hurdles to the wide adoption of IoT ZTA is the absence of a single, global standardization framework. Although a set of guidelines exist at the macro level as part of the enterprise zero trust tenets [23], there are no specific technical limitations for the embedded computing layers. This leaves deeply sliced protocol architectures of both implementing enterprise heavy requirements, yet lightweight edge token configuration, where custom policy engines step in to fill the space [1].

Decentralized Security Architecture Scalability Limitations:

Decentralized ZTA architectures – in particular, the ones leveraging blockchains for identity ledger or federated learning (FL) networks of threat intelligence – often fall short when scaling horizontally aggressively. Sending consensus information persistently or syncing large amounts of FL gradient weights across the LPWANs and mesh topologies results in serious collisions and increased storage costs, as well as time-outs on the network level [24]. Additionally, as the size of a network grows to millions of endpoints, filtering operations encounter computational limitations [38].

Continuous Verification Mechanism's Computational Overhead:

The key principle of the Zero Trust architecture—continuous and runtime verification—is oddly enough the most harmful thing for the battery life cycle of the edge node, providing another example of the Energy-Security Paradox. Making edge nodes perform continuous cryptographical verifications, graph matching topologies in real time [29], complex attribute-based access control checks [23], or fuzzy error correction [13] is very burdensome from the computational standpoint. The continuous verification process significantly decreases the autonomous lifetime of the edge nodes, thus posing an impossible choice between network security and device life [6].

V. Open Problems & Future Directions

While progress has been made in the last three years (2024–26), there are still some important challenges that have not been solved (Table III).

TABLE III
COMPARATIVE ANALYSIS OF SELECTED STUDIES

Author / Year	Proposed Methodology / Framework	Working of Methodology	Applications	Key Findings	Limitations
El-Hajj et al. (2023)	Lightweight Cryptographic Hardware Priming	Involves direct evaluation of lightweight cryptographical operations within physical IoT microcontroller environment to avoid costly software loops.	Constrained IoT Microcontrollers.	Has demonstrated that hardware-based primitives are effective in avoiding lengthy and battery-intensive cryptographic operations.	Limited to devices with built-in hardware support; no cross-layer protocol interoperability.
Maes et al. & Vincent et al. (2009 / 2012)	Soft-Decision Fuzzy Extractors for SRAM-PUF	Involves error-correction and helper data retrieval algorithms to optimize mathematical fuzzy processes and stabilize noise during startup of the chip.	Silicon Root-of-Trust (RoT) / Secure Key Generation.	Reduces the complexity of error correction to a sufficient degree for SRAM-PUF key generation to be practical on low-power devices.	Silicon fingerprints are very sensitive to environmental fluctuations (voltage and temperature fluctuations).
Pinto et al. (2019)	ARM TrustZone-M Micro-segmentation	Ensures hardware isolation by separating processing cycles into structurally determined "secure" and "non-secure" realms.	Runtime Context & Policy Enforcement Protection.	Ensures that enforcement of ZTA policies is totally uncompromised even in case of a compromise in the host operating system.	Universal scaling is seriously hindered by expensive, fragmented, and proprietary chip manufacturing processes.
Khan et al. (2024)	Ascon-128 Micro-segmentation	Customizes NIST-approved Ascon encryption algorithm suite into network microsegmentation	Data-Plane Network Micro-segmentation.	Provides the same level of mathematical security as conventional AES, but with	Addresses mostly the primitive alone and not the entire lifecycle of

Author / Year	Proposed Methodology / Framework	Working of Methodology	Applications	Key Findings	Limitations
		framework for data plane encryption.		substantially greater energy and computation efficiency. .	the protocol.
Chen et al. (2024)	Identity-Based Signcryption Scheme	Combines encryption with digital signature in a single cryptographic handshake process.	Zero-Trust Token Transmission.	Totally avoids expensive Public Key Infrastructure (PKI) certificate checking processes, minimizing payload size and transmission delays.	Very vulnerable to key compromise and latency fluctuations if local session initialization vector (IV)/nonce generation is compromised.
Danish Javeed et al. (2024)	Blockchain-Backed Federated Learning ZTI-IDS	Trains anomaly detection algorithms on IoT devices using raw data and sending only weights updates to Edge-Fog coordinators.	Federated Intrusion Detection Systems (IDS).	Performs intelligent threat coordination on networks while retaining telemetry locally to ensure privacy of the user/device.	Extremely vulnerable to attack by adversaries using poisoning and gradient manipulation.
Ben Saad et al. (2023)	MEC-Assisted Policy Enforcement (PEP)	Shifts costly behavioral validation tasks to local PEP agents, which stream highly-compressed behavior digests to nearby fog PDPs.	Offloaded Edge-Fog Behavioral Analytics.	Prevents individual low-power endpoints from being affected by the excessive battery drain caused by intensive log processing.	Dependent solely on stable proximity architecture; generates additional structure during communication bottlenecks.
Rabbani et al. & Wang et al. (2024 / 2023)	Spatiotemporal Graph Neural Networks (GNNs)	Translates device interactions into dynamic graphs, correlating topological information with actual spatiotemporal asset location.	Multi-Domain Contextual Threat Verification.	Automatically detects any network anomalies and blocks lateral movements whenever a resource changes its footprint.	The graph-matching phase is extremely computationally complex and causes latency while scaling up to thousands of endpoints.
Yao et al. (2025)	Byzantine-Resilient Over-the-Air FL	Applies coordinate-wise Median Mean approach to dynamically eliminate malicious nodes when aggregating the model over-the-air.	Poisoning Defenses in Decentralized Networks.	Facilitates automatic filtering of poisoned updates by the central PDP prior to convergence of the global model.	Unable to overcome scalability challenges when scaling up to millions of heterogeneous endpoints, resulting in high false positives.
Dorri et al. (2019)	Memory-Efficient Decentralized Ledger	Involves optimized DLT or gossiping protocols for decentralized and peer-to-peer trust signaling.	Real-Time, Automated Near-Zero Time Revocation.	Reduces centralized processing burden to a great extent by enabling neighboring nodes to quarantine the infected resource.	Face serious packet delivery and storage challenges when deployed over lossy LPWAN and mesh networks.

A. Poisoning Defenses and Robustness in Federated Anomaly Detection

The study demonstrates the need to prevent poisoned updates before global model convergence by filtering out malicious updates. The problem of creating defense strategies against model poisoning attacks and evasion attacks in decentralized edge computing is still open and poses a huge challenge. While the use of Federated Learning (FL) in detecting attacks is helpful for ensuring privacy, it creates a new type of vulnerability where the centralized PDP component has to trust local model updates in the absence of any way to inspect the local data [37].

A Byzantine-resilient federated learning framework was proposed, which enabled the filtration of malicious nodes during over-the-air aggregation without requiring excessive processing delays on edge equipment, showing that malicious nodes can be adaptively filtered during the over-the-air aggregation process with zero trust [38]. However, one challenge that has not been addressed in these dynamic clustering frameworks is scalability to millions of endpoints with high heterogeneity that are expected to generate numerous false alarms. Future directions should be aimed at developing

self-healing anomaly detection systems that can keep the threat classification clean in unstable and noisy physical vectors.

B. Cross-Layer Co-Design of Hardware and Cryptographic Software

Current research trends remain disjointed: silicon engineers develop hardware without considering software, while cryptographers assume that all physical layers are identical and perfect in execution. A coordinated and cohesive design is essential in closing this divide.Future designs must take the unpredictable analog nature of silicon PUFs and incorporate this directly into the process of generating the seed for encryption using lightweight ciphers such as Ascon. This will result in transient, one-time session keys that are inherently resistant to both side-channel attacks and physical memory dumping. Creating direct communication between hardware RoT and the lightweight cryptographic layer enables the removal of error correction software, resulting in more efficient use of energy.

C. Real-Time, Automated Revocation Logic for Disconnected Edges

Whereas there is significant work done on onboarding, monitoring, and threat scoring mechanisms, little has been proposed in terms of revocation mechanisms which can revoke access within seconds. Once the behavior score of the edge asset falls below the threshold value, communicating the revocation command across the long and lossy connection in LPWAN or mesh networks becomes challenging in itself. There are significant challenges in achieving near-zero-time communication and data storage when enforcing dynamic authentication and attribute-based revocation policies in accordance with the zero-trust concept. These challenges often result in containment delays, allowing lateral movement before the isolation command reaches its destination.

To address this challenge of delay, future research should consider leveraging light weight DLT or gossiping protocols for real-time policy enforcement. A memory-efficient decentralized ledger system optimized for IoT devices was introduced [39][40]. This work proved that local trust assessment and fast core signaling can significantly reduce centralized processing overheads associated with zero-trust principles. In this line, future zero-trust frameworks should ensure that compromised systems are isolated automatically by other nodes in the immediate physical proximity network by eliminating centralized revocation procedures[41].

VI. Conclusion

Switching to a Zero-Trust Architecture from the traditional perimeter-based security approach is crucial for protecting IoT networks, which have become increasingly distributed today. As the survey illustrates, addressing the strict restrictions imposed by the Energy-Security Paradox will require a comprehensive approach across each layer of the system in question. The upcoming years will witness the development of physical security that will begin with unbreakable, hardware-assisted device tracking SRAM-PUF technology, followed by efficient encryption through resource-saving cipher suites such as Ascon. At the same time, decentralized approaches to behavior analysis through federated learning and edge-fog collaboration will ensure that zero-trust security systems provide constant validation without excessive battery usage and data leakage. Tackling research problems in federated attacks prevention, cross-layer design of hardware/software components, and peer-to-peer automatic revocation mechanisms will facilitate the development of secure autonomous edge computing systems in the future decades.

Funding Statement

The authors received no specific funding for this study.

Conflicts of Interest

The authors declare no conflicts of interest to report regarding the present study.

Author Contributions

All authors contributed to the conception, literature review, drafting, and critical revision of this manuscript and approved the final version for submission.

Data Availability Statement

Data is available on reasonable request.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

References

[1] S. Al-Tamimi et al., "Zero-Trust architecture for securing Internet of Things (IoT) networks: A review," in Proc. 2024 Int. Conf. Information, Electronics and Computer Science (CIEES), 2024, pp. 1–6, doi: 10.1109/CIEES62939.2024.10811176.

[2] A. I. Weinberg and K. Cohen, "Zero trust implementation in the emerging technologies era: a survey," Complex Engineering Systems, vol. 4, no. 3, p. 41, Sep. 2024, doi: 10.20517/ces.2024.41.

[3] S. Mushtaq, M. Mohsin, and M. M. Mushtaq, "A Systematic Literature Review on the Implementation and Challenges of Zero Trust Architecture Across Domains," Sensors, vol. 25, no. 19, p. 6118, Oct. 2025, doi: 10.3390/s25196118.

[4] S. Son, D. Kwon, S. Lee, H. Kwon, and Y. Park, "A Zero-Trust Authentication Scheme With Access Control for 6G-Enabled IoT Environments," IEEE Access, vol. 12, pp. 154066–154079, Nov. 2024, doi: 10.1109/ACCESS.2024.3484522.

[5] S. Gnatyuk, B. Akhmetov, D. Zhaxygulova, and Y. Polishchuk, "CLEFIA-based Lightweight Encryption for Resource-Constrained Systems: Design, Algorithms and Security Analysis," International Journal of Computer Network and Information Security, vol. 17, no. 6, pp. 47–56, Dec. 2025, doi: 10.5815/ijcnis.2025.06.04.

[6] M. A. Aleisa, "Blockchain-Enabled Zero Trust Architecture for Privacy-Preserving Cybersecurity in IoT Environments," IEEE Access, vol. 13, pp. 1245–1256, Jan. 2025, doi: 10.1109/ACCESS.2025.3529309.

[7] M. C. Ezeakacha, C. O. M. Ezenwankwo, C. L. Obi, O. P. Osuji, and A. N. Asogwa, "Cryptography for IoT and Smart Grids: Challenges and

Solutions," Iconic Research and Engineering Journals, vol. 9, no. 12, pp. 2083–2088, Jun. 2026, doi: 10.1718/irejournals.1718965.

[8] J. Kolba, "Hardware–Software Co-Design Framework for Secure and Scalable IoT Embedded Systems," Journal of Integrated VLSI, Embedded and Computing Technologies, vol. 3, no. 3, pp. 34–42, Mar. 2026.

[9] A. Ogunbajo, I. Taiwo, A. Q. Abidola, O. F. Adediran, and I. Agbo-Adediran, "Privacy preserving AI models for decentralized data management in federated information systems," GSC Advanced Research and Reviews, vol. 22, no. 2, pp. 104–112, Feb. 2025, doi: 10.30574/gscarr.2025.22.2.0043.

[10] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—A systematic literature review," Information and Software Technology, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

[11] S. Tanwar and A. Kumar, "Extended identity based multi-signcryption scheme with public verifiability," Journal of Information and Optimization Sciences, vol. 39, no. 2, pp. 503–517, Feb. 2018, doi: 10.1080/02522667.2017.1383660.

[12] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," BMJ, vol. 372, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[13] M. El-Hajj, H. Mousawi, and A. Fadlallah, "Analysis of Lightweight Cryptographic Algorithms on IoT Hardware Platform," Future Internet, vol. 15, no. 2, p. 54, Feb. 2023, doi: 10.3390/fi15020054.

[14] M. Maes, P. Tuyls, and I. Verbauwhede, "Low-Overhead Implementation of a Soft Decision Helper Data Algorithm for SRAM PUFs," in Lecture Notes in Computer Science, vol. 5747, pp. 332–347, Sep. 2009, doi: 10.1007/978-3-642-04138-9_24.

[15] Y. Yilmaz, S. R. Gunn, and B. Halak, "Lightweight PUF-Based Authentication Protocol for IoT Devices," in Proc. 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Jul. 2018, pp. 1–5, doi: 10.1109/ISCAS.2018.8494884.

[16] S. Pinto and N. Santos, "Demystifying Arm TrustZone," ACM Computing Surveys, vol. 51, no. 6, pp. 1–36, Jan. 2019, doi: 10.1145/3291047.

[17] S. F. J. J. Ankergård, E. Dushku, and N. Dragoni, "State-of-the-Art Software-Based Remote Attestation: Opportunities and Open Issues for Internet of Things," Sensors, vol. 21, no. 5, p. 1598, Feb. 2021, doi: 10.3390/s21051598.

[18] J. Noorman et al., "Sancus: Low-cost trustworthy extensible networked devices with a zero-software Trusted Computing Base," in Proc. 22nd USENIX Security Symposium, Aug. 2013, pp. 479–494.

[19] S. Khan et al., "Securing the IoT ecosystem: ASIC-based hardware realization of Ascon lightweight cipher," International Journal of Information Security, vol. 23, no. 6, pp. 3653–3664, Aug. 2024, doi: 10.1007/s10207-024-00904-1.

[20] N. G. Zinabu, Y. W. Marye, K. K. Tune, and S. A. Demilew, "Comprehensive Analysis of Lightweight Cryptographic Algorithms for Battery-Limited Internet of Things Devices," International Journal of Distributed Sensor Networks, vol. 2025, no. 1, p. 9639728, Jan. 2025, doi: 10.1155/dsn/9639728.

[21] A. A. Diro, N. Chilamkurti, and N. Kumar, "Lightweight Cybersecurity Schemes Using Elliptic Curve Cryptography in Publish-Subscribe fog Computing," Mobile Networks and Applications, vol. 22, no. 5, pp. 848–858, Apr. 2017, doi: 10.1007/s11036-017-0851-8.

[22] D. Chen, F. Zhou, Y. Liu, L. Li, and Y. Liang, "Secure pairing-free certificateless aggregate signcryption scheme for IoT," Journal of Systems Architecture, vol. 156, p. 103268, Nov. 2024, doi: 10.1016/j.sysarc.2024.103268.

[23] S.-H. Seo, J. Won, S. Sultana, and E. Bertino, "Effective Key Management in Dynamic Wireless Sensor Networks," IEEE Transactions on Information Forensics and Security, vol. 10, no. 2, pp. 371–383, Feb. 2015, doi: 10.1109/TIFS.2014.2375555.

[24] J. Sun, H. Xiong, X. Liu, Y. Zhang, X. Nie, and R. H. Deng, "Lightweight and Privacy-Aware Fine-Grained Access Control for IoT-Oriented Smart Health," IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6566–6575, Jul. 2020, doi: 10.1109/JIOT.2020.2974257.

[25] D. Javeed, M. S. Saeed, M. Adil, P. Kumar, and A. Jolfaei, "A federated learning-based zero trust intrusion detection system for Internet of Things," Ad Hoc Networks, vol. 158, p. 103540, May 2024, doi: 10.1016/j.adhoc.2024.103540.

[26] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273–1282.

[27] J. Canedo and A. Skjellum, "Using machine learning to secure IoT systems," in Proc. 14th Annu. Conf. Privacy, Security and Trust (PST), Dec. 2016, pp. 219–222, doi: 10.1109/PST.2016.7906930.

[28] S. B. Saad, B. Brik, and A. Ksentini, "Toward Securing Federated Learning Against Poisoning Attacks in Zero Touch B5G Networks," IEEE Transactions on Network and Service Management, vol. 20, no. 2, pp. 1612–1624, Jun. 2023, doi: 10.1109/TNSM.2023.3278838.

[29] S. R. Moosavi et al., "SEA: A Secure and Efficient Authentication and Authorization Architecture for IoT-Based Healthcare Using Smart Gateways," Procedia Computer Science, vol. 52, pp. 452–459, 2015, doi: 10.1016/j.procs.2015.05.013.

[30] M. Rabbani, L. Rashidi, and A. A. Ghorbani, "A Graph Learning-Based Approach for Lateral Movement Detection," IEEE Transactions on Network and Service Management, vol. 21, no. 4, pp. 4110–4122, Aug. 2024, doi: 10.1109/TNSM.2024.3414267.

[31] Y. Wang et al., "N-STGAT: Spatio-Temporal Graph Neural Network Based Network Intrusion Detection for Near-Earth Remote Sensing," Remote Sensing, vol. 15, no. 14, p. 3611, Jul. 2023, doi: 10.3390/rs15143611.

[32] A. T. Mozipo and J. M. Acken, "Residual vulnerabilities to power side channel attacks of lightweight ciphers cryptography competition finalists," IET Computers & Digital Techniques, vol. 17, no. 3–4, pp. 75–88, May 2023, doi: 10.1049/cdt2.12057.

[33] V. Rijmen, B. Preneel, and E. De Mulder, "Soft Decision Error Correction for Compact Memory-Based PUFs Using a Single Enrollment," in Lecture Notes in Computer Science, vol. 7371, pp. 268–282, Sep. 2012, doi: 10.1007/978-3-642-33027-8_16.

[34] M. S. Turan, "Ascon-Based Lightweight Cryptography Standards for Constrained Devices," National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA, NIST Special Publication 800-232, Jan. 2025, doi: 10.6028/nist.sp.800-232.

[35] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, and G. Srivastava, "Federated Learning-based Anomaly Detection for IoT Security Attacks," IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2545–2554, Feb. 2022, doi: 10.1109/JIOT.2021.3077803.

[36] M. Neagu, C. M. Serban, A. Hangan, and G. Sebestyen, "Trustworthiness in Resource-Constrained IoT: Review and Taxonomy of Privacy-Enhancing Technologies and Anomaly Detection," Telecom, vol. 7, no. 1, p. 10, Jan. 2026, doi: 10.3390/telecom7010010.

[37] G. Xia, J. Chen, C. Yu, and J. Ma, "Poisoning Attacks in Federated Learning: A Survey," IEEE Access, vol. 11, pp. 10452–10471, Mar. 2023, doi: 10.1109/ACCESS.2023.3241582.

[38] J. Yao, W. Shi, W. Xu, Z. Yang, and D. Niyato, "Byzantine-Resilient Over-the-Air Federated Learning Under Zero-Trust Architecture," IEEE Journal on Selected Areas in Communications, vol. 43, no. 5, pp. 1120–1132, May 2025, doi: 10.1109/JSAC.2025.3560046.

[39] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "LSB: A Lightweight Scalable Blockchain for IoT security and anonymity," Journal of Parallel and Distributed Computing, vol. 134, pp. 180–197, Dec. 2019, doi: 10.1016/j.jpdc.2019.08.005.

[40] I. A. Sumra, H. Hasbullah, I. Ahmad and J. -l. bin Ab Manan, "Forming vehicular web of trust in VANET," *2011 Saudi International Electronics, Communications and Photonics Conference (SIECPC)*, Riyadh, Saudi Arabia, 2011, pp. 1-6, doi: 10.1109/SIECPC.2011.5876941.

[41] I. A. Sumra, I. Ahmad, H. Hasbullah and J. -l. bin Ab Manan, "Behavior of attacker and some new possible attacks in Vehicular Ad hoc Network (VANET)," *2011 3rd International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, Budapest, Hungary, 2011, pp. 1-8.