

Enhancing Cyber Resilience Against Zero-Day Exploits in Saudi Critical Infrastructure Through Advanced Detection, Containment and Rapid Incident Response

Ilyas Mohammad Siddiqui^{1*} 

¹Independent Researcher

***Corresponding Author**
Ilyas Mohammad Siddiqui
Independent Researcher

Article History

Received: 27.07.2026

Accepted: 21.09.2026

Published: 25.09.2026

Abstract: Zero-day exploits pose a challenge to critical infrastructure, as defenders must detect malicious behaviour prior to the existence of a reliable signature, patch, or complete threat profile. This problem is particularly acute in Saudi Arabia where energy, water, transportation, telecommunications, healthcare, government, and smart-city services rely on tightly coupled information technology and operational technology. This review summarises recent evidence about three capabilities that determine whether an unknown exploit becomes a contained security event or a cascading service disruption: advanced detection, rapid containment, and coordinated incident response. The researchers undertook a structured integrative review of the DOI-verified literature, from 2020 to 2025, focusing on zero-day detection, industrial control systems, cyber-physical resilience, zero-trust containment, explainable artificial intelligence, federated learning, security orchestration, and recovery. Thirty sources were included in the thematic synthesis. The evidence suggests that no single detector is robust enough to detect unseen attacks. The combination of behaviour baselining, unsupervised or semi-supervised learning, explainable scoring, asset and process context and cross-domain telemetry leads to greater resilience. Detection must then link to safety-aware containment, including identity restriction, micro segmentation, endpoint or workload isolation, and carefully bounded operational-technology actions. Orchestration and automation enable rapid response, but control actions with high consequences require human authorisation and predefined fallback states. The review introduces a Saudi Zero-Day Cyber Resilience Framework that connects detection confidence, operational criticality, containment authority, response playbooks, validation of recovery, and post-incident learning. The framework shifts the focus from perfect prevention to measurable continuity, controlled degradation and faster restoration under uncertainty. It also discusses research directions in local industrial datasets, adversarial robust models, digital-twin validation, privacy-preserving collaboration, and resilience metrics applicable to the Saudi critical infrastructure.

Keywords: zero-day exploit; cyber resilience; Saudi Arabia; critical infrastructure; operational technology; anomaly detection; zero trust; incident response.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

Citation: Siddiqui, I. M. (2026). Enhancing Cyber Resilience Against Zero-Day Exploits in Saudi Critical Infrastructure Through Advanced Detection, Containment and Rapid Incident Response; *Glob Acad J Econ Buss*, 8(5), 1413-1424.
<https://doi.org/10.36348/gajeb.2026.v08i05.032>

INTRODUCTION

Enterprise networks, cloud services, remote access, industrial controllers, sensors, safety systems and third-party software are sharing data on an ongoing basis, leading to critical infrastructure increasingly being run as a cyber-physical system. This connectivity increases efficiency and observability but also increases the paths by which a previously unknown vulnerability can be exploited. Zero-day exploits are especially hard because signature-based controls have little or no validated knowledge of the malicious code, command sequence or behaviour when first used. Recent reviews thus treat zero-day detection as an inference problem: defenders need to detect abnormal intent from behaviour, context, and deviation rather than matching a known indicator [1]. Likewise, empirical work on unseen data shows that models that seem accurate on familiar attacks can degrade sharply when attack classes are held out, making generalisation a more meaningful criterion than benchmark accuracy alone [2].

In industrial and critical infrastructure environments, the challenge is even greater. Confidentiality may be as important as availability and physical safety, and an aggressive containment action may disrupt production, destabilise a process, or remove visibility from operators. So cyber resilience should not be limited to intrusion prevention. It is the ability to anticipate, endure, contain, recover and adapt while still maintaining critical functionality in adverse conditions. Systematic resilience research notes that assessment needs to relate technical disruption to organisational response and restoration, not just count blocked events [10]. Cyber-physical resilience models also suggest that degradation and recovery take place in both digital and physical domains and require metrics for service loss, time and quality of restoration [26].

A strategically important context for this question is Saudi Arabia. Energy and utility systems at scale, the expanding digital government, industrial automation, and smart-city networks are high-value, highly interconnected targets. Studies focused on Saudi Arabia have pointed out the necessity of better data security management for public digital services [25], and research on resilience from Saudi institutions shows the significance of cyber-physical performance measures for power systems [26] and resilient topology for smart city communication networks [27]. These studies don't solve the zero-day problem, but they do establish why Saudi resilience planning must link cybersecurity decisions to continuity of essential services. There is plenty of literature on individual components. Machine learning research proposes classifiers for unseen

attacks [2-7]; industrial control research considers anomaly detection and protocol-aware monitoring [8,9]; resilience research develops assessment and recovery concepts [10,12-15]; zero-trust research tackles identity-centric restriction and segmentation [19-21]; and explainable artificial intelligence, federated learning, orchestration, and automated response work on trust and speed [17,18,22-24,28,30]. What is less well integrated is the operational chain from uncertain detection to safe containment and rapid recovery in critical infrastructure. This review fills that gap of integration instead of considering detection performance as a final security outcome.

1.1. Aim, objectives and research questions

The review synthesises how advanced detection, safety-aware containment, and rapid incident response can be combined into an operational cyber-resilience model for Saudi critical infrastructure exposed to zero-day exploits. The study is guided by five objectives. First, it summarises behavioural, anomaly-based, machine learning and explainable methods for discovering attacks unseen during training. Second, it compares containment mechanisms that confine lateral movement and cascading effects but do not impose unacceptable operational risk. Third, it reviews response and recovery practices that can shorten the time to detect an incident, contain it, and return to normal operations. Fourth, it identifies Saudi-specific architectural and governance requirements such as IT/OT coordination, critical-service dependencies, data localisation concerns and cross-organizational learning. Fifth, it introduces an integrated framework and research agenda for measurable cyber resilience.

These objectives lead to four research questions. RQ1: What are the most transferable detection approaches to detect unseen attacks in critical-infrastructure environments? RQ2 asks how containment can balance speed, isolation strength, service availability and process safety. RQ3: What is the best combination of automation and human decision making to minimise incident-response and recovery time? RQ4 asks what contextual adaptations are needed for a Saudi implementation in heterogeneous sectors. Adding a lifecycle perspective to the review: detection is measured by the quality of decision it enables, containment by the damage it prevents without creating new hazards, and response by the speed and confidence with which essential service is restored.

RESEARCH METHODS

A structured integrative review design was chosen due to the heterogeneity of evidence regarding the research problem which includes machine learning experiments, industrial intrusion

detection studies, cyber resilience frameworks, zero trust architectures, incident response research, and Saudi cyber-physical applications. Because the studies used different data sets, threat models, outcome measures, and system boundaries, a statistical meta-analysis was not appropriate. Instead, the review uses transparent eligibility criteria and thematic synthesis to enable evidence of technical performance to be interpreted within the framework of operational and governance constraints.

To improve transparency and reproducibility, this review was conducted following the principles of the PRISMA 2020 framework. The review process consisted of four main stages: identification, screening, eligibility assessment, and final inclusion. Relevant studies were identified from selected scientific databases, duplicate records were removed, and remaining studies were assessed according to predefined inclusion and exclusion criteria. The final evidence corpus was selected for thematic synthesis based on methodological relevance and contribution to zero-day cyber resilience in critical infrastructure environments.

2.1 Systematic Review Protocol and PRISMA Framework

Searches were structured around five concept blocks: zero-day or unseen attacks; critical infrastructure, industrial control systems, SCADA, operational technology or cyber-physical systems; anomaly detection, machine learning, explainable AI or federated learning; zero trust, segmentation, containment or isolation; and incident response, orchestration, recovery or cyber resilience. The sources for target discovery were Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect, SpringerLink, and publisher databases. The publication window was from January 2020 to December 2025. Eligible items had to provide a DOI, fall within the date window and contribute evidence directly related to detection, containment, response, recovery or resilience. Preference was given to peer-reviewed journal and conference literature, but the

DOI-bearing NIST Zero Trust Architecture was retained as a standards source due to its foundational role in containment design [19].

Exclusion criteria were work published before 2020, papers without a verifiable DOI, purely offensive exploit-development studies, duplicate versions, generic enterprise-security discussions with no transferable relevance to critical infrastructure, and studies reporting only conventional known-signature detection. The screening was not aimed at maximising the number of records but on the conceptual relevance and methodological usefulness. For the final evidence corpus we deliberately limited the sources to the thirty DOI-verified ones in the reference section. Each source was coded according to threat novelty, data type, detection logic, explainability, deployment domain, containment action, automation level, recovery relevance, and resilience implication. The findings were then synthesised into three operational themes – detect, contain, respond/recover – and one cross-cutting theme – Saudi context, governance, learning.

Appraisal of quality was qualitative but explicit. Detection studies were assessed for unseen-class testing, risk of data leakage, class imbalance, implications of false positives, and deployment realism. We analysed critical-infrastructure studies for process context, awareness of protocols, physical consequence, and constraints on availability. Response studies were assessed for decision latency, automation limits, analyst participation and restoration logic. Instead of being equivalent to controlled experiments, we used review and framework papers to define constructs and identify convergent principles. The weighting prevents a high laboratory accuracy being taken as evidence of operational resilience. The synthesis also dissociates evidence and recommendation: the paper offers the adaptation as a proposed design requirement, not a proven national outcome, where Saudi-specific empirical validation is lacking.

Table 1: Review protocol and evidence coding

Element	Review rule	Purpose
Period	January 2020-December 2025	Use contemporary evidence
Sources	Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink and publisher databases	Cross-disciplinary coverage
Include	DOI verified; zero-day, CI/ICS, containment, incident response or cyber-resilience relevance	Direct answer to RQs
Exclude	Pre-2020; no DOI; duplicate; purely offensive; known-signature-only studies	Reduce scope drift
Coding	Threat novelty, data, model, explainability, OT context, containment, automation, recovery	Enable thematic comparison
Synthesis	Thirty DOI-verified sources grouped into detect, contain, respond/recover and context/governance	Build integrated framework

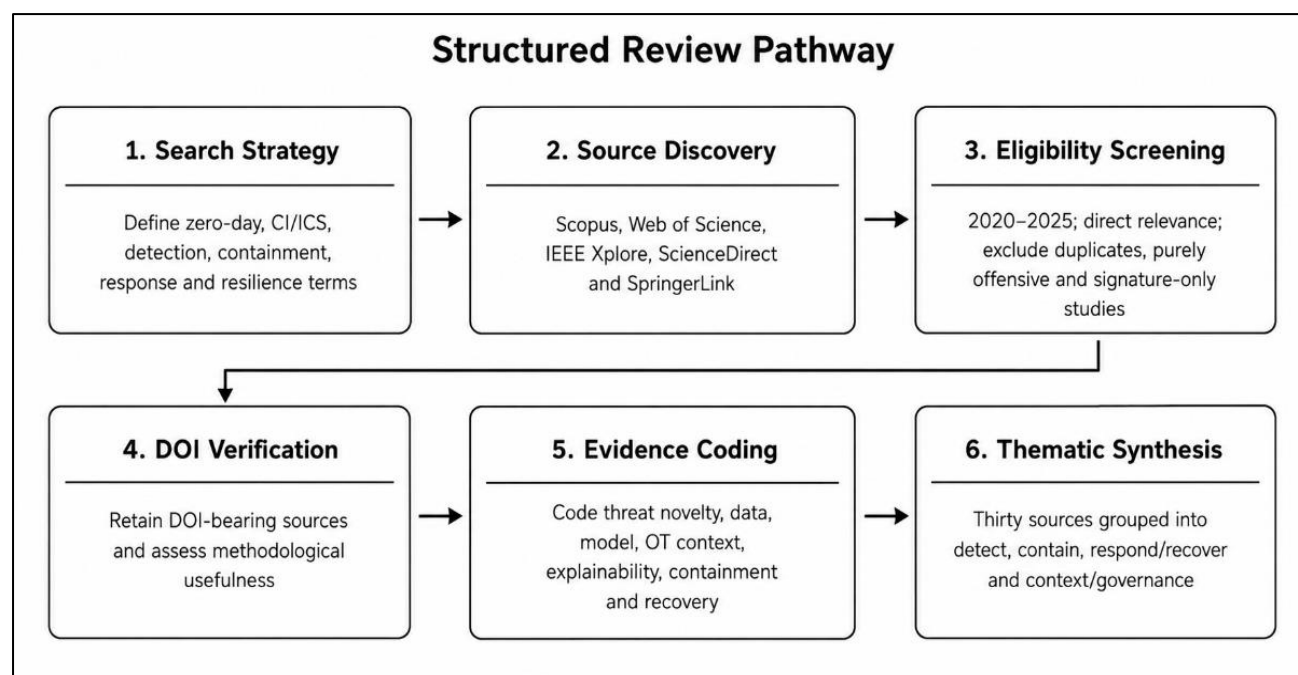


Figure 1: Structured review pathway

2.2 Formal review protocol and Saudi applicability

The study uses a structured integrative review because the evidence combines machine-learning experiments, industrial intrusion-detection studies, cyber-resilience frameworks, zero-trust architecture, incident-response research and Saudi cyber-physical applications with heterogeneous outcomes. Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect, SpringerLink and publisher databases were searched for 2020-2025 evidence using five concept blocks: zero-day/unseen attacks; CI/ICS/SCADA/OT; anomaly detection/ML/XAI/federated learning; zero trust/segmentation/containment; and incident response/SOAR/recovery/resilience. Inclusion required direct relevance and a verifiable DOI for the thirty-source evidence corpus. Saudi NCA publications were used as authoritative contextual requirements and are not counted as DOI-verified research evidence. Appraisal considered unseen-class testing, data leakage, class imbalance, false positives, deployment realism, OT/process context, decision latency, automation limits and restoration logic.

2.3 Saudi regulatory mapping.

The framework is operationalized against NCA requirements for essential cybersecurity, operational technology, cloud and data protection. Governance and asset/dependency context map to ECC and OTCC governance and asset-management

expectations; cloud telemetry and isolation align with cloud cybersecurity controls; data classification and protection align with data cybersecurity controls; safety-aware segmentation and controlled engineering access align with OTCC; and incident management, continuity and recovery map to ECC/OTCC response expectations. This alignment is a requirements traceability exercise, not evidence that a particular control is empirically effective. Implementation evidence should therefore include approved asset maps, segmentation rules, privileged-access logs, detection cases, containment approvals, recovery tests and post-incident corrective actions.

2.4 Validation status.

The Saudi Zero-Day Cyber Resilience Framework is proposed rather than nationally validated. Strong validation should test it in cyber ranges, digital twins or hardware-in-the-loop environments using withheld attack families, realistic benign change, degraded telemetry and process-safety constraints. Success criteria should include generalisation to unseen attacks, acceptable false-positive burden, explainable analyst decisions, bounded propagation, safe containment, continuity of essential function, verified restoration and documented learning. Expert review by Saudi SOC, OT, safety and continuity practitioners can complement technical testing by assessing feasibility, decision authority and compatibility with sector-specific operating constraints.

Table 1A: Saudi requirements-to-framework traceability

Saudi requirement	Framework layer	Implementation evidence
NCA ECC 2-2024	Governance / mission assurance	Risk ownership, asset dependencies, continuity responsibilities
NCA OTCC-1:2022	Safety-aware containment	OT zones, controlled engineering paths, safe-state rules
NCA Cloud Cybersecurity Controls	Infrastructure / detection	Cloud logging, privileged activity, workload isolation, recovery
NCA Data Cybersecurity Controls	Data/privacy rail	Classification, minimisation, protection, retention, recovery
Incident management expectations	Response / recovery	SOC/SOAR playbooks, approvals, timelines, restoration evidence
Threat intelligence practice	Adaptive detection	CTI enrichment, behavioural analytics, hunting and prioritisation

ZERO-DAY RISKS IN SAUDI ARABIA'S CRITICAL INFRASTRUCTURE

Zero-day risk is an asymmetric risk. Attackers need only one exploitable path, while defenders need to interpret sparse and ambiguous signals across many legitimate behaviours. Signature systems are still useful for known malware, but the epistemic weakness is clear when the exploit, payload, or technique has not yet been catalogued. Guo's review explains why zero-day detection with machine learning has shifted towards generalisation, representation learning, and anomaly inference [1]. Kumar and Sinha show the capability of intelligent techniques in detecting unknown attacks, but also highlight the dependence of the results on the quality of features and experimental assumptions [3]. Likewise, Hindy *et al.*, [4] advocate deep-learning approaches that learn traffic representations instead of relying solely on hand-crafted signatures.

Critical infrastructure introduces a second asymmetry: the cost of a false positive can be significantly higher than in ordinary enterprise IT. Blocking a business workstation can be inconvenient and isolating an engineering workstation, historian, or control path during a sensitive process state can impact production or safety. Protocol-aware SCADA research is important as it relates anomaly detection to expected operational exchanges rather than considering all statistical deviations as equally suspicious [8]. Pinto *et al.*, in their broader survey confirmed that machine learning can improve intrusion detection for critical infrastructure, but the quality of deployment is dependent on datasets, realistic validation, interpretability, and adaptation to changing traffic [9].

The third problem is propagation. Modern infrastructure relies on shared identity services, remote maintenance, cloud platforms, data links and supply-chain software. So a zero-day foothold can be lateral movement or cascading consequences before the vulnerability is known. Cyber-resilience research focuses on the key issues of interdependency,

degradation and recovery [10,12]. Digital-twin research offers a helpful perspective: a virtual model of critical assets can support monitoring, scenario testing, continuity analysis, and safer assessment of defensive actions before implementation on live infrastructure [29]. So the key risk question is not just whether an unknown exploit can be detected but whether the operational consequences can be bounded quickly enough to preserve essential service.

ADVANCED DETECTION OF INVISIBLE ATTACKS

The best detection strategy for a zero-day is layered behavioural evidence. Supervised models can discriminate well when training examples look like operational events, but the ultimate test is unseen-class performance. Dai *et al.*, specifically assess unseen attacks and show that feature selection and model design can improve detection of attacks withheld from training [2]. Sohi, Seifert and Ganji use recurrent neural networks to model sequential network behaviour, and show how the temporal structure can reveal malicious activity that static rules may miss [5]. Barros *et al.*, extend the idea to zero-shot learning, where semantic relations are used to discover malware classes with no direct training examples [6]. Such approaches are conceptually attractive for zero-day resilience because they do not rely on exact signatures.

When labels are scarce, unsupervised and semi-supervised methods are especially relevant. Anomaly detection models learn a baseline of normal behaviour and then rank deviations for investigation. Hairab *et al.*, [7] suggest the use of convolutional neural networks with regularisation for improved anomaly detection of zero-day attack. Industrial studies mostly focus on learning the process-consistent patterns from the SCADA traffic [8]. The advantage of resilience is early visibility. The disadvantage of operations is alert uncertainty. Normal maintenance, software changes, failover and unusual production loads or emergency procedures can all cause abnormal behaviour. Therefore, for a

Saudi operator, a pragmatic design should fuse statistical novelty with asset identity, process state, user privilege, network zone, vulnerability context, and threat intelligence rather than trigger containment from anomaly score alone.

Explainability is a critical bridge between detection and action. Patil *et al.*, [22] demonstrate that explainable artificial intelligence can make intrusion-detection decisions very interpretable. Zhang *et al.*, survey explainable cybersecurity applications and the need for exposing reasons behind automated judgements [23]. Almuqren *et al.*, use explainable intrusion detection for cyber-physical systems, directly supporting the importance of interpretable models where physical consequences are of concern [28]. Practically, a model should not just report 'malicious', but report the features and sequence changes, the assets impacted, the confidence and counterfactual conditions that drove the alert. This speeds analyst validation and provides an auditable basis for high impact containment. Another useful layer is federated learning, where organisations cannot centralise sensitive operational telemetry. Ruzafa-Alcázar *et al.*, introduce privacy-preserving federated intrusion detection for Industrial IoT [24]. Federated designs for Saudi critical infrastructure could enable sector participants or geographically dispersed facilities to improve models while keeping

raw data locally. But federation does not mean security by default. Poisoned updates, non-i.i.d data, model drift, and inconsistent asset semantics can all compromise reliability. Therefore, a practical programme would need signed model updates, robust aggregation, local validation gates, common feature definitions, and the ability to reject a global model that performs poorly against local process behaviour.

Thus, the detection architecture should be multi-source, confidence-aware. Network telemetry can reveal scanning, protocol abuse, command-and-control patterns, or lateral movement. Endpoint telemetry can reveal abnormal process trees, memory activity, credential access and persistence. Analysis of OT data can discover unanticipated command sequences, timing deviations, set-point anomalies, and cyber-physical process response mismatches. The identity and access logs give you information about the origin of a session and privilege. Cyber threat intelligence is still valuable against zero-days because campaigns often reuse infrastructure, tools, tactics or identities. The detector must fuse these observations into a risk score that reflects both malicious likelihood and operational importance. This reframes detection as decision support for containment, not a free-standing classification contest.

Table 2: Comparative synthesis of resilience controls for zero-day incidents

Capability	Primary value	Main limitation	CI deployment requirement	Saudi-oriented use
Behavior/anomaly detection	Finds deviations without exact signature	False positives and drift	Process and asset context	SOC/OT cross-domain triage
Explainable AI	Supports analyst trust and auditability	Explanations may not equal causality	Feature transparency and validation	High-consequence alert approval
Federated learning	Shares model learning without raw-data pooling	Poisoning and non-IID data	Secure aggregation and local gates	Cross-site/sector collaboration
Zero trust + segmentation	Limits lateral movement and blast radius	Legacy compatibility and policy errors	Pretested zones and identities	Critical-service containment
SOAR + human authority	Reduces response latency	Unsafe over-automation	Action tiers and safety gates	Rapid evidence collection and bounded isolation
Digital twin / cyber range	Tests controls and restoration safely	Model fidelity and upkeep	Representative dependencies	Exercise and recovery validation

CONTAINMENT WITH SAFETY AWARENESS

Containment is the moment when cyber resilience becomes a reality. When an alert is deemed credible, the aim is to avoid privilege escalation, lateral movement, command propagation, data destruction and physical disruption, while preserving enough system functionality to safely operate. The Zero Trust Architecture provides a valuable

foundation in that it replaces implicit trust based on network location with continuous verification of identity, device, resource, and policy [19]. The broader zero-trust literature confirms that correctly implemented identity centric access, least privilege, continuous evaluation, and fine-grained policy enforcement can reduce attack movement [20,21]. In critical infrastructure, though, the controls must be

designed around deterministic communications, legacy devices, safety functions and recovery requirements.

Micro segmentation and network zoning contain the blast radius of a zero-day. Instead of allowing a compromised host to access large parts of the environment, policy can limit communication to only necessary services and expected peers. Research on network segmentation provides formal ways to reason about allowed paths and reduce unnecessary connectivity. The greatest containment benefit is achieved when the segmentation is prepared prior to an incident and an emergency segmentation that is induced by a crisis can result in outages. Therefore, Saudi operators should identify critical communication flows, establish fail-safe isolation zones and test fallback routes during exercises. High-criticality OT zones should have documented rules for what endpoints can be isolated automatically, what endpoints require control-room approval, and what endpoints should never be disconnected without an engineered transition.

Not binary, graduated containment. For low-confidence alerts, it may make sense to add additional logging, session reauthentication, or rate limiting or temporary restriction of credentials. Medium confidence events can trigger endpoint quarantine, workload isolation, token revocation, blocking of anomalous east-west traffic, and tighter egress control. High-confidence events impacting a critical process may require the coordinated isolation of an engineering station, disabling of a remote access pathway, switching to a redundant control channel, or moving the process to a safe operating mode. This graduated model maintains proportionality: the stronger the cyber evidence and the greater the consequence of inaction, the more aggressive the permitted containment (within safety constraints). Resilience also hinges on pre-authorized decision rights. In a zero-day, teams can't wait for a new governance process to be invented. Playbooks should identify the incident commander, OT authority, safety representative, communications lead, legal/regulatory interface and vendor escalation path. The operational-technology case study of Perrett and Wilson explains why resilience analysis has to be based on real industrial dependencies and operational constraints [16]. Saudi organisations can operationalize that principle through the use of containment matrices that combine asset criticality with authorised actions, maximum tolerable isolation time, manual override conditions and evidence thresholds. This matrix becomes a conduit between policy and machine-executable response.

QUICK INCIDENT RESPONSE AND RECOVERY

Fast incident response is not just fast automation, it is fast, coordinated reduction of uncertainty and consequence. Security orchestration, automation, and response can enrich alerts, correlate telemetry, perform routine queries, open cases, gather evidence, and enact predefined containment actions [17]. CRUSOE shows how integrated situational awareness and decision support can improve incident handling by fusing heterogeneous information into a coherent operational view [18]. The practical value in a zero-day case is analyst cycle time compression. The platform can gather host state, identity context, recent network flows, threat intelligence matches, process criticality, and past alerts before a human analyst begins manual investigation.

Automation should be limited by consequence. Automating evidence collection, sandboxing, IOC searches, scoring credential risk or isolating a noncritical enterprise endpoint is reasonable. It is far less reasonable to permit a black-box model to stop a turbine-related controller or alter a safety-critical configuration without engineered safeguards. Critical-infrastructure SOAR thus requires explainable outputs [22,23,28], explicit confidence thresholds, and human-in-the-loop authorisation. Alevizos describes how unambiguous policy encoding can decrease response delay via automatic policy enforcement and dynamic control adjustment [30]. The same idea should hold for Saudi critical infrastructure, with an extra safety gate: machine recommendations may be fast, but any actions that go past a certain physical-consequence threshold require operational approval.

Recovery has to start while the containment is still in place. Teams should identify clean restoration points, verify configuration integrity, preserve forensic evidence, and determine if the exploit path is still reachable. A recovered server placed back into production without validating credentials, dependencies and adjacent assets may just bring the adversary right back. Staves *et al.*, have developed an incident response and recovery framework for industrial control systems that emphasises the specific requirements of operational environments and the need to link response to restoration [13]. Ahmad *et al.*, show [11] situation awareness is a management practice and a technical capability. To respond effectively, teams need to understand business context and threat evolution, not just follow isolated alerts.

Post-incident learning closes the resilience loop. The organisation must translate forensic findings into behavioural detections, segmented access rules, updated playbooks, recovery tests, and

architectural changes. And resilience is enhanced the next time an unknown exploit “meets” a system that has previously learned from ambiguity. Tzavara and Vassiliadis refer to the development of cyber resilience as a progressive concept based on preparation, absorption, recovery and adaptation [14]. The authors of [15] also associate resilience of critical infrastructure with governance, standards and risk assessment. These views support a Saudi operating model where incident closure is not the end point, but part of closure criteria is measurable control improvement.

PROPOSED CYBER RESILIENCE FRAMEWORK (SAUDI ZERO-DAY)

The synthesis supports a closed operational loop Saudi Zero-Day Cyber Resilience Framework. Layer 1 defines asset and dependency context. Organisations have authoritative inventories, communication baselines, data classifications, operational criticality, recovery dependencies, and safe-state information. Layer 2 uses network, endpoint, identity, cloud and OT telemetry for adaptive detection. Zero-day evidence can be weak in any one source, so multiple analytic methods are combined. Layer 3 combines model confidence, asset criticality, process state, threat intelligence and explainable evidence to validate risk. The output is not a generic severity label, but an actionable decision score with the associated containment authority.

Layer 4: containing with safety awareness. Playbooks that have been pretested define the zero-trust policy, segmentation, identity restriction, endpoint isolation, egress control and alternate operating paths. Layer 5 is rapid response and recovery. Here, SOAR automates evidence gathering and low-risk actions, with high-consequence interventions approved by human incident and OT authorities. Restoration involves validation of integrity, reset of credentials where needed, controlled reconnection, higher levels of monitoring

and validation of performance of essential service. Layer 6 is learning and measuring resilience What detection signals were decisive? What controls slowed propagation? How long were critical functions degraded? Was the restored system stable? The organisation logs this information. All layers are controlled by three rails. The first is operational safety: cyber actions must respect process hazards, the failover design and emergency procedures. The second is accountability: models, playbooks, and automated actions need owners, logs, change control, and periodic validation. Third, collaboration and privacy: organisations should share indicators, behavioural patterns, lessons and model improvements without unwarranted exposure of sensitive operational data. Federated techniques [24] and digital twins [29] are particularly interesting here. Federated learning can facilitate distributed improvement. Digital twins can generate controlled environments to test segmentation, recovery sequences, and response playbooks before deployment.

The Saudi adaptation should focus on the heterogeneity of sectors rather than a single universal technical stack. Power utility, water facility, refinery, hospital, telecom operator, and smart city platform have different safety constraints, latency requirements, legacy technologies and tolerances for isolation. So the commonality at national level should be in the resilience outcomes and the decision logic: time to detect, time to validate, time to contain, percentage of essential service retained, maximum depth of propagation, time to recover, and percentage of lessons converted into tested controls. For instance, AlMuhaini’s cyber-physical resilience quantification [26] and Alenazi’s resilient network design [27] demonstrate the usefulness of measuring continuity and structural robustness, while Alharbi *et al.*, illustrate the governance dimension of Saudi data security [25]. The proposed framework concentrates those concerns in a single incident lifecycle.

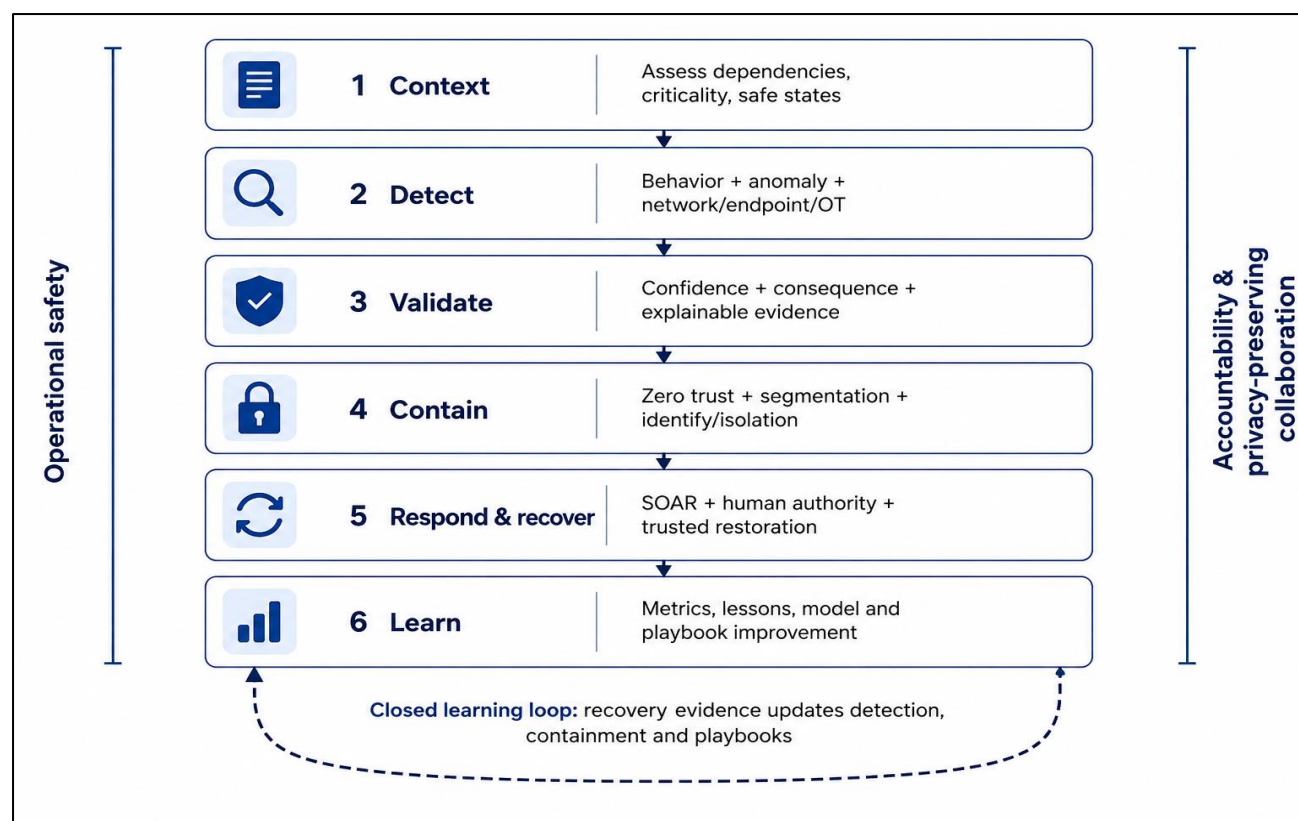


Figure 2: Saudi Zero-Day Cyber Resilience Framework

DISCUSSION AND IMPLEMENTATION PRIORITIES

The literature indicates four implementation priorities. First, organisations should optimise to generalisation to unknown attacks, not headline benchmark accuracy. Evaluation sets should include withheld attack families, realistic benign changes, and operationally important rare events. Models need to be tested for false positive burden and drift, not just precision and recall. Second, detection engineering must be well integrated with asset and process context. The same anomaly on a low-criticality workstation and a privileged engineering station do not pose the same threat. Risk scoring has to reflect consequence.

Third, containment controls should be developed and rehearsed prior to an incident. Zero trust and segmentation are best as architectural properties, not emergency improvisations. They only succeed if they have accurate identity, dependency mapping and known recovery paths. Fourth, response automation should aim to delay rather than delegate uncontrolled authority to algorithms. Good automation candidates include evidence collection, enrichment, ticketing, correlation, and low-risk isolation. Changes which have the potential to change a physical process, disable safety visibility or interrupt an essential service require human authorisation unless an engineered fail-safe

condition has been explicitly approved. The framework also changes what to measure. Mean time to detect is still useful, but not enough. Zero-day programme should measure time from first anomaly to validated incident, validated to effective containment, time to safe service restoration, proportion of critical function retained during disruption, number of zones or dependencies reached before containment, rate of unnecessary high-impact actions, percentage of lessons converted into controls. Resilience metrics force teams to own the entire consequence curve, not just feel good about detection while recovery is slow.

Local evidence is a high priority for Saudi critical infrastructure. Imported datasets rarely capture Arabic-language lures, local operating practices, local vendor mixes, industry protocols, maintenance patterns, or sector-specific threat models. The most valuable research investment may thus be in governed datasets and cyber ranges that reproduce realistic IT/OT dependencies without exposing sensitive production data. With a common evaluation methodology, Saudi organisations could uniformly compare generalisation, explainability, containment latency, and recovery performance across sectors.

Another implementation problem is the design of the exercise Tabletop scenarios are helpful for governance, but technical teams also need

repeatable cyber-range exercises that expose analysts to incomplete evidence, conflicting alerts, degraded telemetry, and operational pressure. Exercises should challenge the decision of when is the evidence sufficient to contain, when process safety is more important than immediate isolation, and what minimum information is needed before restoration. The goal is not to repeat a single forecast exploit, but to practice uncertainty management. A mature exercise programme can also reveal hidden dependencies such as authentication of vendors, time synchronisation, certificate services, backup communications, engineering laptops, and cloud-based monitoring. Such dependencies may be the difference between a feasible containment plan in theory and practice. Metrics collected during exercises should directly inform playbook updates, staffing decisions, architectural improvements, and future training requirements. Infrastructure is constantly changing and can invalidate assumptions about detection coverage, containment permissions, recovery readiness, and dependencies. This means that operational maturity also needs revalidation.

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Limitations of this review Rather than an exhaustive bibliometric or meta-analytic review, it is an integrative synthesis of a deliberately bounded thirty source corpus. Because the studies included used heterogeneous datasets and metrics, numerical performance values are not pooled. A lot of the zero-day literature is evaluated in laboratory network datasets, rather than live critical infrastructure which limits direct transferability. There is relatively little peer-reviewed evidence specific to Saudi Arabia and, therefore, several elements of the framework are rational adaptations of international research on critical-infrastructure and cyber-resilience rather than validated national effects. The 2020-2025 date window is current, but not older foundational industrial-security work.

Future work should validate the approach in Saudi cyber ranges and digital twins on a sector basis, publish privacy-preserving benchmark datasets capturing realistic industrial behaviour, test for adversarial robustness and model drift, compare centralised and federated detection under non-identically distributed data, and evaluate human factors in high-consequence SOAR decisions. Controlled exercises should measure not only if an exploit is found, but also how quickly teams can validate it, if containment maintains safe operation, how far the event propagates and how quickly essential service is restored. Longitudinal studies should also examine if learning from incidents leads to measurable reductions of consequence in later events.

CONCLUSION

The issue with security mechanisms relying on past knowledge is that they have an inherent flaw – zero-day exploits. So, resilience for Saudi critical infrastructure means the ability to make safe decisions in the face of uncertainty. The reviewed evidence supports a layered approach where behavioural and anomaly-based detection, explainable analytics, privacy-preserving collaboration, zero-trust containment, orchestration, human operational authority and recovery engineering operate as a single life cycle. No machine-learning model, segmentation product or response platform is enough by itself. The critical capability is integration: weak signals must be converted into credible risk judgements; risk judgements must activate proportionate containment; containment must protect both cyber assets and physical operations; and recovery must restore trusted service, and provide lessons back into architecture and detection.

The proposed Saudi Zero-Day Cyber Resilience Framework operationalizes that logic through six layers and three governance rails. Its main thesis is pragmatic: critical-infrastructure security should be assessed in terms of the preservation and restoration of essential services when prevention is imperfect. By creating local baselines of behaviour, evaluating generalisation to unseen attacks, pre-approving safe containment, automating low-risk investigative work, validating response in digital twins or cyber ranges, and measuring service continuity in addition to detection performance, organisations in Saudi can improve readiness. This shifts the security goal from knowing all threats before they occur to controlling consequence, reducing disruption, and learning faster than foes can exploit uncertainty.

REFERENCES

1. Guo, Y. (2023). A review of machine learning-based zero-day attack detection: Challenges and future directions. *Computer Communications*, 198, 175-185. <https://doi.org/10.1016/j.comcom.2022.11.001>
2. Dai, Z., Por, L. Y., Chen, Y. L., Yang, J., Ku, C. S., *et al.*, (2024). An intrusion detection model to detect zero-day attacks in unseen data using machine learning. *PLOS ONE*, 19(9), e0308469. <https://doi.org/10.1371/journal.pone.0308469>
3. Kumar, V., & Sinha, D. (2021). A robust intelligent zero-day cyber-attack detection technique. *Complex & Intelligent Systems*, 7, 2211-2234. <https://doi.org/10.1007/s40747-021-00396-9>
4. Hindy, H., Atkinson, R., Tachtatzis, C., Colin, J.-N., Bayne, E., & Bellekens, X. (2020). Utilising deep learning techniques for effective zero-day attack

- detection. *Electronics*, 9(10), 1684. <https://doi.org/10.3390/electronics9101684>
5. Sohi, S. M., Seifert, J.-P., & Ganji, F. (2021). RNNIDS: Enhancing network intrusion detection systems through deep learning. *Computers & Security*, 102, 102151. <https://doi.org/10.1016/j.cose.2020.102151>
6. Barros, P. H., Chagas, E. T. C., Oliveira, L. B., Queiroz, F., & Ramos, H. S. (2022). Malware-SMELL: A zero-shot learning strategy for detecting zero-day vulnerabilities. *Computers & Security*, 120, 102785. <https://doi.org/10.1016/j.cose.2022.102785>
7. Hairab, B. I., Aslan, H. K., Elsayed, M. S., Jurcut, A. D., & Azer, M. A. (2023). Anomaly detection of zero-day attacks based on CNN and regularization techniques. *Electronics*, 12(3), 573. <https://doi.org/10.3390/electronics12030573>
8. Lin, C.-Y., & Nadjm-Tehrani, S. (2023). Protocol study and anomaly detection for server-driven traffic in SCADA networks. *International Journal of Critical Infrastructure Protection*, 42, 100612. <https://doi.org/10.1016/j.ijcip.2023.100612>
9. Pinto, A., Herrera, L.-C., Donoso, Y., & Gutierrez, J. A. (2023). Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure. *Sensors*, 23(5), 2415. <https://doi.org/10.3390/s23052415>
10. Sepúlveda Estay, D. A., Sahay, R., Barfod, M. B., & Jensen, C. D. (2020). A systematic review of cyber-resilience assessment frameworks. *Computers & Security*, 97, 101996. <https://doi.org/10.1016/j.cose.2020.101996>
11. Ahmad, A., Maynard, S. B., Desouza, K. C., Kotsias, J., Whitty, M. T., & Baskerville, R. L. (2021). How can organizations develop situation awareness for incident response: A case study of management practice. *Computers & Security*, 101, 102122. <https://doi.org/10.1016/j.cose.2020.102122>
12. Salvi, A., Spagnoletti, P., & Noori, N. S. (2022). Cyber-resilience of critical cyber infrastructures: Integrating digital twins in the electric power ecosystem. *Computers & Security*, 112, 102507. <https://doi.org/10.1016/j.cose.2021.102507>
13. Staves, A., Anderson, T., Balderstone, H., Green, B., Gouglidis, A., & Hutchison, D. (2022). A cyber incident response and recovery framework to support operators of industrial control systems. *International Journal of Critical Infrastructure Protection*, 37, 100505. <https://doi.org/10.1016/j.ijcip.2021.100505>
14. Tzavara, V., & Vassiliadis, S. (2024). Tracing the evolution of cyber resilience: A historical and conceptual review. *International Journal of Information Security*, 23, 1695-1719. <https://doi.org/10.1007/s10207-023-00811-x>
15. Aghazadeh Ardebili, A., Lezzi, M., & Pourmadadkar, M. (2024). Risk assessment for cyber resilience of critical infrastructures: Methods, governance, and standards. *Applied Sciences*, 14(24), 11807. <https://doi.org/10.3390/app142411807>
16. Perrett, K., & Wilson, I. D. (2023). A cyber resilience analysis case study of an industrial operational technology environment. *Environment Systems and Decisions*, 43, 178-190. <https://doi.org/10.1007/s10669-023-09895-1>
17. Kinyua, J., & Awuah, L. (2021). AI/ML in security orchestration, automation and response: Future research directions. *Intelligent Automation & Soft Computing*, 28(2), 527-545. <https://doi.org/10.32604/iasc.2021.016240>
18. Husák, M., Sadlek, L., Špaček, S., Laštovička, M., Javorník, M., & Komárková, J. (2022). CRUSOE: A toolset for cyber situational awareness and decision support in incident handling. *Computers & Security*, 115, 102609. <https://doi.org/10.1016/j.cose.2022.102609>
19. Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture. NIST Special Publication 800-207. <https://doi.org/10.6028/NIST.SP.800-207>
20. Syed, N. F., Shah, S. W., Shaghaghi, A., Anwar, A., Baig, Z. A., & Doss, R. (2022). Zero Trust Architecture (ZTA): A comprehensive survey. *IEEE Access*, 10, 57143-57179. <https://doi.org/10.1109/ACCESS.2022.3174679>
21. He, Y., Huang, D., Chen, L., Ni, Y., & Ma, X. (2022). A survey on zero trust architecture: Challenges and future trends. *Wireless Communications and Mobile Computing*, 2022, 6476274. <https://doi.org/10.1155/2022/6476274>
22. Patil, S., Varadarajan, V., Mazhar, S. M., Sahibzada, A., Ahmed, N., Sinha, O., Kumar, S., Shaw, K., & Kotecha, K. (2022). Explainable artificial intelligence for intrusion detection system. *Electronics*, 11(19), 3079. <https://doi.org/10.3390/electronics11193079>
23. Zhang, Z., Hamadi, H. A., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEE Access*, 10, 93104-93139. <https://doi.org/10.1109/ACCESS.2022.3204051>
24. Ruzafa-Alcázar, P., Fernández-Saura, P., Mármol-Campos, E., González-Vidal, A., Hernández-Ramos, J. L., Bernal-Bernabe, J., & Skarmeta, A. F. (2023). Intrusion detection based on privacy-preserving federated learning for the industrial IoT. *IEEE Transactions on Industrial Informatics*, 19(2), 1145-1154. <https://doi.org/10.1109/TII.2021.3126728>

25. Alharbi, A. S., Halikias, G., Rajarajan, M., & Yamin, M. (2021). A review of effectiveness of Saudi E-government data security management. *International Journal of Information Technology*, 13, 573-579. <https://doi.org/10.1007/s41870-021-00611-3>
26. AlMuhaini, M. (2024). Resilience quantification model for cyber-physical power systems. *IET Cyber-Physical Systems: Theory & Applications*. <https://doi.org/10.1049/cps2.12098>
27. Alenazi, M. J. F. (2024). ResiSC: A system for building resilient smart city communication networks. *Expert Systems*, 41(11), e13698. <https://doi.org/10.1111/exsy.13698>
28. Almuqren, L., Maashi, M. S., Alamgeer, M., Mohsen, H., Hamza, M. A., & Abdelmageed, A. A. (2023). Explainable artificial intelligence enabled intrusion detection technique for secure cyber-physical systems. *Applied Sciences*, 13(5), 3081. <https://doi.org/10.3390/app13053081>
29. Lampropoulos, G., Larrucea, X., & Colomo-Palacios, R. (2024). Digital twins in critical infrastructure. *Information*, 15(8), 454. <https://doi.org/10.3390/info15080454>
30. Alevizos, L. (2025). Automated cybersecurity compliance and threat response using AI, blockchain and smart contracts. *International Journal of Information Technology*, 17, 767-781. <https://doi.org/10.1007/s41870-024-02324-9>