DOI https://doi.org/10.32689/maup.it.2026.2.11
UDC 004.056.5:004.94:004.7
JEL Classification: C52, C63, O33, R58

MULTILEVEL EXPERIMENTAL VALIDATION OF A SMART-REGION CYBER-RESILIENCE METHODOLOGY: FROM FORMAL VERIFICATION TO HARDWARE-IN-THE-LOOP TESTING, PILOT DEPLOYMENT, AND EXTERNAL REPLICATION

Yuliia Khokhlachova

БАГАТОРІВНЕВА ЕКСПЕРИМЕНТАЛЬНА ВАЛІДАЦІЯ МЕТОДОЛОГІЇ КІБЕРСТІЙКОСТІ РОЗУМНОГО РЕГІОНУ: ВІД ФОРМАЛЬНОЇ ВЕРИФІКАЦІЇ ДО ТЕСТУВАННЯ АПАРАТНОГО ЗАБЕЗПЕЧЕННЯ В ЦИКЛІ, ПІЛОТНОГО РОЗГОРТАННЯ ТА ЗОВНІШНЬОГО ВІДТВОРЕННЯ

Ю. Є. Хохлачова

Abstract

Object of study. Multilevel validation of a cyber-resilience methodology for interdependent smart-region services. Problem. Results from formal verification, simulation, hardware-in-the-loop testing, pilot deployment, and external replication are not interchangeable; missing claim-to-evidence links encourage unsupported generalization. Purpose. To develop and partially verify a reproducible V0-V5 validation programme. Methods. The programme defines twelve attack and failure scenarios A1-A12, six resource-comparable configurations B0-B5, and paired ablations of the evidence pipeline, digital twin, and policy controller; it uses synchronized traces, common event markers, confidence intervals, effect sizes, and multiple-testing correction. Results. In the available V3 aggregates, B5 versus B1 reduced mean detection time from 43,1 to 8,5 s (80,3%) and recovery time from 399 to 122 s (69,4%), achieved 98,7% integrity-verification accuracy with less than 6% overhead, and increased the normalized resilience index from 0,71 to 0,93. Distinctive features. Unlike single-node assessment, the architecture combines validation levels, claim-to-evidence traceability, and cumulative I0-I4 maturity gates while separating laboratory causality, operational feasibility, and external transferability. The changes are explained by the joint operation of a verified telemetry pipeline, a cross-sector dependency model, and a constrained recovery controller. Practical use is appropriate in controlled regional infrastructure with an equal resource budget, synchronized measurements, verified safety rules, and explicit evidence boundaries. Conclusions. The aggregates support a descriptive proof of concept for the integrated system, but component-level causality requires run-level logs, completed ablations, a V4 pilot, and independent V5 replication.

Key words: *experimental validation, hardware-in-the-loop, external replication, cyber-resilience maturity, attack scenarios, baselines, reproducibility, smart region.*

Анотація

Об'єкт дослідження. Багаторівнева валідація методології кіберстійкості взаємозалежних сервісів розумного регіону. Проблема. Результати формальної перевірки, моделювання, апаратного тестування, пілотного впровадження та зовнішнього відтворення не є взаємозамінними, а відсутність зв'язку між твердженнями й доказами спричиняє необґрунтоване узагальнення ефектів. Мета. Розробити та частково перевірити відтворювану програму валідації від V0 до V5. Методи. Сформовано дванадцять сценаріїв атак і відмов A1-A12, шість ресурсно порівнюваних конфігурацій B0-B5, парні абляції конвеєра доказів, цифрового двійника і контролера політик; передбачено синхронізовані траси, спільні часові маркери, довірчі інтервали, розміри ефектів і корекцію множинних порівнянь. Результати. За агрегатами V3 конфігурація B5 порівняно з B1 зменшила середній час виявлення з 43,1 до 8,5 с (на 80,3 %), час відновлення – з 399 до 122 с (на 69,4 %), забезпечила точність перевірки цілісності 98,7 % за накладних витрат менш як 6 % і підвищила нормалізований індекс стійкості з 0,71 до 0,93. Особливості результату. На відміну від оцінювання окремого вузла, архітектура поєднує рівні валідації, простежуваність «твердження-доказ» і рівні зрілості I0-I4, розмежовуючи лабораторну причинність, операційну здійсненність і зовнішню переносимість. Зміни пояснюються спільною роботою перевіреного конвеєра телеметрії, моделі міжсекторних залежностей і обмеженого контролера відновлення. Практичне використання можливе у контрольованій регіональній інфраструктурі за однакового ресурсного бюджету, синхронізованих вимірювань, перевірених правил безпеки та явного врахування меж доказів. Висновки. Агрегати підтверджують описовий доказ концепції інтегрованої системи, але не визначають причинний внесок компонентів без журналів запусків, завершених абляцій, пілоту V4 і незалежної реплікації V5.

Ключові слова: *експериментальна перевірка, апаратне забезпечення в циклі, зовнішня реплікація, зрілість кіберстійкості, сценарії атак, базові рівні, відтворюваність, розумний регіон.*

1. Introduction. A smart region combines energy, telecommunications, water, transport, emergency, and public digital services through shared sensors, edge nodes, networks, cloud platforms, and control centres. A local cyber incident can therefore alter both an asset and dependent services with different criticality and recovery objectives. Validation must address this interdependence rather than report the uptime of an isolated node.

© Khokhlachova Yu. Ye., 2026
Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0

Evidence does not transfer automatically across experimental levels. Formal safety does not demonstrate controller speed; simulation omits driver, queue, and clock effects; hardware-in-the-loop testing cannot reproduce all organizational delays; and a pilot without control does not establish general causality. Each claim must therefore be linked to a validation level, artifact, and explicit boundary of interpretation.

Bocci et al. [1] combine declarative cloud-edge placement with information-flow analysis and evaluate deployment performance through simulation. This supports secure placement decisions, but does not establish service continuity during recovery. Yang et al. [2] develop multi-replica integrity verification and reputation-based collaborative auditing for edge servers. Their approach strengthens the evidence pipeline, yet verification of stored replicas cannot establish whether a sensor measurement was physically truthful before storage. Placement security and evidence integrity therefore require distinct validation outcomes.

Fett et al. [3] bridge conceptual digital-twin reference architectures and practical cloud, edge, and hybrid implementations. This provides deployment guidance, while attack-specific acceptance criteria must be defined separately. El-Hajj [4] evaluates a digital twin and intrusion detection in a Raspberry Pi test environment, reporting flood detection alongside limitations in scan detection and resource use. That study demonstrates why successful alerting should be distinguished from timely recovery; extending its findings to interdependent regional services requires additional evidence.

Cho et al. [5] propose availability metrics and area-under-the-curve normalization, evaluated using transaction and connection rates under denial-of-service conditions. These measures inform the present resilience index, but cannot substitute for integrity or safety criteria. Tao et al. [6] synthesize industrial digital-twin applications and physical-cyber integration. Their conceptual account helps define the twin component; the resilience of a particular implementation nevertheless remains an empirical question. Consequently, architectural plausibility and measured mission preservation should not be treated as equivalent evidence.

Alcaraz and Lopez [7] organize digital-twin threats by functionality layers and operational requirements, providing a basis for attack coverage and trust boundaries. A threat taxonomy, however, does not demonstrate that a control restores mission capability. Rose et al. [8] define Zero Trust around resource protection and explicit authentication and authorization, without implicit trust based on network location. These principles constrain admissible actions in the proposed controller, while their operational effectiveness must be assessed through recovery, service-degradation, and safety measurements.

Shi et al. [9] explain the rationale for processing near data sources and discuss edge service management, security, privacy, and optimization challenges. This motivates the deployment profile, but actual latency under hostile traffic requires measurement. Kritzinger et al. [10] distinguish digital models, digital shadows, and digital twins by their degree of physical-digital integration. This classification prevents monitoring alone from being equated with a synchronized twin; it does not establish forecast accuracy or secure closed-loop operation.

Sengupta et al. [11] survey attacks, countermeasures, and blockchain approaches across IoT and IIoT settings. Their synthesis supports threat selection, but ledger-based recovery still requires tests of communication cost and trust assumptions. Mozzaquatro et al. [12] combine ontology-based security design with runtime threat classification and adaptation, including an industrial case study. This supports consistent security semantics, whereas cross-sector recovery effects require resource-comparable experiments. Thus, threat coverage and semantic consistency inform the programme without proving its integrated effectiveness.

Dietz and Pernul [13] explain how digital twins support industrial-control security through simulation and replication of physical-system behaviour. Replication inside a twin must be distinguished from independent repetition by another team and site, the objective of V5. Aghazadeh Ardebili et al. [14] relate critical-infrastructure cyber resilience to risk-assessment methods, governance, and standards. This broadens the selection of evaluation criteria, but does not establish the performance of a particular recovery controller or its transferability between infrastructures.

Chandola et al. [15] compare anomaly-detection methods through their assumptions, limitations, and computational demands. This supports explicit detector selection, but discrimination of anomalies alone cannot demonstrate safe recovery. Xu et al. [16] examine industrial IoT through interacting control, networking, and computing subsystems and their reliability requirements. This motivates coupled cyber-physical testing, while industrial requirements need recalibration for regional services with different criticality and recovery objectives. Detector performance must therefore be interpreted within the wider service and control context.

Ranaweera et al. [17] map security and privacy threats to the multi-access edge-computing architecture and examine countermeasures. This informs the edge threat model, while end-to-end recovery gains under fixed resources remain to be tested. De Neira et al. [18] classify DDoS-prediction research by warning time, architecture, and methodology. Prediction before an attack develops differs from detection after onset and subsequent recovery; comparisons therefore require common event markers and equivalent exposure rather than interchangeable timing metrics.

Fuller et al. [19] review enabling technologies and challenges across manufacturing, healthcare, and smart-city digital twins. This establishes application breadth, but outcomes from these domains cannot be pooled into a regional resilience claim without explicit transfer conditions. Minerva et al. [20] derive foundational twin properties and a generic architectural model from technical features and application scenarios. This guides state representation and synchronization, while conformity to an architecture must be complemented by calibrated cascade forecasts and verified intervention outcomes.

Sharma et al. [21] identify incomplete implementation disclosure, domain dependence, and inadequate performance metrics as obstacles to evaluating and comparing digital twins. These findings directly motivate run manifests and common outcomes, while the present programme still requires released run-level artifacts to demonstrate reproducibility. Homaei et al. [22] review cybersecurity risks affecting digital twins and the potential role of artificial intelligence in protection. This supports treating the twin itself as security-critical, with separate tests of uncertainty, state freshness, and action safety.

Taken together, these contributions provide placement, auditing, architectural, detection, and assessment foundations. Their different objectives and validation settings leave a methodological need for a common link between claims, resource-comparable experiments, and level-specific artifacts. The proposed programme addresses that need by separating formal safety, laboratory effects, operational feasibility, and external transfer. It does not assume that combining established components has already demonstrated all four: each requires its own evidence and explicit limits of interpretation.

This article addresses that gap through a V0-V5 programme that separates formal correctness, laboratory causality, operational feasibility, and external transferability. Its contribution is an architecture combining twelve scenarios A1-A12, configurations B0-B5, paired ablations, claim-to-evidence traceability, and the I0-I4 maturity model. Only supplied V3 aggregates are treated as observations; arithmetic changes are derived, whereas V0-V2, V4, and V5 remain protocols where the required evidence bundles are absent.

The integration problem is methodological as well as technical. A detector may improve alert speed while leaving recovery unchanged; an integrity audit may verify stored blocks without establishing whether a sensor reported the physical process truthfully; and an optimizer may preserve availability while violating an authorization or safety constraint. The programme therefore evaluates mission outcome, mechanism quality, safety, resource cost, and transferability separately, while retaining a traceable link between each claim and the evidence that can legitimately support it.

Research aim. The aim of this study is to develop a reproducible V0-V5 validation programme for the cyber resilience of interdependent smart-region services and to assess its integrated proof of concept using the available V3 aggregate evidence. The programme links each claim to the artifacts and interpretation boundaries required at its validation level.

Research objectives. To achieve this aim, the study determines the evidence requirements and admissible claims at each V0-V5 level; quantifies the descriptive B5-versus-B1 changes in the available V3 aggregates within the stated resource-comparable profile; defines paired ablations and criteria for identifying component-level causal contributions; specifies tests of scaling limits and external transferability; translates validation evidence into cumulative I0-I4 maturity gates; and critically compares the findings with prior work to identify limitations and further research needs. These objectives distinguish the aggregate analysis performed here from ablation, pilot, and independent-replication studies that remain to be completed.

2. Materials and methods. The object of study is the multilevel validation process for cyber resilience of interdependent smart-region services. The main hypothesis is that, under an equal resource budget, the integrated B5 configuration improves mission preservation, detection, and recovery relative to B1, while the admissible interpretation remains bounded by the validation level. The analysis assumes identical attack traces, workloads, seeds, and time budgets; synchronized clocks and common event markers; and stable service-criticality weights within each comparison. The principal simplifications are a controlled four-sector profile, emulated sensors, one V3 testbed, aggregate rather than run-level observations, and the absence of completed V4 and V5 studies.

2.1. Research questions and hypotheses. The study asks whether B5 reduces mission loss, detection time, and recovery time under an equal resource budget, and where its scaling and transferability limits occur. Hypotheses and present evidence status are fixed before analysis (Table 1).

ISSN 2786-5460 (Print), ISSN 2786-5479 (Online)

Table 1

Research hypotheses and current evidence status

Code	Hypothesis	Current evidence
H1	B5 reduces mission loss versus resource-comparable baselines	Preliminary aggregate comparison only with B1
H2	The evidence pipeline reduces detection delay without excessive false positives	Run-level B5-E ablation required
H3	The digital twin improves cascade forecasting and localization	Protocol only; numerical series absent
H4	The policy controller reduces recovery time without unsafe actions	Aggregate effect reported; B5-P required
H5	The effect persists to 10^2-10^4 devices within service limits	Scale test planned
H6	V4 and V5 preserve the direction of effect	Not completed

Source: compiled by the author.

Component-level claims require B5-E, B5-D, and B5-P ablations on identical traces; a full-system versus B1 comparison cannot isolate their causal contributions.

2.2. Evidence chain and preregistration. Before the main series, the model version, parameters, metrics, configurations, ablations, exclusion rules, minimum detectable effect, test order, multiplicity families, and disclosure plan are frozen. Any later change creates a new protocol version rather than silently replacing the original hypothesis.

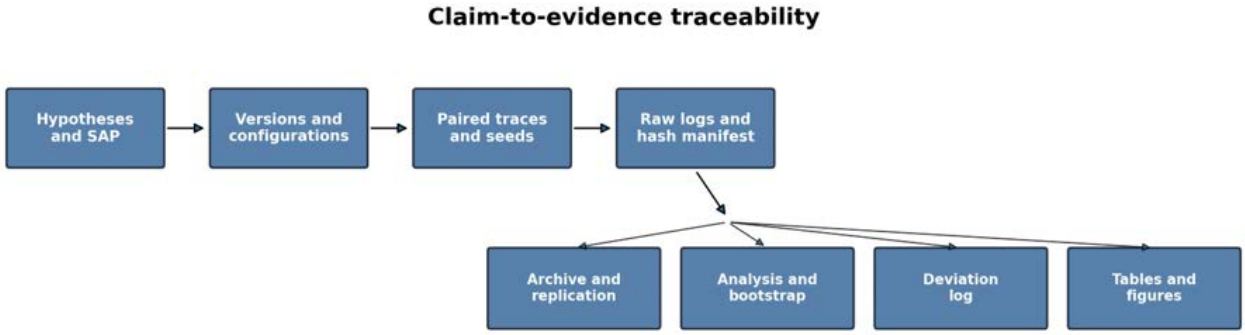


Fig. 1. Claim-to-evidence traceability from preregistration to independent replication

(Source: prepared by the author)

Claim-to-evidence traceability (Figure 1) is implemented through a machine-readable manifest containing artifact hashes, software, firmware and container versions, hardware, seeds, timestamps, synchronization data, and deviations. Hashes establish immutability after acquisition, while authenticity, provenance, transport integrity, and physical plausibility remain separate evidence fields.

2.3. Functional decomposition of the methodology. The evidence pipeline validates identity, timestamps, sequences, provenance, freshness, signatures, and selected replicas. The dependency-aware digital twin synchronizes asset state, estimates cascades, and tests candidate interventions. The constrained policy controller rejects actions that violate authorization, capacity, trust, or safety requirements and selects a reversible action that minimizes expected mission loss.

The test model links energy, telecommunications, water, and public digital services to reproduce the chain physical process – communication – computation – service. It is a controlled cross-sector profile, not a claim of representativeness for every regional asset.

V0 checks authorization, rollback reachability, deadlock freedom, and command consistency. Subsequent levels add simulation, real protocols and queues, hardware limits, operational integration, and independent replication. Figure 2 summarizes the architecture.

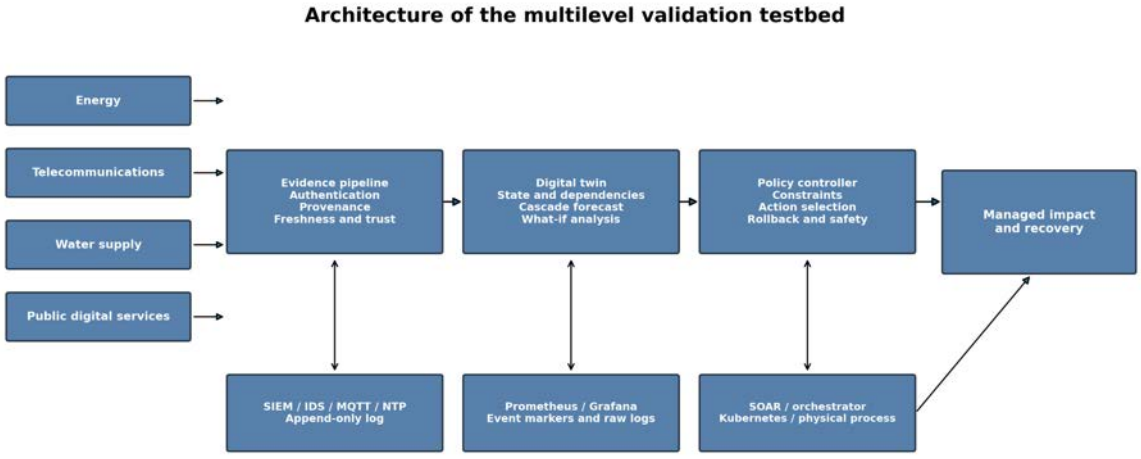


Fig. 2. Architecture of the multilevel testbed and managed recovery loop

(Source: prepared by the author)

2.4. Validation levels V0-V5. Each level answers a distinct evidence question and limits the conclusion that may be drawn (Table 2).

Table 2

Validation levels, evidence purpose, and present status

Level	Evidence purpose	Present status and limit
V0	Formal safety, liveness, authorization, deadlock checks	Protocol; no model-checker report
V1	Simulation of dependencies, extreme cases, factorial series	Protocol; no real network costs
V2	Real protocols, queues, partitions, failover, recovery	Engineering profile only
V3	Hardware and cyber-physical limits, safety interlocks	Reported aggregate evidence in one testbed
V4	Operational fit, data quality, human-in-the-loop performance	Not performed; feasibility only
V5	Independent replication within a frozen profile	Not performed; no universal claim

Source: compiled by the author.

2.5. Experimental platform. The reported V3 profile used 12 emulated Internet-of-Things sensors with Mosquitto 2.0; four Raspberry Pi 5 edge nodes with K3s 1.30, Suricata 7.0, and integrity agents; three 8-vCPU/16-GB virtual machines with Kubernetes 1.30; Eclipse Ditto 3.5; Prometheus 2.53, Grafana 11, and an append-only log; and an isolated Kali Linux host with Hping3 and NMAP. Common event markers identified attack onset, detection, containment, recovery, and the observation-window end. Experiments were confined to an isolated network; pilot data require aggregation or pseudonymization.

Before every run, clocks are synchronized and the environment is restored to a clean snapshot. Attack onset, alert generation, containment, service restoration, and the end of the observation window share common markers so that MTTD, MTTR, and mission loss use the same temporal reference. The supplied description covers four attack profiles with twenty repetitions each; because the individual traces are unavailable, that repetition count cannot be independently audited or used to reconstruct dispersion.

2.6. Catalogue of attacks and failures. A scenario is admissible only when ground truth specifies time, asset, event class, injection profile, expected indicators, permitted responses, and a stop condition. The minimum catalogue in Table 3 covers data, time, gateways, networks, quorum, keys, policies, storage, and physical plausibility.

Each scenario is parameterized by intensity, duration, affected fraction, topology position, and legitimate workload mix. A1 and A2 deliberately separate post-signing tampering from validly signed falsehood; A5 separates hostile flooding from a flash crowd; A6-A10 exercise recovery state machines; and A11-A12 test correlated evidence and cross-sector cascades. The same scenario definition is reused across configurations so that differences are not caused by unequal fault exposure.

ISSN 2786-5460 (Print), ISSN 2786-5479 (Online)

Table 3

Minimum catalogue of attacks and failures

Code	Scenario	Validation objective
A1	Post-signing tampering	Cryptographic integrity
A2	Validly signed false readings	Physical plausibility and sensor-oracle consistency
A3	Replay, timestamp manipulation, clock drift	Freshness and sequence integrity
A4	Compromised edge gateway	Provenance, attestation, redundant routes
A5	DDoS or flash crowd	Admission control with bounded false positives
A6	Partition or quorum loss	Graceful degradation and reconciliation
A7	Byzantine validator or collusion	Quorum trust boundary
A8	Key compromise or revocation	Identity-state recovery
A9	Policy or smart-contract defect	Verification, approval, and rollback
A10	Off-chain storage failure	Restore and integrity control
A11	Coordinated sensor-cluster compromise	Independence-aware trust
A12	Four-sector cascade	Dependency-aware coordinated recovery

Source: compiled by the author.

2.7. Configurations B0-B5 and ablations. Resource-fair comparisons use identical nodes, bandwidth, service demand, attack traces, seeds, time budgets, and recovery rules (Table 4). B5-E removes provenance/freshness/trust, B5-D removes cascade forecasting and what-if analysis, and B5-P removes automated policy selection while preserving observations.

Unused baseline capacity is recorded as a cost rather than reassigned to B5. Configuration order is randomized, and all variants use the same action horizon and stop conditions. Ablations may disable uncertainty handling, fallback, or the safety shield.

Component effects are estimated from paired contrasts and interaction terms, not by dividing the overall B1-B5 improvement into arbitrary percentages.

Table 4

Resource-comparable configurations B0-B5

Code	Configuration	Enabled mechanisms
B0	Standard stack	Local logs, access control, standard failover
B1	Strong detection and manual response	IDS/SIEM, operator decision, manual recovery
B2	Zero Trust and orchestration	Identity, policy rules, playbooks; no cascade forecast
B3	Static SDN/NFV resilience	Rule-based reroute, failover, static chains
B4	Centralized optimizer	Full telemetry and optimization; no provenance-aware trust
B5	Full methodology	Evidence pipeline, digital twin, constrained controller, rollback

Source: compiled by the author.

2.8. Metrics and mathematical model. The primary outcome is mission capability rather than an individual node. For normalized service capability $q_k(t)$ and criticality weight ω_k, the regional mission quality is

$$Q_m(t) = \sum_{k=1}^{K} \omega_k q_k(t), \quad \sum_{k=1}^{K} \omega_k = 1. \tag{1}$$

The area of mission loss and normalized resilience index are

$$L_m = \int_{t_0}^{t_1} \left[1 - Q_m(t)\right] dt, \tag{2}$$

$$\mathrm{NRI} = 1 - \frac{L_m}{t_1 - t_0} = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} Q_m(t)\, dt. \tag{3}$$

Equation (3) measures the preserved capability area across services; the integral is evaluated by the trapezoidal rule over a common comparison window [5].

For source i, dynamic trust combines passed integrity checks $I_i(t)$, provenance consistency $P_i(t)$, and behavioural consistency $B_i(t)$:

$$T_i(t) = \lambda T_i(t-1) + (1-\lambda)\left(w_I I_i(t) + w_P P_i(t) + w_B B_i(t)\right), \tag{4}$$

The weights sum to one and 0 <= lambda <= 1. A valid signature proves origin and post-signing immutability, not the physical truth of a sensor; A2 and A11 therefore test process consistency and evidence-source independence.

Selective replica auditing complements source trust by sampling stored blocks according to a prespecified corruption model and target detection probability. The audit budget is recalculated when replica size, expected damaged share, or target confidence changes and is reported as a cost outcome for each tested profile. Audit evidence detects post-acquisition alteration but does not replace plausibility checks on measurements that were already false when signed.

Local risk propagates through dependency matrix W. The stationary estimate is valid when the spectral radius rho(βW^T) is below one:

$$r = \ell + \beta W^\top r = \left(I - \beta W^\top\right)^{-1} \ell. \tag{5}$$

The controller selects an admissible action by minimizing mission loss, resource cost, disruption, and uncertainty:

$$a_{\text{opt}} = \arg\min_{a \in \mathcal{A}_{adm}} \left[L_m(a) + \eta_c C(a) + \eta_d D(a) + \eta_u U(a) \right]. \tag{6}$$

Any action violating a hard safety predicate is excluded regardless of expected benefit.

Detection and recovery time use common event markers:

$$\text{MTTD} = \frac{1}{n}\sum_{j=1}^{n}\left(t_j^{det} - t_j^{onset}\right), \quad \text{MTTR} = \frac{1}{n}\sum_{j=1}^{n}\left(t_j^{rec} - t_j^{onset}\right). \tag{7}$$

The scaling boundary is the first device count at which decision latency, resource use, safety, or mission-loss limits are violated:

$$N\text{lim} = \min\{N : p99(N) > D_{act} \vee U_{CPU}(N) > Umax \vee L_m(N) > L_{max}\}. \tag{8}$$

The metric hierarchy prevents a secondary improvement from masking mission or safety degradation. Mission loss and NRI are primary; detection quality includes MTTD, false-positive and false-negative rates, calibration, and interval coverage; recovery includes MTTR, recovery-point and recovery-time objectives, slope, and rollback. Network percentiles, cascade size and depth, unsafe actions, processor, memory, traffic, storage, energy, and operator workload are reported as service, mechanism, safety, cost, and V4 outcomes rather than collapsed into one score.

2.9. Statistical analysis plan. The unit of comparison is a paired run: configurations and ablations replay the same trace, seed, service demand, and fault schedule. Order is randomized, and every run starts from a fixed clean snapshot. Sample size is planned from the minimum detectable paired effect and pilot variance:

The design reports differences, effect sizes, and 95% confidence intervals rather than relying on a fixed repetition count.

$$n \geq \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\delta_{min} / \sigma_d} \right)^2. \tag{9}$$

Latency and recovery are summarized by p50, p95, p99 and bootstrap confidence intervals. Paired t, Wilcoxon, or permutation tests are selected after diagnostics; factorial series use a mixed-effects model with configuration, scenario, scale, and sector as fixed effects and trace/seed and site as random effects.

Holm correction is applied within metric families; practical thresholds, negative results, and rejected hypotheses are retained.

Analysis families are defined before results are inspected. Distributional assumptions, missingness, carry-over, time synchronization, and the stability of thresholds under legitimate regime change are checked explicitly. A statistically detectable difference is not treated as useful unless it crosses a preregistered practical threshold, while a safety-predicate violation rejects a configuration even when its average mission metric improves. Exploratory metrics remain labelled as exploratory and are evaluated on a separate validation set before confirmatory use.

2.10. Scaling, pilot deployment, and external replication. Scale tests vary the device count, telemetry rate, graph density, candidate actions, and compromised-source share, reporting p95/p99 latency and compute, memory, storage, traffic, and energy separately for each component. V4 uses approved deidentified or aggregate partner data to test integration, data quality, operator performance, overrides, and rollback readiness; it establishes feasibility, not general causality.

V5 is executed by an independent team with a frozen release, configuration profile, and analysis plan. Transferability is quantified only when the external and laboratory effects have the same direction:

$$\mathrm{TR} = \frac{\Delta^{ext}}{\Delta^{lab}}, \quad \text{provided that } \operatorname{sign}\left(\Delta^{ext}\right) = \operatorname{sign}\left(\Delta^{lab}\right). \tag{10}$$

The ratio is not interpreted when confidence intervals admit opposite directions or the external profile falls outside preregistered bounds.

Production attack code is excluded from V4; faults are introduced only in an approved window or through synthetic markers. V5 receives frozen public and controlled artifacts but not unrestricted operational secrets. A successful external repeat establishes transfer only within the reproduced topology, workload, sector physics, and organizational approval profile; it does not justify universality beyond those bounds.

2.11. Reproducibility, data, and security review. Every run produces a manifest linking the protocol, code or container digest, firmware, hardware, topology, schema, seeds, clock skew, attack parameters, exclusions, raw-log hashes, analysis scripts, outputs, and deviations. Results are regenerated from the frozen manifest rather than copied manually.

Public artifacts include protocols, synthetic traces, schemas, analysis code, aggregate results, and documentation; deidentified detailed traces are controlled, while keys, exact addresses, sensitive rules, and high-resolution incident data remain restricted. Release requires licence, software-bill-of-materials, secret-scanning, vulnerability, and disclosure review.

Failed runs and deviations remain in the manifest. Exclusion is allowed only for predefined synchronization, telemetry, injector, snapshot, or safety-stop failures. The independent package contains deployment instructions, expected intermediate outputs, metric definitions, a data dictionary, and a replication-note template.

The minimum replication package contains a README, threat model, infrastructure-as-code profile, container digests, synthetic seed set, expected intermediate outputs, metric definitions, data dictionary, statistical notebook or script, licence, software bill of materials, disclosure policy, and replication-note template. The package is versioned, archived, and published even when the direction of effect is not preserved, because negative replication is evidence about applicability boundaries.

3. Results and discussion

3.1. Evidence status. The pre-analysis evidence map in Tables 1 and 2 defines a complete V0-V5 programme, but only the V3 topology and aggregate B1-versus-B5 means are available. V0 lacks a model-checker report; V1-V2 remain engineering protocols; and run-level logs, ablations, V4, and V5 are unavailable. The admissible evidence is therefore descriptive and profile-bounded.

3.2. Aggregate hardware-in-the-loop results at V3. The reported series covered a compromised edge device, falsified telemetry, a SYN/ACK flood, and malicious workload migration, with twenty repetitions per scenario according to the supplied description. B1 combined intrusion detection with manual recovery; B5 used the integrated pipeline, twin, and controller. Because run-level distributions are unavailable, Table 5 reports descriptive means and arithmetic changes without confidence intervals or p-values.

Table 5

Reported aggregate V3 results and arithmetic changes

Metric	B1	B5	Change	Interpretation
Mean detection time	43,1 s	8,5 s	-80,3%	Integrated verified evidence and automation
Mean recovery time	399 s	122 s	-69,4%	Faster containment and recovery
Integrity-verification accuracy	Not reported	98,7%	-	B5 only
Additional overhead	Not reported	<6%	-	Small-testbed bound
NRI under DDoS	0,71	0,93	+31,0%	Resilience deficit reduced by 75,9%

Source: compiled by the author.

B5 reduced mean detection time by 80,3% and recovery time by 69,4%. NRI rose by 0,22 (31,0%), while the resilience deficit fell by 75,9%. These values support the integrated proof of concept, but they cannot identify how much improvement came from detection, trust, cascade forecasting, operator removal, enforcement, or post-recovery verification.

The distinction between the 31,0% NRI increase and the 75,9% reduction in resilience deficit is important. NRI describes preserved capability over the whole disruption-and-recovery window, whereas the deficit

isolates the unrecovered fraction relative to perfect service. Both summaries point in the same direction, but neither provides uncertainty or generalization beyond the reported DDoS profile.

3.3. Causal contribution of components. The hypotheses in Table 1 and the resource-comparable configurations in Table 4 define a component contribution as the paired difference between B5 and its matching ablation under the same budget. Because B5-E, B5-D, and B5-P observations are absent, no numerical causal shares or synergies are claimed; the reported improvement belongs to the integrated configuration.

Ablation interpretation must also consider mechanism outcomes: removing the evidence pipeline may alter false positives and calibration, removing the twin may change forecast coverage and cascade size, and removing the controller may shift recovery time, rollback, and unsafe-action counts. A lower average mission loss cannot compensate for a safety violation or a materially higher resource budget.

3.4. Scaling limits and transferability. The V3 profile in Table 2 demonstrates operation with 12 streams, four edge nodes, and three cloud virtual machines, not a limit for 10^2-10^4 devices. A scale claim requires p95/p99 curves, resource profiles, and sensitivity to telemetry rate, graph density, and candidate actions. Transferability is technological, sectoral, and organizational; only an independent V5 repeat and the direction-of-effect condition in equation (10) can test whether the effect survives a new team and site.

3.5. Practical maturity model I0-I4. The maturity model converts validation evidence into implementation gates (Table 6). Technology ownership alone does not justify a level transition.

Table 6

Maturity gates I0-I4

Level	State	Required capabilities	Transition evidence
I0	Baseline	Inventory, synchronized clocks, evidence schema	Complete coverage and stable telemetry
I1	Observable	Provenance, freshness, trust, mission dashboard, audit trail	Calibration and reproducible evidence chain
I2	Predictive	Digital twin, cascade forecast, what-if analysis, uncertainty	Out-of-sample forecast and approved scenarios
I3	Controlled	Reversible actions, safety shield, rollback, orchestration	Formal safety, zero unsafe tests, bounded latency
I4	Adaptive	Federated coordination, local fallback, recalibration	Pilot record, service limits, scale and replication plan

Source: compiled by the author.

An internal weighted score may support gap analysis, but it cannot compensate for missing emergency stop, data minimization, formal safety, or verified rollback.

Transitions are cumulative. I1 requires calibrated cross-sector observability before I2 forecasting can be trusted; I3 requires formal safety, bounded latency, and verified rollback before automated action; and I4 adds pilot evidence, local fallback, service-level monitoring, periodic recalibration, scale tests, and a replication plan. Missing a hard gate lowers the admissible maturity level even when other capability scores are high.

3.6. Discussion, limitations, and development directions. The V3 aggregates in Table 5, interpreted through equations (1)-(3), provide a descriptive comparison of the integrated B5 configuration with B1. Secure placement [1] and collaborative auditing [2] supply relevant component-level foundations, while implementation architectures [3] and twin-assisted intrusion detection [4] motivate the testbed design. Availability-based normalization [5] informs the mission assessment but cannot establish integrity or safe control. Figure 1 links claims to artifacts, and Figure 2 relates the evidence pipeline, twin, and controller. The contribution is their explicit placement within a common validation programme; simultaneous changes in several mechanisms and absent run-level distributions prevent causal attribution or uncertainty estimation from the reported aggregates.

Cryptographic integrity and physical truth must remain distinct. The storage-audit scope of EDI-C [2] should be distinguished from verifying the physical plausibility of data before acquisition: a valid signature can authenticate a false sensor value. The proposed programme therefore combines process models, independent sources, spatial-temporal consistency, and explicit uncertainty. Digital-twin architecture [3] and detection integration [4] provide a basis for implementation, while the proposed extension to cross-sector cascade estimation and counterfactual intervention testing requires its own calibration and validation.

The I0-I4 model likewise separates technology acquisition from evidence of readiness. Negative results define applicability limits: if a simpler B3 or B4 configuration produces lower p99 latency at scale, or V5 preserves direction with a smaller effect, the discrepancy should be reported and explained rather than suppressed.

Operationally, the strongest result is not the isolated percentage change but the disciplined transition from observation to reversible action. The evidence pipeline limits which measurements may influence the twin; the twin exposes likely cross-sector consequences; and the controller enforces resource, authorization, uncertainty, and rollback constraints. This sequence makes automated response auditable and provides concrete failure points for ablation, incident review, and later replication.

Threats to validity and limitations. Internal and statistical validity are limited by an aggregate B1-B5 comparison without randomized run logs, dispersions, confidence intervals, p-values, or effect sizes. Construct validity is limited because NRI measures availability and cannot compensate for confidentiality, integrity, or safety failures. Model validity depends on complete, calibrated dependencies and freshness of the twin state.

External validity is limited by the small synchronized testbed and the absence of V4-V5 evidence for heterogeneous infrastructure, organizational delay, seasonality, and legal constraints. Security disclosure also limits release of exact topologies, keys, rules, and vulnerabilities; replication must use tiered access without weakening the scientific audit trail.

These limitations directly bound the claims: H1 is only a preliminary comparison with B1 rather than the strongest B0-B4 baseline; H2-H4 remain unconfirmed without component ablations; H5 requires scale curves; and H6 requires completed V4 and V5 evidence. The percentages in Table 5 are therefore arithmetic descriptions of supplied aggregates, not estimates of a general population effect.

Twin validity requires incident-based calibration of dependency weights, uncertainty bounds, and a freshness gate that blocks what-if decisions on stale state. Missing dependencies may conceal a cascade, whereas inflated weights may trigger unnecessary intervention. Replication must therefore disclose the modeled topology and calibration procedure at an appropriate access tier, while withholding operational details whose publication would create security risk. Further development should therefore publish auditable run manifests, complete the paired ablation and scale series, conduct the V4 operator pilot, and perform independent V5 replication before broader deployment claims are made.

4. Conclusions

1. A single claim-to-evidence architecture was established for levels V0-V5. Its distinguishing feature is an explicit boundary between formal safety, laboratory causality, operational feasibility, and external transfer; this prevents evidence from one level being generalized to another without the required artifacts.

2. For the available V3 aggregate comparison, B5 reduced mean detection time from 43,1 to 8,5 s (80,3%) and recovery time from 399 to 122 s (69,4%), achieved 98,7% integrity-verification accuracy with less than 6% overhead, and increased NRI from 0,71 to 0,93. The joint evidence pipeline, dependency model, and constrained controller therefore provide a descriptive integrated proof of concept.

3. Component-level causal contributions cannot be assigned from the B1-B5 aggregate contrast. The proposed B5-E, B5-D, and B5-P paired ablations are necessary to separate detection, cascade-forecasting, and policy-control effects under the same traces and resource budget; until they are completed, no numerical causal shares or synergies are justified.

4. The demonstrated operating profile is limited to 12 telemetry streams, four edge nodes, and three cloud virtual machines. Claims for 10^2-10^4 devices require p95/p99 latency and resource curves, while transferability requires a V4 pilot and an independent V5 repeat with the direction of effect preserved inside preregistered bounds.

5. The I0-I4 maturity model converts validation evidence into cumulative implementation gates from baseline inventory to adaptive cross-domain coordination. Its practical advantage is that a high weighted score cannot compensate for a missing hard gate such as formal safety, emergency stop, bounded latency, or verified rollback.

6. The reported improvements are profile-bounded and descriptive because run-level distributions, confidence intervals, completed ablations, and external observations are unavailable. Further development should publish auditable run manifests, complete the ablation and scale series, conduct the V4 operator pilot, and perform independent V5 replication before broader regional deployment claims are made.

Conflict of Interest. The author declares that she has no financial, personal, authorship-related, or other conflict of interest that could have influenced the research or the results presented in this article.

Funding. The research was conducted without financial support.

Data Availability. The manuscript contains aggregate data presented in Table 5. No separate related dataset is available because run-level observations, configurations, and manifests were not supplied; this limitation is stated explicitly in the manuscript.

Use of Artificial Intelligence. ChatGPT (OpenAI, GPT-5.6 Pro) was used in the abstract, Materials and Methods, and Results and Discussion sections to support source-text structuring, mathematical exposition, and diagram generation. The author checked the generated material against the source content and aggregate values, reviewed the equations and figures, and approved the final text. The tool did not generate empirical data or determine the study conclusions; the author formulated and verified the conclusions and assumes full responsibility for the manuscript. During revision, ChatGPT also assisted with bibliographic cross-checking, drafting critical comparisons of the cited literature, and language editing.

Author Contributions. Yuliia Ye. Khokhlachova: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review and editing, Visualization, Supervision, Project administration. The author approved the final manuscript.

References:

1. Bocci, A., Forti, S., Ferrari, G.-L., & Brogi, A. (2023). Declarative Secure Placement of FaaS Orchestrations in the Cloud-Edge Continuum. Electronics, 12, 1332. https://doi.org/10.3390/electronics12061332
2. Yang, F., Sun, Y., Gao, Q., & Chen, X. (2024). EDI-C: Reputation-Model-Based Collaborative Audit Scheme for Edge Data Integrity. Electronics, 13, 75. https://doi.org/10.3390/electronics13010075
3. Fett, M., Kraft, M., Wilking, F., Goetz, S., Wartzack, S., & Kirchner, E. (2024). Medium-Level Architectures for Digital Twins: Bridging Conceptual Reference Architectures to Practical Implementation in Cloud, Edge and Cloud–Edge Deployments. Electronics, 13, 1373. https://doi.org/10.3390/electronics13071373
4. El-Hajj, M. (2024). Leveraging Digital Twins and Intrusion Detection Systems for Enhanced Security in IoT-Based Smart City Infrastructures. Electronics, 13, 3941. https://doi.org/10.3390/electronics13193941
5. Cho, H., Sung, J.-H., Kang, H.-J., Jang, J., & Shin, D. (2025). Quantifying Cyber Resilience: A Framework Based on Availability Metrics and AUC-Based Normalization. Electronics, 14, 2465. https://doi.org/10.3390/electronics14122465
6. Tao, F., Zhang, H., Liu, A., & Nee, A.Y.C. (2019). Digital Twin in Industry: State-of-the-Art. IEEE Trans. Ind. Inform., 15, 2405–2415. https://doi.org/10.1109/TII.2018.2873186
7. Alcaraz, C., & Lopez, J. (2022). Digital Twin: A Comprehensive Survey of Security Threats. IEEE Commun. Surv. Tutor., 24, 1475–1503. https://doi.org/10.1109/COMST.2022.3171465
8. Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture. NIST Special Publication 800-207. National Institute of Standards and Technology: Gaithersburg, MD, USA. https://doi.org/10.6028/NIST.SP.800-207
9. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge Computing: Vision and Challenges. IEEE Internet Things J., 3, 637–646. https://doi.org/10.1109/JIOT.2016.2579198
10. Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital Twin in Manufacturing: A Categorical Literature Review and Classification. IFAC-PapersOnLine, 51, 1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474
11. Sengupta, J., Ruj, S., & Das Bit, S. (2020). A Comprehensive Survey on Attacks, Security Issues and Blockchain Solutions for IoT and IIoT. J. Netw. Comput. Appl., 149, 102481. https://doi.org/10.1016/j.jnca.2019.102481
12. Mozzaquatro, B.A., Agostinho, C., Goncalves, D., Martins, J., & Jardim-Goncalves, R. (2018). An Ontology-Based Cybersecurity Framework for the Internet of Things. Sensors, 18, 3053. https://doi.org/10.3390/s18093053
13. Dietz, M., & Pernul, G. (2020). Unleashing the Digital Twin's Potential for ICS Security. IEEE Secur. Priv., 18, 20–27. https://doi.org/10.1109/MSEC.2019.2961650
14. Aghazadeh Ardebili, A., Lezzi, M., & Pourmadadkar, M. (2024). Risk Assessment for Cyber Resilience of Critical Infrastructures: Methods, Governance, and Standards. Appl. Sci., 14, 11807. https://doi.org/10.3390/app142411807
15. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. ACM Comput. Surv., 41, 15. https://doi.org/10.1145/1541880.1541882
16. Xu, H., Yu, W., Griffith, D., & Golmie, N. (2018). A Survey on Industrial Internet of Things: A Cyber-Physical Systems Perspective. IEEE Access, 6, 78238–78259. https://doi.org/10.1109/ACCESS.2018.2884906
17. Ranaweera, P., Jurcut, A. D., & Liyanage, M. (2021). Survey on Multi-Access Edge Computing Security and Privacy. IEEE Commun. Surv. Tutor., 23, 1078–1124. https://doi.org/10.1109/COMST.2021.3062546
18. de Neira, A. B., Kantarci, B., & Nogueira, M. (2023). Distributed Denial of Service Attack Prediction: Challenges, Open Issues and Opportunities. Comput. Netw., 222, 109553. https://doi.org/10.1016/j.comnet.2022.109553
19. Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital Twin: Enabling Technologies, Challenges and Open Research. IEEE Access, 8, 108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358
20. Minerva, R., Lee, G.M., & Crespi, N. (2020). Digital Twin in the IoT Context: A Survey on Technical Features, Scenarios, and Architectural Models. Proc. IEEE, 108, 1785–1824. https://doi.org/10.1109/JPROC.2020.2998530
21. Sharma, A., Kosasih, E., Zhang, J., Brintrup, A., & Calinescu, A. (2022). Digital Twins: State of the Art Theory and Practice, Challenges, and Open Research Questions. J. Ind. Inf. Integr., 30, 100383. https://doi.org/10.1016/j.jii.2022.100383
22. Homaei, M., Mogollon-Gutierrez, O., Sancho, J.C., Avila, M., & Caro, A. (2024). A Review of Digital Twins and Their Application in Cybersecurity Based on Artificial Intelligence. Artif. Intell. Rev., 57, 201. https://doi.org/10.1007/s10462-024-10805-3

Відомості про авторів

Англ.	Укр.
Yuliia Khokhlachova Candidate of Technical Sciences, Professor, Professor of the Department of Software Engineering and Cybersecurity, State University of Trade and Economics, Doctoral Student, State University of Information and Communication Technologies yuliiakhokhlachova@gmail.com ORCID: 0000-0002-0787-5112	Хохлачова Юлія Євгеніївна кандидат технічних наук, професор, професор кафедри інженерії програмного забезпечення та кібербезпеки, Державний торговельно-економічний університет, докторант, Державний університет інформаційно-комунікаційних технологій yuliiakhokhlachova@gmail.com ORCID: 0000-0002-0787-5112

Дата надходження статті: 21.05.2026
Дата надходження виправленої версії статті: 03.07.2026
Дата прийняття статті: 08.07.2026
Дата публікації статті: 26.10.2026