Detection of Hidden Data Manipulations Using Statistical Steganalysis for Decision Support*

Artem V. Sokolov*,†, Olena Ye. Chepurna†, Stanislav A. Horbachenko† and Olena G. Trofymenko†

National University "Odesa Law Academy", Fontanska Doroha Str. 23, 65009 Odesa, Ukraine

Abstract

Decision support systems (DSS) process numeric data from sensors, transactions, audit logs, and key-performance indicators. The quality of decisions produced by such systems depends on the integrity of these inputs. Hidden data manipulations (statistically plausible alterations that remain within historical 3σ bounds) evade conventional integrity mechanisms (checksums, digital signatures) and bias automated outputs. This paper develops a two-layer detection framework: the first layer monitors distributional drift via CUSUM, EWMA, Kolmogorov–Smirnov, and Benford first-digit tests; the second layer adapts steganalytic features (residual histogram, transition probability matrix, SPA-based asymmetry) to quantised numeric time-series. Outputs of both layers are fused via logistic regression into a Manipulation Probability Index (MPI) with transparent contribution coefficients. Validation on the SWaT public benchmark uses 36 effective attack scenarios, 12 of which satisfy a stealthy 3σ-constrained criterion defined herein, evaluated via 5-fold cross-validation with Wilcoxon significance testing. The proposed method exceeds LSTM-AE, 1D-CNN, and CNN-BiLSTM-Attention baselines on the stealthy subset and operates without GPU. The framework is domain-agnostic and applies to any quantised numeric DSS input.

Keywords

hidden data manipulation, decision support systems, statistical steganalysis, numeric time-series analysis, data integrity, anomaly detection, machine learning algorithms, algorithmic risk assessment, management, governance of digital systems

1. Introduction

Decision support systems (DSS) are used for operational, financial, and analytical decisions in enterprises, government services, and critical infrastructure. They process numeric data (sensor measurements, transaction records, key-performance indicators, and audit logs) drawn from heterogeneous sources and feed automated analytics or human operators [1]. The quality of the resulting decisions depends on the integrity of these inputs. Where input data are silently altered by malicious actors, insider threats, or integration faults, downstream decisions become unreliable, with consequences ranging from financial losses to disruption of services.

Conventional integrity mechanisms (checksums, digital signatures) protect against unauthorised modification at the syntactic level but are blind to semantic-level alterations: fine-grained, statistically plausible perturbations that respect the data's expected ranges and distributions while systematically biasing outputs. We refer to such alterations as hidden data manipulations. Detecting them requires methods that operate not on the integrity envelope of a record but on the statistical and structural properties of the data stream itself. This paper proposes such a method.

RS-2026: II International Scientific and Practical Workshop "Resilient Systems: Secure Digital Technologies and Critical Infrastructure", June 30, 2026, Drohobych, Ukraine

* Corresponding author

† These authors contributed equally.

sokolov@onua.edu.ua (A. V. Sokolov); chepurna@onua.edu.ua (O. Ye. Chepurna); stasgorbachenko@onua.edu.ua (S. A. Horbachenko); trofymenko@onua.edu.ua (O. G. Trofymenko)

0000-0003-0283-7229 (A. V. Sokolov); 0000-0002-1432-0799 (O. Ye. Chepurna); 0000-0001-8442-9581 (S. A. Horbachenko); 0000-0001-7626-0886 (O. G. Trofymenko)

2. Relevance of the Research

The detection of hidden data manipulations gains relevance from three developments. First, false data injection (FDI) attacks against industrial control systems are documented in the literature [2, 3], and adversarial perturbation generators capable of mimicking legitimate sensor distributions while shifting control outputs have been demonstrated [4]. Second, in financial and audit analytics, statistical anomalies in numeric records have been studied via Benford-style approaches [5, 6]; information-geometric methods for analysing complex data structures further widen the methodological base [7]. Third, regulatory and managerial developments, including the EU NIS2 Directive [17], the Ukrainian Law on Critical Infrastructure (No. 1882-IX), and the integration of cybersecurity considerations into management practice [8], require continuous monitoring of the integrity of data feeding decision support systems.

Existing detection methods address the problem only partially. Statistical process monitoring (CUSUM, EWMA) is interpretable and lightweight but limited to distributional signals. Deep-learning anomaly detectors (LSTM-AE, 1D-CNN, hybrid CNN-BiLSTM-Attention [9, 10, 11]) achieve high accuracy on standard benchmarks but require large labelled datasets, lack interpretability, and are vulnerable to adversarial input. Steganalytic techniques originally developed for digital media [12] have been adapted to ICS network traffic [13, 14], and to detect sensor false data injection attacks in ICS environments [19], but have not previously been applied to numeric DSS data streams.

The core difficulty lies in the nature of the threat itself. A stealthy 3σ-constrained perturbation is, by construction, designed to pass distributional tests: its statistical moments, Benford first-digit distribution, and seasonal profile all remain within the historical bounds of the source. Deep learning detectors face a related but distinct problem. A model trained on a specific perturbation distribution generalises poorly when the attacker adjusts the attack to remain within the training-distribution envelope, since the adversarial input is then statistically indistinguishable from normal operation by any feature the model was trained to use. This shared blind spot of both statistical and deep learning approaches motivates the steganalytic perspective explored in the present work.

3. Statement of the Problem

Given a stream of quantised numeric values produced by a data source (sensor, transaction log, KPI generator) feeding a DSS, the goal is to detect, in real time, whether the stream has been subjected to hidden data manipulation. The detector is required to satisfy three conditions: (i) effectiveness on stealthy manipulations, i.e. alterations that remain within the historical 3σ envelope of the source and do not trigger conventional threshold-based alarms; (ii) interpretability: the output must be explainable to a DSS operator and auditable for regulatory purposes; (iii) lightweight integration: deployment must not require GPU provisioning, large training datasets, or major redesign of existing DSS pipelines.

4. Research Results

The proposed framework consists of two complementary layers and a fusion mechanism. The statistical layer (S) applies CUSUM [20] and EWMA [21] control charts with exponentially forgetting baselines, supplemented by Kolmogorov–Smirnov and chi-square tests and Benford's Law first-digit analysis [5, 6]; individual scores are normalised and aggregated into $S \in [0, 1]$. The steganalytic layer (G) operates on residuals $\varepsilon = x \bmod \Delta q$ where Δq is the sensor-specific quantisation step. Three feature families are extracted: residual histogram, first-order transition probability matrix (TPM), and SPA-based asymmetry index. Outputs are fused via logistic regression: $MPI = \sigma(\beta_0 + \beta_1 S + \beta_2 G)$, with coefficients estimated as $\beta_0 = -0.41 \pm 0.07$, $\beta_1 = 2.73 \pm 0.18$, $\beta_2 = 1.94 \pm 0.22$ (mean ± SD across folds). The relative magnitudes of β_1 and β_2 admit a direct managerial reading for DSS operators: each unit increase in the statistical anomaly score S contributes roughly 1.4 times as much to the alert log-odds as the same increase in the steganalytic score G; the negative intercept β_0 encodes the

prior that, on a normal-operation baseline, the framework should not fire. The decision threshold MPI > 0.60 was selected on the calibration split as the value that maximises Youden's J statistic (true-positive rate − false-positive rate) on the receiver-operating-characteristic (ROC) curve; the calibration ROC reaches AUC = 0.93 ± 0.02 across folds and the selected threshold corresponds to TPR = 0.86 at FPR = 0.07. The threshold is not assumed universal: for deployment on a new system it must be re-calibrated from a stream-specific normal-operation period using the same ROC/Youden criterion, and may be shifted deliberately to favour either alert sensitivity or operator trust. Figure 1 illustrates the contrast between normal operation and a stealthy 3σ-FDI attack on a representative flow meter (FIT-101).

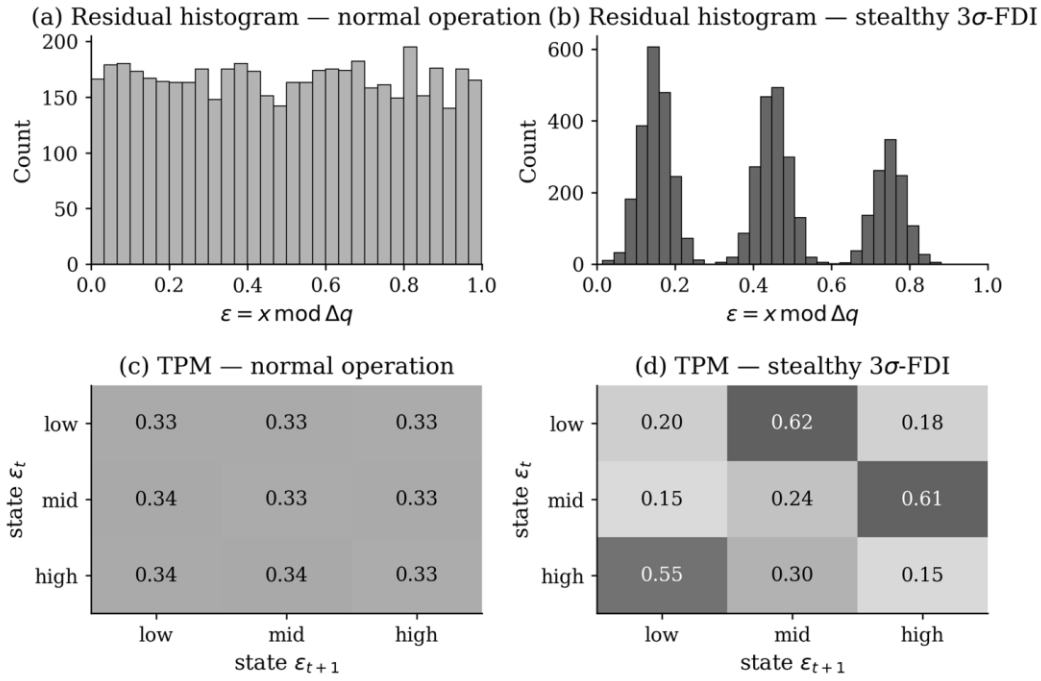

Figure 1: residual histogram (top) and 3×3 transition probability matrix (bottom) under normal operation (left) versus a stealthy 3σ-FDI attack (right) on flow meter FIT-101 (Stage 1, Δq = 0.05 L/min, σ_n = 0.012 L/min).

4.1. Dataset and experimental protocol

Validation is based on data from the SWaT (Secure Water Treatment) benchmark, documented as an ICS security testbed in the survey [16], using the publicly released dataset and analytical reports. Of the 41 attack runs documented in those reports, 36 involve measurable perturbations with unambiguous ground-truth labels; the remaining 5 produce no observable change in the recorded values and are excluded, in line with [10, 11]. Among the 36, 12 satisfy the stealthy 3σ-constrained criterion: maximum absolute deviation below 3σ of the 30-day rolling mean throughout the attack window, attack duration ≥10 min (run IDs 7, 8, 14, 18, 19, 21, 22, 26, 28, 31, 33, 35 in the published numbering). All experiments (implementation of the proposed framework, retraining of the baselines, and benchmarking) were performed on the Cyber Range facilities of the Faculty of Cybersecurity and Information Technologies, National University “Odesa Law Academy”. Five-fold cross-validation is applied at the scenario level; the normal-operation period is split 70/30, stratified by process stage to avoid temporal data leakage. All baselines were re-implemented and trained from scratch under identical conditions on Intel Xeon E5-2680 hardware (single-thread CPU inference, no GPU). Pairwise F1-score differences are tested with the Wilcoxon signed-rank test.

4.2. Detection performance

Table 1 reports the macro-averaged metrics on SWaT for the proposed MPI and four reproduced baselines. F1-score (stealthy) refers to the 12-scenario subset under the stealthy criterion. The p-value column reports the Wilcoxon signed-rank test of MPI versus each baseline on the stealthy subset across folds. This non-parametric test was selected because it does not assume normality of performance distributions and provides a robust assessment of whether the observed differences between methods are statistically significant.

Table 1
Detection performance on SWaT, mean ± SD over 5 folds. Δ = difference; p-value: Wilcoxon vs MPI on stealthy F1-score.

Method	Precision (%)	Recall (%)	F1-score (%)	F1-score stealthy (%)	p-value
CUSUM/EWMA	78.4 ± 2.1	71.2 ± 3.0	74.6 ± 2.4	61.2 ± 3.8	—
Adapted steganalysis (G only)	69.3 ± 2.8	82.1 ± 2.3	75.2 ± 2.5	70.4 ± 3.1	—
LSTM-AE [9]	83.2 ± 1.9	79.6 ± 2.2	81.4 ± 2.0	72.8 ± 2.9	0.038
1D-CNN [10]	86.1 ± 1.7	82.4 ± 2.0	84.2 ± 1.8	75.3 ± 2.6	0.021
CNN-BiLSTM-Att. [11]	91.3 ± 1.2	94.8 ± 0.9	93.0 ± 1.0	78.6 ± 2.1	0.044
Proposed MPI (S + G)	**91.7 ± 1.1**	**88.5 ± 1.4**	**90.1 ± 1.2**	**84.3 ± 1.8**	—

MPI achieves the highest F1-score on the stealthy subset (84.3 ± 1.8 %), with all differences against baselines significant ($p \leq 0.044$). On the overall F1-score, MPI (90.1 %) is below CNN-BiLSTM-Attention (93.0 %): deep-learning models achieve higher recall on overt attacks but generalise poorly to 3σ-constrained FDI that, by construction, remains within the training distribution. The steganalytic layer captures bit-level structural regularities in quantised residuals that are invariant to the attack's statistical plausibility: the complementary signal that DL models miss. Inference latency on the same hardware: MPI 187 ± 12 ms, 1D-CNN 312 ± 18 ms, LSTM-AE 428 ± 24 ms, CNN-BiLSTM-Attention 547 ± 31 ms per 512-sample window. Resident memory: MPI ≈34 MB versus ≈318 MB for CNN-BiLSTM-Attention.

4.3. Ablation

Table 2 isolates the contribution of each component on the stealthy subset. Removing the steganalytic layer causes the largest drop (−23.1 pp); within the steganalytic layer, the TPM provides the largest individual gain (−6.2 pp when removed).

Table 2
Ablation on stealthy F1-score; Δ = difference vs full MPI in percentage points; p-value: Wilcoxon vs full MPI.

Configuration	F1-score stealthy (%)	Δ (pp)	p-value
MPI full (S + G)	**84.3 ± 1.8**	—	—
MPI − steganalytic layer	61.2 ± 3.8	−23.1	<0.001
MPI − statistical layer	70.4 ± 3.1	−13.9	<0.001
MPI − TPM features	78.1 ± 2.2	−6.2	0.003
MPI − SPA asymmetry features	81.6 ± 1.9	−2.7	0.041

4.4. Steganalytic layer coverage

The steganalytic layer is activated for a sensor only when the noise floor σ_n is smaller than half the quantisation step Δq. Table 3 summarises coverage across SWaT sensor groups.

Table 3

Steganalytic layer activation across SWaT sensor groups under the condition $\sigma_n < \Delta q/2$.

Sensor group	Sensors (n)	$\sigma_n < \Delta q/2$	Steganalytic layer
Flow meters (Stages 1–2)	14	11/14	Active
Pressure sensors (Stages 3–4)	12	11/12	Active
Level sensors (Stages 1, 3, 5)	8	7/8	Active
Turbidity / conductivity (Stage 6)	7	4/7	Partial
Binary actuators (all stages)	10	n/a	Inactive
Overall	**51**	**33/51**	**64.7 % active**

4.5. Per-stage performance

Table 4 reports F1-score broken down by SWaT process stage. Performance correlates with steganalytic-layer coverage: stages where the layer is fully active (Stages 1–3) achieve F1-score (stealthy) ≥ 85 %, whereas Stage 6 (turbidity-dominated, with 2 of 10 sensors active) drops to 79.5 %. The relative deviation Δ in Table 4 is computed against the overall stealthy mean (84.3 %).

Table 4
Per-stage detection performance on SWaT; Δ = deviation from the overall stealthy F1-score mean (84.3 %).

SWaT stage	Active sensors (steg.)	F1-score (%)	F1-score stealthy (%)	Δ (pp)
Stage 1 (raw water)	7/8	91.4 ± 1.0	86.2 ± 1.5	+1.9
Stage 2 (pre-treatment)	6/8	90.7 ± 1.3	85.1 ± 1.7	+0.8
Stage 3 (ultra-filtration)	9/10	92.0 ± 1.1	87.4 ± 1.6	+3.1
Stage 4 (dechlorination)	5/6	89.8 ± 1.4	82.9 ± 1.9	−1.4
Stage 5 (RO unit)	4/9	88.5 ± 1.6	81.7 ± 2.2	−2.6
Stage 6 (distribution)	2/10	86.9 ± 1.8	79.5 ± 2.5	−4.8
Overall	**33/51**	**90.1 ± 1.2**	**84.3 ± 1.8**	—

4.6. Limitations and concept drift

Two limitations bound the applicability of the proposed framework. First, in process stages dominated by high-noise analogue sensors (SWaT Stage 6, where 2 of 10 sensors satisfy $\sigma_n < \Delta q/2$), the steganalytic layer deactivates for the majority of channels and the framework reduces to its statistical layer alone, with the corresponding loss of stealthy-attack F1-score (Table 4, −4.8 pp). For such stages, MPI should be regarded as a complementary signal rather than a primary detector.

Second, both layers are vulnerable to legitimate concept drift, i.e. gradual shifts in the underlying process (seasonality, equipment ageing, and operating-mode changes) that may be misclassified as manipulations. CUSUM and EWMA in the statistical layer use an exponentially forgetting baseline with smoothing parameter $\alpha = 1 - \exp(-\ln 2 / \tau)$, where τ is the half-life expressed in samples. We set τ to seven days (the SWaT operational week, equivalent to approximately 6×10^5 samples at the 1 Hz sampling rate). This value was selected by sweeping τ over {1, 3, 7, 14, 30} days on the calibration split: shorter half-lives (1–3 days) increased the false-positive rate above 12 % under benign weekend/weekday cycles, while longer half-lives (14–30 days) suppressed the response to genuine fast attacks. The seven-day setting yielded the best trade-off and matches the natural operational rhythm of the testbed; for deployments with different cycles (financial intra-day, monthly audit), τ should be re-tuned using the same sweep.

Drift adaptation in the proposed framework operates on two timescales. Slow continuous drift is absorbed by the exponentially forgetting baselines as part of normal scoring. Abrupt legitimate transitions (commissioning of a new sensor, planned setpoint change, scheduled maintenance windows) are handled by an explicit operator-issued recalibration command: it resets the CUSUM/EWMA baseline statistics and, optionally, retrains the steganalytic one-class SVM on a fresh

calibration window. Fully automatic adaptation, combining a Bayesian online change-point detector to flag suspected legitimate transitions, a confidence-weighted hold-off period during which MPI alerts are downgraded to advisory status, and automatic recalibration if the change is confirmed by an external indicator, is the subject of further work, together with a quantitative study of the trade-off between drift tolerance and detection responsiveness.

Organisations subject to continuous monitoring obligations under the NIS2 Directive [17] or NIST SP 800-82 guidance for industrial control systems [18] may additionally require that recalibration events be logged with timestamps, operator identifiers, and before/after baseline statistics for audit purposes. The explicit operator-command model of the proposed framework accommodates this requirement without modification: each recalibration event is a discrete, attributable action that produces a versioned snapshot of the baseline parameters, suitable for inclusion in a system-level security event log.

A third consideration, orthogonal to the two limitations above, concerns the interaction between the steganalytic layer and quantisation precision. The residual-based features ($\varepsilon = x \bmod \Delta q$) depend critically on the accuracy of the known quantisation step Δq. In practice, sensors with adaptive or non-uniform quantisation, common in modern smart meters and industrial IoT devices, may report an effective Δq that drifts over time due to firmware updates or sensor-specific calibration cycles. If Δq is misspecified by more than 10 %, the residual histogram loses its structural regularity even under normal operation, producing inflated steganalytic scores and elevated false-positive rates. Practitioners deploying MPI on such sensors are advised to estimate Δq empirically from a calibration window using the mode of the inter-sample absolute differences, and to re-estimate it after any sensor firmware update. When a reliable Δq estimate cannot be obtained, the steganalytic layer should be disabled for that channel and the sensor documented as statistical-layer-only in the system risk register, consistent with the guidance in Section 5.1(ii).

4.7. Computational complexity

The computational budget of the proposed framework is dominated by the Kolmogorov–Smirnov test, which requires $O(n \log n)$ sorting of the window sample; all other components run in $O(n)$ or $O(1)$ time. Table 5 breaks down the per-component contributions to the 187 ± 12 ms total wall-clock latency measured on the Intel Xeon E5-2680 testbed (512-sample window, single-thread). Resident memory is dominated by the steganalytic one-class SVM model (≈28 MB of 34 MB total); the statistical layer requires fewer than 2 kB of state per sensor channel. The complexity profile confirms that MPI is deployable on embedded industrial hardware without dedicated accelerators, a property that is important for critical-infrastructure contexts governed by NIS2 [17] and NIST SP 800-82 [18], where third-party cloud processing may be precluded by data-sovereignty constraints.

Table 5
Per-component computational profile of the MPI framework ($n = 512$ samples per window, $k \approx 20$ quantisation bins).

Component	Time complexity	Space complexity	Latency (ms)
CUSUM / EWMA	$O(n)$	$O(1)$ per channel	12 ± 2
KS test	$O(n \log n)$	$O(n)$	38 ± 4
Benford analysis	$O(n)$	$O(10)$	8 ± 1
Residual histogram	$O(n)$	$O(k)$	11 ± 1
TPM construction	$O(n)$	$O(k^2)$	71 ± 6
SPA asymmetry	$O(n/k)$	$O(k)$	47 ± 5
LR fusion	$O(1)$	$O(1)$	1 ± 0.2
MPI (S + G)	**$O(n \log n)$**	**$O(n + k^2)$**	**187 ± 12**

The latency profile in Table 5 has a direct implication for alert responsiveness. At a 1 Hz sensor sampling rate, the 187 ms per-window latency of the full MPI pipeline is well within the 1 000 ms inter-sample interval, leaving 813 ms of slack for downstream alert routing, logging, and operator

notification. At higher sampling rates (10 Hz or 100 Hz, common in power-grid and vibration-monitoring ICS), the pipeline must be parallelised across sensor channels; the O(1) memory footprint of the CUSUM/EWMA state and the independence of per-channel steganalytic computation make this straightforward. In contrast, the CNN-BiLSTM-Attention baseline requires 547 ms per window and 318 MB of resident memory: at 10 Hz with 51 channels, its aggregate memory requirement would exceed 16 GB, ruling out deployment on standard industrial edge hardware. The MPI framework, at 34 MB total, fits comfortably within the memory budget of entry-level ARM Cortex-A53 class devices widely used in ICS edge nodes.

5. Conclusions

A two-layer detection framework for hidden data manipulations in DSS data streams was developed and evaluated. The framework adapts LSB-class steganalytic features to quantised numeric time-series and combines them with statistical process monitoring through an interpretable logistic-regression fusion. On the SWaT benchmark (36 effective scenarios, 12 stealthy, 5-fold cross-validation, all baselines reproduced under identical conditions), the proposed Manipulation Probability Index reaches stealthy F1-score = 84.3 % ($p \leq 0.044$ versus all baselines, Wilcoxon signed-rank), at 187 ms latency without GPU. The method is domain-agnostic and applies to any quantised numeric DSS input (financial transactions, audit logs, KPI metrics), subject to availability of a calibration period and a known quantisation step. Further work will address (i) automatic concept-drift handling to distinguish legitimate process changes from manipulations and to trigger baseline recalibration without manual intervention; (ii) self-tuning of the MPI decision threshold across deployments; (iii) cross-domain validation on financial transaction logs and audit data streams; (iv) adversarial robustness against residual-targeting perturbations; and (v) a Bayesian fusion variant for uncertainty-aware MPI.

5.1. Practical deployment recommendations

Based on the experimental results and the complexity analysis (Section 4.7), the following recommendations apply to DSS operators deploying the MPI framework. (i) Calibration period: a minimum of one full operational cycle (at least seven days for process-industry systems, or one intra-day cycle for financial analytics) is required to establish reliable CUSUM/EWMA baselines and train the steganalytic one-class SVM; shorter windows increase false-positive rates above acceptable levels (Section 4.6). (ii) Sensor pre-assessment: before deployment, verify the noise-floor condition $\sigma_n < \Delta q/2$ for each input channel using at least one week of historical data; channels that fail the criterion should be monitored by the statistical layer only, with the corresponding reduction in stealthy-attack detection power documented in the system risk register. (iii) Threshold calibration: apply the ROC/Youden criterion on a calibration split specific to the deployment environment; the default MPI > 0.60 threshold derived from SWaT should not be transferred to a new domain without recalibration, as the optimal operating point shifts with the base rate of manipulation attempts and the cost asymmetry between false positives and false negatives in the target organisation. (iv) SIEM/SOAR integration: the MPI score is a continuous value in [0, 1] and can be forwarded as a numeric alert attribute to any SIEM platform, enabling threshold-based rules, temporal correlation with other event sources, and retrospective forensic analysis without requiring a binary alert output. (v) Alert escalation policy: given the interpretable structure of the MPI score, organisations are advised to define a three-tier response protocol rather than a single binary threshold: MPI ≤ 0.40: normal operation, no action; 0.40 < MPI ≤ 0.60: advisory status, log the event and increase sampling frequency for the flagged channel; MPI > 0.60: alert, notify the security operations team and initiate the defined incident-response procedure. The advisory band is particularly useful during the first weeks after deployment, when baseline statistics may not yet have converged, and reduces operator alarm fatigue relative to a hard binary threshold.

Periodic model health checks should be scheduled independently of operator-triggered recalibration. A recommended procedure is to run the framework in shadow mode (scoring but not alerting) on a

held-out normal-operation window at the end of each operational cycle, and to verify that the resulting MPI distribution remains centred below 0.30 with standard deviation under 0.10. Persistent upward drift in the shadow-mode score without confirmed attacks is a signal of unresolved concept drift and should trigger a scheduled recalibration. The framework design is consistent with the data-integrity monitoring principles of NIST SP 800-82 [18] and supports audit-trail generation required under Article 21 of the NIS2 Directive [17], which mandates continuous risk-management and incident-reporting procedures for operators of essential services.

Acknowledgments

The experiments were conducted using the Cyber Range facilities of the Faculty of Cybersecurity and Information Technologies, National University “Odesa Law Academy”. The SWaT-related materials used in this study were obtained from publicly available analytical reports and the open dataset description.

Author Contributions

Conceptualization: A. V. Sokolov and O. Ye. Chepurna; methodology: A. V. Sokolov, S. A. Horbachenko, and O. Ye. Chepurna; software and implementation: A. V. Sokolov; experimental validation: A. V. Sokolov and O. G. Trofymenko; formal analysis: A. V. Sokolov and O. Ye. Chepurna; writing – original draft preparation: A. V. Sokolov and O. Ye. Chepurna; writing – review and editing: all authors; supervision: S. A. Horbachenko. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding. The Cyber Range infrastructure used for experimental validation is operated by the Faculty of Cybersecurity and Information Technologies, National University “Odesa Law Academy”, and is maintained as an institutional resource for applied research in cybersecurity.

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration on Generative AI

During the preparation of this work, the authors used a large language model in order to: grammar and spelling check, paraphrase and reword. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. Soori, F. Karimi Ghaleh Jough, R. Dastres, B. Arezoo, AI-based decision support systems in Industry 4.0, a review, J. Econ. Technol. (2024) doi:10.1016/j.ject.2024.08.005.

[2] S. Khandekar, et al., Network-based real-time detection of data manipulation attacks in industrial control systems, International Journal of Information Security 24 (2025) 1–18. doi:10.1007/s10207-025-01172-3.

[3] S. Mokhtari, A. Abbaspour, K. K. Yen, A. Sargolzaei, A machine learning approach for anomaly detection in industrial control systems based on measurement data, Electronics 10 (2021) 407. doi:10.3390/electronics10040407.

[4] M. Choraria, A. Chattopadhyay, U. Mitra, E. G. Ström, Design of false data injection attack on distributed process estimation, IEEE Trans. Inf. Forensics Secur. 17 (2022) 670–683. doi:10.1109/TIFS.2022.3146078.
[5] J. Petráš, A. Hyseni, J. Zbojovský, M. Pavlík, Detecting Benford's law effectiveness threshold differences according to affecting operation, Axioms 14(4) (2025) 273. doi:10.3390/axioms14040273.
[6] M. Ivanova, E. Feckova Skrabulakova, A. Jandera, Z. Sarosiova, T. Skovranek, Benford's law and transport infrastructure: The analysis of the main road network's higher-level segments in the EU, ISPRS Int. J. Geo-Inf. 14(11) (2025) 450. doi:10.3390/ijgi14110450.
[7] Y. Cherevko, O. Chepurna, Y. Kuleshova, On information geometry methods for data analysis, Geometry, Integrability and Quantization 29 (2024) 11–22. doi:10.7546/giq-29-2024-11-22.
[8] S. A. Horbachenko, The place of cybersecurity management in modern managerial science and practice, Sustainable Development of Economy 1(48) (2024) 144–149. doi:10.32782/2308-1988/2024-48-19.
[9] C. Poutré, D. Chételat, M. Morales, Deep unsupervised anomaly detection in high-frequency markets, J. Finance Data Sci. 10 (2024) 100129. doi:10.1016/j.jfds.2024.100129.
[10] S. Jaradat, et al., Cyberattack detection on SWaT plant industrial control systems using machine learning, Artif. Intell. Auton. Syst. 2024(2) (2024) 0006. doi:10.55092/aias20240006.
[11] M. M. Aslam, et al., An optimized anomaly detection framework in industrial control systems through grey wolf optimizer and autoencoder integration, Sci. Rep. 15 (2025) 27579. doi:10.1038/s41598-025-12775-0.
[12] E. Kaziakhmedov, E. Dworetzky, J. Fridrich, Limits of data-driven steganography detectors, in: Proceedings of the 11th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '23), ACM, New York, NY, 2023, pp. 67–76. doi:10.1145/3577163.3595094.
[13] T. Neubert, et al., An analysis framework for steganographic network data in industrial control systems, in: Proceedings of SECURWARE 2024, IARIA, 2024, pp. 130–139.
[14] M. Karpinski, A. Sokolov, A. Tokkuliyeva, V. Radush, N. Kazakova, A. Shaikhanova, N. Zagorodna, A. Korchenko, Development of high-quality cryptographic constructions based on many-valued logic affine transformations, Electronics 14(10) (2025) 2094. doi:10.3390/electronics14102094.
[15] N. Kazakova, R. Shevchuk, A. Sokolov, A. Aliyev, A. Yarmilko, K. Veres, Adaptive code-controlled steganography with enhanced robustness to JPEG compression, Symmetry 18(4) (2026) 632. doi:10.3390/sym18040632.
[16] M. Conti, D. Donadel, F. Turrin, A survey on industrial control system testbeds and datasets for security research, IEEE Commun. Surv. Tutor. 23(4) (2021) 2248–2294. doi:10.1109/COMST.2021.3094360.
[17] European Parliament and Council, Directive (EU) 2022/2555 of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS2 Directive), Official Journal of the European Union L 333, 27.12.2022, pp. 80–152.
[18] K. Stouffer, V. Pillitteri, S. Lightman, M. Abrams, A. Hahn, Guide to Industrial Control Systems (ICS) Security, NIST Special Publication 800-82 Revision 3, National Institute of Standards and Technology, Gaithersburg, MD, 2023. doi:10.6028/NIST.SP.800-82r3.
[19] M. Elnour, M. Noorizadeh, M. Shakerpour, N. Meskin, K. Khan, R. Jain, A machine learning based framework for real-time detection and mitigation of sensor false data injection cyber-physical attacks in industrial control systems, IEEE Access 11 (2023) 86977–86998. doi:10.1109/ACCESS.2023.3303015.
[20] M. Riaz, H. Alshammari, N. Abbas, T. Mahmood, Navigating process drift: the power of CUSUM in monitoring air quality processes and maintenance operations, Arab. J. Sci. Eng. 50(14) (2025) 11145–11166. doi:10.1007/s13369-024-09453-0.
[21] C. L. Jones, A.-S. G. Abdel-Salam, D. A. Mays, Novel Bayesian CUSUM and EWMA control charts via various loss functions for monitoring processes, Qual. Reliab. Eng. Int. 39(1) (2023) 164–189. doi:10.1002/qre.3229.