https://doi.org/10.51885/3107-2755_JISC_2026_2_3
SRSTI 81.93.29

BEHAVIORAL RISK SCORING FOR INSIDER THREAT MITIGATION IN ZERO TRUST ENTERPRISE NETWORKS

ZERO TRUST АРХИТЕКТУРАСЫ БАР КОРПОРАТИВТІК ЖЕЛІЛЕРДЕГІ ИНСАЙДЕРЛЕРДІҢ ҚАУІПТЕРІН БЕЙТАРАПТАНДЫРУ ҮШІН МІНЕЗ ҚҰЛЫҚ ТӘУЕКЕЛДЕРІН БАҒАЛАУ

ОЦЕНКА ПОВЕДЕНЧЕСКИХ РИСКОВ ДЛЯ НЕЙТРАЛИЗАЦИИ УГРОЗ СО СТОРОНЫ ИНСАЙДЕРОВ В КОРПОРАТИВНЫХ СЕТЯХ С АРХИТЕКТУРОЙ ZERO TRUST

Tamerlan Bodambayev [1]*, **Saule Amanzholova** [1]
1 School of Cybersecurity, Astana IT University, Astana, Kazakhstan
* Corresponding author: Tamerlan Bodambayev, 255210@astanait.edu.kz

Keywords:
Zero Trust; Zero Trust Network Access; enterprise networks; insider threat; behavioral analytics; risk scoring; adaptive access control; top-K ranking

ABSTRACT
Insider threats remain difficult to mitigate because authenticated users can misuse legitimate privileges while enterprise environments continuously produce heterogeneous identity, endpoint, application, and network telemetry. This study develops an applied behavioral risk-scoring pipeline that converts windowed events from the CERT Insider Threat Test Dataset into a time-indexed signal intended for Zero Trust access decisions and analyst triage. The method combines metadata-only feature aggregation, Logistic Regression and HistGradientBoosting baselines, user-window scoring, top-K ranking, and an explicit mapping from risk levels to allow, step-up multi-factor authentication, or restrictive actions. Daily and weekly aggregation are evaluated with 30 stratified 70/30 train/test splits under severe class imbalance. Ranking metrics are emphasized because they correspond to a fixed analyst review budget more directly than a single classification threshold. Weekly aggregation achieved Recall@10 = 1.0000 and Recall@1000 = 1.0000 in the reported case study, whereas daily aggregation achieved Recall@10 = 0.2667 ± 0.1729 and Recall@1000 = 0.4667 ± 0.2249. The contribution is an operational integration rather than a new classifier: behavioral evidence is transformed into a continuous, reviewable risk state and connected to reversible Zero Trust controls, while the temporal trade-off between ranking stability and response latency is made explicit.

Түйінді сөздер:
Zero Trust (сенімсіздік қағидаты); Zero Trust желілік қолжетімділік; корпоративтік желілер; ішкі қауіп; мінез-құлықтық талдау; тәуекелді бағалау; бейімделгіш қолжетімділікті басқару; top-K рейтингі

ТҮЙІНДЕМЕ
Инсайдерлік қауіптерді азайту әлі де қиын, себебі аутентификацияланған пайдаланушылар заңды артықшылықтарды теріс пайдалануы мүмкін, ал корпоративтік орталар гетерогенді сәйкестікті, соңғы нүктені, қолданбаны және желілік телеметрияны үздіксіз жасайды. Бұл зерттеу терезедегі оқиғаларды CERT Инсайдерлік Қауіп-қатер Сынағы Деректер Жинағынан Нөлдік Сенімге қол жеткізу шешімдері мен талдаушыларды сұрыптауға арналған уақыт индекстелген сигналға түрлендіретін қолданбалы мінез-құлық тәуекелдерін бағалау жүйесін әзірлеуде. Бұл әдіс тек метадеректерге арналған мүмкіндіктерді біріктіруді, Логистикалық Регрессияны және Гистограмманы Жоғарылатудың негізгі бағыттарын, пайдаланушы

терезесінің ұпайларын, top-K рейтингін және тәуекел деңгейлерінен нақты салыстыруды біріктіреді.рұқсат ету, көп факторлы аутентификацияны күшейту немесе шектеу әрекеттері. Күнделікті және апталық агрегация 70/30 пойыздардың/сынақтардың 30 стратификацияланған бөлінуімен бағаланады, бұл сыныптағы елеулі теңгерімсіздік жағдайында. Рейтингтік көрсеткіштерге баса назар аударылады, өйткені олар бір жіктеу шегінен гөрі аналитиктердің тұрақты шолу бюджетіне тікелей сәйкес келеді. Апта сайынғы біріктіру@10 = 1.0000 Еске түсіруге және@1000 = 1.0000 Еске түсіруге қол жеткізді, ал күнделікті біріктіру@10 = 0.2667 ± 0.1729 Еске түсіруге және@1000 = 0.4667 ± 0.2249 Еске түсіруге қол жеткізді. Бұл үлес жаңа жіктеуіш емес, операциялық интеграция болып табылады: мінез-құлық деректері тұрақты, қайта қаралатын тәуекел жағдайына айналады және нөлдік Сенімділіктің қайтымды бақылауымен байланысты, ал рейтингтің тұрақтылығы мен жауап берудің кешігуі арасындағы уақытша компаға келу айқын көрсетілген.

Ключевые слова:

Zero Trust; доступ к сети по модели Zero Trust; корпоративные сети; угрозы со стороны инсайдеров; поведенческая аналитика; оценка рисков; адаптивное управление доступом; ранжирование top-K

АННОТАЦИЯ

Внутренние угрозы остаются сложной проблемой, поскольку авторизованные пользователи могут злоупотреблять законными привилегиями, а корпоративные среды постоянно генерируют разнородную телеметрию, касающуюся идентификации, конечных точек, приложений и сети. В этом исследовании разработан прикладной конвейер оценки поведенческих рисков, который преобразует события, ограниченные временными рамками, из набора данных CERT Insider Threat Test Dataset в сигнал с временным индексом, предназначенный для принятия решений о доступе в рамках концепции «нулевого доверия» и сортировки аналитиками. Метод сочетает в себе агрегацию признаков только на основе метаданных, базовые модели логистической регрессии и HistGradientBoosting, оценку по временным окнам пользователей, ранжирование по K лучшим вариантам и явное сопоставление уровней риска с разрешением, усилением многофакторной аутентификации или ограничительными действиями. Ежедневная и еженедельная агрегация оценивается с использованием 30 стратифицированных разбиений на обучающую и тестовую выборки в соотношении 70/30 при сильном дисбалансе классов. Особое внимание уделяется метрикам ранжирования, поскольку они более непосредственно соответствуют фиксированному бюджету аналитиков на проверку, чем единый порог классификации. В представленном тематическом исследовании еженедельная агрегация показала значения Recall@10 = 1,0000 и Recall@1000 = 1,0000, тогда как ежедневная агрегация показала значения Recall@10 = 0,2667 ± 0,1729 и Recall@1000 = 0,4667 ± 0,2249. Вклад заключается скорее в операционной интеграции, чем в создании нового классификатора: поведенческие данные преобразуются в непрерывное, подлежащее проверке состояние риска и связываются с обратимыми механизмами контроля «нулевого доверия», при этом явно указывается временной компромисс между стабильностью ранжирования и задержкой реакции.

INTRODUCTION

Zero Trust has shifted enterprise security from perimeter-based assumptions toward continuous verification, explicit policy evaluation, least privilege, and context-aware access decisions (Ahn et al., 2024; Huang & Zhu, 2024). In this operating model, an authenticated identity is not treated as a permanent indication of trust. Identity attributes, device posture,

session context, and observed behavior can all contribute to a decision that is recalculated during the lifetime of a session. This approach is important for insider-threat mitigation because a malicious employee, a negligent employee, or an attacker operating through a compromised account can possess valid credentials and therefore pass an initial authentication step while still performing actions that increase organizational risk.

The difficulty is not limited to detecting anomalous events. Enterprise environments generate heterogeneous records from authentication systems, endpoints, file services, web activity, e-mail, and network flows. Each source describes a partial view of behavior, uses its own timestamps and identifiers, and has a different baseline of normal activity. A useful insider-risk mechanism therefore has to transform these heterogeneous observations into a common representation, retain enough temporal context to distinguish transient variation from sustained change, and produce an output that can be consumed by a policy engine or an analyst. Access-control research has demonstrated the value of contextual attributes and deception for improving visibility of misuse (Alohaly et al., 2022), while organizational studies emphasize both behavioral indicators and the operational costs associated with false alarms and user friction (Rahman et al., 2022; Whitelaw et al., 2024).

Learning-based insider-threat detection provides multiple ways to process this evidence. Unsupervised ensembles, one-class streaming methods, and density-based approaches are attractive when positive labels are scarce, delayed, or expensive to construct (Al-Shehari et al., 2024; Le & Zincir-Heywood, 2021; Peccatiello et al., 2023). Supervised and deep-learning methods exploit richer labeled structure or sequential representations (Dong et al., 2025; He et al., 2024; Kim & Kim, 2026; Lavanya et al., 2024; Nasir et al., 2021). Graph approaches model interactions among users, resources, hosts, and time contexts (Li et al., 2023; Li et al., 2024; Xiao et al., 2023; Xiao et al., 2025). Cyber-human models add behavioral and contextual dimensions (Roy & Chen, 2024), while sequence-based and masquerader-oriented methods explicitly address temporal dependence and impersonation behavior (Gonzales et al., 2025; Huang et al., 2025). Enterprise-storage anomaly detection (Alsalim, 2025), explainability-oriented surveys (Alketbi & Mehmood, 2025), digital-twin and attention-based methods (Wang & Saddik, 2023), adversarial autoencoder approaches (Zhu et al., 2024), and real-time adaptive-neighbor methods (Cai et al., 2024) further demonstrate the breadth of the technical landscape. The recent risk-decay perspective (Greitzer, 2026) is especially relevant to Zero Trust because it reinforces the idea that insider risk should evolve as evidence becomes older rather than remain a permanent binary label.

Despite this breadth, there is a practical integration gap. A detector may return an anomaly probability, ranking, or node-level score, but the operational path from that output to a continuously evaluated access policy is often not made the primary object of evaluation. Detection quality is commonly reported with threshold-based classification metrics even when a security operations center (SOC) can realistically review only a limited number of cases per hour or day. In such settings, the key question is not simply whether a positive case can be separated from benign activity, but how quickly the highest-risk cases appear near the top of a review queue and how that ranking can be connected to reversible security actions.

This paper addresses that integration problem through an applied case study. It constructs a user-window behavioral risk score from metadata-level CERT telemetry, compares daily and weekly temporal aggregation, uses Logistic Regression (LR) and HistGradientBoosting (HGB) as transparent supervised baselines, and evaluates the resulting score primarily with Precision@K and Recall@K. The study then maps score ranges to three Zero Trust responses: allow, step-up multi-factor authentication (MFA), and restrict or block. The design intentionally separates the learning problem from the policy problem so that model outputs can be analyzed as a continuous signal and policy thresholds can be adjusted independently.

The study is organized around three research questions. RQ1 asks how temporal aggregation changes ranking quality under severe insider-threat class imbalance. RQ2 asks how a behavioral score can be expressed as an operational input to a Zero Trust policy decision point and policy enforcement point without assuming that the score itself is a final security verdict. RQ3 asks which methodological limitations must be made explicit before results from a synthetic, scenario-driven dataset can inform enterprise deployment. These questions are deliberately practical: the paper does not claim a fundamentally new classifier, but rather an experimentally evaluated risk-scoring loop that connects detection evidence, analyst prioritization, and access control.

The practical contributions are fourfold. First, the paper provides a reproducible transformation from heterogeneous CERT events to a user-window feature representation. Second, it makes the literature comparison explicit across anomaly, deep-learning, graph, cyber-human, sequence, real-time, and explainability-oriented families rather than treating insider-threat detection as a single homogeneous problem. Third, it evaluates temporal granularity with fixed review budgets using top-K metrics that are directly interpretable by analysts. Fourth, it defines a small set of reversible Zero Trust actions and explains the trade-off between faster reaction and more stable evidence when the temporal window changes. The remainder of the article follows the journal template by presenting the materials and methods, results and discussion, conclusion, and publication statements.

Research context and comparison with existing approaches

The literature contains several complementary perspectives on insider risk. Zero Trust work emphasizes that authorization should remain contextual and adaptive rather than being granted indefinitely after network entry or an initial login (Ahn et al., 2024; Huang & Zhu, 2024). This system-level perspective establishes why a risk signal is useful, but it does not dictate which behavioral representation or learning algorithm should create that signal. Attribute-based access control integrated with cyber deception (Alohaly et al., 2022) addresses the visibility of suspicious activity and demonstrates how access policy can consume richer context. Readiness studies of unintentional insiders (Rahman et al., 2022) and practitioner reviews from financial services (Whitelaw et al., 2024) further indicate that operational adoption depends on more than model accuracy: monitoring must fit organizational processes, acceptable user friction, and the responsibilities of security analysts.

Unsupervised and one-class methods (Al-Shehari et al., 2024; Le & Zincir-Heywood, 2021; Peccatiello et al., 2023) are especially suitable when labeled malicious behavior is sparse. Their principal advantage is that the detector can learn a representation of normal behavior and flag deviations without requiring a large set of known incidents. Their operational output is commonly an anomaly score or rank, which naturally supports prioritization but does not automatically determine a policy response. A score can also be sensitive to population drift, data-volume changes, and the choice of neighborhood or density assumptions. The present study retains the useful ranking interpretation of these approaches while using supervised baselines to take advantage of the scenario labels available in the CERT case study.

Supervised and deep-learning approaches (Dong et al., 2025; He et al., 2024; Kim & Kim, 2026; Lavanya et al., 2024; Nasir et al., 2021) can represent more complex relationships and sequences. Deep models may detect nonlinear interactions that a compact feature vector cannot capture, and specialized autoencoder or diffusion-based architectures can model distinctive structures in insider activity. The trade-off is implementation complexity, larger data requirements, and a more demanding explainability problem. The present case study therefore uses LR and HGB as baseline models rather than attempting to compete with the most sophisticated architecture in the literature. The purpose is to test the operational value of the score-to-policy loop under controlled conditions, not to make a model-family superiority claim.

Graph-based methods (Li et al., 2023; Li et al., 2024; Xiao et al., 2023; Xiao et al., 2025) provide an additional layer of context by representing interactions among users, resources, hosts, and time. Cyber-human modeling (Roy & Chen, 2024) broadens this context further by introducing human and organizational relationships, while contextual sequence and masquerader-focused methods (Gonzales et al., 2025; Huang et al., 2025) target temporal structure and impersonation. These approaches strengthen the argument that insider activity should not be represented as an isolated event. However, greater representational richness does not remove the need for a policy interface. A Zero Trust system still needs a stable and auditable way to convert evidence into an access decision, to explain the reason for a challenge, and to decide when a high-risk state should decay as evidence becomes stale (Alketbi & Mehmood, 2025; Greitzer, 2026).

Recent work also illustrates adjacent directions. Efficient anomaly detection for enterprise storage (Alsalim, 2025) focuses on a specific technical surface, whereas digital-twin and self-attention approaches (Wang & Saddik, 2023) aim at richer environmental representation. Review studies (Alzaabi & Mehmood, 2024) map the broader machine-learning landscape, while adversarial autoencoder work (Zhu et al., 2024) indicates that robustness against adaptive behavior is itself a concern. Real-time learning approaches such as LAN (Cai et al., 2024) demonstrate the importance of responsiveness at activity level. The gap addressed here is therefore not the absence of advanced models. It is the need for a simple, testable operational chain in which heterogeneous behavior is aggregated over time, converted to a score, ranked under a defined review budget, and explicitly consumed by Zero Trust enforcement.

Table 1. Comparison of representative insider-threat approaches

Approach	Signals / scope	Typical output	Gap for Zero Trust risk scoring
Unsupervised / one-class [6]-[8]	Multi-source or streaming behavior	Anomaly rank / score	Useful with scarce labels; policy mapping is typically implicit.
Deep / temporal [9]-[13]	Sequences and rich learned representations	Class probability / score	Higher complexity and stronger deployment, data, and explanation requirements.
Graph / cyber-human [14]-[18]	Relational, temporal, human, or interaction context	Node / event anomaly or risk	Strong context; conversion into a stable access-control state is not the primary output.
Sequence / masquerader [19], [20]	Temporal context and impersonation patterns	Sequence anomaly / attack indication	Useful timing information; operational thresholds and actions remain deployment-specific.
Explainability / adjacent methods [21], [22], [24]-[27]	Storage, interpretability, digital-twin, adversarial, or real-time context	Anomaly, explanation, or adaptive score	Addresses important dimensions, but no common score-to-policy operational loop.
Proposed case study	Identity, endpoint, web/email, and flow metadata	Continuous score + top-K rank	Explicit score-to-policy mapping and 1D vs 7D evaluation under review budgets.

MATERIALS AND METHODS

1. Research design and threat model

The study follows an applied simulation design. The unit of analysis is a user-window rather than an individual event. Each window contains aggregated observations associated with one user over a fixed interval, and the same construction is evaluated under daily (1D) and weekly (7D) temporal aggregation. The model output is a continuous score in the interval [0, 1] that is interpreted as relative behavioral risk within the evaluated population. The analysis is intentionally organized around ranking and actionability because a SOC typically operates under a finite review budget and because score thresholds can change with organizational tolerance for friction.

The threat model covers three operational cases already represented by the notion of an insider in enterprise security: a malicious legitimate user, an otherwise legitimate user whose behavior becomes risky through negligence or misuse, and an external attacker operating through a compromised account. In all three cases, authentication can be valid. The security problem is therefore framed as post-authentication monitoring: determine whether the observed combination of identity, endpoint, resource, communication, and temporal behavior is consistent with the user baseline and with the broader enterprise population. The score is not treated as proof of malicious intent; it is treated as a control signal that can trigger additional verification or temporary restriction.

This distinction is important for policy design. A risk score is evidence about behavior, whereas an access decision is an organizational control action. Treating the score as an automatic accusation would make false positives more costly and would blur the boundary between detection and authorization. The proposed design keeps these functions separate: the scoring layer ranks risk, while the policy layer applies configurable thresholds and reversible actions. The same score can consequently be reviewed by analysts, logged for audit, or consumed automatically by a policy decision point.

2. Dataset and label construction

The study uses the CERT Insider Threat Test Dataset as a controlled proxy for enterprise telemetry (Lindauer, 2020). The dataset contains synthetic background activity and synthetic malicious-actor scenarios across enterprise-style logs. Its value for this case study is the ability to join multiple behavioral surfaces while retaining scenario-level ground truth. At the same time, its synthetic and scenario-driven nature is a central limitation, and results are interpreted as a methodological case study rather than as a prevalence estimate for real organizations.

The primary entity is the user. Raw records are normalized by user identifier and timestamp and then aggregated into either one-day or seven-day windows. A window receives a positive label when its time interval overlaps an insider-scenario interval in the dataset answer key. This interval-overlap rule converts scenario annotations into a continuous-monitoring formulation: instead of labeling only a particular event as malicious, the method asks whether the user was in a risk-bearing state during the evaluated time window. The approach also exposes a boundary effect. An activity immediately before or after a scenario interval may be behaviorally similar but receive a different binary label. Consequently, ranking and early-warning interpretation are emphasized rather than assuming that the label is a perfect statement of intent.

For reproducibility, the source CERT r4.2 release represents 1,000 synthetic employees, including 70 simulated insider users, over approximately 17 months of activity. The experiment evaluates 1-day (1D) and 7-day (7D) user-windows. In the reported 7D evaluation, each test split contains exactly one positive user-window. This is implied by the reported Recall@10 = 1.0000

and Precision@10 = 0.1000, and is independently consistent with Recall@1000 = 1.0000 and Precision@1000 = 0.00100: the single positive window is recovered in every split, while 999 of the first 1,000 ranked cases are negative. The full negative test-set size is not separately retained in the reported run summary, so no larger exact negative count is asserted here. For the 1D results, the reported mean Precision@10 and Recall@10 are consistent with four positive test windows per split under the stated per-split top-K evaluation.

The data processing pipeline removes payload content and retains metadata-level counts, durations, novelty indicators, and relational statistics. This design limits the study to behavioral signals that can be computed without inspecting message or file contents. It also makes the feature representation easier to reproduce across environments where content collection is restricted. The resulting representation is compatible with a Zero Trust setting in which a policy engine receives summarized telemetry rather than raw content.

3. Behavioral feature engineering

Feature engineering maps heterogeneous events into compact behavioral descriptors. Identity and authentication features include logon counts, the proportion of activity occurring outside ordinary hours, host dispersion, and novelty relative to a user-specific baseline. These indicators capture changes in authentication intensity and environmental context without assuming that a login is malicious by itself. Endpoint and data-movement features include file-access intensity, the number of unique files touched, rare-directory access, removable-media activity, and burst-like copying behavior. Such features represent changes in the breadth or concentration of a user's interaction with enterprise data.

Web and communication features include unique domains, domain novelty, distinct recipients, external-recipient ratios when derivable, and temporal changes in communication patterns. For selected hosts, domains, and shared drives, inverse-frequency and user-to-population divergence measures capture organizational rarity. The intention is to combine two related viewpoints: deviation from the user's own recent history and deviation from the organization's broader behavior. A behavior that is common for a user but rare for the organization can therefore remain distinguishable from a behavior that is unusual for both.

Table 2 summarizes the main feature groups and their operational interpretation. The table also clarifies that the method is not dependent on a single signal type. Instead, it fuses complementary metadata streams so that a score can reflect identity context, data movement, communication, network behavior, and relative rarity at the same time.

4. Feature groups and operational semantics

Table 2. Behavioral feature groups used in the case study

Feature group	Examples	Operational interpretation
Identity and authentication	Logon count, after-hours ratio, host dispersion, novelty	Change in access rhythm and authentication context.
Endpoint and data movement	Unique files, access intensity, rare directories, removable media, burst copying	Breadth and intensity of data interaction.
Web and e-mail	Unique domains, domain novelty, distinct recipients, external-recipient ratio	Communication expansion and unusual destinations.
Network / flow metadata	Connection counts, flow concentration, source/host dispersion	Changes in communication topology without payload inspection.

Feature group	Examples	Operational interpretation
Population rarity	Inverse-frequency and user-to-population divergence	How unusual an object or pattern is relative to peers.
Temporal change	Current activity versus moving historical average	Persistence, acceleration, and transition into a new behavioral state.

5. Scoring models and decision policy

Two supervised baseline models are used: Logistic Regression (LR) and HistGradientBoosting (HGB). LR provides a compact and interpretable linear baseline, while HGB captures nonlinear interactions among engineered behavioral features without requiring a deep sequence architecture. The imbalance problem is handled with class weighting for LR and sample weighting for HGB. Their outputs are converted into a probability-like risk score r(u,t) in the interval [0,1]. The term probability-like is intentional: the case study does not claim that the raw model output is fully calibrated as a probability of maliciousness. Calibration is identified as a future extension. Figure 1 illustrates the resulting detection-to-policy loop, while Figure 2 shows the corresponding reversible action mapping.

For LR, the scoring function can be expressed conceptually as r(u,t) = sigmoid(beta_0 + beta^T x(u,t)), where x(u,t) is the feature vector for user u in time window t. HGB estimates a nonlinear score from the same representation. The operational design does not require that every classifier use the same internal mechanism; it requires only that the model expose a comparable monotonic risk scale so that ranking and policy thresholds can be applied. This separation makes the proposed framework compatible with an alternative anomaly detector when labels are unavailable.

The Zero Trust policy layer uses two thresholds tau_1 and tau_2, with tau_1 < tau_2. Scores below tau_1 are mapped to allow, intermediate scores to step-up MFA, and scores at or above tau_2 to restrict or block. Table 5 summarizes this mapping. In the present case study, tau_1 and tau_2 are configurable policy parameters rather than learned model parameters, and they were not optimized from the reported top-K results. For a deployment-oriented calibration, tau_1 should be selected on a validation set to satisfy a target challenge or false-acceptance constraint, while tau_2 should satisfy a stricter restriction/blocking criterion; both should be reviewed against expected user friction and security cost. Because such threshold calibration was not experimentally performed here, the manuscript does not claim an empirically optimal threshold pair. The risk score can nevertheless be logged and reviewed independently even when a policy decision is automated.

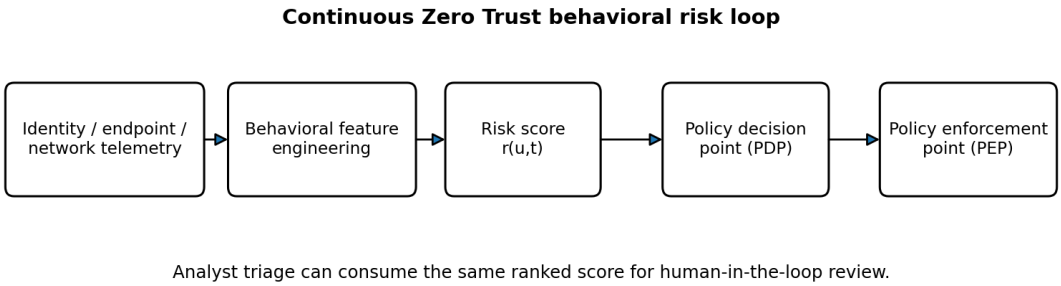


Figure 1. Behavioral risk scoring integrated into a Zero Trust policy enforcement loop

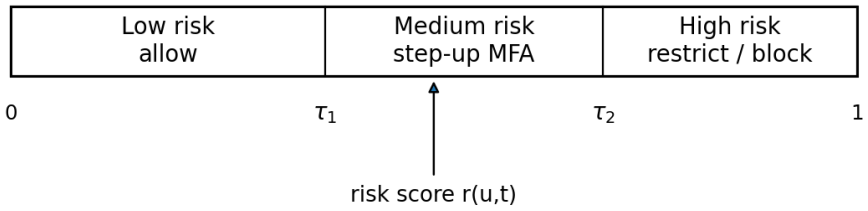


Figure 2. Mapping of a continuous behavioral risk score to reversible Zero Trust actions

6. Evaluation protocol

The default case-study settings are summarized in Table 3. The dataset is evaluated with both daily and weekly windows. Earlier windows can be used for training and later windows for testing in a time-aware formulation; however, the reported robustness results in this manuscript use 30 stratified 70/30 train/test splits so that the positive class remains represented across repeated experiments. This distinction matters when interpreting the results: stratification provides a useful sensitivity analysis, but it is not equivalent to a fully time-forward production validation.

Because insider labels are highly imbalanced, Precision@K and Recall@K are treated as the primary indicators. Here K represents the number of ranked user-windows that an analyst is assumed to inspect. Recall@K asks how much of the positive population is recovered within the first K cases, while Precision@K indicates the fraction of reviewed cases that are positive under the study label. The metrics therefore connect directly to a review budget. Conventional Precision, Recall, F1, ROC-AUC, and PR-AUC remain useful secondary measures, but they do not by themselves express how many cases a SOC must examine before finding a meaningful portion of the positive population.

The operational evaluation also tracks false-acceptance risk, challenge rate, detection delay, and score stability as candidate deployment measures. These indicators are not all estimated from the present synthetic experiment. They are included because a production Zero Trust control loop must balance detection against latency and user friction. In other words, a scoring method is operationally incomplete if it detects a large number of positives only by generating an unacceptable number of challenges for benign users.

Table 3. System settings and case-study parameters

Parameter	Value
Dataset	CERT Insider Threat Test Dataset r4.2 [28]
Source population	1,000 synthetic users; 70 simulated insider users
Activity period	Approximately 17 months
Positive test windows	1D: 4 per test split (consistent with reported mean top-K metrics); 7D: 1 per test split

Parameter	Value
Negative test cases	7D: 999 negative cases within the top-1000 review set; full test-set count not retained in the run summary
Windows	1 day (1D) and 7 days (7D)
Repeated evaluation	30 stratified 70/30 train/test splits
Models	Logistic Regression; HistGradientBoosting
Imbalance handling	Class weights for LR; sample weights for HGB
Policy thresholds	tau_1 / tau_2 are configurable; not tuned in the reported experiment
Primary ranking metrics	Precision@K; Recall@K
Secondary / operational metrics	Precision, Recall, F1, ROC-AUC, PR-AUC; false-acceptance risk, challenge rate, detection delay, score stability

7. End-to-end implementation flow

Figure 3 summarizes the computational workflow used to produce the reported results. First, source logs are extracted and normalized. Second, events are grouped into the selected 1D or 7D windows and aligned with scenario intervals. Third, feature vectors are generated from counts, novelty, rarity, relational statistics, and temporal changes. Fourth, LR and HGB are trained with the specified imbalance handling. Fifth, each user-window receives a score that can be ranked across the evaluation population. Sixth, top-K metrics are computed. Finally, the score is mapped to the Zero Trust policy layer. This sequence establishes traceability from raw telemetry to the reported performance values.

The implementation flow also clarifies a key reproducibility principle. Feature generation, model training, scoring, and policy mapping are separate stages. A future implementation can therefore replace the classifier while leaving the surrounding data model and policy interface unchanged. Similarly, a security team can modify the policy thresholds without retraining the detector if the score remains on a stable scale. Such modularity reduces the risk that a model update becomes inseparable from an authorization-policy update.

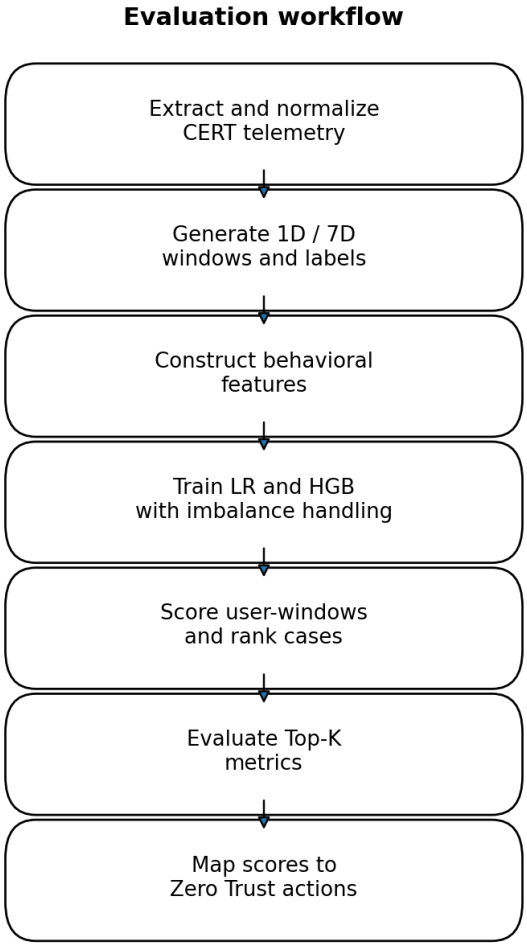


Figure 3. End-to-end evaluation workflow used in the case study

RESULTS AND DISCUSSION

1. Detection and ranking results

The reported results come from the developed simulation pipeline using the 1D and 7D feature windows and 30 stratified train/test splits. Table 4 summarizes the selected operating points. Weekly aggregation achieved Recall@10 = 1.0000 and Recall@1000 = 1.0000. The corresponding Precision@10 was 0.1000 and Precision@1000 was 0.00100. Daily aggregation achieved Recall@10 = 0.2667 +/- 0.1729 and Recall@1000 = 0.4667 +/- 0.2249, with Precision@10 = 0.1067 +/- 0.0691 and Precision@1000 = 0.00187 +/- 0.00090. These values indicate that the weekly representation consistently places the positive window near the top of the ranking in the reported simulation, whereas the daily representation requires a larger review budget to recover a comparable portion of positive cases.

Figure 4 visualizes the same operating points in a common scale. The most visible difference is not in Precision@10, which is similar for the two windows, but in Recall@10 and Recall@1000. Under the weekly representation, the reported recall remains at 1.0 at both review budgets. Under the daily representation, recall is lower and more variable. This pattern is consistent with the idea that repeated weak signals can become more informative when

aggregated across a longer interval, while short windows preserve local variation that can dilute the ranking signal.

The result should not be interpreted as evidence that weekly aggregation is universally superior. The observed advantage is conditional on the CERT scenarios, the feature construction, the labeling rule, and the evaluation design. A seven-day window can suppress short-lived benign fluctuations, but it can also delay the response to a rapidly developing incident. Conversely, a one-day window can respond sooner but may require more aggressive feature normalization or smoothing to prevent short-term noise from dominating the score. Temporal granularity is therefore an operational design choice rather than a property that can be optimized independently of the organization's response requirements.

Table 4. Selected top-K performance (mean +/- standard deviation over 30 splits)

KPI	1D window	7D window
Precision@10	0.1067 +/- 0.0691	0.1000 +/- 0.0000
Recall@10	0.2667 +/- 0.1729	1.0000 +/- 0.0000
Precision@1000	0.00187 +/- 0.00090	0.00100 +/- 0.00000
Recall@1000	0.4667 +/- 0.2249	1.0000 +/- 0.0000

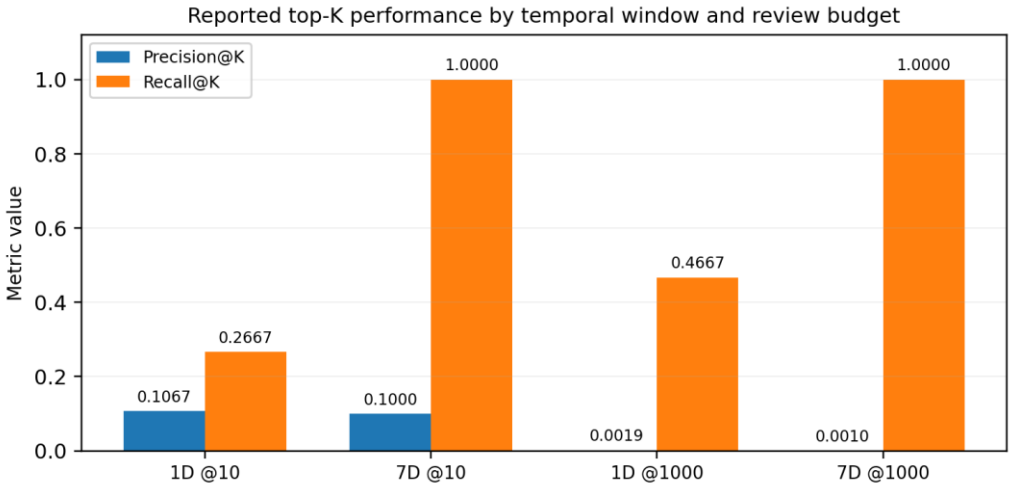


Figure 4. Reported Precision@K and Recall@K at K = 10 and K = 1000 for daily and weekly aggregation.

2. Interpretation in terms of analyst review budgets

Top-K evaluation changes the interpretation of an insider-threat detector. In a balanced classification problem, it is often natural to ask whether a threshold produces acceptable precision and recall. In a SOC, however, the number of cases that can actually be investigated may be fixed by staffing, shift length, ticket volume, or escalation rules. A ranked queue is therefore more operationally meaningful than an abstract threshold alone. Recall@10 indicates whether the highest-priority ten cases contain the positive case under the study label; Recall@1000 asks the same question with a much larger review budget. The difference between these two operating points exposes how quickly useful signal rises to the top of the queue.

The reported results also show why Precision@K becomes extremely small at K = 1000 under severe imbalance. Even when a detector successfully places positive cases near the top, the denominator of a large review budget can contain many benign windows. This is not necessarily a failure of the ranking mechanism; it reflects the prevalence structure of the evaluation set. A practical deployment should therefore define different review queues for different operational consequences. A small high-risk queue can feed immediate analyst attention or step-up

authentication, while a larger lower-priority queue can be retained for retrospective hunting and correlation.

The score-to-policy mapping adds another dimension. A top-ranked case can be investigated without necessarily being blocked, whereas a threshold crossing in a high-assurance application may justify step-up MFA before an analyst has completed manual review. This separation supports reversible escalation: the system can demand a second factor, restrict a session to a lower-privilege resource set, or temporarily block an action while preserving the original identity and session context for investigation. Such measures are more consistent with the idea of continuous authorization than a single permanent trust decision.

3. Operational policy mapping and reversibility

The policy states in Table 5 are intentionally reversible. An elevated score may decrease when subsequent evidence becomes benign or when the scoring window moves beyond the activity that triggered the alert. This design is consistent with the risk-decay perspective discussed in (Greitzer, 2026), but the current case study does not estimate a decay parameter empirically. Instead, it defines decay as an architectural property: risk is repeatedly recomputed from a moving evidence window rather than stored indefinitely as a static label.

Table 5. Illustrative score-to-policy mapping

Risk state	Policy action	Operational rationale
Low: r(u,t) < tau_1	Allow / continue session	Behavior is within the current risk tolerance; monitor normally.
Medium: tau_1 <= r(u,t) < tau_2	Step-up MFA or additional verification	Increase assurance before permitting sensitive activity; preserve session continuity when possible.
High: r(u,t) >= tau_2	Restrict, isolate, or block selected actions	Limit exposure while preserving evidence for investigation and avoiding unnecessary permanent account disablement.

Reversibility also reduces the operational cost of uncertainty. A detector trained on synthetic scenarios cannot determine intent with certainty, especially when the label is derived from interval overlap. A temporary challenge or targeted restriction acknowledges that uncertainty and buys time for additional evidence. In contrast, an irreversible disciplinary or account action would require a stronger evidentiary process than the score alone provides. The framework therefore treats automated controls as security friction mechanisms, not as substitutes for incident investigation or organizational judgment.

4. Comparison with prior research

The findings fit the broader literature in several ways. First, the importance of multiple behavioral signals is consistent with anomaly-detection studies that use heterogeneous activity rather than a single event source (Al-Shehari et al., 2024; Le & Zincir-Heywood, 2021; Peccatiello et al., 2023). Second, the temporal aggregation effect is directionally compatible with sequence- and interaction-based research (Gonzales et al., 2025; Huang et al., 2025; Li et al., 2023; Li et al., 2024; Roy & Chen, 2024; Xiao et al., 2023; Xiao et al., 2025), which treats insider behavior as an evolving pattern rather than an isolated action. The present contribution is to make the temporal choice observable at the level of an operational review budget: the analyst can see how many cases must be examined under each windowing scheme.

The baseline choice also clarifies the scope of the contribution. LR and HGB are not expected to outperform every deep, graph, or specialized model described in (Alzaabi & Mehmood, 2024; Cai et al., 2024; Dong et al., 2025; Gonzales et al., 2025; He et al., 2024; Huang et

al., 2025; Kim & Kim, 2026; Lavanya et al., 2024; Li et al., 2023; Li et al., 2024; Nasir et al., 2021; Roy & Chen, 2024; Wang & Saddik, 2023; Xiao et al., 2023; Xiao et al., 2025; Zhu et al., 2024). Instead, they provide a controlled test of whether a relatively simple score can be integrated into a continuous access-control loop. The result is particularly useful for an applied study because model complexity can otherwise obscure the operational question. A more sophisticated detector can be inserted later without changing the surrounding data model, review-budget logic, or policy interface.

Explainability is another point of connection with the literature. Survey work on explainable artificial intelligence for malicious-insider detection (Alketbi & Mehmood, 2025) emphasizes that an output must remain understandable enough to support analyst action. The proposed framework does not yet provide a formal explanation engine, but its feature groups and policy states are deliberately human-readable. A future deployment can attach feature contributions, baseline deviations, and recent supporting events to the score so that a policy challenge is accompanied by an auditable reason code. This direction is important because the difference between “high risk” and “high risk because of a new host, unusual file access, and after-hours activity” is operationally significant.

The case study is also consistent with the emphasis on robustness in recent work. Adversarial autoencoder methods (Zhu et al., 2024) and real-time adaptive methods (Cai et al., 2024) illustrate that attackers may change their behavior and that detection latency matters. The current pipeline does not claim robustness to deliberate evasion. It does, however, make the relevant interface explicit: if a future model improves robustness or latency, its output can still be converted into the same risk state and policy mapping. This modular perspective is one reason the paper frames its contribution as operational integration rather than classifier novelty.

5. Detection quality versus response latency

A central implication of the daily-versus-weekly comparison is that detection quality and response latency cannot be optimized independently. The weekly window produced higher recall at both reported review budgets, but it aggregates seven days of activity. A real incident can unfold more quickly than that interval. The daily window has the opposite property: it can expose change sooner, yet the reported ranking is noisier and more variable. In a production system, the answer may therefore be a multi-timescale design rather than a single universal window.

A multi-timescale design can preserve the simplicity of the current framework. A short window can serve as a rapid-change detector, while a longer window can provide behavioral context and stabilize an otherwise volatile score. The policy engine can consume both signals through a composite rule or a confidence-aware aggregation layer. Such an extension should be evaluated with the same review-budget logic used in this paper, because adding more detectors without controlling alert volume can simply increase analyst workload.

The present results do not provide a measured multi-timescale model, so no claim is made about its performance. The methodological point is narrower: temporal granularity is a first-class design variable. A system that reports only a classifier threshold hides the fact that the same user can look benign in a short snapshot and anomalous over a longer sequence of interactions. Explicitly reporting the window makes the operational behavior of the score easier to reason about and reproduce.

6. Privacy, data minimization, and governance considerations

The feature design intentionally excludes payload content. This reduces the amount of sensitive information that must be processed by the detector and makes the pipeline centered on metadata, context, and temporal behavior. Even metadata can be sensitive, however, because access patterns, communication destinations, and working hours can reveal information about

employees. A production deployment should therefore define retention periods, role-based access to monitoring data, purpose limitations, and audit procedures appropriate to the organization and jurisdiction. These governance requirements are outside the scope of the synthetic dataset experiment but are directly relevant to operational adoption.

The score should also be treated as a security signal rather than a personnel-performance score. A high-risk state can result from unusual work schedules, emergency response activity, project changes, or a compromised endpoint. Without contextual review, the same feature may have multiple benign explanations. Accordingly, a production policy should distinguish between security enforcement and employee evaluation workflows, provide a documented escalation path, and allow analysts to attach contextual explanations to a challenged session. This is another reason the present framework keeps automated actions limited to access assurance and session restriction rather than disciplinary conclusions.

7. Limitations and threats to validity

The first limitation is dataset realism. The CERT Insider Threat Test Dataset is synthetic and scenario-driven (Lindauer, 2020). It is valuable for controlled experimentation but does not reproduce every enterprise environment, identity structure, logging policy, workload pattern, or attacker strategy. The reported scores therefore should not be interpreted as an estimate of real-world incident rates or as evidence that a 7D window will produce the same recall in a production network.

The second limitation concerns evaluation design. The primary robustness results use 30 stratified 70/30 train/test splits. Stratification ensures that the positive class is represented, which is useful for repeated sensitivity analysis, but it can reduce the degree to which the evaluation mimics a fully forward-looking deployment. In real operations, models often learn from earlier behavior and then encounter later behavior under possible distribution shift. Strict time-forward validation is therefore required before any production conclusion can be drawn.

The third limitation is model coverage. LR and HGB are useful baselines, but they are not a complete survey of modern deep, graph, self-supervised, or real-time detectors. The literature contains more expressive alternatives (Alzaabi & Mehmood, 2024; Cai et al., 2024; Dong et al., 2025; Gonzales et al., 2025; He et al., 2024; Huang et al., 2025; Kim & Kim, 2026; Lavanya et al., 2024; Li et al., 2023; Li et al., 2024; Nasir et al., 2021; Roy & Chen, 2024; Wang & Saddik, 2023; Xiao et al., 2023; Xiao et al., 2025; Zhu et al., 2024). The study does not claim that the baselines are technically dominant. Their purpose is to establish an interpretable reference point for the operational integration of scoring, ranking, and policy enforcement.

The fourth limitation concerns calibration and explainability. The current study uses probability-like scores for ranking and thresholding but does not report a formal calibration analysis. Consequently, a score of 0.8 should not be interpreted as a literal 80% probability of malicious behavior. Future work should evaluate calibration across models and data segments. Explainability should also be developed so that each challenge or restriction can be traced to a compact set of behavioral evidence and a well-defined policy rule.

Finally, the study does not estimate all production costs. Challenge rate, user friction, analyst workload, detection delay, and recovery outcomes are named as operational metrics but are not fully measured on the synthetic benchmark. These are necessary next steps because a Zero Trust control loop must be judged not only by detection quality but also by whether its actions remain proportionate and sustainable in daily operation.

8. Reproducibility and future extension

The study can be reproduced as a sequence of deterministic data transformations followed by repeated model fitting and metric calculation. The essential inputs are the CERT logs, the scenario answer key, the window definition, the feature-generation rules, and the split procedure.

Figure 3 makes the dependency chain explicit. A code-level implementation can export intermediate user-window tables, model scores, and evaluation figures so that changes in features or thresholds can be compared without rewriting the full workflow.

Future extensions can proceed in several directions. A strict time-forward evaluation should be added first to estimate performance under realistic temporal drift. Score calibration should then be measured so that policy thresholds have a more stable interpretation across model families. Explanation mechanisms can be attached to the same feature groups used by the risk model. Additional datasets should be used to test whether the observed 1D-versus-7D pattern persists outside CERT. Finally, real-time activity-level methods such as LAN (Cai et al., 2024) and richer sequence or graph models can be inserted behind the same policy interface to determine whether better detection quality can be obtained without exceeding an acceptable review and challenge budget.

CONCLUSION

This paper presented an applied behavioral risk-scoring framework for insider-threat mitigation in Zero Trust enterprise networks. The framework aggregates heterogeneous identity, endpoint, communication, and network metadata into time-windowed user representations, produces a continuous risk score with supervised baseline models, ranks cases under an explicit analyst review budget, and maps the score to reversible Zero Trust actions. The design keeps detection and authorization conceptually separate: the model supplies evidence and ranking, while the policy layer decides how much additional assurance is required.

The case study demonstrated a clear effect of temporal granularity. Weekly aggregation achieved Recall@10 = 1.0000 and Recall@1000 = 1.0000 in the reported evaluation, whereas the daily representation produced lower and more variable recall at the same review budgets. This finding supports the use of 7D aggregation as an empirically supported operating point for the studied CERT scenarios, but it does not establish a universal rule. The main trade-off is explicit: longer windows can stabilize weak repeated signals but may increase response latency for rapidly developing incidents.

The scientific contribution is therefore primarily methodological and operational rather than algorithmic. The paper formalizes a chain from heterogeneous behavioral evidence to a temporal score, from the score to a ranked review queue, and from the ranked signal to configurable Zero Trust enforcement. This structure makes it possible to evaluate detection quality and policy relevance within the same framework. It also creates a clear insertion point for more advanced detectors, calibration methods, graph or sequence representations, and real-time learning systems without changing the surrounding policy logic.

Before deployment, the framework requires validation on additional and preferably real-world datasets, strict time-forward testing, formal score calibration, explanation support, and quantitative measurement of detection delay, analyst workload, false challenges, and user friction. Privacy and governance controls are also necessary because even metadata-level behavioral monitoring can expose sensitive employee activity patterns. With these extensions, the proposed pipeline can serve as a reproducible intermediate step between an offline insider-threat experiment and a production-oriented Zero Trust control loop.

PUBLICATION STATEMENTS

CONFLICT OF INTEREST: The authors declare no conflict of interest.

FUNDING: This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT: Not applicable. The study is based on a publicly available synthetic cybersecurity dataset and does not involve human participants, animals, or direct collection of personal data.

INFORMED CONSENT STATEMENT: Not applicable. No human participants were recruited or directly studied.

DATA AVAILABILITY STATEMENT: The study uses the CERT Insider Threat Test Dataset, which is publicly available through the dataset record cited as (Lindauer, 2020). No new human-subject dataset was generated.

ACKNOWLEDGEMENTS: The authors thank colleagues and reviewers whose methodological comments helped clarify the operational framing of the study.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES: Generative artificial intelligence was used to support editorial restructuring, language refinement, and formatting of the manuscript. The authors verified the references, technical content, reported metrics, and final manuscript. The model used for editorial assistance is cited as (OpenAI, 2026).

REFERENCES

Ahn, G., Jang, J., Choi, S., & Shin, D. (2024). Research on improving cyber resilience by integrating the zero trust security model with the MITRE ATT&CK matrix. IEEE Access, 12, 89291–89309. https://doi.org/10.1109/ACCESS.2024.3417182.

Al-Shehari, T. A., Rosaci, D., Al-Razgan, M. S., Alfakih, T., Kadrie, M., Afzal, H., & Nawaz, R. (2024). Enhancing insider threat detection in imbalanced cybersecurity settings using the density-based local outlier factor algorithm. IEEE Access, 12, 34820–34834. https://doi.org/10.1109/ACCESS.2024.3373694.

Alketbi, K. S., & Mehmood, A. (2025). A comprehensive survey of explainable artificial intelligence techniques for malicious insider threat detection. IEEE Access, 13, 121772–121798. https://doi.org/10.1109/ACCESS.2025.3587114.

Alsalim, A. S. (2025). Improving the security of enterprise storage systems with efficient anomaly detection. IEEE Reliability Magazine, 2(2), 35–39. https://doi.org/10.1109/MRL.2025.3556806.

Alzaabi, F. R., & Mehmood, A. (2024). A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. IEEE Access, 12, 30907–30927. https://doi.org/10.1109/ACCESS.2024.3369906.

Alohaly, M., Balogun, O., & Takabi, D. (2022). Integrating cyber deception into attribute-based access control (ABAC) for insider threat detection. IEEE Access, 10, 108965–108978. https://doi.org/10.1109/ACCESS.2022.3213645.

Cai, X., Wang, Y., Xu, S., Li, H., Zhang, Y., Liu, Z., & Yuan, X. (2024). LAN: Learning adaptive neighbors for real-time insider threat detection. IEEE Transactions on Information Forensics and Security, 19, 10157–10172. https://doi.org/10.1109/TIFS.2024.3488527.

Dong, J., Wei, J., Hu, X., Dong, Z., Chen, F., Hu, X., & Qi, J. (2025). DDCC: Synergizing denoising diffusion probabilistic models and curriculum-based complexity control for insider threat detection. IEEE Transactions on Industrial Informatics, 21(11), 8351–8361. https://doi.org/10.1109/TII.2025.3574417.

Gonzales, O., Yang, K., & Huang, S. (2025). Contextual sequence-based user behavior anomaly detection. IEEE Access, 13, 35539–35554. https://doi.org/10.1109/ACCESS.2025.3543500.

Greitzer, F. L. (2026). When an insider threat is no longer an insider threat: A look at risk decay. IEEE Engineering Management Review. https://doi.org/10.1109/EMR.2026.3661857.

He, D., Lv, X., Xu, X., Chan, S., & Choo, K.-K. R. (2024). Double-layer detection of internal threat in enterprise systems based on deep learning. IEEE Transactions on Information Forensics and Security, 19, 4741–4751. https://doi.org/10.1109/TIFS.2024.3372771.

Huang, L., & Zhu, Q. (2024). ZETAR: Modeling and computational design of strategic and adaptive compliance policies. IEEE Transactions on Computational Social Systems, 11(3), 4001–4015. https://doi.org/10.1109/TCSS.2023.3323539.

Huang, Z., Li, X., Cao, X., Chen, K., Wang, L., & Liu, L. B.-Y. (2025). IDU-Detector: A synergistic framework for robust masquerader attack detection. IEEE Internet of Things Journal, 12(8), 9653–9670. https://doi.org/10.1109/JIOT.2024.3503057.

Kim, Y., & Kim, Y.-G. (2026). ISFL-AE: Insider-specific feature learning autoencoder for lightweight insider threat detection. IEEE Transactions on Information Forensics and Security, 21, 1638–1653. https://doi.org/10.1109/TIFS.2026.3659000.

Lavanya, P., Anila Glory, H., & Shankar Sriram, V. S. (2024). Mitigating insider threat: A neural network approach for enhanced security. IEEE Access, 12, 73752–73768. https://doi.org/10.1109/ACCESS.2024.3404814.

Le, D. C., & Zincir-Heywood, N. (2021). Anomaly detection for insider threats using unsupervised ensembles. IEEE Transactions on Network and Service Management, 18(2), 1152–1164. https://doi.org/10.1109/TNSM.2021.3071928.

Li, X., Li, X., Jia, J., Li, L., Yuan, J., Gao, Y., & Yu, S. (2023). A high accuracy and adaptive anomaly detection model with dual-domain graph convolutional network for insider threat detection. IEEE Transactions on Information Forensics and Security, 18, 1638–1652. https://doi.org/10.1109/TIFS.2023.3245413.

Li, X., Li, L., Li, X., Cai, B., Jia, J., Gao, Y., & Yu, S. (2024). GMFITD: Graph meta-learning for effective few-shot insider threat detection. IEEE Transactions on Information Forensics and Security, 19, 7161–7175. https://doi.org/10.1109/TIFS.2024.3430106.

Lindauer, B. (2020). Insider threat test dataset [Data set]. Carnegie Mellon University. https://doi.org/10.1184/R1/12841247.v1.

Nasir, R., Afzal, M., Latif, R., & Iqbal, W. (2021). Behavioral based insider threat detection using deep learning. IEEE Access, 9, 143266–143274. https://doi.org/10.1109/ACCESS.2021.3118297.

OpenAI. (2026, September 19). ChatGPT (GPT-5.6 Luna): Editorial assistance for restructuring and language refinement of a cybersecurity manuscript [Large language model]. OpenAI. https://chatgpt.com/.

Peccatiello, R. B., Gondim, J. J. C., & Garcia, L. P. F. (2023). Applying one-class algorithms for data stream-based insider threat detection. IEEE Access, 11, 70560–70573. https://doi.org/10.1109/ACCESS.2023.3293825.

Rahman, M. M. H., Naeem, M. A. A., & Abubakar, A. (2022). Threats from unintentional insiders: An assessment of an organization's readiness using machine learning. IEEE Access, 10, 110294–110308. https://doi.org/10.1109/ACCESS.2022.3214819.

Roy, K. C., & Chen, G. (2024). GraphCH: A deep framework for assessing cyber-human aspects in insider threat detection. IEEE Transactions on Dependable and Secure Computing, 21(5), 4495–4509. https://doi.org/10.1109/TDSC.2024.3353929.

Wang, Z. Q., & Saddik, A. E. (2023). DTITD: An intelligent insider threat detection framework based on digital twin and self-attention based deep learning models. IEEE Access, 11, 114013–114030. https://doi.org/10.1109/ACCESS.2023.3324371.

Whitelaw, F., Riley, J., & Elmrabit, N. (2024). A review of the insider threat, a practitioner perspective within the U.K. financial services. IEEE Access, 12, 34752–34768. https://doi.org/10.1109/ACCESS.2024.3373265.

Xiao, F., Chen, S., Chen, S., Ma, Y., He, H., & Yang, J. (2025). SENTINEL: Insider threat detection based on multi-timescale user behavior interaction graph learning. IEEE Transactions on Network Science and Engineering, 12(2), 774–790. https://doi.org/10.1109/TNSE.2024.3519155.

Xiao, J., Yang, L., Zhong, F., Wang, X., Chen, H., & Li, D. (2023). Robust anomaly-based insider threat detection using graph neural network. IEEE Transactions on Network and Service Management, 20(3), 3717–3733. https://doi.org/10.1109/TNSM.2022.3222635.

Zhu, X., Dong, J., Qi, J., Zhou, Z., Dong, Z., Sun, Y., & Wang, M. (2024). AUTH: An adversarial autoencoder based unsupervised insider threat detection scheme for multisource logs. IEEE Transactions on Industrial Informatics, 20(9), 10954–10965. https://doi.org/10.1109/TII.2024.3393491.

Авторлар туралы мәліметтер
Информация об авторах
Information about authors



Бодамбаев Тамерлан Кайратулы – қауіпсіз бағдарламалық қамтамасыз етуді әзірлеу магистрі, Astana IT University, Астана қ., Қазақстан

Бодамбаев Тамерлан Кайратулы – магистр разработки безопасного программного обеспечения, Astana IT University, г. Астана, Казахстан

Bodambayev Tamerlan Kairatuly– Master of Secure Software Engineering, Astana IT University, Astana, Kazakhstan,
e-mail: 255210@astanait.edu.kz,
ORCID: https://orcid.org/0009-0006-6331-5810



Аманжолова Сауле Токсановна– ғылым кандидаты, Astana IT University, Астана қ., Қазақстан

Аманжолова Сауле Токсановна – кандидат наук, Люблинский Технологический Университет Astana IT University, г. Астана, Казахстан

Amanzholova Saule – Candidate of science, Astana IT University, Astana, Kazakhstan,
e-mail: s.amanzholova@astanait.edu.kz,
ORCID: https://orcid.org/0000-0002-6779-9393