Action-to-Authorization Ratio

A Continuous Contextual Metric for Detecting Human Insider Threats

Stephen M. Layne
Red Vector, Inc.
steve@redvector.ai

September 14, 2026

***Abstract*—Human insider threats remain the most persistent and costly security challenge facing enterprise organizations and government agencies. Despite decades of investment in Security Information and Event Management (SIEM) systems, Data Loss Prevention (DLP) tools, and User and Entity Behavior Analytics (UEBA) platforms, detection rates for insider threats remain low while false positive rates remain prohibitively high. This paper introduces the Action-to-Authorization Ratio (AAR), a continuous contextual metric that quantifies the relationship between the actions a human user attempts and the actions for which the user holds explicit authorization within a defined temporal window. Unlike binary access control enforcement or signature-based detection, AAR provides a continuously computable, role-normalized measure of contextual alignment that captures the gradual scope expansion characteristic of insider threat kill chains. We present the formal definition, mathematical properties, and operational characteristics of AAR, together with a formalization of Temporal Access Pattern Deviation (TAPD), a complementary metric that captures anomalies in when and how users access resources relative to their established contextual baselines. We describe an integrated detection framework in which AAR and TAPD combine with organizational context to produce a dynamic risk signal rather than a static point-in-time score, evaluate the framework against five incumbent detection approaches across five dimensions of detection capability, and illustrate its operation through three constructed human insider scenarios. The framework is designed to detect the contextual precursors of data exfiltration, privilege abuse, and unauthorized access that precede material harm, shifting the detection window from post-incident forensics to pre-incident intervention. In decision-cycle terms, it closes a continuous observe–orient–decide–act loop against an insider progression that unfolds over weeks, keeping the defender's decision tempo ahead of the insider's progression tempo.**

Keywords*—*Insider threat, insider risk, contextual analytics, action-to-authorization ratio, contextual drift, risk signals, data exfiltration, privilege abuse, user and entity behavior analytics (UEBA), continuous monitoring

I. Introduction

The insider threat problem is as old as the concept of organizational trust. Employees, contractors, and trusted partners operate within the security perimeter by design, possessing legitimate credentials and authorized access to the systems and data necessary to perform their roles [1], [2]. When these trusted insiders act against organizational interests, whether through malicious intent, negligence, or coercion, the resulting damage is disproportionately severe. The 2026 Ponemon Institute Cost of Insider Risks Global Report found that the average annualized cost of insider risk reached $19.5 million per organization [3]. Although the mean time to contain an insider incident has improved to 67 days, incidents that persist beyond 90 days drive annualized costs above $21 million, underscoring the disproportionate value of earlier detection [3].

The difficulty of insider threat detection stems from a fundamental asymmetry: insiders operate with legitimate access, making their malicious activity indistinguishable from normal work when examined through the lens of traditional access control. A financial analyst who routinely accesses customer account data as part of their job function generates the same access log entries whether performing legitimate analysis or staging data for exfiltration. The discriminating signal is not the action itself but the pattern, context, and scope of the action relative to the individual's established contextual norms.

Current detection approaches have evolved through three generations [4]. First-generation systems relied on rule-based audit log analysis, flagging specific prohibited actions such as copying files to removable media or accessing systems outside business hours. Second-generation approaches introduced statistical anomaly detection through UEBA platforms, building behavioral baselines from historical activity and alerting on deviations. Third-generation approaches incorporated machine learning to identify complex, multi-dimensional patterns of user activity [5], [6]. Despite this evolution, detection remains fundamentally reactive, focused on identifying known threat indicators rather than measuring the continuous contextual state of each user. Throughout this paper we use the term contextual analytics for this measurement discipline, historically termed behavioral analytics in the UEBA literature; the shift in terminology reflects the central role that organizational, human, and technical context plays in the framework we propose.

The operating environment has also shifted in ways that stress indicator-based detection further. Hybrid work has dissolved many of the temporal and network regularities on

which detection baselines have historically depended. Sanctioned generative AI platforms have created high-bandwidth, business-approved channels through which sensitive data can leave organizational control, often at the hands of well-intentioned human users [3], [7]. Agentic AI systems are simultaneously expanding the population of trusted identities beyond the human workforce; this paper deliberately confines its scope to the human insider, and we discuss extension to non-human identities as future work in Section X. These shifts do not change the fundamental character of the human insider problem, but they widen the gap between what indicator-based systems were built to detect and how insider activity actually manifests.

This paper proposes a paradigm shift from indicator-based detection to continuous contextual measurement through the introduction of the Action-to-Authorization Ratio (AAR). AAR provides a quantifiable, continuously computable metric that captures the alignment between a human user's observed activity and their authorized operational scope. The contributions are fourfold. First, we give the formal definition of AAR together with scope-weighted and role-normalized variants (Section III). Second, we formalize Temporal Access Pattern Deviation (TAPD), a complementary metric quantifying divergence in when and how users access resources (Section IV). Third, we describe an integrated framework in which AAR and TAPD combine with organizational context to produce a composite, trajectory-aware risk signal mapped to insider threat kill chain phases and graduated response tiers (Section V). Fourth, we evaluate the framework against five incumbent detection approaches across five dimensions of detection capability (Section VI) and illustrate its operation through three constructed human insider scenarios (Section VII). Sections VIII and IX address implementation and governance considerations; Section X states limitations and future work.

II. The Insider Threat Kill Chain

Understanding how insider threats progress is essential to designing effective detection metrics. Research by the Carnegie Mellon Software Engineering Institute (SEI) and the CERT Insider Threat Center has established that insider threats do not emerge instantaneously but develop through a recognizable progression of observable conditions [8]. This progression is consistent with the critical-pathway model of insider risk, in which personal predispositions and stressors precede observable concerning behaviors, which in turn precede the hostile act [9]. We term this progression the insider threat kill chain; it typically follows five phases.

A. Grievance or Motivation Formation

The initiating phase involves the development of a motivating factor: personal financial distress, workplace grievance, ideological motivation, or external coercion [10]. While this phase is primarily psychological and largely invisible to technical monitoring, it establishes the preconditions for subsequent changes in observable activity.

B. Reconnaissance and Boundary Testing

The insider begins exploring the boundaries of their access, often subtly. This may manifest as accessing systems or data repositories outside their normal work scope, querying databases with unusual search parameters, or exploring administrative interfaces they have not previously accessed. Critically, these actions may remain entirely within the user's authorized permissions while representing a significant departure from their established contextual baseline.

C. Privilege Accumulation

The insider may seek to expand their access, requesting additional permissions, retaining access to systems associated with previous roles, or discovering and exploiting overly permissive access controls. In organizations with weak access governance, this phase can be remarkably successful, as access review processes are often perfunctory or infrequent.

D. Data Staging and Collection

The insider begins systematically collecting or staging the target data. This phase often involves increased volume and velocity of data access, access to data outside the user's immediate project scope, and patterns of bulk download or aggregation that differ from legitimate analytical workflows.

E. Exfiltration or Destructive Action

The terminal phase involves the actual removal of data from organizational control or the execution of a destructive action. By the time traditional detection systems identify activity in this phase, material harm has typically already occurred.

The critical insight from this progression is that each phase from B through D is characterized by measurable expansion of operational scope: the user's actions progressively diverge from their established contextual baseline, even when those actions remain technically authorized. AAR is designed to capture precisely this progressive divergence. The progression also has a tempo: phases B through D typically unfold over days to weeks [8]. Any detection approach that observes at discrete intervals therefore concedes entire phases to the insider between observations—a structural disadvantage that Section V-E formalizes in decision-cycle terms.

III. The Action-to-Authorization Ratio

A. Formal Definition

We define the Action-to-Authorization Ratio (AAR) as the ratio of discrete actions attempted by a user to the actions for which the user holds explicit, current authorization within a defined temporal window. Formally,

$$AAR(u, t) = A_{total}(u, t) / A_{auth}(u, t) \quad (1)$$

where u identifies the user, A_{total}(u, t) denotes the total count of discrete actions attempted by user u during

temporal window t, and A_{auth}(u, t) denotes the count of those actions covered by a valid authorization policy at the time of execution. An AAR of 1.0 indicates perfect contextual alignment with the user's authorized scope. Values exceeding 1.0 indicate the user is attempting unauthorized actions. Values below 1.0 but trending upward within authorized scope indicate increasing utilization of permissions, a subtler but equally important signal.

B. Extended AAR: Scope-Weighted Formulation

The basic AAR formulation treats all actions equally. In practice, not all unauthorized actions carry equal risk. Accessing a colleague's shared drive differs materially from querying a classified database. We therefore introduce a scope-weighted variant,

$$AAR_w(u, t) = \Sigma_i w_i a_i / \Sigma_i w_i a_i^{auth} \quad (2)$$

where w_i is a sensitivity weight assigned to each resource or action class based on data classification, regulatory status, or organizational criticality, a_i is the count of attempted actions in class i, and a_i^{auth} is the count of authorized actions in class i. This weighted formulation ensures that AAR inflation driven by access to high-sensitivity resources produces a stronger signal than equivalent inflation from low-sensitivity access.

C. Role-Normalized Baseline

A distinguishing feature of AAR for human insider threat detection is role normalization. Human users occupy organizational roles that create expected patterns of activity [11]. A software engineer's normal AAR profile differs fundamentally from a financial analyst's. The framework establishes per-role baseline distributions during a supervised learning period, enabling the detection of anomalies that are statistically significant relative to role-appropriate activity, not merely relative to global population norms.

Role normalization addresses a critical weakness of many UEBA implementations: the tendency to generate false positives for users whose legitimate work patterns are inherently different from the organizational mean. A security operations center analyst who legitimately accesses a wide range of systems should not be penalized for having a higher baseline AAR than a human resources coordinator. By establishing role-specific baselines, the framework detects meaningful deviation within the context of each user's legitimate function.

D. Key Properties

Continuous Measurement. AAR is computed continuously across sliding temporal windows rather than evaluated at discrete checkpoints. This continuity enables detection of gradual scope expansion that unfolds over days or weeks, the characteristic pattern of sophisticated insider threats that deliberately avoid triggering point-in-time threshold alerts.

Pre-Incident Sensitivity. Because AAR measures operational scope rather than specific malicious indicators, it is sensitive to the reconnaissance and boundary-testing phases of the insider threat kill chain, phases that precede material harm. A user exploring the boundaries of their access will exhibit AAR inflation before any data exfiltration occurs.

Policy Independence. AAR does not require enumeration of specific prohibited behaviors. Organizations need not anticipate every possible insider threat scenario. The metric detects deviation from authorized scope regardless of the specific form that deviation takes, providing coverage against novel attack patterns.

Graduated Signal. The continuous nature of AAR provides a naturally graduated signal that supports proportionate response rather than binary alerting; Section V-D develops this graduation into explicit response tiers.

IV. Temporal Access Pattern Deviation

A. Definition and Rationale

Human work activity follows temporal patterns shaped by organizational schedules, personal habits, project cycles, and collaboration rhythms. These patterns, while more variable than those of automated systems, are nevertheless statistically characterizable for individual users. We define Temporal Access Pattern Deviation (TAPD) as a composite metric quantifying the divergence between a user's observed temporal-sequential activity and their established contextual baseline across three dimensions. Formally,

$$TAPD(u, t) = \Sigma_d \beta_d \Delta(P_{u,t}(d), B_u(d)) \quad (3)$$

where d indexes the three dimensions described below, $P_{u,t}$(d) denotes the observed pattern of user u in dimension d during temporal window t, B_u(d) denotes the user's established baseline in that dimension, Δ is a statistical divergence measure such as the Jensen–Shannon divergence, and β_d are nonnegative dimension weights that sum to one.

B. Three Dimensions of Temporal Activity

Temporal Distribution. The distribution of a user's activity across hours, days, and weeks. A user who typically works from 8 AM to 6 PM on weekdays and suddenly begins accessing sensitive systems at 2 AM on weekends presents a clear temporal anomaly. The framework models this as a probability distribution over time slots and measures deviation using statistical divergence metrics.

Access Sequencing. The order in which a user accesses resources and systems. Work tasks create characteristic access sequences: a financial analyst may predictably access the ERP system, then the reporting platform, then the data warehouse in the course of monthly close activities. A disrupted sequence, particularly one that introduces previously unaccessed systems into the workflow, may indicate reconnaissance or data staging.

Velocity and Cadence. The rate and rhythm of user actions. Insiders engaged in data staging often exhibit increased action velocity as they systematically collect target data, creating cadence anomalies relative to their normal work patterns. Conversely, an insider exercising unusual caution may exhibit artificially reduced velocity on sensitive systems, another detectable deviation.

C. Human-Specific Baseline Considerations

Baseline construction for human users must account for legitimate variability that does not exist in automated systems. The framework incorporates several mechanisms: seasonal and cyclical adjustments for predictable work pattern changes such as quarter-end financial activities or annual review cycles; project-aware windowing that adjusts baseline expectations when a user's project assignment changes; and decay-weighted learning that emphasizes recent activity while retaining long-term pattern memory to detect slow-developing threats.

V. Integrated Detection Framework

A. Complementary Signal Design

AAR and TAPD are designed as orthogonal detection signals that together provide coverage across the full spectrum of insider threat activity. AAR captures what the user is doing relative to their authorization, measuring contextual scope. TAPD captures how and when the user is operating relative to their established patterns, measuring contextual consistency. The combination addresses a critical limitation of single-signal detection: an insider who carefully operates within authorized scope to avoid AAR inflation will likely alter their temporal patterns, and an insider who maintains temporal consistency while expanding their operational scope will inflate their AAR.

B. Composite Risk Signal

The framework combines AAR and TAPD with contextual factors to produce a composite insider risk signal,

$$R(u, t) = w_A A^*(u, t) + w_T T^*(u, t) + w_C C(u, t) \quad (4)$$

where $A^*(u, t)$ denotes the role-normalized, scope-weighted AAR mapped to the unit interval, $T^*(u, t)$ the normalized TAPD, $C(u, t)$ a composite of contextual factors, and w_A, w_T, w_C nonnegative weights that sum to one. Contextual factors include the sensitivity classification of accessed resources, the user's proximity to known risk indicators such as announced resignation or performance issues, organizational events such as announced layoffs or restructuring, and peer comparison within the user's role group [12]. Structured taxonomies of sociotechnical and organizational indicators provide a principled basis for constructing $C(u, t)$ [13].

Unlike a static score evaluated at discrete intervals, the composite signal is a dynamic quantity that carries magnitude, direction, and velocity: the framework reports both $R(u, t)$ and its discrete-time trajectory $\Delta R/\Delta t$, expressing not only the current level of risk but the direction and rate along which risk conditions are moving. The composite signal is designed to be operationally actionable within existing security operations workflows, feeding directly into case management systems and investigation queues.

C. Kill Chain Phase Mapping

A distinctive capability of the integrated framework is its ability to map detected anomalies to phases of the insider threat kill chain. Reconnaissance and boundary testing manifests as mild AAR inflation combined with TAPD sequencing anomalies, as the user explores unfamiliar systems. Privilege accumulation manifests as an expanding AAR baseline, as newly acquired permissions are exercised. Data staging manifests as both AAR inflation on data repositories and TAPD velocity anomalies, as the user's data access volume and cadence increase. This phase mapping enables security analysts to assess not only that anomalous activity is occurring but where the insider is in the threat progression, informing the urgency and nature of the response.

D. Graduated Response

The continuous nature of the composite signal supports proportionate, graduated response rather than binary alerting, as summarized in Table II. Minor signal elevation warrants passive monitoring and trend logging, and can often be resolved through non-punitive access governance review. Moderate elevation triggers enhanced monitoring or automated workflow actions. Significant elevation warrants active investigation with protective controls. This graduation directly addresses alert fatigue, one of the most significant operational challenges in security operations centers, and ensures that the earliest organizational responses are corrective rather than accusatory.

TABLE II

Graduated Signal Levels and Proportionate Response

Signal level	Characteristic pattern	Proportionate response
Baseline	R(u, t) within role-normalized norms	Passive trend logging
Minor elevation	Sustained low-level AAR inflation	Passive monitoring; access governance review
Moderate elevation	AAR inflation with TAPD corroboration	Enhanced monitoring; automated workflow triggers
Significant elevation	Multi-signal convergence with rising trajectory	Active investigation; protective controls

E. The Framework as a Continuous Defender Decision Cycle

The integrated framework can be read as an instantiation of Boyd's observe–orient–decide–act (OODA) loop [14], [15], run continuously against an insider progression that unfolds over days to weeks. Continuous computation of AAR and TAPD is the observe step: unlike periodic review, the observation window never closes. Kill chain phase mapping (Section V-C) is orientation, the step Boyd identified as the one on which every subsequent decision depends [15]: it converts raw deviation into an assessment of where the insider stands in the progression. The graduated tiers of Table II are the decide step made ahead of time—proportionate responses designed once, deliberately, and encoded as playbooks, so that decisions need not be improvised under time pressure alert by alert. Proportionate response is the act step, and its effects feed the next observation cycle.

This reading makes the comparative claim of Section VI precise. A quarterly access review is a defender decision loop that closes four times a year against an insider who can complete reconnaissance, privilege accumulation, and staging inside a single review interval. A SIEM correlation rule closes its loop only when a pre-imagined indicator fires. The proposed framework closes the full loop continuously, so that the defender's decision cycle is always shorter than the insider's progression. Pre-incident detection is, in this reading, a tempo property: the defender need not predict intent, only complete observation-to-action cycles faster than the insider completes kill chain phases. The deterrence effect discussed in Section IX-B follows from the same structure: continuous cycling leaves no predictable interval in which the defender's loop stands open.

VI. Comparative Analysis

We evaluate the AAR/TAPD framework against five incumbent detection approaches across five dimensions of detection capability: (D1) pre-incident phase coverage, the ability to detect reconnaissance, boundary testing, and privilege accumulation before material harm occurs; (D2) continuity of measurement, whether the approach measures continuously or at discrete points in time; (D3) coverage of standard-privilege users, the majority population from which most insider incidents originate; (D4) signal interpretability, whether an analyst can readily understand what change in activity produced a detection; and (D5) false positive control, the degree to which the approach normalizes for legitimate individual and role variation. Table I summarizes the assessment; the following subsections discuss each incumbent approach.

TABLE I
Detection Capability Across Five Dimensions

Approach	D1	D2	D3	D4	D5
SIEM	○	◐	◐	◐	○
DLP	○	◐	●	●	◐
UEBA	◐	●	●	○	◐
PAM	◐	●	○	●	◐
Static access reviews	○	○	●	●	◐
AAR/TAPD framework	●	●	●	●	●

● full capability; ◐ partial capability; ○ minimal or absent. The assessment of the proposed framework reflects design intent; empirical validation remains future work (Section X).

A. SIEM-Based Detection

SIEM platforms aggregate log data and correlate events against predefined detection rules. Their primary value in insider threat detection lies in retrospective forensic analysis rather than proactive detection. SIEM rules must be authored by analysts who anticipate specific threat behaviors, creating a structural blind spot for novel or evolving insider tactics. A SIEM rule that flags bulk file downloads may detect unsophisticated data exfiltration but will miss an insider who systematically accesses individual records over weeks, a pattern that AAR's continuous measurement is designed to capture. Additionally, SIEM alert volumes in large enterprises often exceed the capacity of security operations teams to investigate, contributing to the high mean time to detection that characterizes insider threat incidents.

B. Data Loss Prevention

DLP systems monitor data movement and enforce policies on data transfer, email attachments, cloud uploads, and removable media. While effective at blocking known exfiltration channels, DLP operates at the point of exfiltration, the terminal phase of the insider threat kill chain. By the time DLP triggers, the insider has already completed reconnaissance, accumulated privilege, and staged data. Furthermore, DLP is inherently channel-specific: it monitors the exfiltration vector rather than the pattern of activity leading to exfiltration. An insider who identifies an unmonitored channel, or who leverages their legitimate access to achieve exfiltration through a business-approved path, can circumvent DLP entirely. The proliferation of sanctioned generative AI platforms has compounded this limitation: sensitive material submitted to public AI services traverses business-approved channels that channel-specific DLP policies struggle to cover [3], [7]. AAR detects the upstream contextual pattern regardless of the downstream exfiltration method.

C. Traditional UEBA

UEBA platforms represent the closest incumbent technology to the proposed framework. However, most commercial UEBA implementations suffer from three

limitations in the insider threat context. First, they rely heavily on peer group comparison, which generates false positives for users whose legitimate roles require atypical access patterns. Second, their behavioral models are typically trained on the full user population, creating baselines that are too broad to capture the subtle, individualized deviations that characterize sophisticated insider threats. Third, they are often configured as scoring engines that produce opaque, static, point-in-time scores, making it difficult for analysts to understand what specific change in activity triggered an alert. AAR's explicit, interpretable formulation, the ratio of attempted to authorized actions, provides analysts with an immediately comprehensible signal that accelerates investigation triage.

D. Privileged Access Management

PAM solutions monitor and control privileged account usage. While essential for managing administrative access, PAM addresses a subset of the insider threat population: users with elevated technical privileges. The majority of insider threat incidents involve standard business users abusing their routine access rather than administrators exploiting technical privileges. An accounts payable specialist who manipulates vendor records, a research scientist who copies proprietary formulations, or a customer service representative who harvests personal information are all standard-privilege insider threats that fall outside PAM's monitoring scope. AAR applies uniformly across all privilege levels.

E. Static Access Reviews

Periodic access reviews and certification campaigns are a compliance staple but provide no continuous monitoring capability. Conducted quarterly or annually, they capture a point-in-time snapshot of access entitlements without visibility into how those entitlements are actually being used. The gap between an access review cycle and the speed at which insider threats develop, often a matter of weeks, renders this approach ineffective as a detection mechanism. AAR provides the continuous contextual visibility that static access reviews structurally cannot.

VII. Illustrative Detection Scenarios

The following constructed scenarios illustrate the framework's operation across representative human insider threat patterns documented in the CERT insider threat corpus [8]. They are illustrative compositions rather than empirical case studies; empirical validation is addressed in Section X.

A. Data Staging by a Departing Analyst

A financial analyst submits notice of resignation, an organizational event that raises the contextual term $C(u, t)$. Over the following two weeks, the analyst's scope-weighted AAR inflates as queries extend to customer records outside the assigned portfolio, while TAPD registers velocity and cadence anomalies consistent with systematic aggregation. The composite signal crosses the moderate tier with a rising trajectory, triggering enhanced monitoring; convergence of all three terms elevates it to the significant tier, and the organization intervenes through a coordinated security, human resources, and legal review before any boundary-crossing transfer occurs. Channel-specific DLP never triggers, because no monitored exfiltration channel is ever used.

B. Privilege Accumulation Following Reorganization

A software engineer retains legacy entitlements after a reorganization. Months later, dormant permissions begin to be exercised: administrative interfaces are accessed for the first time, and repository queries extend beyond the current project scope. AAR remains below 1.0, but utilization of the authorization envelope expands steadily relative to the role-normalized baseline, and TAPD registers sequencing anomalies as unfamiliar systems enter the workflow. The signal reaches the minor tier, prompting a non-punitive access governance review that revokes the legacy entitlements and tightens the role definition. The risk is resolved without a formal investigation, illustrating both graduated response and the framework's value to identity governance.

C. Fraud Within Authorized Scope

An accounts payable specialist begins modifying vendor banking records to redirect payments. Every action is authorized, and AAR remains near 1.0. TAPD, however, registers sequence and cadence anomalies: vendor-master edits occur outside the monthly close cadence, after hours, and in a sequence that bypasses the reporting platform ordinarily consulted first. Sensitivity weighting on vendor-master data amplifies the contribution, and the composite signal reaches the significant tier. Active investigation confirms fraudulent redirection. The scenario illustrates the orthogonality of the two metrics: an insider who operates entirely within authorized scope still alters temporal and sequential patterns.

VIII. Implementation Considerations

A. Data Source Integration

Computing AAR for human users requires integration with multiple enterprise data sources. Identity and Access Management (IAM) systems provide the authorization baseline, defining each user's current role assignments and permission sets. Application audit logs, endpoint telemetry, network flow data, and cloud platform audit trails provide the action stream. The framework consumes these as normalized event streams, requiring a common schema that maps each action to an entity identifier, a target resource, a timestamp, and an authorization evaluation result.

B. Privacy and Proportionality

Continuous contextual monitoring of human users raises significant privacy considerations that must be addressed

through policy, architecture, and governance. The framework is designed to operate on metadata, specifically action-resource-authorization tuples, rather than content. It does not require inspection of email bodies, document contents, or communication text. Organizations deploying the framework must establish clear policies governing data retention, access to contextual analytics by HR and management, and the thresholds at which monitoring transitions from passive observation to active investigation. The graduated nature of AAR signaling supports proportionate response: low-level AAR inflation can be addressed through process review and access governance improvements without initiating a formal investigation.

C. Baseline Establishment and Calibration

We recommend a minimum baseline establishment period of 30 days of active user activity before AAR alerting is enabled for each monitored user. During this period, the framework constructs the per-user and per-role contextual distributions that define normal operations. New employees, role changers, and users returning from extended absence require baseline recalibration, which the framework handles through adaptive windowing. The calibration period must account for organizational cyclicality: a baseline established entirely during a quiet operational period will generate false positives during peak activity periods unless the framework incorporates cyclical awareness.

D. Regulatory and Compliance Alignment

Multiple regulatory frameworks require or recommend continuous monitoring of user activity for insider threat detection. The U.S. National Insider Threat Policy (Executive Order 13587) and the National Insider Threat Task Force minimum standards [16], NIST SP 800-53 controls for personnel security and audit [17], CISA's insider threat mitigation guidance [18], CMMC continuous monitoring requirements, and GDPR's mandate for appropriate technical measures to protect personal data all create regulatory drivers for continuous monitoring capabilities. The AAR framework provides a demonstrable, measurable control that satisfies continuous monitoring requirements while maintaining the metadata-only analysis approach that supports privacy compliance.

E. Operational Workflow Integration

The value of contextual detection is realized only through effective operational response. The framework is designed to integrate with existing Security Operations Center (SOC) workflows and case management platforms. AAR and TAPD signals feed into investigation queues with contextual enrichment: the analyst receives not only the alert but the user's AAR trend, the specific actions contributing to the inflation, the kill chain phase mapping, and relevant contextual factors. This enrichment dramatically reduces the time from alert to informed investigative decision, addressing the triage bottleneck that undermines many UEBA deployments.

IX. Organizational and Governance Implications

A. Bridging Security and Human Resources

Human insider threats exist at the intersection of information security and human capital management [19]. AAR-based detection has implications beyond the SOC. Persistent low-level AAR inflation across multiple users within a department may indicate systemic access governance failures, overly broad permissions, or inadequate role definitions rather than individual malicious intent. The framework can inform identity governance programs by identifying roles with consistently high or variable AAR baselines, signaling opportunities to tighten access through least-privilege refinement. Because elevated signals may ultimately implicate employment decisions, we recommend that moderate- and significant-tier responses be reviewed by a multi-disciplinary governance function spanning security, human resources, legal, and privacy, ensuring that responses remain proportionate and legally defensible.

B. Deterrence Through Awareness

Research on insider threat deterrence demonstrates that the perceived probability of detection is a more effective deterrent than the severity of consequences [10]. Organizations that communicate the existence of continuous contextual monitoring, without revealing specific detection mechanisms, establish a deterrent effect that may prevent some insider threats from progressing beyond the ideation phase. AAR's continuous nature supports this deterrence posture: unlike periodic audits, which create predictable windows of reduced monitoring, continuous contextual measurement leaves no exploitation window.

X. Limitations and Future Work

Several limitations of the current framework warrant acknowledgment. First, AAR is dependent on the quality and completeness of authorization policy data. Organizations with poorly maintained IAM systems, stale role definitions, or extensive exceptions will experience degraded signal quality. Second, the framework is least sensitive to insiders who operate entirely within their authorized scope using only their normal patterns of activity, such as a user who has legitimate access to the exact data they intend to exfiltrate and does so through their normal work processes at their normal cadence; the fraud scenario in Section VII illustrates partial, but not complete, mitigation of this case. Third, baseline construction for new users with limited activity history requires interpolation from role-group baselines, which may be less accurate than established per-user baselines. Fourth, the scenarios presented in Section VII are constructed illustrations; the framework has not yet been validated against empirical incident data.

Future work will address several extensions. We intend to develop empirical validation using anonymized enterprise telemetry across multiple industry verticals. We will explore

the integration of AAR with organizational risk indicators, including human resources data, to test whether models that fuse activity telemetry with organizational and human context outperform telemetry-only approaches. We will investigate the application of AAR to contractor and third-party populations, where baselines are shorter and access patterns less predictable. Although this paper has deliberately confined its scope to the human insider, we will extend the framework to AI agents, service accounts, and other non-human identities, whose better-defined authorization envelopes suggest that AAR will transfer directly, with TAPD baselines adapted to machine-speed cadences. There the decision-cycle asymmetry of Section V-E inverts: an agent's observe–orient–decide–act loop closes at machine tempo, faster than any human defender's, and the defender must out-position rather than out-pace it—a progression model for that setting is developed in companion work [20]. Finally, we intend to explore adversarial robustness, specifically whether insiders who are aware of AAR-based monitoring can systematically evade detection and what countermeasures are effective.

XI. Conclusion

The human insider threat remains one of the most consequential and difficult-to-detect security challenges facing modern organizations. The inherent tension between granting trusted users the access they need to perform their work and monitoring for the abuse of that access has not been resolved by decades of investment in perimeter-focused and signature-based security technologies.

The Action-to-Authorization Ratio offers a fundamentally different approach: continuous, quantifiable measurement of contextual alignment between a user's actions and their authorized scope. By shifting from retrospective indicator matching to continuous contextual measurement, AAR detects the progressive scope expansion that characterizes the insider threat kill chain in its earlier, pre-harm phases.

When integrated with Temporal Access Pattern Deviation in a comprehensive detection framework, AAR enables security organizations to detect insider threats earlier, investigate them more efficiently, and respond to them more proportionately. The framework's role-normalized baselines reduce false positives, its graduated signaling mitigates analyst fatigue, its interpretable metrics accelerate investigative triage, and its evaluation across five dimensions of detection capability demonstrates coverage that incumbent approaches structurally cannot provide. Read as a decision cycle, the framework's contribution is tempo: it closes the defender's observe–orient–decide–act loop continuously, establishing the structural condition—a defender cycle shorter than the insider's progression—on which pre-incident intervention depends.

As organizations navigate an environment of expanding digital footprints, remote workforces, proliferating AI systems, and increasing regulatory scrutiny, the need for effective, privacy-conscious human insider threat detection will only intensify. The AAR framework provides a principled, practical, and measurable foundation for meeting this challenge.

Declarations

Conflict of interest: The author is the founder and Chief Executive Officer of Red Vector, Inc., which develops commercial insider-risk and AI-agent governance technology, including implementations of the Action-to-Authorization Ratio and Temporal Access Pattern Deviation described in this paper. The author therefore has a financial interest in the adoption of the constructs presented here. This disclosure is made in the interest of transparency; the framework and its evaluation are presented on their technical merits, and the reader is invited to assess them accordingly.

Funding: This work received no external funding.

Data availability: The detection scenarios in Section VII are constructed illustrations; no proprietary, customer, or personal data were used, and no dataset is associated with this manuscript.

Ethics: No human or animal subjects were involved; institutional review was not applicable.

References

[1] M. Bishop and C. Gates, "Defining the insider threat," in Proc. 4th Annu. Workshop on Cyber Security and Information Intelligence Research, 2008.

[2] J. Hunker and C. W. Probst, "Insiders and insider threats: An overview of definitions and mitigation techniques," J. Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, vol. 2, no. 1, pp. 4–27, 2011.

[3] Ponemon Institute, "2026 Cost of Insider Risks: Global Report," Ponemon Institute/DTEX Systems, 2026.

[4] M. B. Salem, S. Hershkop, and S. J. Stolfo, "A survey of insider attack detection research," in Insider Attack and Cyber Security. Springer, 2008, pp. 69–90.

[5] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," in Proc. AAAI Workshop on Artificial Intelligence for Cyber Security, 2017.

[6] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici, and M. Ochoa, "Insight into insiders and IT: A survey of insider threat taxonomies, analysis, modeling, and countermeasures," ACM Comput. Surv., vol. 52, no. 2, pp. 1–40, Apr. 2019.

[7] Thales, "2026 Data Threat Report," Thales Group, 2026.

[8] D. Cappelli, A. Moore, and R. Trzeciak, The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes, 2nd ed. Boston, MA, USA: Addison-Wesley, 2012.

[9] E. Shaw and L. Sellers, "Application of the critical-path method to evaluate insider risks," Studies in Intelligence, vol. 59, no. 2, pp. 1–8, Jun. 2015.

[10] E. D. Shaw, K. G. Ruby, and J. M. Post, "The insider threat to information systems: The psychology of the dangerous insider," Security Awareness Bulletin, vol. 2, no. 98, pp. 1–10, 1998.

[11] R. Sandhu, E. Coyne, H. Feinstein, and C. Youman, "Role-based access control models," IEEE Computer, vol. 29, no. 2, pp. 38–47, Feb. 1996.

[12] F. L. Greitzer, D. A. Frincke, and M. Zabriskie, “Social/behavioral modeling for insider threat detection,” in Proc. IEEE Symp. Computational Intelligence in Cyber Security, 2009.

[13] F. L. Greitzer, J. Purl, Y. M. Leong, and D. E. Becker, “SOFIT: Sociotechnical and organizational factors for insider threat,” in Proc. IEEE Security and Privacy Workshops (SPW), 2018, pp. 197–206.

[14] J. R. Boyd, “The essence of winning and losing,” unpublished briefing, Jun. 1996.

[15] F. P. B. Osinga, Science, Strategy and War: The Strategy of John Boyd. Abingdon, U.K.: Routledge, 2007.

[16] National Insider Threat Task Force, “Insider Threat Guide: A Compendium of Best Practices to Accompany the National Insider Threat Minimum Standards,” Office of the Director of National Intelligence, 2017.

[17] NIST, “Security and Privacy Controls for Information Systems and Organizations,” NIST Special Publication 800-53, Rev. 5, 2020.

[18] Cybersecurity and Infrastructure Security Agency, “Insider Threat Mitigation Guide,” CISA, Nov. 2020.

[19] A. P. Moore, D. L. McIntire, and D. Mundie, “Building an Insider Threat Program: A Step-by-Step Guide for Building an Insider Threat Mitigation Program,” Software Engineering Institute, Carnegie Mellon University, Technical Note CMU/SEI-2014-TN-012, 2014.

[20] S. M. Layne, “The agentic kill chain: A condition-driven model for governing autonomous AI agents in the enterprise,” preprint, 2026.