Behavior Before Identity

CoreBond Research Series — CBRS-006
August 2026
Todd Kirton
DOI: 10.5281/zenodo.22224598

Abstract

Security architectures are often described as though identity must be resolved before meaningful security decisions can begin. Operational practice is less orderly. Session monitoring, Zero Trust policy engines, intrusion detection, and industrial monitoring all evaluate behavioral or contextual evidence while activity is occurring. This paper examines that ordering problem and argues that observed behavior can support detection and proportionate, reversible responses before the attribution needed for the next security decision is complete. Behavior also remains relevant after an identity has been authenticated or otherwise recognized.

Behavior, however, is not an identity mechanism. Anomaly is not proof of compromise, familiar-looking activity is not proof of legitimacy, and behavioral evidence is often meaningful only when combined with role, asset, resource, process, historical, or threat context. Behavioral references also differ materially: explicit rules and allowlists, statistical profiles, and learned models do not share identical failure modes. This paper therefore proposes a graduated-response principle in which the evidence burden rises with uncertainty, consequence, irreversibility, and the need for actor-specific accountability. The resulting architecture is one of evidence ordering: security may sometimes need to answer what is happening before it can fully answer who is responsible.

Keywords: behavioral evidence; identity; attribution; anomaly detection; Zero Trust; graduated response

1. Which Question Comes First?

Security architecture favors clean sequences: identify the subject, authenticate it, authorize it, then monitor what follows. That order is useful because it makes systems easier to reason about, document, and audit. It is also incomplete. Real systems encounter events while they are still resolving what those events mean. A process starts. A device opens an unexpected connection. A user session changes character. A controller produces values outside an expected relationship. The first useful security question may not be who is this? It may be what is happening?

In this paper, behavior before identity refers specifically to evidence ordering when the attribution needed for the next security decision is not yet complete. It does not mean that no identity evidence exists, nor does it propose behavior as a replacement credential. A session may already be authenticated while the system still lacks enough attribution or context to explain a new event. A device may be known while the responsible process is not. Security can preserve evidence, enrich context, challenge a session, or cautiously constrain activity while that uncertainty is resolved.

Authentication, attribution, and behavioral interpretation answer related but different questions. A successful authentication process does not establish that every later action is appropriate. Unusual behavior can justify attention without establishing which actor is responsible or whether the activity is

malicious. The practical question is which conclusion must be available before the action under consideration.

2. Identity and Behavior Do Different Jobs

NIST's current Digital Identity Guidelines define authentication as establishing that a claimant controls one or more authenticators associated with a subscriber account. The same guidance separately describes session monitoring as ongoing evaluation of session characteristics for possible fraud and identifies usage patterns, timing, behavioral biometrics, device and browser characteristics, geolocation, and IP-address characteristics as possible inputs. When potential fraud is detected, the relying party may respond through reauthentication, session termination, or notification (NIST, 2025a).

Those signals also carry a cost. NIST notes that most characteristics used for session monitoring have privacy implications and requires organizations to address them in privacy risk assessment (NIST, 2025a). Collecting more behavioral evidence is therefore not automatically better security; the monitoring architecture creates governance obligations of its own.

Identity and behavior should not be treated as perfectly independent layers. Behavior often acquires meaning from context. A command that is routine for a maintenance role may be inappropriate elsewhere. A network scan may be expected from a security tool and suspicious from another endpoint. Role, asset class, resource, process state, history, and other information can change the interpretation of identical observable activity.

The useful distinction is evidentiary. Identity-related evidence helps establish which subject, device, role, or relationship is under consideration. Behavioral evidence helps evaluate what is happening now and whether the activity fits an expected or permitted pattern. Operational systems can fuse those inputs without treating them as interchangeable.

3. Security Already Acts on Behavior

Security already uses behavior before decision-relevant attribution is complete. In digital identity systems, NIST permits post-authentication session monitoring and allows behavioral and contextual characteristics to trigger reauthentication, session termination, or notification when potential fraud is detected (NIST, 2025a).

Zero Trust architecture uses behavior differently. NIST SP 800-207 describes a policy engine that can draw from subject attributes and roles, historical subject behavior patterns, threat intelligence, and other metadata when making access decisions (Rose et al., 2020). NIST's Zero Trust implementation work similarly combines identity and role information with dynamic information such as device health, resource sensitivity, and access-pattern anomalies (NIST NCCoE, n.d.). These are multi-input decisions, not behavioral identity systems.

Industrial control environments provide a machine-oriented example. NISTIR 8219 documents behavioral anomaly detection in manufacturing control systems, including network, device, and process observations used to identify anomalous conditions that may threaten operational data or physical processes (McCarthy et al., 2020). Here, the immediate security value can lie in recognizing that an expected relationship has changed before the cause is fully attributed.

Across these examples, current activity changes what the security system does next. The specific action varies by architecture: monitor, alert, request fresh authentication, adjust an access decision, or begin a

bounded response. The common point is narrower than any one response chain: behavioral evidence can become actionable before every attribution question relevant to the event has been resolved.

4. A Known Identity Can Still Behave Wrong

Authentication answers an important question, but it does not settle every subsequent security decision. NIST's session-monitoring guidance explicitly assumes that characteristics observed after authentication can justify reauthentication or termination (NIST, 2025a). Zero Trust likewise treats historical behavior and dynamic information as continuing inputs rather than assuming that a prior identity check settles future risk (Rose et al., 2020).

Identity evidence and behavioral evidence can therefore diverge. A system can have strong evidence about which subject or device is involved while current activity still violates policy, expected relationships, or risk thresholds. Conversely, suspicious activity does not by itself show that the identity evidence was wrong. Keeping those conclusions separate allows monitoring to add information instead of merely repeating authentication.

5. Anomaly Is Not Guilt

Behavioral security becomes dangerous when deviation is treated as a verdict. An anomaly is a difference from an expectation, rule, profile, or baseline. Legitimate maintenance, software changes, unusual workloads, and rare operating conditions can all produce deviations without implying an attacker.

NIST intrusion-detection guidance has long recognized this problem. Anomaly-based approaches can produce false alarms because legitimate behavior varies, and their effectiveness depends on the quality of the profile used to represent normal activity (Bace & Mell, 2001; NIST, 2007). Current NIST adversarial-machine-learning guidance shows that the tradeoff has not disappeared: cybersecurity anomaly detection may require both very low false-negative and very low false-positive rates, a difficult combined requirement (Vassilev et al., 2025).

The reference also has a lifecycle. NIST SP 800-137 requires continuous-monitoring strategies to remain aligned with changing missions, architectures, risks, threats, vulnerabilities, and observed trends (Dempsey et al., 2011). That does not make every behavioral reference an adaptive model. It means yesterday's operational assumptions may not remain sufficient indefinitely.

The boundary is simple: deviation can justify attention; it does not establish compromise. The cost of being wrong remains part of the response decision.

6. Normal Is Not Innocent

The opposite shortcut fails as well. CISA and partner agencies describe living-off-the-land techniques in which threat actors abuse legitimate, built-in tools and processes to blend with ordinary administrative activity and reduce obvious malicious artifacts (CISA et al., 2024). Joint guidance on state-sponsored activity likewise describes adversaries using built-in network-administration tools in ways that make malicious behavior harder to distinguish from legitimate system and network activity (CISA et al., 2023).

These examples do not show that behavioral monitoring is useless. They show why conformity cannot be treated as proof of benign intent. Legitimate tools and expected administrative mechanisms can serve malicious objectives. A patient adversary may deliberately minimize conspicuous deviation.

Behavioral monitoring is therefore evidence, not an oracle. Unusual does not automatically mean malicious, and familiar-looking activity does not automatically mean safe.

7. The Baseline Becomes Part of the Security Argument

Most behavioral judgments used for anomaly or policy decisions depend on a reference: an explicit allowlist, a threshold or rule, a statistical profile, or a learned model. These mechanisms are often grouped together under behavioral detection, but they are not the same security object.

An explicit rule can be narrow and explainable: a controller should not initiate a certain outbound connection, or a service role should not invoke a particular function. A statistical profile instead asks whether an observation is unusual relative to historical distributions. A learned model may infer more complex boundaries from data. Each approach can be useful, but its failure modes differ. Rules can become stale when architecture changes. Statistical profiles can misclassify rare legitimate behavior. Learned models add security concerns associated with their training data and model behavior.

NIST's adversarial-machine-learning guidance documents poisoning and evasion as attack classes affecting machine-learning systems, including cybersecurity contexts (Vassilev et al., 2025). That evidence is scope-limited. A deterministic allowlist is not a machine-learning model and should not inherit ML-specific claims merely because both are used to evaluate behavior.

Once a rule, profile, or model influences a security decision, defenders need to understand the reference behind the judgment: what expectation is being applied, how it was defined or learned, how it is maintained, and whether it still fits the operating state.

8. Context Gives Behavior Meaning

A clean formulation would say identity answers who, behavior answers what, and the two are orthogonal. Operational systems are messier. The same observable action can be acceptable or unacceptable depending on the actor, device, resource, process state, or surrounding history.

Zero Trust makes this context dependence explicit. NIST places historical subject behavior beside subject attributes, roles, threat intelligence, and other metadata rather than treating behavior as a self-sufficient signal (Rose et al., 2020). Implementation guidance similarly combines identity and role information with dynamic device and resource context (NIST NCCoE, n.d.).

The strong orthogonality thesis therefore fails, but the evidentiary distinction survives. A system can know which entity is acting and still need behavioral evidence to decide whether the current action is acceptable. It can also observe activity that warrants attention before attribution is complete, then use identity and contextual information to reduce ambiguity. The useful architecture is evidence fusion without category collapse.

9. Response Must Match Evidence

If behavioral evidence is useful but ambiguous, response becomes the central design problem. A low-confidence anomaly may be sufficient to preserve logs or alert an analyst. The same evidence may be nowhere near sufficient to shut down a safety-critical process, revoke a credential, or permanently isolate a production system.

This paper proposes a response ladder as an analytical framework: observe, enrich, challenge, limit, isolate, and remediate or revoke. It is not a NIST, CISA, or industry standard. The principle is

straightforward: the evidence burden should rise as an action becomes more consequential, less reversible, or more actor-specific.

Under this framework, observation records the event without deciding what it means. Enrichment adds identity, role, asset, resource, process, historical, and threat context. A challenge seeks fresh proof when behavior creates doubt. Limitation narrows scope, rate, privilege, or session capability when evidence is stronger but not final. Isolation carries greater operational cost and should generally require stronger corroboration unless policy defines the observed behavior as categorically unsafe. Durable remediation or revocation should demand the strongest basis because the action persists and often depends on identifying what must be corrected.

This avoids two bad extremes: refusing to act until attribution is perfect and allowing every anomaly to become a denial decision. Graduated response keeps uncertainty visible instead of pretending it has already been resolved.

10. Behavior Before Identity — But Not Instead of It

Evidence does not always arrive in the order shown on an architecture diagram. A detector may see an unexpected communication before it knows the responsible process. A monitoring system may see a suspicious session pattern before it knows whether an account has been taken over. An industrial system may observe a process deviation before anyone knows whether the cause is malicious, accidental, mechanical, or environmental.

In those cases, a bounded response can begin while attribution continues: preserve evidence, increase scrutiny, enrich context, request fresh proof, or cautiously constrain exposure. As the desired action becomes durable or actor-specific, attribution matters more. Accountability requires an actor. Credential revocation requires knowing which credential matters. Remediation requires identifying what must change.

> **The distinction is between beginning a security response and completing a security conclusion. Behavioral evidence can often support the first; it rarely supplies the second by itself.**

11. What Deserves to Be Discovered

The remaining questions concern boundaries rather than existence. What evidence should be required for an alert, a challenge, a temporary limitation, an isolation decision, or a durable revocation? How should reversibility and physical consequence alter those thresholds? How should systems communicate uncertainty so operators do not mistake a behavioral score for a factual conclusion?

Constrained machine environments deserve particular attention because expected relationships can sometimes be defined more narrowly than human behavior. A deterministic violation of an approved communication path may carry different evidentiary weight than a statistical deviation in a human session. That difference should be studied rather than flattened into a single category called anomaly detection.

Baseline lifecycle is another open problem. Explicit policies, statistical profiles, and learned models require different forms of maintenance. For learned or data-centric models, adaptation and retraining can also introduce model- and data-specific security questions. Future work should examine how identity, behavior, device state, process state, and threat intelligence can be fused without allowing one uncertain signal to masquerade as proof.

The objective is not to make behavior win an argument against identity. It is to design decisions that ask the right question at the right time and demand evidence proportionate to the action that follows.

Conclusion

Security does not always encounter complete attribution first. Sometimes it encounters motion: a command, a connection, a sequence, a timing change, a process deviation, or a session that no longer resembles the one that began. Existing guidance and operational practice show that such evidence can matter while attribution remains incomplete.

The conclusion is deliberately narrow. Behavioral evidence can begin detection and proportionate response, but anomaly is not guilt, familiar-looking activity is not innocence, and stronger or more durable actions require stronger grounds. The question is not whether security should choose behavior or identity. It is which question must be answered first for the action being considered.

References

Bace, R., & Mell, P. (2001). Intrusion detection systems (NIST Special Publication 800-31). National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-31.pdf

Cybersecurity and Infrastructure Security Agency, National Security Agency, Federal Bureau of Investigation, & international partners. (2023, May 24). People's Republic of China state-sponsored cyber actor living off the land to evade detection. Cybersecurity and Infrastructure Security Agency. https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-144a

Cybersecurity and Infrastructure Security Agency, National Security Agency, Federal Bureau of Investigation, & partner agencies. (2024, February 7). Identifying and mitigating living off the land techniques. https://www.cisa.gov/resources-tools/resources/identifying-and-mitigating-living-land-techniques

Dempsey, K., Chawla, N. S., Johnson, L., Johnston, R., Jones, A. C., Orebaugh, A., Scholl, M., & Stine, K. (2011). Information security continuous monitoring (ISCM) for federal information systems and organizations (NIST Special Publication 800-137). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-137

McCarthy, J., Powell, M., Stouffer, K., Tang, C., Zimmerman, T., Barker, W., Ogunyale, T., Wynne, D., & Wiltberger, J. (2020). Securing manufacturing industrial control systems: Behavioral anomaly detection (NISTIR 8219). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8219

National Institute of Standards and Technology. (2007). Guide to intrusion detection and prevention systems (IDPS) (NIST Special Publication 800-94). https://doi.org/10.6028/NIST.SP.800-94

National Institute of Standards and Technology. (2025a). Digital identity guidelines: Authentication and authenticator management (NIST Special Publication 800-63B-4). https://doi.org/10.6028/NIST.SP.800-63B-4

National Institute of Standards and Technology, National Cybersecurity Center of Excellence. (n.d.). Implementing a zero trust architecture: Project overview. https://pages.nist.gov/zero-trust-architecture/VolumeA/ProjectOverview.html

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero trust architecture (NIST Special Publication 800-207). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-207

Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). Adversarial machine learning: A taxonomy and terminology of attacks and mitigations (NIST AI 100-2e2025). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-2e2025