MINDVAULT · WHITE PAPER

The Session Is Legitimate. The User Is Not.

Why Authentication Is No Longer Enough, and What Comes Next

PRIVATE AND CONFIDENTIAL
© 2026 MindVault Technologies LLC. All rights reserved.
info@mindvaultinc.com · **mindvaultinc.com**

Contents

MINDVAULT

Executive Summary

Every organization that has deployed multi-factor authentication, invested in endpoint detection, or adopted a Zero Trust framework has done so on the basis of a shared assumption: **that a successfully authenticated session represents a legitimate user**.

That assumption is no longer reliable.

The most damaging identity-based attacks in today's threat landscape do not defeat authentication; they bypass it. Adversary-in-the-middle phishing captures session tokens after MFA succeeds. Cookie theft replays valid credentials from attacker infrastructure. Remote access tools allow threat actors to operate a legitimate session in real time, from anywhere in the world, while the account holder remains unaware. In each case, the session is technically valid. The authorization is real. Every existing control has fired correctly. And an attacker is operating freely within an environment that has no mechanism to know they are there.

The gap is not a failure of implementation. It is a failure of assumption. The security industry built its identity stack around the authentication event and left the entirety of the post-authentication session unmeasured.

Continuous identity confidence addresses that gap directly. By learning the behavioral persona of each individual user: how they type, how they navigate, how they interact with the systems they use every day, and evaluating active sessions against it in real time, the system produces an ongoing Human Confidence Signal. This allows the evaluation of whether the person behind the keyboard is the person the session was issued to. When behavior aligns, confidence remains high, and sessions proceed without friction. When behavior drifts in ways consistent with a remote operator, a stolen session, or an unattended device, confidence degrades, and the response scales in proportion.

Done well, this is what stands to change the outcomes that matter to security leadership: compressed dwell time for session-based attacks, friction-right access that serves both assurance and user experience, richer audit trails for compliance, and a behavioral signal that makes existing detection platforms more precise.

The technology is designed for any organization where employees perform meaningful work from a network-connected terminal. It is indifferent to industry, workforce model, or network architecture. The exposure it addresses is universal. So is the question it answers.

> *Is the person operating this session still the person who started it?*

Today, in most environments, that question has no reliable answer. It should.

Section 1: The Problem

The Assumption That Broke

For three decades, cybersecurity operated on a foundational premise: a successfully authenticated session represents a legitimate user. This assumption was reasonable when gaining access required physically possessing credentials, when the only path to a valid session ran through the person who owned it.

That premise no longer holds.

The modern threat landscape has not broken authentication. It has bypassed it entirely. Today's most sophisticated attacks do not attempt to defeat MFA, crack passwords, or exploit authentication infrastructure. Instead, they wait. They watch. And then they walk in through the front door using credentials that are, by every technical measure, completely valid.

© 2026 MindVault Technologies LLC. All rights reserved.

Credentials Are No Longer Proof of Identity

The rise of adversary-in-the-middle (AiTM) phishing has fundamentally altered the calculus of identity-based attacks. Where traditional phishing sought to steal a password, AiTM frameworks intercept the authenticated session itself, capturing the token that is issued *after* a user successfully completes MFA. The attacker never needs the password. They never need to defeat the second factor. By the time the security stack has declared the session legitimate, the real user has been removed from the equation entirely. This is not a fringe technique: Microsoft's threat intelligence documented a single AiTM campaign that targeted more than 10,000 organizations, and reported a 146% year-over-year rise in these attacks, growth driven precisely by their ability to defeat accounts already protected by MFA.[2]

Cookie theft operates on the same principle. A single stolen session token, harvested from an endpoint via infostealer malware, grants an attacker an authenticated, authorized session with no authentication required. The identity provider has no visibility into what happens after the token is issued. As far as the system is concerned, the user is still there. The scale this enables is not theoretical. In the 2024 compromise of Snowflake customer environments, the group tracked as UNC5537 authenticated to roughly 165 organizations using credentials harvested by infostealer malware, some dating back to 2020, and exfiltrated data on more than 500 million individuals, without exploiting any vulnerability in Snowflake itself.[3] The 2025 Salesloft Drift incident showed the same pattern with OAuth tokens: stolen refresh tokens handed attackers valid, pre-authenticated sessions into the Salesforce data of more than 700 organizations, with no password prompt and no MFA challenge.[4]

Remote access tooling has introduced a third vector that is particularly difficult to detect because it requires no theft at all. Applications like ScreenConnect, AnyDesk, TeamViewer, and RustDesk are legitimate enterprise tools, trusted by IT departments, whitelisted by security policies, and invisible to most detection stacks. When a threat actor convinces a user to install one of these tools, or deploys one silently following initial access, they do not hijack the session. They *share* it. The user's credentials remain valid. The session remains active. A second party is now in full control of everything that session is authorized to do. This pattern is well documented: CISA, the NSA, and the MS-ISAC issued a joint advisory after identifying a campaign that used phishing to deploy legitimate RMM software, portable executables that need no administrator rights and rarely trigger endpoint defenses, and renewed the warning in 2025 as the abuse spread.[5]

The Visibility Gap

What these attacks share is not a technical sophistication that outpaces defenses. What they share is a fundamental exploitation of a gap in how security has historically been designed: the moment authentication completes, the question of *who is operating this session* disappears from the security model entirely.

Identity and access management platforms verify the user at the door. Endpoint detection tools watch for malicious process behavior. Network monitoring flags anomalous traffic patterns. But no layer in the conventional stack is asking a continuous question: *is the person behind this keyboard still the person who signed in?*

This is not a gap in implementation. It is a gap in assumption. The security industry has invested billions in answering the authentication question with increasing precision, while leaving the post-authentication period, the entirety of a user's working day, as a blind spot.

The consequences are measurable. Attacks that ride valid credentials and live sessions now account for a substantial and growing share of enterprise breaches. Verizon's *2024 Data Breach Investigations Report* found stolen credentials present in roughly a quarter of breaches analyzed that year, and in almost a third over the preceding decade, among the most common entry paths in the entire dataset.[1] They are so difficult to surface precisely because they produce no authentication anomalies, trigger no credential-based alerts, and leave behind logs that show nothing but normal, authorized activity. Incident responders reconstructing these breaches face the same fundamental problem the security stack faced in real time: everything looked legitimate, because technically, it was.

© 2026 MindVault Technologies LLC. All rights reserved.

The Vectors in ATT&CK Terms

None of these patterns is novel in the abstract. Each maps cleanly to a documented technique in the MITRE ATT&CK framework[8], the industry's shared vocabulary for adversary behavior, and several succeed precisely because they operate *after* the MFA challenge has already been satisfied.

Observed pattern	MITRE ATT&CK technique
Phishing as the initial delivery vector	T1566: Phishing
Proxy interception of an authenticated session (AiTM)	T1557: Adversary-in-the-Middle
Theft of session cookies, including infostealers reading the browser store	T1539: Steal Web Session Cookie; T1555.003: Credentials from Web Browsers
Theft of OAuth / application access tokens	T1528: Steal Application Access Token
Replaying stolen material to act as the user	T1550: Use Alternate Authentication Material (.004 Web Session Cookie; .001 Application Access Token)
Abuse of legitimate remote access tooling	T1219: Remote Access Tools
Operating inside the session as a valid principal	T1078: Valid Accounts

The mapping sharpens the structural point. Every technique in that list resolves, by design, to activity that looks authorized: a valid token, a legitimate tool, an authenticated account. That is exactly why they are so difficult to detect from inside the session, and exactly the gap the rest of this paper examines.

The Operator Is No Longer Always Human

The discussion so far has assumed the entity operating a session is a person, legitimate or not. That assumption is eroding too. Inside the modern enterprise, non-human identities now vastly outnumber human ones: CyberArk's 2025 Identity Security Landscape put the ratio at more than 80 to 1, driven by cloud automation and the rapid adoption of AI agents.[9] A growing share of the activity inside authenticated sessions is initiated not by a user at a keyboard but by an automated identity acting with delegated permissions.

For the post-authentication gap, this widens the scope of the question rather than changing its nature. An AI agent is issued a token or an OAuth grant and then operates, often with the same entitlements as the human who delegated to it, and nothing continuously verifies that it is still doing what it was authorized to do. The Salesloft Drift compromise was an early demonstration: a chatbot integration's stolen OAuth tokens, not a human's password, became the path into the Salesforce data of more than 700 organizations.[4]

Agents also carry a failure mode humans do not. Prompt injection, ranked the number-one risk in the OWASP Top 10 for LLM Applications[10], exploits the fact that a language model cannot reliably separate trusted instructions from untrusted data it is asked to process. A legitimate, fully authenticated agent can be redirected against its task by content it merely reads, with no credential stolen and no policy violated. OpenAI, describing its own browser agent in December 2025, concluded that prompt injection is "unlikely to ever be fully 'solved.'"[11] If the threat cannot be eliminated at the model layer, detecting a subverted agent comes to depend on the same thing detecting a hijacked human session does: noticing that the entity operating the session has begun behaving in a way its established pattern does not predict.

The question this paper poses therefore widens by a word. It is no longer only *is the person operating this session the person it was issued to?*, but *is the entity, human or agent, operating this session still the one it was authorized for, and still behaving as intended?*

© 2026 MindVault Technologies LLC. All rights reserved.

What Is Actually Missing

The industry has a mature vocabulary for the moment of authentication and a growing vocabulary for the behavior of processes and network traffic. What it lacks is a vocabulary, and a capability, for the continuous confirmation of who, or what, is operating the session.

Authenticating a user is a point-in-time event. Confirming a persona is an ongoing one. The distinction matters because the attack surface is not the authentication event. The attack surface is every minute that follows it.

Section 2: Why Existing Approaches Fall Short

Built for the Door, Not the Room

The security industry's response to identity-based threats has been substantial. Over the past decade, organizations have deployed multi-factor authentication at scale, adopted single sign-on platforms, invested in identity governance, and embraced Zero Trust as an architectural principle. These investments were not misguided. They addressed real vulnerabilities and meaningfully raised the cost of certain attacks.

But they were designed around a specific threat model, one in which the attacker's goal is to defeat authentication. Today's most damaging identity-based attacks do not share that goal. They assume authentication will succeed, and they build their strategy around what happens after.

Measured against that reality, the existing stack has a structural limitation that no amount of incremental improvement will resolve: it stops asking questions at the moment a session begins.

Multi-Factor Authentication

MFA remains one of the highest-value controls an organization can deploy. Against credential stuffing, password spray attacks, and traditional phishing, it is highly effective. The problem is not that MFA fails at what it was designed to do. The problem is that what it was designed to do is no longer sufficient.

MFA is a point-in-time verification mechanism. It answers one question, at the moment of login, can this user prove possession of a second factor?, and then it is finished. The session that follows exists entirely outside MFA's visibility. A session token stolen after a successful MFA event carries full authorization indefinitely. An attacker who gains remote access to a machine following a legitimate MFA login inherits every permission that authentication granted, for as long as the session remains active.

The attack does not defeat MFA. It simply waits for MFA to finish.

Endpoint Detection and Response

EDR platforms have transformed the ability to detect and respond to malicious activity at the process and file system level. They are essential infrastructure for any mature security program. But EDR is fundamentally a tool for answering questions about software behavior, what processes ran, what files were touched, what network connections were made.

It is not designed to answer questions about human behavior. When a threat actor operates through a legitimate remote access tool, they produce no malicious process activity. ScreenConnect is not malware. TeamViewer does not trigger behavioral detections. The actions taken through those tools, navigating a file system, exfiltrating documents, moving laterally through legitimate applications, are indistinguishable at the process level from the actions of the legitimate user who owns the session.

EDR tells you the "what happened" portion of the story. It was never built to tell you who was doing it.

Identity and Access Management

IAM platforms govern what an authenticated identity is permitted to do. They enforce role-based access policies, manage entitlements, and provide audit trails of resource access. They are the right answer to the question of authorization. They are not designed to answer the question of continuous identity confirmation.

© 2026 MindVault Technologies LLC. All rights reserved.

An IAM platform cannot distinguish between a legitimate user accessing a sensitive resource and an attacker operating through that user's hijacked session, because both present the same authenticated identity, carrying the same entitlements, performing actions that are within the defined scope of that role's permissions. The policy engine sees a valid principal making an authorized request. The fact that a different human being is making that request is outside the platform's visibility entirely.

User and Entity Behavior Analytics

UEBA tools represent the closest existing approximation to the problem this paper addresses. They analyze patterns of user activity over time and flag statistical anomalies, access outside normal hours, unusual resource requests, atypical data volumes. For detecting compromised accounts through aggregate behavioral signals, they provide genuine value.

The limitation is scope and latency. UEBA operates at the level of access events, what systems were touched, when, by which identity. It answers questions about what an account did over time. It does not answer questions about who was operating that account in a given session, in real time, at the keyboard level. The signals it consumes are too high-level and too delayed to detect an attacker who has taken over a session mid-day, operates within normal access patterns, and exits before aggregate anomalies accumulate.

By the time a UEBA platform surfaces an alert, the session in question may have concluded hours ago.

The Common Thread

Each of these tools is well-designed for the problem it was built to solve. The gap is not in their execution, it is in the assumption that runs beneath all of them: that a verified identity at login remains the relevant question throughout the life of a session.

That assumption made the existing stack coherent. It also made it blind to an entire category of attack. When the threat actor's strategy is not to impersonate a user but to operate *as* one, sharing a session, riding a token, controlling a screen, every control designed around the authentication event becomes a control that has already fired by the time the attack begins.

The gap here isn't a weak door; the door worked. What's missing is detection and monitoring of the room after someone's been let in.

Section 3: The Solution Framework

Continuous Identity Confidence

Addressing the post-authentication blind spot requires a different model, one that does not replace existing identity infrastructure but extends it into the dimension that has gone unmeasured: time.

The concept is straightforward. Every person who operates a computer does so in ways that are, to a meaningful degree, uniquely their own. The rhythm of their typing. The way they move and use a mouse. The applications they open, in what order, at what times. The vocabulary they use. The pace at which they read and scroll. Individually, none of these signals is definitive. Collectively, they form a behavioral persona, a signature unique to each individual, that reflects not just *that* someone is using a system, but *who* is using it.

Continuous identity confidence is the practice of measuring that persona in real time and expressing the result as a dynamic, ongoing signal: not a binary authorized/unauthorized state, but a Human Confidence score that rises and falls based on how well the observed behavior matches that user's established persona.

How the Model Works

When a user authenticates, they do not simply open a session. They begin contributing to, and being evaluated against, their own behavioral persona. The system observes passively, without friction, across the signals available in a normal working session. Over time, that persona takes shape as a baseline specific to the individual: not a population average, not a role-based policy, but a model of how *this person* actually behaves.

That baseline becomes the reference point for every subsequent session. As long as observed behavior aligns with the established persona, confidence remains high and the session proceeds without interruption. When behavior drifts, suddenly, gradually, or in patterns consistent with known attack techniques, confidence degrades, and the response scales in proportion.

The response is not binary. A modest confidence drop might trigger a passive step-up, a re-authentication prompt in the background, invisible to a legitimate user who simply confirms presence and continues working. A significant drop, or one that matches a high-risk behavioral pattern, escalates the response: session suspension, IT notification, forced re-authentication, or integration with SIEM and SOAR platforms to trigger broader investigation workflows.

What Behavioral Drift Looks Like

The practical value of this model becomes clear when mapped against the attack vectors that have outpaced conventional defenses.

A remote access tool taking control of an active session produces an immediate and detectable behavioral shift. Mouse movement patterns change, remote operators tend to move cursors differently than local users, with latency, acceleration curves, and path patterns that can diverge from the established persona. Typing cadence changes. Application interaction patterns change. The session, technically legitimate, begins producing behavioral signals inconsistent with the person who opened it. Confidence degrades, and a control built on that signal can respond while the deviation is still early, rather than after the damage is done.

A session operating from a stolen cookie, replayed from an attacker's infrastructure, arrives with no behavioral history and immediately begins producing interactions inconsistent with the legitimate user's persona. The authentication token says one thing. The behavior says another. That contradiction, invisible to every other layer of the security stack, is precisely what continuous identity confidence is designed to surface.

An insider threat, a legitimate user taking actions outside their normal pattern, produces subtler drift, but drift nonetheless. Access to resources never previously visited. Application usage patterns that deviate from the established persona. Behavioral signals that, taken individually, might be unremarkable, but in aggregate represent a meaningful departure from the known persona.

Fitting Into the Existing Stack

Continuous identity confidence is not a replacement for authentication infrastructure. MFA, SSO, and identity providers remain the mechanism by which sessions are opened. What this model provides is what happens after, the ongoing confirmation that the authenticated session continues to be operated by the person it was issued to.

For organizations operating within a Zero Trust framework, this maps directly onto the principle of continuous verification. Zero Trust has established the policy architecture; continuous identity confidence provides the Human Confidence signal that policy engines have lacked. Trust is not granted at login and held indefinitely. It is earned, session by session, minute by minute, by behavior that confirms the person behind the credentials is still the person who presented them.

For security operations teams, the output is not noise, it is context. A Human Confidence signal integrated into existing SIEM infrastructure transforms alert triage. Instead of investigating whether a process was malicious, analysts gain visibility into whether the *person* initiating that process was the legitimate account holder. That distinction is often the difference between a false positive and a confirmed incident.

© 2026 MindVault Technologies LLC. All rights reserved.

Section 4: What Closing This Gap Would Change

From Authentication Events to Living Assurance

Security investment has always been justified through one of two lenses: risk reduction or operational efficiency. Continuous identity confidence is aimed at both, but the more consequential point is that it changes the *nature* of what security teams can see, not just how quickly they can see it.

The Dwell-Time Window

Dwell time, the interval between an attacker gaining access and that access being detected, remains one of the most consequential metrics in enterprise security. The headline numbers have improved: Mandiant's *M-Trends 2024* put global median dwell time at roughly 10 days, down from 16 the year before.[6] But that aggregate masks the problem this paper describes. Breaches involving stolen or compromised credentials remain the slowest of any category to identify and contain, an average of 292 days, according to IBM's *Cost of a Data Breach Report 2024*.[7] The reason is structural: when an attacker operates through a legitimate session, they produce no signals that existing detection tools are designed to surface. The clock runs silently.

Closing this gap means moving identity detection from after-the-fact log review to continuous, real-time evaluation. A control that scores behavioral alignment as the session unfolds, rather than reconstructing it from logs days later, would compress the relevant window from weeks or months to the time it takes observed behavior to diverge from the established persona. The objective is concrete: an attacker who takes control of a session at 2pm should not have until the end of the day, or until a UEBA platform surfaces an aggregate anomaly the following week. That is the standard against which any post-authentication control should be measured.

Friction-Right Access

One of the persistent tensions in enterprise security is the tradeoff between assurance and usability. Controls that increase security confidence tend to increase user friction. Blanket step-up authentication policies frustrate legitimate users. Overly aggressive session timeouts interrupt productive work. Security teams face consistent pressure to loosen controls that impede the business.

Continuous identity confidence inverts this dynamic. Because a real-time signal of behavioral alignment is available, step-up challenges can be reserved for the moments that actually warrant them, when confidence has dropped below a meaningful threshold, not on a fixed timer or arbitrary policy trigger. A legitimate user who has been at their desk all morning, typing in their established pattern, interacting with familiar applications, would see no additional friction. The same session, suddenly behaving like a remote operator rather than the enrolled user, would warrant a response proportional to the confidence drop.

The aim is to let security assurance rise while average user friction falls, two objectives usually in opposition. Moving both at once is the central design promise of a Human Confidence signal, and the bar any implementation has to clear in practice.

Richer Audit Trails for Compliance

Regulated industries face an audit reality that existing identity infrastructure handles imperfectly. Access logs confirm that an authenticated identity accessed a resource. They do not confirm that the person associated with that identity was the one who did so. For compliance frameworks built around the assumption that authenticated access equals accountable access, this gap is currently papered over rather than solved.

A continuous identity confidence layer would add a dimension to the audit record that does not exist today: a real-time Human Confidence signal associated with every access event. Auditors and compliance teams would gain the ability to distinguish not just *what* was accessed and *when*, but *how confident the system was* that the legitimate account holder was present at the time of access. For industries operating under HIPAA, SOX, PCI-DSS, or emerging AI governance frameworks, this represents a meaningful step toward access accountability that reflects reality rather than assumption.

© 2026 MindVault Technologies LLC. All rights reserved.

Insider Threat Detection Without Surveillance

Insider threat programs carry a persistent organizational tension: the controls that are most effective at detecting malicious insider activity are also the ones most likely to be perceived as invasive by the legitimate workforce. Keylogging, screen recording, and continuous activity monitoring programs create legal exposure, erode employee trust, and generate significant noise that security teams must filter.

Continuous identity confidence approaches insider threat through a fundamentally different lens. The model is not built to record what users do, it is designed to characterize *how* they do it, and to flag when that pattern changes in ways consistent with risk. A legitimate employee whose behavior drifts gradually toward high-risk resource access, unusual data movement, or interaction patterns outside their established persona would surface as a confidence anomaly without any individual action being surveilled or recorded.

The distinction matters both ethically and operationally. Security teams get a signal that warrants investigation. Employees operate in an environment that monitors for anomaly rather than cataloguing activity. Legal and HR teams have a defensible framework for how the signal was generated. The tension between effective insider threat detection and employee trust does not disappear, but it becomes significantly more manageable.

A Signal That Makes the Existing Stack Smarter

Perhaps the most immediate practical value of continuous identity confidence is not what it detects independently, but what it adds to detections the existing stack is already making.

A SIEM alert that fires on an anomalous access event carries one level of investigative weight. The same alert, correlated with a simultaneous confidence drop indicating behavioral divergence from the legitimate user's persona, carries a fundamentally different one. The alert is no longer asking whether the access was unusual. It is asking whether the person who performed it was the account holder at all. That is a different investigation, with a different urgency, and a much clearer path to response.

Security operations teams are not short on alerts. They are short on context that tells them which alerts represent genuine human-layer compromise versus process-level anomalies that resolve to legitimate activity. A Human Confidence signal is intended to provide that context continuously, without requiring analysts to manually correlate session data, review access logs, or interview end users to establish whether a legitimate human was present.

The intended effect on SOC operations is a reduction in mean time to triage and an improvement in the signal-to-noise ratio for identity-related alerts, two metrics security leadership tracks closely. Whether a given implementation delivers them is an empirical question, and the right one for any evaluation to ask.

Section 5: Who This Is Built For

Any Organization Where Identity Is an Assumption They Can No Longer Afford to Make

Continuous identity confidence is not a solution built for a specific industry, a particular network architecture, or a defined workforce model. It is built around a single criterion: does your organization have employees who perform meaningful work from a network-connected terminal?

If the answer is yes, the exposure this technology addresses is present.

The Workforce, Not the Vertical

Healthcare organizations protecting patient records and financial institutions safeguarding transaction infrastructure face different regulatory environments and different threat profiles. What they share is a workforce that authenticates into systems each morning and operates within those sessions for hours, sessions that, under the current security model, are trusted implicitly from the moment they open until the moment they close.

The same is true of a logistics company managing supply chain systems, a law firm accessing case management platforms, a government contractor operating within a classified environment, or a technology company with engineers committing to production infrastructure. The vertical changes. The exposure does not.

Continuous identity confidence is horizontal infrastructure, a layer that sits beneath industry-specific controls and addresses the human-layer blind spot that every organization shares regardless of what those controls are protecting.

In-Office and Distributed Workforces Alike

The shift toward distributed work expanded the attack surface for session-based threats meaningfully. Remote employees authenticate over VPN or zero-trust network access, operate from home networks, and interact with IT support remotely, creating conditions where social engineering attacks involving remote access tools are both easier to execute and harder to detect. The unattended laptop at a coffee shop, the employee who clicks a support scam and hands a stranger their screen, the session left open on a shared home computer, these are distributed workforce problems, and they are real ones.

But the threat is not exclusive to remote environments. In-office workforces face their own version of the same exposure. The unattended workstation whose session was never locked. The shared terminal in an operations center. The social engineering call that convinces an on-site employee to install a remote support tool. Physical presence in a corporate office does not eliminate the gap between authenticated session and confirmed persona, it only changes the specific scenarios through which that gap gets exploited.

This technology is designed to be indifferent to where the terminal sits. The persona is specific to the individual, not the environment. Confidence is evaluated against the person, not the location.

The Two Scenarios It Was Built to Address

At opposite ends of the threat spectrum sit two scenarios that, despite their differences, share the same underlying vulnerability.

The first is active exploitation, a threat actor who has gained remote access through social engineering, a support tool, or a stolen session token, and is now operating a legitimate session in real time. This is an attacker at a keyboard, somewhere else in the world, doing things the account holder did not authorize and is likely unaware of. Speed of detection is everything. Every minute of undetected access is a minute of potential data exfiltration, lateral movement, or credential harvesting.

The second is passive exposure, a legitimate session left unattended, on an unlocked device, in a location where physical access is not guaranteed. No attacker has compromised infrastructure. No credentials have been stolen. A person simply walked away, and someone else sat down. The session is valid. The authorization is real. The person operating it is not the account holder.

Both scenarios are addressed by the same underlying capability: continuous evaluation of whether the behavior of the active session matches the established persona of the person it was issued to. The attacker halfway around the world and the opportunist at the next table in a café produce the same signal, behavior that does not match the persona, and trigger the same response.

What This Means for Security Leadership

For a CISO evaluating this technology, the relevant question is not whether their organization fits a predefined profile. It is whether they can currently answer, with confidence, the following question about any active session in their environment:

> *Is the person operating this session the person it was issued to?*

© 2026 MindVault Technologies LLC. All rights reserved.

If the honest answer is no, if the best available answer is "we assume so, unless something else fires an alert", then the gap this technology addresses is present, it is active, and it is being exploited in organizations that look exactly like theirs.

Section 6: The Hard Problems

What an Honest Accounting Requires

A definition of a problem is only as credible as its treatment of what makes the problem hard. Continuous identity confidence is a probabilistic control operating on noisy human signals, and it introduces real engineering and operational challenges. Naming them is not a hedge, it is the line between a defensible security control and a marketing claim. Any serious implementation has to answer the questions below, and any serious evaluation should ask them.

The Cold-Start Problem

A behavioral baseline is only meaningful once there is enough interaction history to characterize an individual. That requirement creates an unavoidable gap at the edges: new hires, contractors, infrequent users, and anyone whose role has just changed have little or no established pattern to measure against. The same is true on the first day the approach is deployed in any environment. This forces a design decision with no free answer, during the learning period, does the control fail open, accepting a temporary blind spot, or fail closed, imposing friction on legitimate users who have done nothing wrong? Shared, kiosk, and role-based accounts complicate the picture further, because there may be no single person whose baseline can be built at all. A credible approach has to state how long a baseline takes to establish, how it behaves before one exists, and how it handles identities that never produce a stable signal.

False Positives and the Cost of Being Wrong

Behavioral signals are probabilistic, not deterministic. A sore wrist, a new keyboard, an injury, fatigue, an unfamiliar task, or a second monitor can all shift how a person interacts without any compromise having occurred. The operational cost of acting on that shift is real: interrupting a legitimate user mid-task, or worse, terminating their session, erodes trust and adoption faster than almost any other security measure. This creates a precision-recall tradeoff that cannot be engineered away. Tuned for sensitivity, the system generates friction and alert fatigue, the very problems behavioral signals were meant to reduce. Tuned for specificity, it risks missing the subtle, careful takeover. Multi-signal consensus and graduated response narrow the tradeoff but do not eliminate it; the threshold is ultimately an environment-specific policy decision. Two requirements follow directly: any such control must be measurable on its false-positive rate, and its default response to uncertainty must be proportionate and reversible, step-up before suspend, suspend before terminate.

Evasion and the Adaptive Adversary

The moment behavioral verification matters, adversaries adapt to it. A patient operator who moves deliberately can stay closer to a baseline; recorded or synthesized interaction patterns can be replayed; an attacker who mostly observes and acts rarely offers little behavior to flag; and automated agents can increasingly be tuned to imitate human cadence. Behavioral signals raise the cost and complexity of impersonation, but they do not make it impossible. The defensible claim is narrow, and should be stated plainly: this is a probabilistic control that shifts attacker economics, not a deterministic gate that cannot be beaten. Its value comes from being one layer among several, hard to evade at the same moment as endpoint, network, and identity controls, rather than from any pretense of standing alone.

Coverage, Edge Cases, and Fairness

Not all legitimate work looks the same at the keyboard, and a model trained to expect a typical pattern can penalize people who simply interact differently. Assistive technologies, screen readers, switch access, voice input, and other accessibility needs can read as anomalous, raising efficacy and fairness concerns at the same time. Low-interaction sessions such as reading, monitoring, or supervising long-running jobs produce little

© 2026 MindVault Technologies LLC. All rights reserved.

signal to evaluate. Remote desktop chains, virtualized and VDI environments, and roaming across devices distort the very timing and movement characteristics the model depends on. A credible approach has to define where it works well, where its confidence degrades, and how it avoids disadvantaging legitimate users whose behavior is atypical by necessity rather than by compromise.

Privacy and the Proportionality of Response

Even a content-free model that captures rhythm rather than text raises governance questions that cannot be deferred. What is retained, for how long, who is permitted to query the signal, and what prevents a control built for security from drifting into productivity surveillance? Legitimacy depends on tight scoping, flagging anomaly, not cataloguing activity, on transparency with the workforce, and on a defensible basis for any automated action that affects an individual. Getting this wrong carries more than legal exposure: it erodes the workforce trust the control quietly depends on to function at all.

None of This Is a Reason Not to Close the Gap

These are not arguments against continuous identity confidence. They are the terms on which the gap must be closed, and the questions any technical evaluation should put to a solution that claims to close it. A control that cannot answer them honestly is not ready for the environments this paper is addressed to. A problem worth defining this precisely deserves to be solved with the same precision.

Conclusion

The Question Security Forgot to Keep Asking

Authentication solved a hard problem. It gave organizations a reliable mechanism for verifying identity at the threshold, a way to ensure that the person requesting access could prove they had the right to it. Decades of investment made that mechanism faster, stronger, and increasingly resistant to the attacks designed to defeat it.

In doing so, it created an assumption so foundational that the industry stopped questioning it: that the verified identity at the door remains the relevant fact for the life of everything that follows.

That assumption is now a liability.

The most consequential identity-based attacks in the current threat landscape do not break authentication. They inherit it. They wait for the door to open legitimately, and then they walk through it, through a shared screen, a stolen token, a hijacked session, a remote tool installed by a user who thought they were talking to IT. By the time any of that happens, every control designed around the authentication event has already fired and gone quiet. The session is active. The authorization is valid. And no layer in the conventional stack is asking the only question that matters anymore.

Is the person behind this session still the person who started it?

Continuous identity confidence exists to answer that question, not once, at login, but continuously, for the life of every session, across every terminal in the environment. It does not replace the controls organizations have already built. It extends them into the dimension they have always left unmeasured: the ongoing confirmation that authenticated access remains under authorized control.

The attack surface has moved. It moved from the credential to the session, from the authentication event to the post-authentication period, from defeating identity infrastructure to simply waiting for it to succeed. The security model has to move with it.

© 2026 MindVault Technologies LLC. All rights reserved.

The organizations that recognize this shift earliest will not just reduce their exposure to a growing class of attacks. They will be the first to close a gap that, right now, exists silently in every environment that has ever assumed a valid session means a legitimate user.

That assumption has been wrong for longer than most organizations know.

References

- Verizon. *2024 Data Breach Investigations Report (DBIR).* 2024. https://www.verizon.com/business/resources/reports/dbir/
- Microsoft Threat Intelligence. "From cookie theft to BEC: Attackers use AiTM phishing sites as entry point to further financial fraud." Microsoft Security Blog, July 12, 2022; and "Detecting and mitigating a multi-stage AiTM phishing and BEC campaign," June 8, 2023. https://www.microsoft.com/en-us/security/blog/2022/07/12/from-cookie-theft-to-bec-attackers-use-aitm-phishing-sites-as-entry-point-to-further-financial-fraud/
- Mandiant / Google Cloud Threat Intelligence. "UNC5537 Targets Snowflake Customer Instances for Data Theft and Extortion." 2024. https://cloud.google.com/blog/topics/threat-intelligence/unc5537-snowflake-data-theft-extortion
- Google Threat Intelligence Group (GTIG). "Data Theft from Salesforce Instances via Salesloft Drift (UNC6395)." 2025. https://cloud.google.com/blog/topics/threat-intelligence/data-theft-salesforce-instances-via-salesloft-drift
- CISA, NSA, and MS-ISAC. *Joint Cybersecurity Advisory AA23-025A: Protecting Against Malicious Use of Remote Monitoring and Management Software.* January 25, 2023 (updated July 2025). https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-025a
- Mandiant / Google Cloud. *M-Trends 2024.* April 2024. https://cloud.google.com/blog/topics/threat-intelligence/m-trends-2024
- IBM Security. *Cost of a Data Breach Report 2024.* July 2024. https://www.ibm.com/reports/data-breach
- MITRE ATT&CK®, Enterprise Matrix. The MITRE Corporation. Techniques referenced: T1566, T1557, T1539, T1555.003, T1528, T1550 (.001/.004), T1219, T1078. https://attack.mitre.org/
- CyberArk. *2025 Identity Security Landscape.* April 2025 (machine identities outnumber human identities by more than 80 to 1). https://www.cyberark.com/press/machine-identities-outnumber-humans-by-more-than-80-to-1-new-report-exposes-the-exponential-threats-of-fragmented-identity-security/
- OWASP Gen AI Security Project. *OWASP Top 10 for LLM Applications 2025*, LLM01: Prompt Injection. https://genai.owasp.org/llmrisk/llm01-prompt-injection/
- OpenAI. "Continuously hardening ChatGPT Atlas against prompt injection attacks." December 2025. https://openai.com/index/hardening-atlas-against-prompt-injection/

To learn more or to request a briefing, contact info@mindvaultinc.com or visit mindvaultinc.com. Early access partnerships are available for organizations ready to pilot continuous identity confidence in their environment.

© 2026 MindVault Technologies LLC. All rights reserved.