Governing the Human Layer

A Dynamic Zero Trust Framework with Out-of-Band Human Sovereignty

Prepared by: Jose A. Cedeño, Principal Advisor
Organization: Cofresí Consulting Services
Date: March 26, 2026
Classification: Confidential — Advisory Distribution Only

I. Executive Summary

As the threat landscape evolves, traditional "Castle and Moat" security architectures have proven insufficient for protecting complex environments — particularly within Critical Infrastructure and Operational Technology (OT). This paper introduces a behavioral-centric analogy for Zero Trust (ZT): the Autonomous Security Vehicle. By shifting focus from entry-point verification to continuous, policy-driven behavioral governance, organizations can achieve a resilient posture that effectively governs what this framework calls the Human Layer.

This paper extends that model with a critical and underexplored insight: autonomous security systems can themselves be compromised — not by brute force, but by the corruption of their own baseline. When that happens, the system reports green while the real world burns. The answer is not more automation. The answer is a doctrine of Out-of-Band Human Sovereignty: a structured, pre-validated human override capability that exists entirely outside the automated system's chain of trust.

Visualization tools such as VStrike serve as the advanced situational awareness platform through which authorized personnel exercise that sovereignty — transforming raw data into the high-fidelity "Cyber Canvas" required for assertive decision-making under pressure.

II. The Limitation of the Static Perimeter

For decades, cybersecurity relied on the assumption that anything inside the network was inherently trusted. In high-stakes environments — such as those managed by the Department of Energy (DOE) or modern automated production lines — this lateral trust is a catastrophic vulnerability. Once a single credential or device is compromised, the attacker has unfettered access to the entire kingdom.

The industry has correctly identified that we must move beyond asking "Who are you?" as the only security question. Zero Trust reframes the challenge: Where are you allowed to go, and how are you behaving while getting there? This paper argues that we must add a third question, one the current literature has largely neglected:

> *What happens when the system answering those questions can no longer be trusted to answer them correctly?*

III. The Zero Trust Autonomous Vehicle Model

To solve the core challenge of Zero Trust, we re-envision the network not as a static fortress, but as a fleet of Autonomous Vehicles governed by a central intelligence. Each element of this analogy maps precisely to a technical control.

1. Defined Policy Boundaries — The 10-Mile Radius

In a ZT framework, every user and device is assigned a strict Radius of Operation. This is the digital equivalent of an autonomous car programmed to never leave a specific 10-mile zone. The car is not a broken machine because it cannot cross that boundary — it is correctly programmed.

Technical Application: This represents the Policy Enforcement Point (PEP). If an identity attempts to access data or systems outside their specific mandate, the system executes an Implicit Deny. There is no negotiation, no exception path, and no door to knock on.

2. Micro-Segmentation — The Secure Neighborhoods

Even within the allowed radius, the system further restricts destinations. The vehicle is authorized only for specific neighborhoods — workstations, HMIs, or databases. It cannot stop at an unauthorized location simply because it is on an approved street. Each stop must be a designated one.

Technical Application: Micro-segmentation prevents lateral movement. A compromise in the General Office segment cannot migrate to Grid Control Systems because they are not on the same route — they are different neighborhoods with separate access credentials and separate path authorizations.

3. Verified Entitlements — The Credentialed Passenger

Access is never granted based on the vehicle (the device) alone. It is granted to the Identity. The car remains locked until the authorized passenger provides multi-factor, cryptographically sound authentication. The device is a vessel; the human identity and behavior are the true variables to be governed.

Technical Application: This is where the Human Layer becomes the primary security variable. It acknowledges that no technical control is more important than the integrity of the identity using it — and that identity must be verified continuously, not just at the point of entry.

IV. Dynamic Incident Response: The Safety Protocol

The most critical evolution in this model is the shift from static access control to Continuous Adaptive Trust. The system does not stop monitoring once the session begins.

Behavioral Monitoring — Erratic Driving

If a vehicle begins swerving or accelerating at dangerous speeds, a safety protocol is triggered. In network terms, this is User and Entity Behavior Analytics (UEBA). A user who typically accesses five files per day and suddenly attempts to retrieve five thousand has triggered a behavioral anomaly. The system recognizes this in real-time and escalates accordingly.

The Hijack Protocol — Active Containment

If a passenger attempts to seize control of the steering wheel — the analogy for malware attempting to gain administrative privileges — the system initiates an immediate shutdown sequence. The vehicle pulls over, removes itself from the flow of traffic (Network Isolation), and all digital keys are instantly revoked (Session Termination). No human decision is required at this layer; the response is automated and instantaneous.

V. The Sovereign Human Layer — The Framework's Critical Pillar

This is the most important section of this paper — and the most underrepresented concept in the existing Zero Trust literature.

Every autonomous system described above operates on a principle of internal coherence: it compares current behavior against an established baseline to detect anomalies. This is its strength. It is also its fundamental vulnerability.

The Poisoned Baseline Problem

Consider the following scenario. A sophisticated adversary does not attack the cars. They attack the traffic control system — the baseline itself. They corrupt the reference point against which all behavior is measured. The ZT platform now reports green across all dashboards. The UEBA sees no anomalies. The 3D visualization canvas looks clean. From the system's perspective, everything is functioning optimally.

Meanwhile, at the Security Operations Center, human analysts are receiving a completely different picture. Video feeds from inside vehicles show passengers behaving erratically. Street cameras capture vehicles moving in patterns that should not be possible. Phone calls come in from customers and authorities reporting dangerous behavior. The humans have access to out-of-band signals — sensor inputs that exist entirely outside the compromised system's chain of observation.

> *The system cannot see the problem because it is the problem. It is experiencing confident wrongness: internally coherent, externally catastrophic.*

This failure mode applies equally to AI-assisted security systems, where a model operating on corrupted training data or a poisoned context window will generate authoritative-sounding outputs that are fundamentally incorrect — with no internal signal to indicate the compromise. The confidence remains. The accuracy does not.

Out-of-Band Human Sovereignty — The Doctrine

The response to this failure mode is a formal doctrine: Out-of-Band Human Sovereignty. This principle holds that no automated system shall be the sole authority on its own health, and that human override capability must be maintained on a separate signal path that the governed system cannot access, influence, or corrupt.

Out-of-Band Human Sovereignty consists of four mandatory elements:

- Independent Observation Channels: Human operators must receive situational data from sensor networks that are architecturally separate from the monitored system. These may include physical surveillance, direct end-user communications, independent network taps, and cross-domain intelligence feeds.
- Pre-Validated Override Protocols: The human response to a compromised system must be documented, rehearsed, and pre-authorized before any incident occurs. In the moment of crisis, operators must execute a known protocol — not improvise one. This is

critical because a compromised system may attempt to influence human decision-making through its own outputs.

- Authority That Cannot Be Delegated Back: The override authority must rest with human operators whose judgment cannot be overridden by the system being evaluated. The hierarchy must be absolute: the system governs the network; the humans govern the system.
- Predetermined Shutdown and Recovery Sequences: When the override is invoked, the steps must be clear, fast, and independent of the compromised infrastructure. The "pull the car over" protocol must not rely on the car's own navigation system to find the shoulder.

VI. VStrike: The Sovereign Command and Control Center

If Zero Trust provides the autonomous vehicle fleet and its safety protocols, VStrike serves as the platform through which Out-of-Band Human Sovereignty is exercised. It transforms abstract network telemetry into a high-fidelity Cyber Canvas, giving human operators the situational clarity to make assertive decisions under uncertainty.

1. Restoring Context to the Authorities

In a traditional SOC, when an alert fires, analysts are buried in text-based logs. They understand what happened, but not where, how, or why in any visual or spatial sense. VStrike fuses disparate data streams into a unified 3D visualization. Authorities do not see an IP address — they see the specific neighborhood, the exact vehicle, and the precise route the threat has traveled. This compresses time-to-understanding from hours to seconds.

2. Assertive Decision-Making Through Storyboarding

When a hijack attempt occurs, VStrike's network flow mapping allows analysts to model the downstream consequences of their response before executing it. If a segment must be isolated, VStrike shows which adjacent systems will be affected — ensuring that the containment response does not inadvertently take down critical infrastructure. The authorities can see their decision before they make it.

3. The Out-of-Band Validation Layer

Most critically, VStrike serves as the independent observation platform described in the Out-of-Band Human Sovereignty doctrine. It continuously validates that policy boundaries remain intact, that behavioral baselines reflect reality, and that the autonomous system's self-reported health aligns with independently observed network behavior. When those two pictures diverge — when the system says green but VStrike's independent feeds say otherwise — that divergence is itself the alert.

> *VStrike does not merely show what the system sees. It shows what the system might be missing.*

VII. Framework Summary: The Four Pillars

The complete framework introduced in this paper rests on four integrated pillars:

- Behavioral Governance: Zero Trust continuous verification, micro-segmentation, and UEBA govern the movement and behavior of all identities within the network. The

autonomous car does not cross the boundary because it lacks the programming to do so.

- Active Containment: Automated incident response protocols — the hijack shutdown, network isolation, session revocation — execute at machine speed without requiring human authorization at the moment of trigger.
- Out-of-Band Human Sovereignty: A formal doctrine establishing that humans govern the automated system through independent signal channels and pre-validated override protocols. The humans do not depend on the cars to tell them the cars are malfunctioning.
- Sovereign Visualization: A platform such as VStrike that serves as the human operators' independent window into the network — separate from the system under governance, capable of surfacing the divergence between what the system reports and what is actually occurring.

VIII. Conclusion: The Hierarchy Must Hold

Zero Trust is not merely a collection of software. It is a philosophy of governance. The Autonomous Vehicle model makes that philosophy accessible to leadership: the car enforces the rules so that humans do not have to intervene in every transaction. The system scales. The humans focus on what only they can do.

But governance without a governor is not governance. It is automation. And automation, no matter how sophisticated, cannot certify its own integrity. A poisoned system will defend its corruption with the same confidence it once defended its health.

The doctrine of Out-of-Band Human Sovereignty closes this gap. It establishes, formally and without ambiguity, that the hierarchy runs in one direction: the autonomous system governs the network; the humans govern the autonomous system. That chain of authority must be protected as zealously as any cryptographic key.

The autonomous car provides safety. VStrike provides the vision. Out-of-Band Human Sovereignty provides the chain of command. Together, they ensure that the Human Layer is not just monitored, but mastered — and that when the system fails, the humans are already in position to take the wheel.

About the Author

Jose A. Cedeño is Principal Advisor at **Cofresí Consulting Services**, specializing in Zero Trust architecture, Critical Infrastructure protection, and Operational Technology security. He advises government and enterprise clients on behavioral security frameworks, human-centered governance models, and the intersection of AI systems and cybersecurity resilience.