

NOVAK PROTOCOL STANDARDS SERIES (NTM-1)




NOVAK Threat Model

Adversarial Baseline for Proof-Before-Action Systems (PBAS)

Version 1.0 (Final Draft) - Nov 2025

The Foundational Premise

The NOVAK Protocol establishes a new class of computing: **Proof-Before-Action (PBAS)**. No action executes without a cryptographically-bound, verifiable proof of correctness.

-  **Integrity:** Proof of correctness precedes execution.
-  **Determinism:** Output must be reproducible and predictable.
-  **Forensic Grade:** Every step is immutable and traceable.



NTM-1 is the formal baseline for all NOVAK defenses.





Six Adversary Classes

NTM-1 assumes adversaries have high capabilities and insider access.

- | | | |
|----------|-------------------------------------|--|
| A | Network Adversary (Dolev-Yao++) | <i>Intercepts, alters, and replays messages.</i> |
| B | Internal Privilege Adversary | <i>Insider threat, admin override abuse.</i> |
| C | Human PS-X Adversary | <i>Fraud, coercion, adversarial inputs.</i> |
| D | Physical PL-X Adversary | <i>Voltage drift, clock skew, timing injection.</i> |
| E | Automation / AI Adversary | <i>Self-modifying AI, malicious LLMs, deceptive outputs.</i> |
| F | Regulatory/Jurisdictional Adversary | <i>Conflicting laws, contradictory interpretations.</i> |

Adversary Goals

The primary goal of every NTM-1 adversary is to achieve ****undetected execution deviation****.

-  Bypass Proof-Before-Action controls.
-  Forge Execution Identity Receipts (EIR).
-  Reorder or delete Recursive Global Audit Chain (RGAC) entries.
-  Exploit timing, social, or physical flaws.

Core Vulnerability

****Non-Deterministic Execution****

If the same rule and input can produce two different outputs, the adversary wins.

The Six-Tuple Attack Surface

Every automated action is a 6-tuple that the adversary targets: ** **.

R

Rule Integrity (HR)

D

Input Attestation (HD)

O

Output Determinism (HO)

I

Execution Identity

T





Timestamp (Temporal)

E

Execution Receipt (EIR)




Adversary Class A, B, & C Attacks

Network & Privilege (A & B)

-  ****Rule Injection:**** Change rule logic silently (Class B).
-  ****EIR Forgery:**** Create fake pre-execution receipts (Class A).
-  ****Timestamp Tampering:**** Falsify T to confuse ordering (Class B).
-  ****Admin Override:**** Attempt to disable Safety Gate (Class B).





Human & Social (C - PS-X)

Attacks focused on exploiting cognitive biases and ambiguous phrasing.

-  ****Adversarial Input:**** Crafting input that fools the system.
-  ****Coercion/Fraud:**** Tricking operators into manual bypasses.
-  ****Misrepresentation:**** Altering context to influence rule interpretation.




Adversary Class D, E, & F Attacks

Physical & AI (D & E)

-  ****Voltage/Jitter:**** Fault injection to cause non-determinism (Class D).
-  ****Thermal Drift:**** Hardware instability causing bit errors (Class D).
-  ****Model Mutation:**** AI changes its output to evade detection (Class E).
-  ****Semantic Deception:**** AI generates outputs that are true but misleading (Class E).

Regulatory/Jurisdictional (F)

Exploiting the lack of rule consistency across governments and departments.

-  ****Conflicting Laws:**** Forcing contradictory execution paths.
-  ****Ambiguity:**** Exploiting vague statutory definitions to yield an adversarial outcome.
-  ****Interpretation Drift:**** Changing policy meaning over time.

Defense Matrix: Adversary vs. Defense

NOVAK Defense Component

****HVET**** (HR, HD, HO)

****EIR**** (Identity, Signature)

****RGAC**** (Chain Linkage)

****Safety Gate**** (Enforcement)

****PL-X**** (Physical Integrity)

****PS-X**** (Human Integrity)

Threat Class Blocked

****R, D, O Tampering****

****Token Forgery (E)****

****History Rewrite****






****PBA Bypass****

****Voltage/Timing****

****Fraud/Coercion****

Formal Security Goals (The Guarantees)

NOVAK systems must provide mathematical guarantees for all execution processes. Failure of any goal triggers an immediate block.

-  ****Rule Purity:**** Rules cannot change dynamically.
-  ****Input Integrity:**** Inputs must match the attested HD hash.
-  ****Output Integrity:**** Output must match the deterministic HO hash.
-  ****Identity Binding:**** Every action is tied to a verifiable actor.
-  ****Temporal Integrity:**** No timestamp falsification is possible.



Formal Security Claims

NOVAK Defends Against...

- ✓ **Silent Data Corruption**
- ✓ **Internal Admin Tampering**
- ✓ **Audit-Chain Rewrites**
- ✓ **AI-Driven Exploits**
- ✓ **Physical Adversarial Drift**

NOVAK Does NOT Defend Against...

- ✗ Broken Crypto Primitives (e.g., SHA-256)
- ✗ Fully Compromised Hardware Root-of-Trust
- ✗ Jurisdictions Lacking Rule-of-Law

Summary & Conclusion

A large, stylized blue number 6 with a slight glow effect, centered within a dark blue square.

**Adversary
Classes Modeled**

First-in-Class Protocol

NTM-1 ensures NOVAK is the world's first protocol where every computational action is protected by a unified layer spanning cryptography, hardware, and social integrity.

Outcome: A scientifically testable and cryptographically defensible execution environment.

Questions?

NOVAK Protocol Standards Series

Category: NTM-1 Threat Model