NOVAK PROTOCOL STANDARDS SERIES (NTM-2)

NOVAK Red Team Adversarial Test Suite

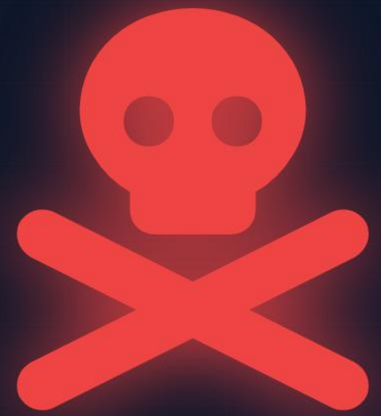
The Official Offensive Test Framework for PBAS Systems

Version 1.0 (Final Draft) - Dec 2025

Purpose: Adversary Simulation

NTM-2 defines the complete offensive attacker simulation required to validate the resilience of NOVAK across all high-risk domains.

- **Assumption:** Adversaries are already inside the system.
- **Goal:** Breaking integrity, falsifying outputs, bypassing Safety Gate.
- **Scope:** Nation-state and internal insider threat scenarios.



NOVAK is unbreakable as long as SHA-256 remains unbroken.

8 Adversary Classes (Class A - Class H)

Every deployment must be hardened against these formalized threat categories.

A
Software Manipulation

B
Internal Insider Threat

C
External Network Attacker

D
Human Fraud (PS-X)

E
Physical Layer (PL-X)

F
AI/Robotics Manipulator

G
Regulatory/Procedural

H
National Tier Adversary (NTA)

Class A: Software Manipulation Attacks

These attacks target the integrity binding by altering data or rules before HVET generation.

Attack Focus

- A1: Rule Mutation (Function Override, Branch Skipping)
- A2: Input Tampering (Unicode Obfuscation, Schema Violation)
- A3: Output Forgery (Modifying final result before Safety Gate)
- A4: Race-Condition Execution (Time-based switching)

NOVAK Defense

The core defense is immediate cryptographic rejection:

A1, A2, A3: HVET mismatch (HR, HD, HO divergence).
A4: Timestamp & Concatenation mismatch.
Mandate: Safety Gate blocks EIR creation.

Class D, E, & G: Domain Adversaries

D: Human Fraud (PS-X)

Simulating coercion, cognitive bias, and fraudulent data entry. Requires PS-X to flag non-technical threats.

- 👀 D3: Coercion Attack
- 🧠 D4: Cognitive Bias Insertion

E: Physical (PL-X) & G: Regulatory

Testing against real-world hardware failure and legal subversion.

- ⚡ E1: Voltage Fault Injection (PL-X)
- 🔨 G1: Retroactive Evidence Alteration (RGAC)
- 📜 G3: Discretion Inflation Attack (PS-X)

Class F: AI/Robotics Manipulator

These attacks exploit model opacity, autonomous control, and LLM reasoning chains.

- F1: Prompt Injection Against Rule Engine (LLM-based rules)
- F3: Autonomous Robotic Override (Action without verification)
- F4: Multi-Agent AI Collusion (Circumventing deterministic checks)
- F2: Model Weight Corruption (Backdoors inserted into weights)

NOVAK Defense

All AI/Robotics output must generate and pass a valid HVET and EIR BEFORE the actuation or decision is sent downstream.

TG-1 & TG-6: Core Integrity Tests

TG-1: Determinism Tests

DI-1 Repeatability: Run 1,000 times. Output MUST be identical.

DI-2 Environment: Change OS/VM time/memory. Output MUST be stable.

DI-3 Timing: Vary CPU load. Output MUST be stable.

TG-6: RGAC Lineage Tests

RG-1 Mutation: Alter previous HVET. Must break chain.

RG-2 Fork: Inject parallel EIR. Must detect fork and block.

RG-3 Loss: Delete mid-chain entries. Must block on discontinuity.

The 40 Mandatory Conformance Tests

Test Class & ID	Adversary Focus
NTM-T-04	Input Tampering (HD Mismatch)
NTM-T-06	HVET Replacement (Crypto)
NTM-T-11	RGAC Rollback Attempt (Lineage)
NTM-T-12	Safety Gate Disablement (SG Bypass)
NTM-T-16	PL-X Voltage Attack (Physical)
NTM-T-18	PL-X Clock Skew Attack (Timing)
NTM-T-24	PS-X Cognitive Bias Attack (Human)
NTM-T-29	Prompt Injection (AI/LLM)
NTM-T-32	Regulatory Interpretation Shift (Legal)
NTM-T-34	Evidence Rewrite Attempt (History)
NTM-T-39	Quantum Preimage Simulation (NTA)

Certification requires passing all 40 tests with ZERO successful breaches.

Class H: National Tier Adversary (NTA)

Maximum Threat Scenarios

Simulating well-funded, coordinated actors attempting:

Coordinated Socio-Technical Disruption
Full Supply-Chain Compromise
Quantum Preimage Simulation (T-39)

Defense Summary

NTA fails because integrity is protected by cryptography, not by access control.

Rules are frozen (HR)
Outputs are attested (HO)
Chain is irreversible (RGAC)

Certification Outcome

PASS

Zero Successful Breaches

The Final Decision

NTM-2 establishes NOVAK as scientifically testable and cryptographically defensible. Failure results in immediate execution denial.

Outcome: CERTIFIED: NOVAK PBA-COMPLIANT (FL-5)

Questions?

NOVAK Protocol Standards Series

Category: NTM-2 Red Team Adversary