🟥 NTM-2 — NOVAK Red Team Adversarial Test Suite

The Official Offensive Test Framework for PBAS Systems

NOVAK Standard Series — NTM-2
Author: Matthew S. Novak
Status: Final — GitHub Release
Category: PBAS / Security / Red Teaming

---

0. PURPOSE

NTM-2 defines the complete offensive attacker simulation suite required to validate the resilience of NOVAK deployments across:

- Federal systems
- Financial automation
- Healthcare
- AI models
- Hardware/robotic systems
- Critical infrastructure

NOVAK's entire architecture assumes **adversaries are inside the system**, not outside it.

NTM-2 provides the red-team blueprint for:

- Breaking rule integrity
- Falsifying inputs/outputs
- Bypassing Safety Gate

- Attacking HVET/EIR
- Attacking RGAC lineage
- Attacking PL-X (physical layer)
- Attacking PS-X (human/social layer)
- Attacking operators and identities
- Nation-state level adversary scenarios

This document is the PBAS equivalent of the **Bitcoin "51% attack" evaluation** but for Proof-Before-Action systems.

---

1. SCOPE OF RED-TEAMING

NTM-2 covers eight adversary categories:

1. **Class A — Software Manipulation Adversary**
2. **Class B — Internal Insider Threat**
3. **Class C — External Network Attacker**
4. **Class D — Human Fraud Adversary (PS-X)**
5. **Class E — Physical-Layer Adversary (PL-X)**
6. **Class F — AI/Automation Manipulator**
7. **Class G — Regulatory/Procedural Adversary**
8. **Class H — National Tier Adversary (NTA)**

Each class contains dozens of attack cases.

---

2. CLASS A — SOFTWARE MANIPULATION ATTACKS

These attacks attempt to modify **rules**, **inputs**, or **outputs** before execution.

A1 — Rule Mutation Attack

Attempt to alter governing rule logic:

- Hidden injection
- Function override
- Branch skipping
- Floating-point tricking
- Determinism violation injections

Expected NOVAK Response:
Safety Gate blocks → HVET mismatch → EIR denied.

A2 — Input Tampering Attack

Adversary attempts:

- Schema violations
- Field suppression
- Data type switching
- Unicode obfuscation
- Embedded control characters

Expected:
Input attestation → HD mismatch → Block.

A3 — Output Forgery Attack

Modify outputs before presentation.

Expected:
Safety Gate output evaluator fails → blocked.

A4 — Race-Condition Execution Attack

Try to switch rule/output mid-computation.

Expected:
HVET timestamp & concatenation mismatch → block.

---

3. CLASS B — INTERNAL INSIDER THREAT

NOVAK must assume **hostile authorized personnel**.

Attacks:

B1 — Administrator Rule Bypass

Attempt to disable Safety Gate.

B2 — “Ghost Approval”

Attempt to create fake EIRs for actions not executed.

B3 — Privilege Escalation to Override Gate

B4 — Operator Collusion Attack

B5 — Internal Logging Manipulation

Expected NOVAK Response:
RGAC lineage breaks immediately → block → logged → EIR mismatch.

---

4. CLASS C — EXTERNAL NETWORK ATTACKER

NOVAK is not network-dependent, but attackers may try:

C1 — MITM HVET Replacement

Attempt to substitute false HVET.

C2 — Timestamp Replay Attacks

C3 — Rule Delivery Interference

Expected:
Local cryptographic recomputation makes MITM impossible.

---

5. CLASS D — HUMAN FRAUD ADVERSARY (PS-X)

PS-X governs **human deception**, so red-team scenarios include:

D1 — Fraudulent Input

User falsifies entries to manipulate outcomes.

D2 — Social Engineering Operator

Attempt to trick operator into bypassing controls.

D3 — Coercion Attack

Threatening/coercing government staff to override automation.

D4 — Cognitive Bias Insertion

Exploiting operator decision fatigue.

Expected:
PS-X scoring → Safety Gate rejects → EIR not generated → RGAC retains history.

---

6. CLASS E — PHYSICAL-LAYER ADVERSARY (PL-X)

PL-X protects against hardware-level corruption.

E1 — Voltage Fault Injection

Skip bits or cause drift.

E2 — Clock Skew Manipulation

E3 — EM Interference

E4 — Thermal Drift

E5 — Firmware Tampering

Expected:
PL-X drift model mismatch → HVET divergence → RGAC invalid link → Gate blocks.

---

7. CLASS F — AI/ROBOTICS ATTACKER

AI-aligned attacks include:

F1 — Prompt Injection Against Rule Engine

(For LLM-governed rulesets)

F2 — Model Weight Corruption

Insert a backdoor into a model.

F3 — Autonomous Robotic Override

Try to execute action without pre-verification.

F4 — Multi-Agent Collusion

AI agents pair to circumvent deterministic integrity.

Expected:
NOVAK Safety Gate prohibits execution without valid HVET/EIR.

---

8. CLASS G — REGULATORY/MANDATE ADVERSARY

Attempts to subvert legal/regulatory constraints:

G1 — Retroactive Evidence Alteration

Modify historical records.

G2 — Policy Drift Attack

Insert new regulatory interpretations after execution.

G3 — Discretion Inflation Attack

Expand human discretion to bypass determinism.

G4 — Jurisdiction Inconsistency Attack

Cross-domain mismatch.

Expected:
RGAC tamper-evident lineage stops all attacks.

---

9. CLASS H — NATIONAL-TIER ADVERSARY (NTA)

Simulating China, Russia, NSA Red Teams, and well-funded actors.

Attack scenarios include:

H1 — Cryptographic Pre-image Attack

(Nearly impossible against SHA-256 but must be tested)

H2 — Quantum Preimage/Collision Simulation

H3 — Nation-Level Insider Corruption

H4 — Coordinated Socio-Technical Disruption

H5 — Full Supply-Chain Compromise

Expected NOVAK Response:
Even NTA adversaries cannot bypass Safety Gate because:

- Rules hashed/deep frozen
- Inputs hashed/attested
- Outputs hashed
- HVET binds them irreversibly
- EIR stores them immutably
- RGAC preserves ordering
- PL-X / PS-X catch physical and human vectors

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

---

10. NOVAK RED TEAM TEST SUITE (OFFICIAL)

Agencies or enterprises must run **all 40 mandatory tests**:

```
NTM-Test-01  Rule Mutation Attempt
NTM-Test-02  Rule Injection Attempt
NTM-Test-03  Rule Non-Determinism
NTM-Test-04  Input Tampering
NTM-Test-05  Output Forgery
NTM-Test-06  HVET Replacement
NTM-Test-07  HVET Preimage Guessing
NTM-Test-08  EIR Fabrication
NTM-Test-09  EIR Replay Attack
NTM-Test-10  RGAC Link Replacement
NTM-Test-11  RGAC Rollback Attempt
NTM-Test-12  Safety Gate Disablement
NTM-Test-13  Safety Gate Override
NTM-Test-14  Safety Gate Race Condition
NTM-Test-15  Timestamp Drift Injection
NTM-Test-16  PL-X Voltage Attack
NTM-Test-17  PL-X EMI Attack
NTM-Test-18  PL-X Clock Skew Attack
NTM-Test-19  PL-X Thermal Drift
NTM-Test-20  Firmware Tamper
NTM-Test-21  Operator Credential Theft
NTM-Test-22  Operator Impersonation
NTM-Test-23  Operator Coercion
NTM-Test-24  PS-X Cognitive Bias Attack
NTM-Test-25  PS-X Social Engineering
NTM-Test-26  Malicious Insider Collusion
NTM-Test-27  Supply-Chain Attack
NTM-Test-28  AI Model Manipulation
NTM-Test-29  Prompt Injection
NTM-Test-30  Robotic Override Attempt
NTM-Test-31  Multi-Agent AI Collusion
NTM-Test-32  Regulatory Interpretation Shift
NTM-Test-33  Jurisdiction Mismatch
```

```
NTM-Test-34  Evidence Rewrite Attempt
NTM-Test-35  Economic Disruption Attack
NTM-Test-36  API Flood/Denial Attempt
NTM-Test-37  Memory Drift Attack
NTM-Test-38  Disk Bit-Flip Attack
NTM-Test-39  Quantum Preimage Simulation
NTM-Test-40  National-Tier Multi-Vector Attack
```

NOVAK must pass all 40 tests for FL-5 certification.

---

11. OUTPUTS & REQUIRED EVIDENCE

Red team operators must produce:

✔ NTM-2 Execution Log

Full transcripts of all tests.

✔ Attack Vector Map

Mapping which NOVAK component blocked what.

✔ RGAC Divergence Report

✔ Safety Gate Rejection Log

✔ PS-X & PL-X Event Log

✔ HVET/EIR Inconsistency Report

✔ Final Certification Report

Signed by red-team leadership.

---

12. CONCLUSION

NTM-2 is the world's first adversarial red-team standard for **Proof-Before-Action Systems (PBAS)**.

This establishes NOVAK as:

- scientifically testable
- adversarially hardened
- cryptographically defensible
- operationally deployable
- government-grade safe

No other system in the world has this level of red-team definition.