NTM-1 — NOVAK THREAT MODEL

Adversarial Model, Attack Surface, and Defensive Guarantees

(PBAS-Class Execution-Integrity System Threat Specification)
NOVAK Protocol Standards Series — November 2025
Author: Matthew S. Novak

---

0. INTRODUCTION

The NOVAK Protocol defines a new class of computational systems:
PBAS — Proof-Before-Action Systems.

In such systems, *no action may execute* unless a verifiable, deterministic, cryptographically-bound proof demonstrates:

- rule purity
- input attestation
- output determinism
- identity binding
- timestamp correctness
- audit-chain continuity
- physical-layer stability
- psycho-social legitimacy

This threat model specifies the adversaries NOVAK must withstand and describes the attacks it must detect, reject, or make cryptographically impossible to execute.

NTM-1 is equivalent in purpose to:

- The Dolev–Yao adversary model (cryptographic protocols)
- Satoshi's threat model (Bitcoin)
- NIST SP-800 threat specifications
- FIPS-level cryptographic adversary frameworks

This is the formal adversarial baseline against which NOVAK's defenses (HVET, EIR, RGAC, Safety Gate, PL-X, PS-X, and NOVAK Laws L0–L15) are evaluated.

---

1. SYSTEM MODEL

NOVAK treats every computational action as a 6-tuple:

```
A = ⟨R, D, O, I, T, E⟩

R = Governing Rule
D = Attested Input
O = Deterministic Output
I = Execution Identity
T = Timestamp
E = Execution Receipt (EIR)
```

An action is permitted *if and only if*:

```
HVET(R, D, O, T) is valid
EIR is complete
RGAC continuity is maintained
Safety Gate returns TRUE
PL-X physical stability holds
PS-X human legitimacy validated
```

The threat model defines *how each component may be attacked* and *what NOVAK must guarantee*.

---

2. ADVERSARY MODEL OVERVIEW

NOVAK faces **six primary adversary classes**, each with distinct privileges, goals, and capabilities:

1. **Network-capable adversary (Dolev–Yao++)**
2. **Internal privilege adversary (insider threat)**
3. **Human/social adversary (PS-X)**
4. **Physical-layer adversary (PL-X)**
5. **Automation/AI adversary**
6. **Regulatory/jurisdictional adversary**

Each is described in detail below.

---

3. ADVERSARY CLASS A: DOLEV–YAO++ ADVERSARY

NOVAK extends the classical Dolev–Yao model.

The attacker can:

- intercept traffic
- alter messages
- reorder or replay messages
- delete messages
- spoof identities
- modify fields

- compromise networked intermediaries
- simulate nodes
- inject malformed rules

Assumption: No cryptographic primitive is breakable.
But: The adversary may exploit logic, determinism, timing, or input manipulation.

Attack Goals

- Bypass proof-before-action
- Modify D, R, or O before hashing
- Insert forged EIR
- Reorder RGAC entries
- Inject phantom identities
- Tamper with timestamps

NOVAK Defenses

- HVET binds R, D, O, T into a single tamper-evident vector
- EIR pre-execution receipt prevents post-action fraud
- RGAC links each execution to the previous (immutable chain)
- Laws L0–L15 disallow undetectable modification
- Deterministic rules prevent adversarial branching

---

4. ADVERSARY CLASS B: INTERNAL PRIVILEGE ADVERSARY

This class mirrors real-world fraud, insider tampering, administrative override abuse, and privileged manipulation.

The attacker may be:

- database admin
- developer
- SRE / operator
- cloud engineer
- contractor
- privileged system user

They may have partial or high access to:

- logs
- rules
- input sources
- output channels
- system state

Attack Goals

- Change a rule silently
- Inject bias into output
- Modify records or decisions retroactively
- Overwrite audit entries
- Circumvent deterministic checks
- Tamper with timestamps

- Hide traces of wrongdoing

NOVAK Defenses

- HVET breaks if any component changes
- EIR time-locks pre-execution truth
- RGAC prevents silent reordering or deletion
- Safety Gate blocks non-deterministic rules
- PL-X reveals impossible timing or drift
- Cross-domain attestation catches insider fraud

---

5. ADVERSARY CLASS C: HUMAN PS-X ADVERSARY

This class covers:

- fraud
- deception
- misrepresentation
- cognitive manipulation
- misleading inputs
- ambiguous statements
- intent obfuscation

The attacker may be:

- claimant
- citizen
- employee
- doctor
- judge
- reviewer
- AI content generator
- adversarial user

Attack Goals

- Fool the system into producing illegal or incorrect outputs
- Craft adversarial input text
- Misrepresent context to influence rule interpretation
- Bypass execution integrity through social means

NOVAK Defenses

- PS-X detects linguistic manipulation patterns
- Rule purity (L1–L4) prevents interpretation drift
- EIR locks pre-action semantics
- HVET ensures that *exact* wording is preserved
- Safety Gate rejects manipulated intent gradients

---

6. ADVERSARY CLASS D: PHYSICAL PL-X ADVERSARY

PL-X defines physical-layer threats:

- clock drift
- metastability
- voltage fluctuations
- bit-rot
- partial hardware failure
- timing jitter injection
- sensor ambiguity
- thermal noise
- race condition exploitation

Attack Goals

- Cause nondeterminism in rule execution
- Shift internal timing to confuse ordering
- Introduce errors into D or O
- Create hash mismatches below application layer

NOVAK Defenses

- PL-X monitors stability domains
- Time-bounded determinism checks

- Hash-anchored lineage prevents unnoticed corruption
- Safety Gate rejects execution if PL-X anomalies are detected
- RGAC prevents inconsistent ordering

---

7. ADVERSARY CLASS E: AUTOMATION / AI ADVERSARY

Automation adversaries include:

- self-modifying AI
- autonomous robots
- algorithmic financial systems
- autonomous fraud systems
- malicious ML models
- AI-generated rule bypass strategies

Attack Goals

- Modify rules on the fly
- Mutate outputs while preserving superficial validity
- Compute adversarial hash collisions (computationally impossible but attempted)
- Generate deceptive semantic inputs undetectable to humans
- Optimize toward bypassing Safety Gate

NOVAK Defenses

- Deterministic rules disallow adaptive branching
- EIR must match expected outputs
- HVET binds the full computation, not the text
- Safety Gate enforces semantic purity
- PS-X detects adversarial linguistic patterns
- PL-X detects timing irregularities typical of algorithmic adversaries

---

8. ADVERSARY CLASS F: REGULATORY / JURISDICTIONAL ADVERSARY

This class includes:

- conflicting laws
- inconsistent definitions
- overlapping jurisdictions
- contradictory interpretations
- adversarial statutory construction
- intentional ambiguity

Attack Goals

- Create rule ambiguity
- Force an inconsistent decision
- Construct contradictory rule paths

- Create a scenario where deterministic execution becomes impossible

NOVAK Defenses

- Law purity requirements (L11–L15)
- Regulatory determinism enforcement
- Mandatory rule canonicalization
- EIR records *exact legal foundation*
- RGAC preserves legal lineage
- Safety Gate rejects internally inconsistent laws

---

9. ATTACK SURFACE MATRIX

Target	Possible Attacks	NOVAK Defense
Input (D)	corruption, spoofing, replay, manipulation	HVET-HD, PS-X, PL-X
Rule (R)	tampering, injection, bias	HVET-HR, Safety Gate
Output (O)	modification, fraud, nondeterminism	HVET-HO, Rule determinism
Identity (I)	impersonation, token misuse	EIR identity binding
Timestamp (T)	drift, override, reordering	HVET timestamp, RGAC continuity
Receipt (EIR)	forgery, deletion, editing	chain linkage, digital signature
RGAC	reordering, trimming, insertion	hash-linked lineage

---

10. FORMAL SECURITY GOALS

NOVAK MUST guarantee:

10.1 Execution Non-Deviation

Output must equal the deterministic result of R(D).
Any deviation → BLOCKED.

10.2 Rule Purity

Rules may not change dynamically.
Any impurity → BLOCKED.

10.3 Input Integrity

Inputs must be cryptographically identical to the attested version.
Any mismatch → BLOCKED.

10.4 Output Integrity

Output hash must match the deterministic computation.
Any mismatch → BLOCKED.

10.5 Identity Binding

Every action must be tied to an identity and timestamp.

10.6 Irreversible Global Ordering

No silent reordering or deletion of events is possible.

10.7 PL-X Physical Stability

No timing, voltage, clock, or metastability manipulation may pass undetected.

10.8 PS-X Human Legitimacy

Fraud, manipulation, and adversarial inputs must be detected and prevented.

---

11. FORMAL SECURITY CLAIMS

NOVAK defends against:

✔ silent data corruption
✔ internal admin tampering
✔ rule manipulation
✔ fraudulent output modification
✔ timestamp falsification
✔ audit-chain rewrites
✔ AI-driven exploitation strategies
✔ physical adversarial drift
✔ social-engineered manipulation
✔ regulatory inconsistency attacks

NOVAK DOES NOT attempt to defend against:

✘ broken cryptographic primitives (SHA-256, etc.)
✘ fully compromised hardware root-of-trust
✘ jurisdictions where rule-of-law does not exist

---

12. CONCLUSION

NTM-1 establishes NOVAK as a **first-in-class execution-integrity protocol**.
This threat model is foundational for:

- SP-4 (Implementation Requirements)
- SP-5 (Certification Standards)
- The 90-page scientific paper
- ResearchGate publication
- Government evaluations (VA, DoD, OMB, DHS, Treasury, CMS)
- Investor due-diligence
- Academic peer review

NTM-1 completes the NOVAK foundational suite.