Verified programming against cybernukes

Ben Smyth · 30 April 2026

BSI-recommended encryption spotted not implementing requirements for identity checks: Client emerges off the street, wants entry into embassy filing room, guard demands identification — only the system doesn't. Anyone can walk in, remove, manipulate, even add classified material. No checks. Our pipeline caught it: 1 of 453 causes for concern — AI is building tactical cybernukes against critical infrastructure.

Pipelining internet security

Look at the scale of the thing.

```
content delivery     HTTP · DNSSEC · SMTP · MQTT · OAuth · Kerberos · LDAP
 ↳ Apache · nginx · BIND · Unbound · Postfix · Exim · Mosquitto · Keycloak · krb5 · OpenLDAP ...
transport security   TLS · QUIC · SSH · DTLS · IKEv2 · OpenVPN
 ↳ BoringSSL · Botan · dropbear · GnuTLS · JSSE · libreswan · LibreSSL · libssh · mbedTLS ·
   msquic · neqo · ngtcp2 · NSS · OpenIKED · OpenSSH · OpenSSL · OpenVPN · paramiko ·
   picoquic · quiche · quic-go · rustls · s2n · Schannel · strongSwan · wolfSSL ...
host-to-host comms   Ethernet · IP · TCP · UDP · ARP · NDP · ICMP · VLAN · eBPF
 ↳ Linux · FreeBSD · OpenBSD · NetBSD · macOS · Windows · Zephyr · DPDK · QNX · lwIP · seL4 ...
```

Governed by the Internet Engineering Task Force, driven by working groups drafting open standards, overwhelmingly implemented as open-source software maintained by small teams. This community is responsible for reliable host-to-host packet delivery, establishing security on top, and delivering content through those guarantees. Three thousand mandatory clauses across twenty-one stacks, one hundred and eleven codebases, fourteen languages, one-hundred million lines of code. Nobody ever checked they're right. They're not. Billions spent, decades of effort, specifying, auditing, deploying these protocols. Four hundred and fifty-three causes for concern raised. You break it, you fix it.

Brute-force search space dwarfs anything ever attempted — Bitcoin's lifetime mining surface is a rounding error in comparison — opening gambits span 2^64 possibilities, responses similar, joint handshake taking 2^256 shapes. Throwing the world's mining capacity at this, using state-of-the-art fuzzing, it wouldn't find a single bug before the sun burns out. The challenge is to compress into something tractable; we reduce exponential search to linear.

```
┌──── L12 · adjudicate ─────┐
│                           │
│ 🤖  ai hacking swarm       │
│     drives wire messages  │
│                           │
│ 🌐  critical infrastructure│
│     air · seas · ground   │
│                           │
│ 🎤  judge                  │
│     RFC violation forced ✓│
│                           │
└───────────────────────────┘
```

PROTOCOL

SECURITY ARENA

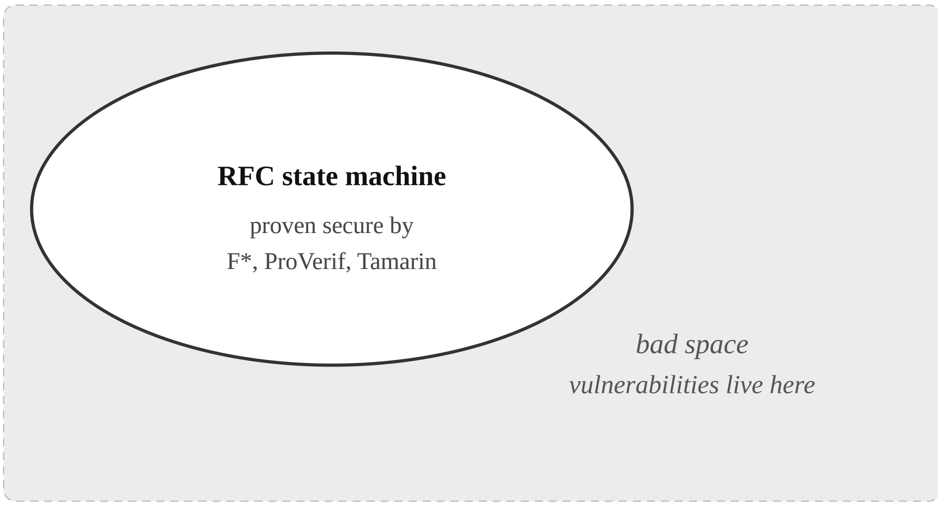
don't trust, verify

Pipeline proven on TLS and QUIC, fourteen zero days responsibly disclosed, three confirmed (CVE-2026-6772/32883/35582), six under my review, four hundred and thirty-three unopened.

AI hacking swarm

Anthropic's near-zero success at autonomous exploits showcases difficulty (Red Team blog, 6 March'26): Agents frazzle on exponential search. Pipeline compresses to linearity (over clauses).

Compliance is our vulnerability proxy. Formal methods have proven RFC state machine secure; inside, attacks cannot exist. Outside is where vulnerabilities live. Implementations duped by non-compliant wire messages are targets. Not all are insecurities, CVSS scoring skips the harmless. The rest are weaponised, agent holds defect-with-reach, exponential complexity erased.



Pipeline hands agent a wire probe forcing non-compliance. Discovery problem is gone: Anomalous behaviour observed, vulnerability site located, impact classified. Attack construction next.

Let's play a game; agent picks up probe, walks into arena, self-adjudication against catalogue drawn from published cryptography: Match-security (Dowling et al.), Forward Secrecy under Corrupt Party and 0-RTT replay (Fischlin-Günther), downgrade-resilience (Bhargavan et al.), connection-state-integrity (Davis-Günther). Each game names its win condition. The agent drives implementation to break.

A R E N A

🤖 swarm v botan 🌐

judge loaded match-security

Can swarm subvert guards:
Fake & steal classified?

Yes, decorative authentication, CVSS 10

Judge watches throughout. Agent iterates toward what the judge predicate accepts (guidance); false claims dismissed (filter). Six zero days in six days for a fiver, that's how the project began, currently triaging over four hundred causes for concern.

Cybernukes and verified programming

An adversary can blow up critical infrastructure with our tooling; it's dual use, verified programming, checking compliance, that's defensive, finding non-compliance, that's adversary foothold. Beyond changing beliefs of Germany, we've also spotted:

- **GHOSTSEAL —** Botan never verifies the seal on revocation responses. The verifier exists, the unit tests pass, the production path skips the call. Any TLS server with a key that ever chained to a public CA can forge "still good" responses for revoked certificates; MitM with a revoked-but-trusted cert intercepts traffic. CVE-2026-32883, disclosed 2026-03-29, fixed Botan 3.11.0.
- **PALEFIRE** — Twenty-two CertificateRequests crash any Firefox client with post-handshake auth enabled. 2.9 KB on the wire, under 100 ms, no client certificate required. Two attack surfaces: *extinction-switch* — a malicious server terminates every Firefox session that reaches it; *sustained-via-ads* — TLS endpoints embedded in ad networks bombard users until they migrate to a different browser. CVE-2026-6772, disclosed 2026-03-27, fixed NSS 3.123 / Firefox 150 (2026-04-21).
- **ORPHAN —** Adjacent attacker on the local link sends a Router Advertisement; kernel reads one byte past a slab allocation, leaks a neighbour's route metadata. Default Debian Trixie, RHEL 10, Alpine, Google COS, Oracle UEK, plus ~100M Android devices on the GKI 6.12 branch. Upstream-fixed in mainline 2026-02-19, never backported to stable. Rediscovered by our framework's host environment passively — we weren't even probing for it.

Top-tier TLS, QUIC, SSH currently in the Superleague. Kernel vulnerability surfaced where we weren't even looking — architecture itself needs smoke-testing; autonomous patching is about to become of unprecedented importance. AI deploying tactical cybernukes, taking out critical infrastructure:

Capabilities against protocols for comms-secure-delivery demonstrated, spinning up local infrastructure to bring lower tiers in superleague into play — entire internet is painfully unsafe.