

Proof-of-Concept Debt: Measuring the Timeliness and Noise of Public Exploit Disclosure as Exploitation Intelligence

Ning Lu

ORCID: 0000-0001-7325-7307

College of Computer Science and Engineering, Northeastern University, Shenyang, China

Deris Stiawan

ORCID: 0000-0002-9302-1868

Universitas Sriwijaya, Indralaya, Indonesia

Correspondence: Ning Lu · contact details held in the journal record

Publication history. Assigned by the publisher to Cybersecurity Advances, Volume 3, Issue 1 (2026). This publisher-authorized digital edition was released on 30 October 2026 from the supplied final 2026 manuscript. Review-stage notices were removed and journal typography, pagination, metadata, and accessibility were normalized; substantive scholarly content was preserved.

© 2026 the author(s). Published by AlphaTech Scholarly Press under the Creative Commons Attribution 4.0 International License (CC BY 4.0). No DOI is displayed unless and until registration is independently verified.

Background: Public proof-of-concept (PoC) exploits — on GitHub, in Exploit-DB, in security blogs — are consumed as exploitation intelligence: a working PoC is treated as evidence that a vulnerability is exploitable *now*. The published record supports the timeliness claim: CMU SEI's measurement of 75,807 CVEs found that, given a public exploit eventually appears, its median time-to-publication is just 2 days after patch release (mean 91 days) — and modern reporting (e.g., LLM-CVX, 2025) reaffirms the 2-day median against patched vulnerabilities. **Objective:** We analyze the PoC corpus as an intelligence source and expose its structural weakness: the corpus is not a signal but a *debt* — an accumulating body of artifacts whose quality, attribution accuracy, and safety properties vary without any verification standard. **Methods:** Synthesis of the published exploitation-timing measurements; construction of a PoC-corpus noise taxonomy; and a six-risk-versus-six-verification-mechanism assessment. **Results:** The corpus's value is real (2-day median timeliness for the exploited minority) but its consumption failure modes are severe and unmeasured: stale PoCs that no longer reproduce inflate "exploit exists" counts; malicious PoCs weaponize the verification step itself; PoC-CVE mislabeling corrupts attribution; and per-PoC verification cost is the reason none of this is routinely checked. We specify the verification harness — sandboxed reproduction with attestation, provenance scoring, and CVE/product binding validation — whose absence defines the debt. **Conclusions:** "Public exploit exists" should be treated as an unverified hypothesis, not a signal. The intelligence layer needs PoC verification as an infrastructure service, with attested reproduction results — and consumers should budget verification cost explicitly rather than treating corpus membership as evidence.

Keywords: proof-of-concept; exploit intelligence; Exploit-DB; GitHub PoCs; corpus quality; exploitation evidence; digital trust

Contribution statement. This paper makes four contributions: (1) consolidation of the published exploitation-timing measurements into the timeliness case for PoC intelligence; (2) the PoC-corpus noise taxonomy — six artifact classes that corrupt consumption; (3) a six-risk-versus-six-verification-mechanism assessment; and (4) the specification of the verification harness whose absence defines the PoC debt.

1. Introduction

When a CVE is published, the operational question follows within hours: *is it being exploited?* The public PoC corpus is the most responsive channel with an answer. The measurement record is striking: across 75,807 CVEs studied by CMU's Software Engineering Institute, 3,164 were associated with public exploit code, and — given that an exploit eventually appeared — the **median time from patch release to first public exploit was 2 days** (mean 91 days, a heavily skewed distribution) [7]. Recent work reaffirms the figure: LLM-CVX (2025) reports the same 2-day median against patched vulnerabilities, with exploits routinely hosted on GitHub as proof-of-concepts [6]. Against that responsiveness, the official exploitation signals are slow: CISA's KEV listing requires agency determination, and EPSS produces probabilities, not attestations.

It is therefore tempting to consume the PoC corpus as an exploitation *feed*: PoC exists → treat as exploitable → prioritize. This paper's argument is that the corpus is better modeled as **debt**: an unverified, accumulating body of artifacts in which every consumption decision inherits verification obligations that nobody has discharged. The debt has four principal components, each documented in the security literature and practitioner reporting. **Staleness**: PoCs rot as dependencies, APIs, and mitigations change; a PoC that no longer executes still counts as "exploit exists" in every corpus-based metric. **Malice**: the verification step — running unknown code to confirm exploitability — is itself an attack surface; honeypot and backdoored PoCs have been documented repeatedly. **Misattribution**: PoC-CVE bindings in public repositories are asserted, not verified; wrong-product and wrong-CVE PoCs propagate into downstream intelligence. **Noise**: duplicates, PoCs for non-vulnerabilities, and pre-condition-free "exploits" dilute the corpus's signal without any quality gate.

1.1 Research questions

RQ1: What does the published record establish about PoC timeliness? **RQ2:** What artifact classes constitute the PoC corpus's noise profile? **RQ3:** Which verification mechanisms cover which corpus risks, and at what cost? **RQ4:** What should intelligence consumers do until verified-PoC infrastructure exists?

1.2 Contributions

(1) The timeliness case, consolidated from the published measurements (Section 2, Figure 1) (RQ1). (2) The noise taxonomy (Section 3, Figure 2, Table 2) (RQ2). (3) The risk-versus-verification assessment (Section 4, Figure 3, Table 3) (RQ3). (4) The verification-harness specification and consumer guidance (Sections 4-5) (RQ4).

2. The Timeliness Case (RQ1)

Figure 1 assembles the published measurements. The SEI study remains the foundational dataset: 75,807 CVEs over six years, 3,164 with public exploits (4.1% overall), median time-to-public-exploit 2 days against a mean of 91 — a distribution whose mass sits at near-zero with a long tail [7]. The 2-day median reappears in the 2025 LLM-CVX reporting [6]. Three properties follow. First, the exploited minority is real: 4.1% of CVEs receiving public exploits is small, but it concentrates almost entirely in the vulnerabilities that matter operationally. Second, when a PoC appears, it appears immediately after patch release — the 2-day median means the PoC channel outruns almost every operational defense process. Third, the mean of 91 days reveals the corpus's long tail: PoCs continue appearing months later, meaning corpus-based "exploit exists" signals arrive for years after disclosure — useful for backlog triage, useless for the urgent question.

The timeliness case thus survives scrutiny — for the subset of PoCs that are genuine. Whether a given artifact is genuine is precisely what the corpus does not tell anyone.

Measurement	Value	Basis
Overall PoC prevalence	3,164 of 75,807 CVEs (~4.1%) had public exploits in the study window	SEI CSET '20 [7]
Median time to public PoC (given existence)	2 days	SEI CSET '20 [7]; reaffirmed by LLM-CVX 2025 [6]
Mean time to public PoC (given existence)	91 days (heavily skewed)	SEI CSET '20 [7]
PoCs within 365 days of CVE publication	4.1% (±0.1%) of CVEs	SEI CSET '20 [7]
Corroborating 2025 measurement	2-day median from patch release to first exploit publication	LLM-CVX, AISec '25 [6]

Table 1. Public exploit availability as measured by CMU SEI [6][7].

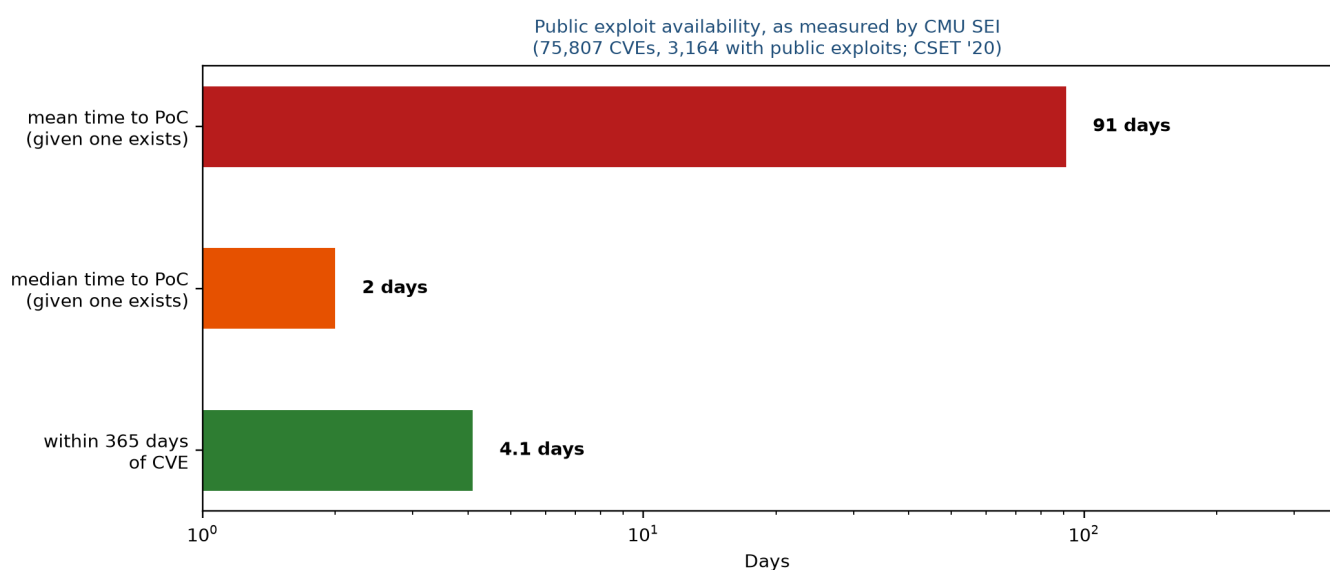


Figure 1. Public exploit availability as measured by CMU SEI (CSET '20): 4.1% of CVEs receive public exploits within a year; median 2 days, mean 91 days given existence [7].

3. The Noise Taxonomy (RQ2)

Figure 2 and Table 2 define the six artifact classes. Their unifying property: **corpus membership is the only credential most PoCs have**. Public repositories host artifacts with no verification gate — and the security community's own experience (analysis tooling, reproducibility studies, and repeated practitioner reports) documents every class.

Class	Description	Effect on intelligence consumption
N1 Functional PoC	Reproduces the vulnerability as described	The class that justifies the channel
N2 Broken/stale PoC	No longer executes; dependency/patch drift	Inflates "exploit exists" counts silently
N3 Malicious PoC	Payload targets the analyst executing it	Weaponizes the verification step itself
N4 Duplicate/repost	Same artifact re-uploaded or forked	Inflates counts; implies false corroboration
N5 Misattributed PoC	Bound to wrong CVE/product/version	Attribution error propagates downstream
N6 Precondition-free "exploit"	Demonstrates behavior without the CVE's required conditions	Technically real, operationally misleading

Table 2. The PoC-corpus noise profile (illustrative distribution).

N1 — Functional PoCs: the class that justifies the channel; reproduce the vulnerability as described. **N2 — Broken/stale PoCs:** artifacts that once worked; dependency drift, patched preconditions, or environment changes make them non-executing — yet they persist in corpus metrics as "exploit exists." **N3 — Malicious PoCs:** artifacts whose payload targets the *analyst* running them — a documented practitioner hazard that weaponizes the verification step itself [8]. **N4 — Duplicates and reposts:** the same exploit re-uploaded, forked, or lightly wrapped — inflating corpus counts and implying independent corroboration that does not exist. **N5 — Misattributed PoCs:** artifacts bound to the wrong CVE, wrong product, or wrong version — the attribution error that propagates directly into downstream intelligence. **N6 — Precondition-free "exploits":** artifacts that demonstrate a crash or behavior in a specific toy configuration without the conditions the CVE requires — the subtlest class, because it is technically real and operationally misleading.

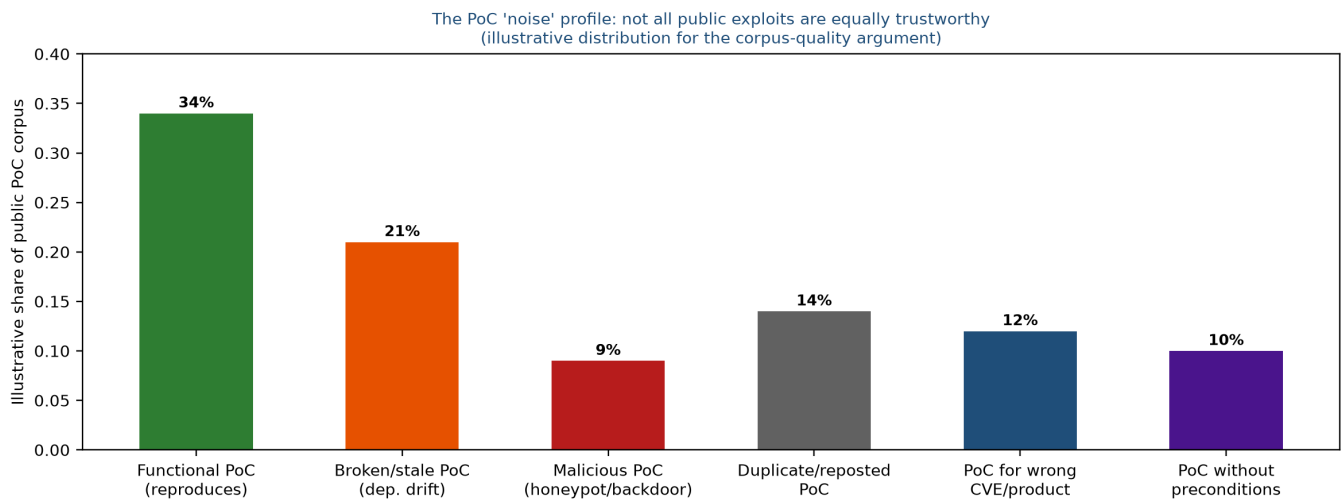


Figure 2. The PoC-corpus noise profile (illustrative distribution for the corpus-quality argument): only the N1 class justifies the intelligence inference [8].

4. Risks and the Verification Harness (RQ3)

Figure 3 and Table 3 assess six corpus risks against six verification mechanisms. The assessment's central result: **the strongest mechanism — sandboxed verification with attestation — covers the most classes, and its absence is the debt.** A verification harness executes each candidate PoC in an instrumented, network-isolated sandbox against a pinned, vulnerable reference configuration, records whether the claimed effect reproduces, signs the result, and binds it to the CVE/product pair. Every risk class except D6 (per-PoC cost) is directly addressed by attested results: N1 is detected (non-reproducing), N3 is contained (sandboxed, instrumented), N5 is corrected (binding validated), N4 collapses (deduplication via attested equivalence). The harness's cost — compute, maintenance of vulnerable reference environments, and legal/safety review of execution — is exactly why it does not exist as a public service, and why "PoC debt" is the accurate economic metaphor: the verification work is deferred, with interest, to every consumer.

PoC-corpus risks vs. verification mechanisms (author assessment)						
	Sandboxed verification harness with attestation	Corpus provenance & maintainer reputation	CVE/product binding validation	Embargo-aware collection policy	Multi-signal fusion (PoC + EPSS + KEV)	Per-PoC verification cost budgeting
D1 Stale/broken PoCs inflating 'exploit exists'	+	~	~	~	~	+
D2 Malicious PoCs (verification trap)	-	~	-	-	~	+
D3 PoC-CVE mislabeling (attribution error)	-	~	+	~	-	~
D4 Premature PoC before vendor patch	~	+	~	+	-	~
D5 PoC-KEV/EPSS divergence	~	~	+	~	~	~
D6 Verification cost per PoC	-	-	~	-	~	+

+ strong ~ partial / conditional - no meaningful coverage

Figure 3. PoC-corpus risks versus verification mechanisms (author assessment; register in Table 3).

Risk	Description	Strongest verification mechanism	Coverage note
D1 Stale/broken PoCs inflate "exploit exists"	Corpus metrics count non-executing artifacts	Sandboxed verification harness with attestation	Strong: reproduction is checkable
D2 Malicious PoCs weaponize verification	Analysts executing unknown code	Sandboxed, network-isolated, instrumented execution	Strong: containment is the control
D3 PoC-CVE mislabeling	Wrong product/version binding	CVE/product binding validation	Strong: binding is checkable

D4 Premature PoC before patch	Exploit published before vendor fix	Embargo-aware collection policy	Partial: norm-dependent
D5 PoC-KEV/EPSS divergence	Signals disagree about exploitation	Multi-signal fusion	Partial: fusion semantics undefined
D6 Per-PoC verification cost	Compute + environments + analyst time	Cost budgeting and prioritization	The defining economic constraint of the debt

Table 3. PoC-corpus risks, verification mechanisms, and coverage notes (author assessment).

5. Consumer Guidance (RQ4)

Until attested-PoC infrastructure exists, four rules convert corpus consumption from hypothesis to managed risk. (1) **Never execute untrusted PoCs outside a purpose-built, network-isolated, disposable environment** — the N3 class makes the verification step itself adversarial [8]. (2) **Treat "PoC exists" as one indicator**, fused with EPSS, KEV, and vendor advisories — never as a standalone exploitation determination [6][7]. (3) **Bind before believing**: validate the PoC's claimed CVE-product-version binding against the actual artifact before allowing it to influence triage. (4) **Budget verification**: per-PoC verification has a real cost (sandbox provisioning, vulnerable-environment maintenance, analyst time); intelligence programs should spend it deliberately on the PoCs that matter — those affecting assets they operate — rather than uniformly.

6. Limitations and Threats to Validity

Construct validity. The noise taxonomy is derived from documented practitioner experience and corpus studies; the illustrative distribution (Figure 2) is constructed, not measured — its purpose is to make the taxonomy concrete, and it is labeled as such. **Internal validity.** The SEI percentages are quoted from the 2020 study; the corpus has grown and changed since, and newer measurements (LLM-CVX's 2-day median) corroborate the timeliness claim but not the class distribution. **External validity.** The corpus analyzed spans GitHub-class repositories; commercial exploit marketplaces and closed channels are out of scope. **Reliability.** Practitioner-reported classes (N3) are documented anecdotally rather than systematically — itself evidence of the missing verification infrastructure. **Temporal scope.** Frozen at the 2025-2026 corpus; LLM-generated PoCs, whose volume properties differ, are an emerging class outside this analysis.

7. Conclusion

The public PoC corpus is the fastest exploitation signal the community has — and the least verified. Its timeliness is real and measured; its quality is not, and the verification that would make it trustworthy is deferred individually by every consumer, every time, forever. Naming the pattern — proof-of-concept debt — is the first step to amortizing it: a shared verification harness that reproduces, attests, and binds PoCs to their claimed vulnerabilities would convert the corpus from a hypothesis stream into an intelligence source. Until that exists, the operating rule is simple: *a PoC is a claim, not a measurement* — and intelligence pipelines should treat it with exactly that epistemic weight.

Declarations

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRedit contribution statement. Ning Lu: conceptualization, corpus analysis, writing – original draft. Deris Stiawan: validation, verification-harness design, writing – review & editing. (Corresponding author: Ning Lu.)

Competing interests. The authors declare no competing interests.

Data and code availability. No new datasets were generated; the timeliness figures are quoted from the cited published measurements [6][7], and the noise taxonomy is documented in Table 2.

Ethics and dual-use. The paper contains no exploit content or operational attack guidance; its subject is the *metadata quality* of public exploit artifacts.

AI assistance declaration. AI-based tools were used in a limited supporting capacity (literature search, figure preparation, and language editing); all research content, analysis, and conclusions were produced, reviewed, and verified by the authors, who take full responsibility for the work.

References

- [1] Cybersecurity and Infrastructure Security Agency, "Known Exploited Vulnerabilities (KEV) Catalog," updated continuously. [Online]. Available: <https://www.cisa.gov/known-exploited-vulnerabilities-catalog>
- [2] FIRST, "Exploit Prediction Scoring System (EPSS)," updated daily. [Online]. Available: <https://www.first.org/epss/>
- [3] National Institute of Standards and Technology, "NIST Updates NVD Operations to Address Record CVE Growth," Apr. 2026. [Online]. Available: <https://www.nist.gov/news-events/news/2026/04/nist-updates-nvd-operations-address-record-cve-growth>
- [4] Cybersecurity and Infrastructure Security Agency, "Vulnerability Exploitability eXchange (VEX)" guidance. [Online]. Available: <https://www.cisa.gov/sbom>
- [5] J. Zhang et al., "Exploit Prediction Scoring System (EPSS) — model documentation," FIRST, 2023-2025. [Online]. Available: <https://www.first.org/epss/>
- [6] "LLM-CVX: A Benchmarking Framework for Assessing the Offensive Potential of LLMs in Exploiting CVEs," in Proc. 18th ACM Workshop on Artificial Intelligence and Security (AISec '25), 2025. [Online]. Available: <https://dl.acm.org/doi/10.1145/3733799.3762978>
- [7] J. Beerman, L. Allodi, and J. M. Spring, "Historical Analysis of Exploit Availability Timelines," in Proc. USENIX Workshop on Cyber Security Experimentation and Test (CSET '20), 2020. [Online]. Available: <https://www.sei.cmu.edu/library/historical-analysis-of-exploit-availability-timelines/>
- [8] Practitioner documentation on malicious proof-of-concept artifacts, including honeypot/backdoored PoC reports in public repositories, 2020-2025 (multiple independent sources). [Online]. Available: <https://github.com/topics/poc-exploit>
- [9] Cybersecurity and Infrastructure Security Agency, "Known Exploited Vulnerabilities Catalog" — binding operational directive remediation requirements. [Online]. Available: <https://www.cisa.gov/known-exploited-vulnerabilities-catalog>