IKriS Institut für Kriminalitäts- und Sicherheitsforschung

POLIZEIAKADEMIE NIEDERSACHSEN

Research Brief

Forschung an der
POLIZEIAKADEMIE NIEDERSACHSEN

2026, # 10

Nutzung von Künstlicher Intelligenz zur Analyse von Certificate-Transparency-Logs zur Erkennung von Phishing-Domains

Andreas Knüttel

Abstract:
Phishing-Angriffe bleiben eine zentrale Bedrohung, wobei Angreifer zunehmend täuschend echt gestaltete, TLS-gesicherte Domains nutzen. Certificate-Transparency-(CT)-Logs stellen als öffentlich verfügbare Datenquelle ein vielversprechendes Mittel dar, um solche Domains frühzeitig zu erkennen. Dieser Beitrag untersucht daher den Einsatz von Künstlicher Intelligenz und Python zur automatisierten Analyse von CT-Logs mit dem Ziel, potenziell phishingverdächtige Domains unmittelbar nach Zertifikatsausstellung zu identifizieren.

Eine Pipeline verarbeitet CT-Daten in Echtzeit, extrahiert relevante Domain-Features und trainiert ein Machine-Learning-Modell zur Klassifikation. In einer prototypischen Evaluation erreicht der Ansatz eine Präzision von über 90 % und identifiziert Domains, die zum Erkennungszeitpunkt noch nicht in Sicherheitsdatenbanken gelistet sind. Die Ergebnisse verdeutlichen das Potenzial für proaktive Phishing-Abwehr und Brand-Monitoring, zeigen jedoch auch Herausforderungen wie False Positives und adaptive Angreiferstrategien.

Einführung

Phishing zählt weiterhin zu den erfolgreichsten Angriffsmethoden und verursacht erhebliche wirtschaftliche Schäden. Aktuelle Berichte der Anti-Phishing Working Group zeigen für 2023 und 2024 nahezu fünf Millionen Angriffe jährlich, insbesondere im Umfeld von Finanzdienstleistern und SaaS-Anbietern (APWG, 2024, 2025). Gleichzeitig hat sich die technische Qualität der Angriffe deutlich erhöht: Realistische

Webseiten, TLS-Zertifikate und flächendeckendes HTTPS erschweren eine rein visuelle Erkennung (Haraldsdóttir et al., 2024; Mohammad et al., 2014).
CT-Logs wurden zur Stärkung des Vertrauens in das Web-PKI-Ökosystem eingeführt und liefern eine öffentliche, unveränderliche Chronik ausgestellter TLS-Zertifikate (Laurie et al., 2013). Da viele Zertifikate nur noch mit CT-Eintrag als vertrauenswürdig gelten, ermöglichen sie eine nahezu vollständige Sicht auf neu ausgestellte Zertifikate (Holz et al., 2016). Diese Transparenz eröffnet neue Möglichkeiten zur frühzeitigen Erkennung von Phishing-Domains.

Vorarbeiten wie Phish-Hook zeigen, dass CT-Daten für ML-basierte Phishing-Erkennung geeignet sind (Fasllija et al., 2019). Der vorliegende Beitrag greift dies auf und untersucht eine KI-gestützte Analyse von CT-Logs zur frühzeitigen Identifikation und forensischen Dokumentation verdächtiger Domains.

Vorgehen

Systemarchitektur und Datenerfassung

Das System folgt einer Pipeline aus Datenerfassung, Aufbereitung, Analyse und Ergebnisbereitstellung. CT-Daten werden über Schnittstellen wie Certstream oder die Google-API als Echtzeitstrom bezogen (Google, o. J.; Phishbeat, 2020). Die Verarbeitung erfolgt in Python unter Nutzung etablierter Bibliotheken für Webkommunikation und Datenanalyse.

Die CT-Einträge werden in strukturierte Formate überführt (z. B. JSON) und enthalten unter anderem Zertifikatssubjekt, SANs, Aussteller und Zeitstempel. Im Fokus stehen die extrahierten Domainnamen als zentrale Analyseeinheit.

Datenaufbereitung und Feature-Engineering

Aus den Domain- und Zertifikatsdaten werden strukturelle, linguistische und zeitliche Merkmale abgeleitet. Dazu zählen u. a. Domainlänge, Subdomain-Struktur, Zeichenverteilung und Entropie sowie n-Gram-Muster, Markenähnlichkeit und Levenshtein-Distanzen (AlSabah et al., 2022; Xu et al., 2020).

Für das Training werden Domains mittels externer Phishing-Feeds und Whitelists gelabelt. Ein ausgewogenes Verhältnis der Klassen verhindert Verzerrungen im Modell.

KI-gestützte Analyse

Zur Klassifikation werden überwachte Lernverfahren wie Random Forests oder Gradient Boosting eingesetzt, die sich für tabellarische Daten als robust erwiesen haben (Haraldsdóttir et al., 2024). Training und Evaluation erfolgen auf historischen, gelabelten CT-Daten unter Einsatz von Kreuzvalidierung und Hyperparameteroptimierung.

Das Modell liefert Wahrscheinlichkeiten für Phishing-Zuordnungen, die über geeignete Schwellenwerte in operative Entscheidungen überführt werden.

Ergebnisinterpretation und Benachrichtigung

Die Ergebnisse werden als priorisierte Liste verdächtiger Domains ausgegeben, ergänzt um erklärende Merkmale zur Nachvollziehbarkeit.

CT-Logs bieten zudem eine manipulationsresistente Grundlage für die forensische Dokumentation. Zeitstempel und Zertifikatsdaten ermöglichen die Rekonstruktion von Angriffskampagnen und unterstützen Ermittlungsverfahren (Laurie et al., 2013; Pletinckx et al., 2023).

Empirie und Ergebnisse

Die Evaluation basiert auf historischen CT-Daten über mehrere Wochen mit zehntausenden gelabelten Domains. Das Modell erreicht eine Präzision von über 90 % und identifiziert zahlreiche Domains vor deren Aufnahme in Sicherheitsdatenbanken.

Ein besonders relevanter Befund war, dass ein signifikanter Anteil dieser verdächtigen Domains in gängigen Sicherheitsdatenbanken zum Zeitpunkt der Erkennung noch nicht gelistet war. Dies unterstreicht das Potenzial von CT-basierten Verfahren, neu aufgesetzte Phishing-Infrastrukturen zu identifizieren, bevor sie in großem Umfang in Kampagnen eingesetzt oder in Blocklisten und kommerzielle Threat-Feeds übernommen werden.

Die Analyse der Feature-Importances zeigt, dass insbesondere Zeichenstruktur, Markenähnlichkeit und zeitliche Muster entscheidend sind. Damit lässt sich Phishing bereits auf Domain- und Zertifikatsebene erkennen, ohne Inhalte analysieren zu müssen.

Aus polizeilicher Sicht ist hervorzuheben, dass die prototypische Umsetzung neben der Detektion auch die Beweissicherung unterstützt. Durch die Speicherung relevanter CT-Einträge inklusive Zeitstempeln und Modelldiagnosen lässt sich nachvollziehen, wann eine Domain erstmals auffällig wurde. Dies erleichtert die strafprozessuale Beweisführung und unterstützt die Priorisierung von Maßnahmen wie Takedown-Anfragen bei Registraren oder Hosting-Providern.

Implikationen

Die Ergebnisse zeigen, dass die KI-gestützte Analyse von CT-Logs ein vielversprechender Baustein proaktiver Phishing-Abwehr ist. Im Gegensatz zu reaktiven Verfahren setzt der Ansatz bereits bei der Zertifikatsausstellung an und ermöglicht so einen zeitlichen Vorsprung zur Identifikation und potenziellen Blockierung verdächtiger Domains.

Vor dem Hintergrund der fortschreitenden Automatisierung von Phishing, einschließlich KI-generierter Phishing-Mails und dynamischer Angriffsinfrastrukturen, kommt frühzeitigen, datengetriebenen Detektionsmechanismen eine zentrale Rolle zu. Die vorgestellte Nutzung von Python und KI zur Analyse von CT-Logs zeigt, dass ein solcher Ansatz nicht nur technisch umsetzbar ist, sondern auch einen konkreten Mehrwert für Prävention, Incident Response und Beweissicherung liefert und damit einen substantiellen Beitrag zur Bekämpfung von Phishing-Angriffen leisten kann.

Literatur

AlSabah, M., AlMansoori, N., Alketbi, S., & Khader, M. (2022). Content-agnostic detection of phishing domains using certificate transparency logs and passive DNS. In Proceedings of the 2022 ACM Workshop on Artificial Intelligence and Security (AISec) (S. 1–12).
APWG. (2024). Phishing activity trends report, 1st quarter 2024. Anti-Phishing Working Group.
APWG. (2025). Phishing activity trends report, 4th quarter 2024. Anti-Phishing Working Group.
Fasllija, E., Happe, A., & Hof, H.-J. (2019). Phish-Hook: Detecting phishing certificates using certificate transparency logs. In Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES 2019) (Article 29).
Google. (o. J.). Certificate transparency API documentation. https://developers.google.com/certificate-transparency. Aufgerufen am 19.03.2026.
Haraldsdóttir, M., Homayoun, S., Lynge, E., & Jensen, C. D. (2024). Unmasking phishers: ML for malicious certificate detection. Computers and Industrial Engineering, 198, Article 110652. https://doi.org/10.1016/j.cie.2024.110652
Holz, R., Amann, J., Mehani, O., & Kaafar, M. A. (2016). TLS in the wild: An Internet-wide analysis of TLS-based protocols for electronic communication. In Proceedings of the Network and Distributed System Security Symposium (NDSS 2016).
Laurie, B., Langley, A., & Kasper, E. (2013). Certificate transparency. In Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13) (S. 327–337).
Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting phishing websites based on self-structuring neural network. Neural Computing and Applications, 25(2), (S. 443–458).
Pletinckx, S., Amann, J., Gasser, O., & Carle, G. (2023). Using certificate transparency logs for target reconnaissance. In Proceedings of the 31st USENIX Security Symposium (S. 1–16).
Phishbeat. (2020). Phishbeat: Real-time phishing detection from CT logs [Open-source project]. https://github.com/stric-co/phishbeat. Aufgerufen am 19.03.2026.
Xu, F., Chu, B., & Cai, G. (2020). Detecting phishing websites through machine learning techniques. In 2020 IEEE International Conference on Computer Science, Engineering and Applications (ICCSEA) (S. 1–8).

POLIZEIAKADEMIE
NIEDERSACHSEN

IKriS Institut für Kriminalitäts- und Sicherheitsforschung

Management und Redaktion:
PD Dr. Sybille Reinke de Buitrago
sybille.reinke-de-buitrago@polizei.niedersachsen.de

IKriS – Institut für Kriminalitäts- und Sicherheitsforschung
Polizeiakademie Niedersachsen
Bürgermeister-Stahn-Wall 9
31582 Nienburg/Weser
Tel.: (+49) 05021 844-1880
ikris@pa.polizei.niedersachsen.de

Prof. a.d. PA Dr. Andreas Knüttel ist an der Polizeiakademie Nienburg, Standort Hannover, im Studiengebiet 1 tätig und forscht zu Cybercrime-Phänomenen, insbesondere Phishing und Ransomware sowie KI/ML.
Andreas.Knuettel@polizei.niedersachsen.de

Der Beitrag gibt die Auffassung des/der Autor*in wieder.