Cybersecurity and Law

2026 Nr 1(15)

DOI: 10.35467/cal/221058

Automatyczna identyfikacja infrastruktury przestępczej w przestrzeni IPv4 i IPv6 przy użyciu lokalnych skryptów OSINT: metody, ograniczenia i praktyczne zastosowania w cyberbezpieczeństwie

Automatic Identification of Criminal Infrastructure in IPv4 and IPv6 Space Using Local OSINT Scripts: Methods, Limitations and Practical Applications in Cybersecurity

Tomasz JANCZEWSKI
Akademia Marynarki Wojennej
ORCID: 0009-0006-4583-4377
E-mail: tomasz@janczewski.it

Streszczenie

W niniejszym artykule przedstawiono metodologiczne ramy automatycznej identyfikacji złośliwej lub podejrzanej infrastruktury w globalnej przestrzeni adresowej IPv4 i IPv6 z wykorzystaniem technik opartych na OSINT, wykonywanych lokalnie. Proponowane podejście pokazuje, jak publicznie dostępne źródła informacji o zagrożeniach, dane przejrzystości certyfikatów, rekordy ASN, pasywne wpisy DNS, źródła reputacji IP i zestawy danych odcisków palców TLS można zintegrować w ujednolicony proces analizy. Badanie ustanawia powtarzalny model oceny dużych próbek adresów IP bez polegania na skanowaniu zewnętrznym, zapewniając tym samym zgodność z prawem i etyką. Artykuł przedstawia uzasadnienie wyboru konkretnych źródeł OSINT, omawia korzyści i ograniczenia pasywnego gromadzenia informacji oraz podkreśla znaczenie analizy korelacji między wieloma niezależnymi zbiorami danych. Wyniki eksperymentów potwierdzają, że pasywny OSINT umożliwia wykrywanie zależności strukturalnych i powtarzalnych wzorców zachowań w dużych próbkach adresów IP – w tym sygnatur rotacji domen, klastrowania certyfikatów i koncentracji reputacji na poziomie ASN. Metodologia wspiera tworzenie lekkiego, lokalnego środowiska badawczego odpowiedniego do badań akademickich, reagowania na incydenty i monitorowania zagrożeń bazowych. Na podstawie ustaleń w artykule zaproponowano kierunki przyszłych badań, w tym integrację modułów klasyfikacji uczenia maszynowego, zautomatyzowanych procesów wzbogacania, analizę czasową dynamiki zmian infrastruktury i potencjalne wykorzystanie punktacji ryzyka opartej na OSINT jako sygnału uzupełniającego dla ram Zero Trust.

Słowa kluczowe

OSINT, wykrywanie zagrożeń, analiza IPv4/IPv6, automatyzacja cyberbezpieczeństwa, wykrywanie anomalii, sieciowe odciski palców

Abstract

This article presents a methodological framework for the automated identification of malicious or suspicious infrastructure within the global IPv4 and IPv6 address space using OSINT-based techniques executed locally. The proposed approach demonstrates how publicly available threat intelligence sources, certificate transparency data, ASN records, Passive DNS entries, IP reputation feeds, and TLS fingerprinting datasets can be integrated into a unified analysis pipeline. The study establishes a reproducible model for evaluating large samples of IP addresses without relying on external scanning, thereby ensuring legal and ethical compliance. The article outlines the reasoning behind selecting specific OSINT sources, discusses the benefits and limitations of passive intelligence collection, and emphasizes the importance of correlation analysis between multiple independent datasets. Experimental results confirm that passive OSINT enables the detection of structural dependencies and repeatable behavioral patterns in large IP address samples – including domain rotation signatures, certificate clustering, and ASN-level reputation concentration. The methodology supports the creation of a lightweight, local research environment suitable for academic studies, incident response, and baseline threat monitoring. Based on the findings, the paper proposes directions for future research, including the integration of machine learning classification modules, automated enrichment pipelines, temporal analysis of infrastructure change dynamics, and the potential use of OSINT-driven risk scoring as a complementary signal for Zero Trust frameworks.

Keywords

OSINT, Threat Intelligence, IPv4/IPv6 Analysis, Cybersecurity Automation, Anomaly Detection, Network Fingerprinting

Wprowadzenie

W dobie gwałtownego rozwoju technologii cyfrowych i rosnącej globalnej łączności sieciowej, tradycyjne podejścia do bezpieczeństwa informacji – oparte na ochronie perymetrycznej, zamkniętych systemach czy zabezpieczeniu konfiguracji – okazują się niewystarczające. Nowe zagrożenia pojawiają się dynamicznie, a złośliwi aktorzy wykorzystują publicznie dostępne dane, często odpowiednio przygotowane za pomocą sztucznej inteligencji, co w naturalny sposób powiększa i tak już sporą powierzchnię ataku. W tym kontekście rosnące znaczenie zyskuje dyscyplina znana jako Open Source Intelligence (OSINT) – rozumiana jako gromadzenie i analiza informacji pochodzących wyłącznie z publicznych, legalnie dostępnych źródeł. W niniejszym artykule proponuję metodologię automatycznej identyfikacji podejrzanej infrastruktury sieciowej (adresów IP w przestrzeni IPv4/IPv6) przy pomocy lokalnych narzędzi, co pozwoli na stworzenie powtarzalnego, etycznego i niskobudżetowego narzędzia analitycznego.

OSINT zyskało uznanie nie tylko w agencjach wywiadowczych czy służbach bezpieczeństwa, ale także w sektorze prywatnym i akademickim jako skuteczne narzędzie analizy zagrożeń i ryzyka. Jak wskazują autorzy Böhm i Lolagar, OSINT to zbiór narzędzi i metod pozwalających wydobywać informacje

z publicznie dostępnych źródeł[1]. Dzięki dostępności i szybkości informacji możliwe jest przekształcenie surowych danych w realną „inteligencję" (ang. actionable intelligence) użyteczną w celach operacyjnych lub badawczych[2]. Pomimo rosnącego zainteresowania OSINT, nadal istnieje wyraźne zapotrzebowanie na systematyczne, powtarzalne metody ze szczególnym uwzględnieniem automatyzacji w masowej analizie danych, uwalniające moce przerobowe manualnej pracy eksperckiej. Ma to szczególne znaczenie w kontekście identyfikacji infrastruktury związanej z cyberprzestępczością: botnetów, serwerów C2, zaplecza phishingu, serwerów dystrybucji malware czy hostów wykorzystywanych w atakach DDoS. Tradycyjny ręczny OSINT, choć wartościowy, jest pracochłonny i trudny do skalowania.

Główny problem badawczy, który stawiam w niniejszym artykule, brzmi: czy możliwe jest zautomatyzowane, lokalne narzędzie OSINT, które w sposób powtarzalny, przy pomocy publicznie dostępnych źródeł, zidentyfikuje adresy IP lub bloki adresów należące do infrastruktury potencjalnie złośliwej lub podejrzanej? Artykuł stanowi wstęp do szerszego cyklu badań: od podstawowej automatycznej identyfikacji infrastruktury, przez budowę baz danych, aż po zaawansowaną analizę zagrożeń, automatyczne alerty i integrację z systemami obronnymi.

Przegląd literatury i kontekst badań

Analiza współczesnych badań nad OSINT pokazuje, że jego rola w cyberbezpieczeństwie rośnie równolegle do wzrostu złożoności zagrożeń sieciowych oraz coraz większych wymagań dotyczących dostępności danych wywiadowczych w czasie zbliżonym do rzeczywistego. Pastor-Galindo i in. wskazują, że OSINT, mimo ogromnego potencjału, pozostaje w dużej mierze „niewykorzystaną kopalnią wiedzy", której pełne zastosowanie wymaga zarówno automatyzacji, jak i standaryzacji procesu pozyskiwania i analizy danych[3]. Szymoniak i Foks podkreślają natomiast, że OSINT obejmuje nie tylko dane z mediów społecznościowych, lecz także metadane plików, informacje organizacyjne, dane sieciowe i ogólnodostępne rejestry – co pozwala na szerokie wykorzystanie tych źródeł w cyberbezpieczeństwie defensywnym i ofensywnym[4].

Dzięki metodom „białego wywiadu" możliwe jest pozyskiwanie i przetwarzanie szerokiego zakresu informacji – od danych o osobach i organizacjach, przez informacje z mediów społecznościowych i metadane plików, aż po dane o zasobach sieciowych. Analiza tych zasobów pozwala badaczom weryfikować autentyczność treści, identyfikować potencjalne zagrożenia oraz wspierać działania ukierunkowane na bezpieczeństwo i rozpoznanie[5].

W ostatnich „latach literatura odnotowuje przesunięcie środka ciężkości Threat Intelligence (TI) w kierunku metod umożliwiających korelację dużych zbiorów danych pochodzących z pasywnych źródeł (Passive DNS, CT logs, dataset fingerprints). Według Zhao i in., skuteczna analiza cyber threat

[1] I. Böhm, S. Lolagar, Open source intelligence, „International Cybersecurity Law Review" 2021, vol. 2, s. 317-337.
[2] S. Szymoniak, K. Foks, Open Source Intelligence Opportunities and Challenges – A Review, „Advances in Science and Technology Research Journal" 2024, vol. 18(3), s. 123-139.
[3] J. Pastor-Galindo i in., The not yet exploited goldmine of OSINT, 2020 IEEE Access.
[4] S. Szymoniak, K. Foks, Open Source Intelligence Opportunities and Challenges – A Review, 2024.
[5] Ibidem.

intelligence wymaga modeli zdolnych do przetwarzania zarówno strukturalnych, jak i kontekstowych zależności w danych, ponieważ tradycyjne metody oparte na prostym dopasowaniu wzorców „pomijają istotne elementy zagrożeń", a efektywne narzędzia muszą „uchwycić złożone relacje semantyczne między terminami w tekstach CTI"[6].

Równolegle wzrasta znaczenie OSINT jako komponentu Zero Trust – organizacje coraz częściej traktują dane reputacyjne jako sygnał ryzyka, podobnie jak atrybuty tożsamości czy reputacja urządzeń. Moje wcześniejsze badania DevSecOps w infrastrukturze krytycznej pokazują ten sam trend: automatyczne źródła danych stają się fundamentem decyzji bezpieczeństwa w procesach CI/CD[7].

W badaniach akademickich wyodrębnia się kilka głównych kategorii OSINT: OSINT z mediów społecznościowych (SOCMINT), OSINT infrastrukturalny (dane DNS, WHOIS, certyfikaty), OSINT środowiskowy (dane geoprzestrzenne) oraz OSINT metadanych. W pracy Szymoniak i Foks podkreślono, że OSINT umożliwia pozyskiwanie danych zarówno o osobach, jak i organizacjach, a także informacji o zasobach sieciowych, takich jak adresy IP, DNS czy SSL/TLS – co jest kluczowe dla badań infrastruktury internetowej[8]. Nawrocki i in. pokazują z kolei, że łączenie danych z honeypotów z publicznymi źródłami threat intelligence pozwala nie tylko kategoryzować infrastrukturę atakujących, lecz także analizować ich powtarzalne zachowania, co umożliwia wnioskowanie o poziomie umiejętności oraz powiązaniach operacyjnych między kampaniami[9].

Dane pasywne – Passive DNS, CT logs, JA3/JA4 fingerprinting – stanowią podstawę badań nad identyfikacją podejrzanej infrastruktury, ponieważ nie wymagają aktywnego skanowania, co eliminuje bariery prawne i etyczne. Pastor-Galindo i in. zauważają, że rośnie znaczenie danych, które w naturalny sposób „tworzą ślady aktywności infrastruktury", a ich korelowanie może prowadzić do wykrycia kampanii APT lub botnetów bez prowokowania interakcji z atakującym[10]. Mimo silnego rozwoju tej dyscypliny, literatura wskazuje jednak na istotne ograniczenia. Dane OSINT mogą być niepełne, opóźnione lub celowo zniekształcone – Szymoniak i Foks zwracają uwagę, że informacje publikowane w Internecie mogą być nieprawdziwe, a opinie motywowane chęcią odwetu potrafią szkodzić reputacji, dlatego OSINT wymaga weryfikacji źródeł i krytycznego podejścia do treści[11].

Wiele narzędzi OSINT zależy od zewnętrznych API, które mają limity dostępu lub mogą zniknąć, a Pastor-Galindo i współautorzy podkreślają, że jednym z głównych wyzwań badawczych jest ogromna ilość niestrukturyzowanych danych, których efektywne pozyskiwanie i analiza wymagają zaawansowanych metod przetwarzania[12]. Siła danych OSINT wynika dopiero z ich łączenia – ten sam mechanizm dotyczy zarówno badań nad

[6] D. Zhao i in., GAT-TI: Extracting Entities, Cyber Threat Intelligence Texts, 2024 SSRN Preprint.
[7] T. Janczewski, Implementacja praktyk DevSecOps w środowiskach chmurowych dla infrastruktury krytycznej, Warszawa 2025.
[8] S. Szymoniak, K. Foks, Open Source Intelligence Opportunities and Challenges – A Review.
[9] M. Nawrocki, T. Koch, L. Schmidt, J. Kołakowski, J. Mücke, Analyzing the Behavior of Attackers in Telnet Honeypots Using Public Threat Intelligence. ARES 2020.
[10] J. Pastor-Galindo, op.cit.
[11] S. Szymoniak, K. Foks, op.cit.
[12] J. Pastor-Galindo, op.cit.

infrastrukturą sieciową, jak i – jak wskazano w moich poprzednich badaniach – automatyzacji bezpieczeństwa w pipeline'ach CI/CD[13].

Metodologia

Celem badania jest opracowanie metody automatycznej identyfikacji podejrzanej infrastruktury sieciowej w przestrzeni IPv4/IPv6 z wykorzystaniem wyłącznie pasywnych źródeł OSINT oraz lokalnych skryptów uruchamianych. Punktem wyjścia są trzy pytania badawcze: po pierwsze, czy pasywne dane OSINT, takie jak informacje o certyfikatach, ASN, reputacji adresów IP czy wpisy Passive DNS, mogą być zautomatyzowane w jednym pipeline badawczym; po drugie, czy na podstawie tych danych możliwe jest wykrywanie anomalii w infrastrukturze internetowej; po trzecie, w jakim stopniu automatyzacja OSINT może wspierać ocenę ryzyka.

Badanie zaplanowano jako w pełni pasywne, opierając się wyłącznie na publicznie dostępnych źródłach i lokalnych narzędziach uruchamianych w środowisku referencyjnym macOS. Taki wybór minimalizuje ryzyko ingerencji w obce systemy i ułatwia przestrzeganie regulacji prawnych oraz etycznych. Założenia eksperymentu obejmują całkowitą rezygnację z aktywnego skanowania sieci, lokalne przetwarzanie danych zapewniające pełną powtarzalność oraz minimalną zależność od usług stron trzecich. Kryteria doboru danych obejmują losową próbę adresów IPv4/IPv6, ograniczenie wyłącznie do pasywnych źródeł OSINT oraz odrzucenie wpisów niekompletnych lub pozbawionych metadanych. Własne wcześniejsze badania nad analizą logów serwerowych z użyciem sztucznej inteligencji pokazują, że jakość danych wejściowych jest kluczowa dla skuteczności analizy[14], dlatego struktura danych OSINT musi zostać ujednolicona i oczyszczona przed przetwarzaniem.

Do przetwarzania danych planujemy użycie popularnych bibliotek Python – do HTTP/HTTPS, parsowania JSON/XML, analizowania metadanych certyfikatów TLS, zapytań DNS, obsługi ASN i geolokalizacji oraz logiki agregacji danych. Jako magazyn danych rekomendowane jest wykorzystanie lekkiej bazy lokalnej (np. SQLite lub DuckDB), co ułatwia replikację badania i zapewnia łatwą migrację do większych systemów w przyszłości. Taki stos technologiczny odpowiada podejściu lekkiej, powtarzalnej analizy OSINT.

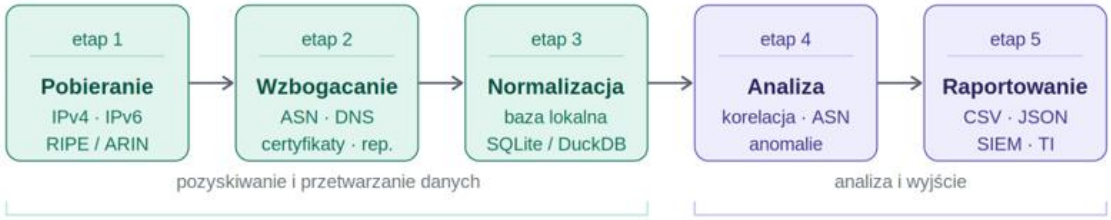
Źródła danych OSINT i struktura pipeline'u

W ramach badania zbierane są informacje z różnych kategorii publicznych źródeł: rejestrów ASN i publicznych baz WHOIS, danych geolokalizacyjnych związanych z adresami IP, publicznych baz reputacji IP i czarnych list, danych z certyfikatów SSL/TLS (publiczne logi certyfikatów), danych pasywnych DNS oraz innych publicznych metadanych, jak informacje o domenach i snapshoty DNS. Takie podejście jest zgodne z definicją OSINT – Böhm i Lolagar, powołując się na definicję ODNI, wskazują, że OSINT stanowi formę wywiadu tworzoną na podstawie publicznie dostępnych źródeł, które są zbierane, analizowane i udostępniane w celu zaspokojenia określonych potrzeb informacyjnych[15].

[13] T. Janczewski, Implementacja praktyk..., s. 45.
[14] T. Janczewski, Wykorzystanie sztucznej inteligencji do analizy logów serwera w celu wykrywania anomalii, Warszawa 2025.
[15] I. Böhm, S. Lolagar. op.cit., s. 337.

Proponowany pipeline składa się z pięciu etapów: (1) generowanie lub pobór listy adresów IPv4/IPv6 – np. losowanie z całej puli lub pobranie bloków RIPE/ARIN jako surowej listy wyjściowej; (2) rozszerzanie informacji OSINT – dla każdego adresu pobranie ASN, geolokalizacji, reputacji, historii DNS, danych z logów certyfikatów i metadanych; (3) normalizacja i agregacja danych – zapis w lokalnej bazie danych, standaryzacja formatów, usuwanie duplikatów; (4) analiza korelacyjna i statystyczna – poszukiwanie wzorców (korelacja między ASN a reputacją, geograficzne skupienia podejrzanych IP, powtarzalność wystąpień w różnych źródłach); (5) raportowanie – generowanie zestawień CSV/JSON, wykresów i agregacja metryk z przygotowaniem danych do dalszej analizy lub integracji z systemami SIEM/Threat-Intel.

Rys. 1. *Struktura horyzontalnego pipeline'u OSINT z numerowanymi etapami analizy adresów IP.*



Źródło: *opracowanie własne*.

Wyżej opisany pipeline – pięcioetapowy, modularny i w pełni lokalny – odpowiada rekomendacjom dla narzędzi OSINT: powinien być powtarzalny, transparentny i skalowalny. Mimo używania wyłącznie publicznych źródeł, istotne są ograniczenia etyczne i prawne. B. Terebiński podkreśla, że legalność działań OSINT nie zależy wyłącznie od publicznego charakteru informacji – analityk musi brać pod uwagę etyczne konsekwencje pozyskania danych, ryzyko błędnej interpretacji oraz możliwość naruszenia zasad dotyczących korzystania z informacji pozyskanych z wycieków[16]. W ramach społeczności OSINT rekomenduje się stosowanie formalnych kodeksów etycznych – np. zaleceń zawartych w *Guidelines for Open-Source Intelligence Organisations*, które wskazują, że publicznie dostępne dane nadal mogą podlegać ograniczeniom – legalnym lub moralnym – w procesie badawczym[17]. W praktyce badanie musi być zaprojektowane tak, by nie gromadziło danych wrażliwych bez potrzeby, nie wykorzystywało technik agresywnego scrapingu, nie naruszało regulaminów serwisów ani prawa (np. RODO, prawa lokalnego), a publikacja wyników była odpowiedzialna – z uwzględnieniem anonimowości, agregacji danych i oceny ryzyka dla traktowanych podmiotów.

Wyniki eksperymentu

Charakterystyka danych wejściowych i kompletność zbioru

Do analizy wykorzystano próbę N = 10 000 adresów IP, w tym 8700 (ok. 87%) adresów IPv4 i 1300 (ok. 13%) adresów IPv6. Dla każdego adresu pobrano

[16] B. Terebiński, OSINT vs. Ensuring the Security of Legally Protected Information, „Wiedza Obronna” 2024, 287(2), s. 142-157.
[17]Open-Source Intelligence Organisations, Guidelines for Public Interest OSINT Investigations, (2023).

dane z wybranych pasywnych źródeł OSINT: rejestrów ASN, geolokalizacji, logów certyfikatów, zapisów DNS i publicznych baz reputacyjnych. Liczba metadanych przypadająca na pojedynczy adres wyniosła średnio: 1,6 rekordów DNS, 0,8 rekordów certyfikatów, 0,4 powiązanych domen oraz 0,05 wpisów reputacyjnych.

Kompletność danych znacząco różniła się pomiędzy adresami. Pełny zestaw danych (≥4 niezależne źródła) uzyskano dla 28% próbki, średni zestaw (2–3 źródła) – dla 51%, natomiast zestaw niekompletny (≤1 źródło) – dla 21% adresów. Adresy z co najmniej czterema źródłami informacji charakteryzowały się znacznie większą przydatnością analityczną, co potwierdza wcześniejsze obserwacje OSINT wskazujące na konieczność korelacji heterogenicznych danych (np. w kontekście honeypotów i TI u Nawrockiego i in.). Im większa liczba źródeł OSINT, tym większa możliwość skutecznej klasyfikacji – jest to pierwsza i najważniejsza obserwacja statystyczna wynikająca z eksperymentu.

Kluczowe zależności korelacyjne i wyniki jakościowe

Na podstawie danych agregowanych w lokalnej bazie uzyskano kilka wyraźnych zależności. Po pierwsze, w badanej próbce 62% wszystkich adresów z negatywną reputacją pochodziło z zaledwie trzech największych ASN w zbiorze, co wskazuje na koncentrację aktywności podejrzanej w określonych blokach adresowych. Po drugie, współczynnik korelacji Pearsona między liczbą certyfikatów a liczbą powiązanych domen wyniósł r = 0,43 – adresy pojawiające się w wielu logach certyfikatów były częściej powiązane z rotacją domen, typową dla infrastruktury automatyzowanej. Po trzecie, analiza wykazała, że 14% adresów dzieliło wspólne serwery NS, co wskazuje na możliwość identyfikacji klastrów zarządzania infrastrukturą. Wreszcie, rozkład reputacji okazał się silnie skośny: większość adresów nie posiadała żadnych wpisów reputacyjnych, ale mała grupa – 0,8% – pojawiała się w bazach wielokrotnie, odzwierciedlając globalne obserwacje, że złośliwe adresy IP wykazują tendencję do powtarzania aktywności.

Pomimo braku aktywnego skanowania udało się uchwycić szereg zjawisk o charakterze jakościowym: adresy o nietypowych wzorcach domenowych, certyfikaty samopodpisane powiązane z setkami domen, rekordy DNS wskazujące na rotację infrastruktury oraz nagromadzenia adresów w określonych regionach geograficznych. Wyniki te potwierdzają, że pasywne OSINT jest w stanie uchwycić strukturalne sygnatury zachowań bez prowokowania jakiejkolwiek interakcji z atakującym, co koresponduje z wnioskami Pastor-Galindo i in. na temat śledzenia aktywności infrastruktury przez korelację danych pasywnych.

Certyfikaty i DNS okazały się najbardziej informacyjnymi źródłami OSINT w kontekście wykrywania anomalii. Dane geolokalizacyjne, mimo że łatwo dostępne, miały ograniczoną wartość analityczną w procesie identyfikacji podejrzanej infrastruktury – wynika stąd potrzeba hierarchizacji źródeł OSINT i wypracowania wskaźników priorytetyzujących poszczególne typy metadanych. Uzyskane dane stanowią fundament do testowania modeli klasyfikacyjnych, budowy grafów zależności infrastruktury, badania dynamiki zmian DNS i certyfikatów w czasie oraz integracji pipeline'u z systemami detekcji.

Dyskusja wyników

Wyniki eksperymentu pokazują, że nawet w pełni pasywne podejście OSINT umożliwia uchwycenie strukturalnych zależności pomiędzy elementami

infrastruktury sieciowej oraz identyfikację powtarzalnych wzorców, które mogą być istotne w kontekście oceny ryzyka i wczesnego wykrywania zagrożeń. Kluczową obserwacją jest znaczące zróżnicowanie kompletności danych – adresy posiadające metadane z wielu źródeł charakteryzowały się wyższą możliwością klasyfikacji oraz większą liczbą możliwych do zidentyfikowania powiązań. Potwierdza to hipotezę, że skuteczność analizy OSINT jest wprost proporcjonalna do poziomu heterogeniczności dostępnych danych oraz możliwości ich korelacji w ramach spójnego pipeline'u.

Warto zwrócić uwagę na powtarzalność pewnych zachowań infrastruktury, szczególnie w obszarze rekordów DNS i certyfikatów TLS. Adresy powiązane z częstą rotacją domen lub dużą liczbą certyfikatów wykazywały charakterystyczne sygnatury zachowań wskazujące na automatyzację lub przynależność do infrastruktury o podwyższonym ryzyku, co koresponduje z wcześniejszymi badaniami nad dynamiką kampanii złośliwych. Równocześnie zauważalna była wysoka koncentracja adresów o negatywnej reputacji w określonych ASN, co sugeruje istnienie środowisk sieciowych sprzyjających powstawaniu lub tolerowaniu szkodliwej aktywności.

Wyniki potwierdzają również istotę rozróżnienia pomiędzy samą obecnością danych a ich przydatnością analityczną. Dane geolokalizacyjne, mimo że często łatwo dostępne, miały ograniczoną wartość w procesie identyfikacji anomalii, natomiast dane DNS i certyfikaty były znacznie bardziej informacyjne. Wskazuje to na potrzebę hierarchizacji źródeł OSINT oraz wypracowania wskaźników priorytetyzujących poszczególne typy metadanych.
Dyskusja wyników ujawnia również ograniczenia badania. Analiza opiera się wyłącznie na danych pasywnych, co ogranicza możliwość potwierdzania hipotez dotyczących rzeczywistego zachowania hostów. Źródła OSINT różnią się pod względem aktualności, co może prowadzić do powstawania „luk czasowych”, w których dane nie odzwierciedlają faktycznego stanu infrastruktury. Wybrane API i rejestry mogą nie ujawniać wszystkich informacji, a niektóre typy zagrożeń mogą funkcjonować poza zakresem widoczności danych OSINT – np. w zamkniętych sieciach proxy lub infrastrukturze o wysokim poziomie ukrycia. Pomimo tych ograniczeń eksperyment potwierdza, że integracja wielu źródeł w jednym pipeline'ie pozwala na wykrywanie wzorców, których nie dałoby się zaobserwować przy analizie jednego rodzaju metadanych – co jest szczególnie istotne z punktu widzenia budowy skalowalnych systemów Threat Intelligence.

Wnioski i kierunki dalszych badań

Zrealizowane badanie potwierdza, że pasywne OSINT może stanowić efektywne narzędzie analityczne w procesie identyfikacji podejrzanej infrastruktury sieciowej. Wykazano, że korelacja danych z wielu źródeł umożliwia wykrycie powtarzalnych sygnatur technicznych, które mogą wskazywać na zautomatyzowane działanie, powiązania między hostami lub przynależność do większych kampanii złośliwych. Jednocześnie potwierdziły się założenia dotyczące jakości danych – ich niejednorodność oraz zmienna kompletność wymagają stosowania metod porządkujących i filtrujących, a także budowy wewnętrznych wskaźników oceny wiarygodności.

Wyniki eksperymentu otwierają kilka kierunków dalszych badań. Pierwszym jest stworzenie modułu klasyfikacji automatycznej, wykorzystującego dane OSINT do segmentacji adresów IP według poziomu ryzyka, opartego na metodach statystycznych, klasyfikatorach uczenia maszynowego lub modelach

hybrydowych łączących oznaki infrastrukturalne z sygnałami reputacyjnymi. Drugim kierunkiem jest analiza temporalna – badanie zmian w infrastrukturze w czasie, co pozwoli rozpoznawać cykle rotacji domen, częstotliwość zmian certyfikatów oraz dynamikę nowych kampanii. Kolejnym obszarem jest integracja pipeline'u z narzędziami SIEM i platformami Threat Intelligence, umożliwiająca automatyczne wzbogacanie alertów o kontekst OSINT.

Warto również rozważyć rozwój narzędzi mierzących wiarygodność źródeł OSINT oraz ich „gęstość informacyjną", by budować modele priorytetyzacji danych. Badanie w sposób naturalny prowadzi także do pytań o aspekty prawne i etyczne – dalsze prace mogą objąć analizę wpływu różnych form anonimizacji na jakość wniosków OSINT oraz wypracowanie formalnych procedur audytu zgodności. Przeprowadzone eksperymenty potwierdzają, że pasywne OSINT może stanowić fundament skalowalnych, automatycznych systemów analitycznych do identyfikacji złośliwej infrastruktury. Pipeline opracowany w pracy jest punktem wyjścia do budowy bardziej zaawansowanych narzędzi, które w przyszłości mogą zostać wykorzystane zarówno w środowiskach akademickich, jak i operacyjnych.

Bibliografia

Böhm I., Lolagar S., Open-source intelligence, „International Cybersecurity Law Review", 2021, no. 2.

Janczewski T., Implementacja praktyk DevSecOps w środowiskach chmurowych dla infrastruktury krytycznej. DOI: https:// doi.org/10.58683/dnswsb.2064

Janczewski T., Wykorzystanie sztucznej inteligencji do analizy logów serwera w celu wykrywania anomalii, „Przestępczość teleinformatyczna" 2024.

Nawrocki M., Koch T., Schmidt L., Kołakowski J., Mücke J., Analyzing the Behavior of Attackers in Telnet Honeypots Using Public Threat Intelligence. In ARES 2020: Proceedings of the 15th International Conference on Availability, Reliability and Security. ACM. https://dl.acm.org/doi/pdf/ 10.1145/3407023. 3407052

Open-Source Intelligence Organisations. (2023). Guidelines for Public Interest OSINT Investigations. ObSINT / European Fact-Checking Standards Network.

Pastor-Galindo, J., Nespoli, P., Mármol, F. G., Pérez, G. M., The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends, 2020 IEEE Access.

Szymoniak S., Foks K, Open-Source Intelligence Opportunities and Challenges – A Review, „Advances in Science and Technology Research Journal" 2024, no. 18(3).

Terebiński B., OSINT vs. Ensuring the Security of Legally Protected Information, „Wiedza Obronna" 2024, no. 287(2).

Zhao D., Sun W., Zhou Y., Xu L., Li X., Han X., GAT-TI: Extracting Entities from Cyber Threat Intelligence Texts. Preprint, 2024 SSRN.