Osintowy Łapacz Cyfrowych Śladów – OŁCyŚ

Czyli „Jak identyfikować i oceniać powiązania pomiędzy cyfrowymi śladami a adresami krypto”

1. Wprowadzenie – opis organizacji, sytuacji i stanu aktualnego

Jednym z wyzwań związanych z opodatkowaniem kryptowalut jest identyfikowanie sytuacji, w których osoby fizyczne osiągają przychody z obrotu krypto, ale nie wykazują ich prawidłowo w rocznym zeznaniu podatkowym PIT-38. Może do tego dojść na przykład wtedy, gdy kryptowaluty zostają wymienione na tradycyjne waluty, takie jak złoty, euro czy dolar. Część podatników nie zgłasza takich operacji albo określa ich wartość niedokładnie, na przykład szacując ją „na oko”.

Wielu uczestników rynku kryptowalut zakłada, że ich aktywność pozostaje anonimowa. W praktyce użytkownicy zostawiają jednak w sieci wiele cyfrowych śladów. Publikują adresy portfeli na forach, w mediach społecznościowych, serwisach ogłoszeniowych, czy przy ofertach sprzedaży produktów i usług.

Pracownicy Krajowej Administracji Skarbowej (KAS) prowadzą czynności sprawdzające, analizy i kontrole dotyczące między innymi nabywania i sprzedaży kryptowalut. W ramach tych działań przeszukują publicznie dostępne źródła internetowe oraz platformy P2P, próbując połączyć adresy kryptowalutowe z konkretnymi podmiotami.

Dziś duża część tej pracy wymaga ręcznego przeglądania rozproszonych danych. Skala rynku krypto i liczba informacji dostępnych w Internecie sprawiają, że proces jest czasochłonny, trudny do powtarzania i mało efektywny. Potrzebne jest rozwiązanie, które pomoże szybciej odnajdywać istotne informacje i skuteczniej „łączyć kropki”.

Rozwiązaniem może być: Osintowy łapacz cyfrowych śladów - OŁCyŚ

Identyfikacja potencjalnych powiązań pomiędzy podmiotem a aktywnością w sieciach blockchain stanowi jeden z elementów czynności analitycznych i weryfikacyjnych prowadzonych przez KAS. Dlatego KAS potrzebuje nowoczesnego narzędzia OSINT, które automatycznie przeszukuje publicznie dostępne zasoby internetowe i identyfikuje możliwe powiązania pomiędzy publicznie dostępnymi śladami cyfrowymi a adresami kryptowalutowymi oraz prezentuje stopień prawdopodobieństwa tych powiązań.

Istotą działania tego narzędzia będzie zautomatyzowane powiązanie dwóch światów:

1. Śladów z otwartego Internetu (Web2): Publicznych wpisów, profili, ogłoszeń i wypowiedzi, w których użytkownik ujawnił adres kryptowalutowy. Analiza takich informacji może pomóc w ustaleniu, z kim dany adres może być powiązany.
2. Danych z blockchaina (Web3): Historii transakcji zidentyfikowanego adresu, analizowanej pod kątem podstawowych wskaźników aktywności. Mogą to być na przykład:
    1. łączny wolumen transakcji,
    2. wymiana pomiędzy kryptowalutami,

3. potencjalne wyjście z rynku krypto do tradycyjnych walut,
4. interakcje z usługami wymiany, w tym usługami typu P2P.

2. Wyzwanie

Zadaniem uczestników jest stworzenie narzędzia analitycznego, które po wprowadzeniu co najmniej jednego identyfikatora, np.:

- imię i nazwisko,
- nick używany na forum,
- identyfikator z komunikatora,
- login z portalu społecznościowego,
- adres e-mail,
- publiczny adres kryptowalutowy.

automatycznie przeszuka publicznie dostępne źródła internetowe i odnajdzie powiązane z nimi adresy kryptowalutowe.

Narzędzie powinno umożliwiać:

- cykliczne wykonywanie analiz,
- zapisywanie wyników kolejnych analiz,
- porównywanie rezultatów,
- pozyskiwanie danych wyłącznie z informacji publicznie dostępnych w Internecie,
- dalszą rozbudowę rozwiązania o nowe źródła danych, sieci blockchain i mechanizmy analityczne,
- prezentowanie wyników w czytelny i zrozumiały sposób.

Podstawa działania (katalog otwarty):

1. Otwarte źródła Web2: Komunikatory internetowe, profile w mediach społecznościowych, fora branżowe, serwisy programistyczne, portale ogłoszeniowe P2P i inne publicznie dostępne zasoby Internetu.
2. Eksploratory łańcucha bloków (API Web3): Publiczne eksploratory oraz API sieci blockchain, na przykład Bitcoin i Ethereum, umożliwiające odczyt historii transakcji oraz identyfikowanie potencjalnych punktów wymiany kryptowalut na waluty tradycyjne.

3. Oczekiwany rezultat

Zaprojektowane przez uczestników narzędzie powinno zawierać:

Kryteria wymagane:

- Analizować i łączyć ślady OSINT:

    - Wyszukiwać możliwe powiązania między danymi podmiotów a publicznymi adresami kryptowalutowymi - na podstawie informacji znalezionych w otwartych źródłach.
    - Kategoryzować znalezione powiązania i obliczać Wskaźnik Pewności (Confidence Score). Uczestnicy powinni przygotować własny algorytm punktacji określający prawdopodobieństwo, że dany adres kryptowalutowy jest powiązany z analizowaną osobą.
    - Automatycznie rozpoznawać adresy kryptowalutowe, prawidłowo wykrywać je w treści oraz wyodrębniać z tekstów opublikowanych w otwartych źródłach internetowych.

- Wykrywać wskaźniki mogące świadczyć o wyjściu z rynku krypto do walut tradycyjnych (Fiat Cash-Out):

    - Mogą to być na przykład interakcje z usługami P2P, bramkami płatniczymi, rozwiązaniami obsługującymi karty krypto-fiat albo wymiana znacznych wartości kryptowalut na stablecoiny.
    - Ograniczać błędne przypisania poprzez rozpoznawanie i odrzucanie adresów giełd, portfeli powierniczych, smart kontraktów, adresów instytucjonalnych i innych obiektów, których nie należy przypisywać bezpośrednio badanej osobie.

Kryteria opcjonalne:

- Porównywać zmiany zachodzące w czasie, na przykład wykrywać podmianę adresów portfeli publikowanych na profilach użytkowników w różnych okresach.

Kryteria akceptacji aplikacji:

W celu zweryfikowania poprawności działania aplikacji uczestnicy przygotowują scenariusze testowe wykorzystujące dane zanonimizowane lub syntetyczne.

Scenariusze powinny umożliwiać sprawdzenie, czy narzędzie:

- prawidłowo odnajduje adresy kryptowalutowe,
- łączy informacje pochodzące z różnych źródeł,
- oblicza Wskaźnik Pewności zgodnie z przyjętym algorytmem,
- przedstawia źródła i przesłanki, na podstawie których utworzono powiązanie,
- wykrywa wybrane wzorce aktywności w blockchainie,
- zapisuje wyniki analizy wraz ze źródłem, adresem URL lub identyfikatorem zasobu, datą i czasem pozyskania oraz metadanymi potrzebnymi do późniejszego odtworzenia i zweryfikowania analizy.

Użytkownikami końcowymi narzędzia będą pracownicy Krajowej Administracji Skarbowej (KAS).

4. Wymagania formalne

Projekt przesyłany do oceny powinien zawierać:

- szczegółowy opis i tytuł projektu, w tym opis wykorzystanych parametrów wyszukiwania,
- prezentację w formacie PDF, zawierającą maksymalnie 10 slajdów,
- film umieszczony w dostępnym, otwartym repozytorium, trwający maksymalnie 3 minuty i prezentujący działanie narzędzia,
- link do repozytorium kodu.

Dodatkowo projekt może zawierać:

- zrzuty ekranu,
- linki do wersji demonstracyjnej,
- materiały graficzne,
- inne elementy związane z projektem.

5. Wymagania techniczne

1. Kod aplikacji powinien znajdować się w dostępnym repozytorium i korzystać z komponentów na licencjach Open Source zgodnych z zasadami wydarzenia. Zespół powinien wskazać w dokumentacji projektu licencje wykorzystanych komponentów.
2. Technologia backendowa: Najlepiej .NET, JAVA, NodeJS. Dopuszczamy inne WIODĄCE technologie jeżeli zostanie to uwarunkowane korzyściami (GO, Python, Rust).
3. Technologie Webowe: Wiodące: Angular, React, Vue. Biblioteki do tych frameworków funkcjonujące na podobnej zasadzie i licencjach jak powyżej (Open Source).
4. Technologie bazodanowe: standardowe narzędzia Open Source, takie jak MySQL, PostgreSQL, MongoDB lub Microsoft SQL Server.
5. Technologie AI: dopuszczalne jest wykorzystanie infrastruktury umożliwiającej zastosowanie różnych modeli, w KROKU 5 załącznika pojęciowo-metodycznego.
6. Architektura rozwiązania musi pozwalać na łatwą zmianę modelu AI na inny model np. przy użyciu uniwersalnego API.
7. Aplikacja powinna spełniać standardy bezpieczeństwa, nie odwoływać się do niezaufanych zasobów oraz nie wysyłać danych testowych, wyników analizy ani informacji o stanie działania do niezatwierdzonych usług zewnętrznych.
8. Aplikacja musi znajdować się w jednym repozytorium (monorepo). Dopuszczalne jest wykorzystanie konteneryzacji i elementów skalowalnej infrastruktury, takich jak mikroserwisy lub mikrofrontendy.
9. Do repozytorium musi zostać dołączony plik README.md, w którym opisane zostaną:
    - Opis projektu – jaki problem rozwiązuje, w jaki sposób,
    - Założenia projektu,
    - Instrukcja instalacji/technologii potrzebnych do uruchomienia,
    - Przykłady (screenshoty/logi) z finalnego wyniku oraz działającej aplikacji,
    - Opcjonalnie: dodatkowa dokumentacja techniczna lub informacje.

6. Sposób testowania i/lub walidacji

1. Prezentacja następuje na przykładowych danych, przygotowanych przez uczestników, niezależnie od prezentacji pdf i krótkiego filmu.

2. Parametry wyszukiwania – prezentacja i omówienie przyjętych założeń, wykorzystanych źródeł oraz sposobu interpretacji uzyskanego wyniku.

Za obowiązkowy scenariusz prezentacji uznaje się pełny przebieg analizy: wprowadzenie danych wejściowych, wyszukanie śladów OSINT, wykrycie i weryfikację adresu kryptowalutowego, wyliczenie Wskaźnika Pewności, odrzucenie błędnych przypisań, analizę wybranych wskaźników aktywności blockchain oraz przedstawienie źródeł i wyniku końcowego. Kryterium poprawności realizacji scenariusza odnosi się do wszystkich wymienionych etapów.

7. Dostępne zasoby

a) Wszystkie zadania w ramach HackNation muszą być realizowane wyłącznie przy użyciu metody OSINT-u (białego wywiadu).
b) Uczestnicy mogą korzystać wyłącznie z publicznie dostępnych, jawnych i legalnych źródeł informacji, w tym np.: z publicznych rejestrów państwowych, otwartych baz danych, ogólnodostępnych profili w mediach społecznościowych, publicznych forów, wyszukiwarek internetowych oraz zarchiwizowanych wersji stron. Korzystanie ze źródła nie może naruszać jego regulaminu ani wymagać obchodzenia zabezpieczeń.
c) W celu zapewnienia legalności działań bezwzględnie zabrania się przełamywania zabezpieczeń, ingerowania w systemy teleinformatyczne, hackingu, phishingu, skanowania podatności itd, np. za pomocą narzędzia nmap, przeprowadzania ataków DDoS oraz naruszania prywatności. Zabronione jest także pozyskiwanie danych wymagających logowania na cudze konta, łamania haseł lub korzystania z wycieków zawierających prawnie chronione dane osobowe albo hasła użytkowników.
d) Organizator nie ponosi odpowiedzialności za działania uczestników wykraczające poza wyznaczone ramy OSINT.

Dane wyjściowe powinny umożliwiać przygotowanie kompletnego raportu OSINT dokumentującego publiczne ślady powiązań podmiotu z rynkiem kryptowalut. Raport powinien zawierać zidentyfikowane ślady transakcyjne, źródła, stopień prawdopodobieństwa powiązań, chronologiczny opis przepływu środków oraz wyraźne rozróżnienie pomiędzy faktami źródłowymi, wynikami obliczeń i wnioskami wymagającymi dalszej weryfikacji. Raport powinien być przejrzysty i czytelny, w tym dla osób nieposiadających wiedzy informatycznej.

8. Kryteria oceny

Kryterium	Waga
Zgodność z wyzwaniem i poprawność realizacji scenariusza	25%
Jakość Wskaźnika Pewności i ograniczania błędnych przypisań	25%
Przejrzystość raportu, źródeł i uzasadnienia wyników	20%
Pomysł i oryginalność rozwiązania	15%
Potencjał wdrożeniowy i architektura techniczna	15%

1. Zgodność z wyzwaniem i poprawność realizacji scenariusza — 25%
Oceniana jest zgodność rozwiązania z zakresem zadania, poprawność identyfikowania śladów OSINT i adresów kryptowalutowych, realizacja wymaganego scenariusza oraz poprawność uzyskanych wyników.

2. Jakość Wskaźnika Pewności i ograniczania błędnych przypisań — 25%
Oceniana jest jakość i przejrzystość algorytmu punktacji, sposób uzasadniania wyniku oraz skuteczność rozpoznawania i odrzucania adresów giełd, smart kontraktów i innych fałszywych trafień.

3. Przejrzystość raportu, źródeł i uzasadnienia wyników — 20%
Oceniane są czytelność raportu, kompletność źródeł i metadanych, możliwość odtworzenia procesu analitycznego oraz rozróżnienie między faktami źródłowymi, wynikami obliczeń i wnioskami wymagającymi dalszej weryfikacji.

4. Pomysł i oryginalność rozwiązania — 15%
Oceniane są kreatywność podejścia do wyszukiwania i łączenia śladów, wartość dodatkowych funkcji, sposób prezentacji danych oraz rozwiązania wykraczające ponad wymagane minimum.

5. Potencjał wdrożeniowy i architektura techniczna — 15%
Oceniane są możliwość wykorzystania rozwiązania w realnych warunkach, skalowalność, zgodność z regulacjami, jakość architektury oraz realność dalszego rozwoju, wdrożenia i utrzymania w środowisku KAS.

9. Dodatkowe uwagi / kontekst wdrożeniowy

Najlepszy model lub aplikacja może zostać poddana szerszym testom na dodatkowym materiale przygotowanym przez KAS. Po uzyskaniu pozytywnej oceny rozwiązanie może być dalej rozwijane i wdrażane we współpracy z podmiotami świadczącymi usługi informatyczne na rzecz KAS. Docelowo powinno działać w środowisku KAS bez zależności od publicznej chmury i z możliwością lokalnego uruchamiania komponentów AI. Na potrzeby hackathonu dopuszczalne jest korzystanie z narzędzi AI do tworzenia kodu, pod warunkiem przestrzegania wymagań określonych w dokumencie. Dane testowe i wyniki analizy nie mogą być przekazywane do niezatwierdzonych usług zewnętrznych.

10. Kontakt

Kontakt z mentorami jest możliwy na kanale Discord oraz podczas wydarzenia przy stoisku w wiosce mentorów.

ZAŁĄCZNIK POJĘCIOWO-METODYCZNY (CHEAT SHEET)

Wyzwanie KAS: „Osintowy Łapacz Cyfrowych Śladów — OŁCyŚ”

TEN ZAŁĄCZNIK NIE OKREŚLA KRYTERIÓW OCENY - JEST ON TYLKO MATERIAŁEM POMOCNICZYM DLA UCZESTNIKÓW WYZWANIA, WYJAŚNIAJĄCYM NIEKTÓRE POJĘCIA UŻYWANE NA RYNKU KRYPTO. WYJAŚNIA SPOSÓB ROZUMIENIA POJĘĆ ANALITYCZNYCH KAS, KATEGORYZACJĘ OBIEKTÓW W KODZIE ORAZ LOGIKĘ BUDOWY WSKAŹNIKA PEWNOŚCI (CONFIDENCE SCORE).

KROK 1. Identyfikatory cyfrowe i kategoryzacja adresów Web3

W ramach wyzwania przez ślad/identyfikator cyfrowy rozumiemy unikalny parametr pozwalający powiązać aktywność użytkownika z określonym zasobem lub usługą w środowisku cyfrowym.

Celem aplikacji jest automatyczne odnajdywanie i identyfikowanie powiązań pomiędzy publicznie dostępnymi śladami/identyfikatorami cyfrowymi w środowisku Web2 a adresami kryptowalutowymi w środowisku Web3 oraz prezentacja stopnia prawdopodobieństwa tych powiązań.

1. Ślady / Identyfikatory cyfrowe Web2

Identyfikatory wykorzystywane w otwartym Internecie, umożliwiające przypisanie aktywności do danej osoby lub profilu, w tym:

- pseudonimy i nazwy użytkowników (nicki)
- loginy
- adresy e-mail
- identyfikatory w komunikatorach (np. Telegram ID, Handle)
- adresy URL profili w mediach społecznościowych, serwisach ogłoszeniowych i forach

2. Adresy kryptowalutowe Web3 — Niehostowane (*Unhosted / Non-custodial*)

Jawne adresy publiczne w łańcuchu bloków, do których klucze prywatne posiada bezpośrednio dana osoba fizyczna (np. Trust Wallet, MetaMask, Ledger, portfel papierowy).

- To główny cel poszukiwań i atrybucji OSINT.

3. Adresy kryptowalutowe Web3 — Giełdy scentralizowane i portfele powiernicze (*CEX / Custodial*)

Adresy publiczne należące do giełd i pośredników. Użytkownik nie posiada do nich klucza prywatnego.

- Zasada w kodzie: Jeśli adres należy do giełdy → nie przypisujesz go jako prywatnego portfela badanej osoby (ochrona przed *False Positive*), ale odnotowujesz fakt interakcji.

4. Adresy kryptowalutowe Web3 — Giełdy zdecentralizowane i smart kontrakty (*DEX / Smart Contracts*)

Zautomatyzowane protokoły wymiany kryptowalut (np. Uniswap, PancakeSwap).

- Weryfikacja w API: Sprawdź pole kodu bajtowego adresu (bytecode). Jeśli bytecode != 0x (adres zawiera kod wykonywalny), jest to smart kontrakt, a nie adres prywatnego użytkownika.

5. Dostawcy usług wymiany i bramki krypto-fiat (*CASP / On-Ramp / Off-Ramp / Kantory*)

Podmioty umożliwiające kupno lub sprzedaż kryptowalut za waluty tradycyjne (np. PLN, EUR, USD), w tym kantory internetowe i stacjonarne (OTC) oraz bramki płatnicze.

KROK 2. Algorytm wyliczania Wskaźnika Pewności (*Confidence Score*)

Aplikacja wylicza procentowy poziom prawdopodobieństwa (0–100%), że znaleziony adres kryptowalutowy jest powiązany z analizowaną osobą, na podstawie źródła pochodzenia śladu z Web2 w przykładowy sposób:

- 80–100% (BARDZO WYSOKIE PRAWDOPODOBIEŃSTWO): Adres kryptowalutowy opublikowany na oficjalnym, autorskim profilu podmiotu, np. na własnym profilu GitHub, osobistej stronie WWW lub zweryfikowanym koncie w mediach społecznościowych.
- 60–79% (PRAWDOPODOBNY): Adres kryptowalutowy w ogłoszeniu sprzedaży (np. oferta P2P, sprzedaż przedmiotu), w stopce e-maila lub podany w oficjalnym komunikatorze w kontekście rozliczenia.
- 40–59% (ŚREDNI): Adres kryptowalutowy rzucony w dyskusji na forum branżowym lub grupie tematycznej (wymaga potwierdzenia drugim śladem).
- 10–39% (NISKI / SZUM): Adres kryptowalutowy w masowych komentarzach (np. posty konkursowe, promocje, giveawaye).
- 0% (WYKLUCZENIE / FALSE POSITIVE): Adres publiczny znanej giełdy CEX, znany adres zbiórki charytatywnej, miksera lub smart kontraktu.

Przykładowa reguła premii (Synergia śladów):

Jeśli ten sam adres kryptowalutowy pojawi się w co najmniej dwóch niezależnych źródłach Web2, np. na profilu GitHub oraz w ogłoszeniu P2P powiązanym z tym samym nickiem lub

numerem telefonu, wynik wyjściowy można zwiększyć o 15 punktów procentowych, nie przekraczając 100%. Za niezależne uznaje się źródła różnego pochodzenia. Wielokrotne wystąpienie adresu w ramach jednej usługi nie stanowi dodatkowego potwierdzenia.

KROK 3. Odrzucanie fałszywych trafień (*False Positives*)

Aplikacja musi być odporna na sytuację, w której użytkownik wkleił w sieci link do adresu giełdy, zbiórki czy smart kontraktu.

1. Sprawdzanie etykiet w API (Address Tags / Name Tags):

   Przed przypisaniem adresu do osoby odpytaj publiczne API (np. Etherscan, Blockchair). Jeśli API zwraca oficjalny tag (np. Kraken: Hot Wallet), adres otrzymuje 0% pewności przy przypisaniu osobowym.

2. Sprawdzanie flagi kontraktu (bytecode):

   Jeśli adres posiada kod wykonywalny (bytecode != 0x), nie jest kontem prywatnym użytkownika.

KROK 4. Zasady wykorzystania sztucznej inteligencji (AI & Ochrona Danych)

Jeśli zespół stosuje modele sztucznej inteligencji (AI / NLP):

- Modele AI mogą pełnić rolę wspomagającą (np. do wykrywania i wyodrębniania adresów z treści opublikowanych w otwartych źródłach internetowych).
- Wynik analizy oraz sposób wyliczenia stopnia prawdopodobieństwa powiązań muszą być dla analityka KAS czytelne, wyjaśnialne i możliwe do odtworzenia. Aplikacja musi jednoznacznie wskazywać źródła, przesłanki i reguły punktacji wykorzystane do utworzenia powiązania.