Avointen lähteiden tiedustelun tuottaman kyberuhkatiedon hyödyntäminen ja jakaminen

Stephen McMenamin

2024 Laurea

Laurea-ammattikorkeakoulu

Avointen lähteiden tiedustelun tuottaman kyberuhkatiedon hyödyntäminen ja jakaminen

Stephen McMenamin
Tietojenkäsittely, kyberturvallisuus
Opinnäytetyö
01, 2024

Laurea-ammattikorkeakoulu **Tiivistelmä**

Tietojenkäsittely, kyberturvallisuus

Tradenomi (AMK)

Stephen McMenamin

Avointen lähteiden tiedustelun tuottaman kyberuhkatiedon hyödyntäminen ja jakaminen

Vuosi 2024 Sivumäärä 49

Avointen tietolähteiden tiedustelu (OSINT) on keskeinen osa organisaatioiden kyberturvallisuutta, jossa julkisesti saatavilla olevaa tietoa hyödynnetään potentiaalisten uhkien proaktiiviseen tunnistamiseen ja niiden torjunnan suunnitteluun.

Tämän opinnäytetyön tavoitteena oli tutkia, minkälaista uhkatietoa avointen lähteiden tiedustelulla voidaan tuottaa, miten kyseistä tietoa voidaan hyödyntää organisaatioissa kyberuhkilta puolustautumiseen, selvittää, mitä tiedon turvallisesti jakamisessa organisaatioiden välillä tulee ottaa huomioon ja pohtia, kuinka avoimen lähdekoodin työkaluja voidaan hyödyntää uhkatiedon jakamisessa.

Työ sitoutui Euroopan Unionin rahoittamaan DYNAMO-hankkeen alakohtaan 4.2, jossa selvitettiin uhkatiedon jakamiseen liittyviä haasteita ja suunniteltiin luottamusympäristöä, jossa tietoa voidaan turvallisesti jakaa eri tahojen välillä. Opinnäytetyön tavoitteena oli tuottaa hyödyllistä tietoa kyseisen kohdan edistämisen tueksi sekä DYNAMO-projektille kokonaisuudessaan.

Opinnäytetyön tutkimusmenetelmänä käytettiin integroivaa kirjallisuuskatsausta. Integroivassa kirjallisuuskatsauksessa tarkastellaan kriittisesti ja monipuolisesti eri lähteitä, tuottaen monipuolisen ja laajan kuvan tutkittavasta aiheesta. Lähteitä analysoitiin systemaattisesti ja objektiivisesti sisällönanalyysi-menetelmällä. Työssä tutustuttiin OSINT työkaluihin ja tekniikoihin, uhkatiedon hyödyntämiseen ja tiedon jakamiseen liittyviin eettisiin ja lainsäädännöllisiin haasteisiin. Tutkimustyön lähdemateriaalina käytettiin alan kirjallisuutta, artikkeleita, blogikirjoituksia, aiempia tutkimuksia sekä organisaatioiden verkkosivuja. Koska kyberturvallisuus on nopeasti kehittyvä ala, tutkimustyössä pyrittiin käyttämään mahdollisimman tuoreita lähteitä.

Työn tuloksena saatiin arvokasta tietoa avointen tietolähteiden tiedustelulla kerättävästä tiedosta ja sen hyödyntämisestä, sekä ohjeita siihen, miten uhkatietoa voidaan tuottaa ja jakaa luottamusympäristössä eettisesti ja lainsäädännölliset vaatimukset huomioon ottaen.

Asiasanat: OSINT, tietoturva, kyberturvallisuus, uhkatiedustelu

Laurea University of Applied Sciences **Abstract**

Bachelor's Degree Programme in Business Information Technology, Cybersecurity

Bachelor of Business Administration

Stephen McMenamin

Utilization and Sharing of Cyber Threat Intelligence Produced by Open Source Intelligence

Year 2024 Pages 49

Open Source Intelligence (OSINT) plays a crucial role in the cybersecurity of organizations by using publicly available information to proactively identify potential threats and plan their mitigation. This thesis aimed to explore the type of threat intelligence that can be generated through open source intelligence, how this information can be used by organizations to defend against cyber threats, to understand the considerations necessary for safe information sharing between organizations, and to examine how open-source tools can be utilized in the sharing of threat information.

The study was part of the European Union-funded DYNAMO project, specifically under section 4.2, which focused on the challenges associated with sharing threat intelligence and designing a trust environment for safe information exchange between different entities. The goal of the thesis was to provide useful information to support the advancement of this particular section of the DYNAMO project and the project as a whole.

The research method used in the thesis was an integrative literature review, which critically and comprehensively examines various sources to provide a broad perspective on the subject studied. The sources were analyzed systematically and objectively using a content analysis method. The study delved into OSINT tools and techniques, the utilization of threat intelligence, and the ethical and legal challenges related to information sharing. The research utilized a variety of sources including industry literature, articles, blog posts, previous studies, and organizational websites. Given the rapidly evolving field of cybersecurity, the study aimed to use the most recent sources available.

The results of the thesis yielded valuable information on the intelligence gathered through open source intelligence and its application, as well as guidelines on how to produce and share threat information ethically and in compliance with legal requirements within a trusted environment.

Keywords: OSINT, information security, cybersecurity, threat intelligence

Sisällys

1 Johdanto 6
2 Opinnäytetyön tutkimusmenetelmä 7
3 Avointen tietolähteiden tiedustelu 7
3.1 Käyttötavat 8
3.2 Avoimen tiedustelutiedon tietolähteet 8
3.3 Tiedustelutekniikat 9
4 Kyberuhkatietoa avointen tietolähteiden tiedustelulla 9
5 Tekniikat ja työkalut 10
5.1 Tekniikat 11
5.2 Monikäyttöiset työkalut 13
5.3 Sosiaalinen media 18
5.4 Hakukoneet 20
5.5 Verkkosivut 24
5.6 Verkkotunnukset 25
5.7 Historia, tietovuodot ja pimeä verkko 28
5.8 Kryptovaluutat 30
5.9 Tiedostot 33
6 Tiedon hyödyntäminen kyberuhkilta puolustautumiseen 35
7 Säännökset, lainsäädäntö ja eettisyys 35
7.1 Uhkatiedon jakamista koskeva lainsäädäntö 37
7.1.1 Yleinen tietosuoja-asetus 37
7.1.2 Tekoälyä koskeva sääntelykehysehdotus 40
7.1.3 NIS2-direktiivi 41
8 Uhkatiedon jakaminen luottamusympäristössä 42
8.1 MISP, TheHive ja Cortex 42
8.2 Luottamusympäristössä jaettavat tiedot 43
9 Yhteenveto, jatkotutkimusaiheet ja pohdinta 43
Lähteet 45
Kuvat 49
Taulukot 49

1 Johdanto

Avointen tietolähteiden tiedustelua (Open Source Intelligence, OSINT) tekevät mm. lainvalvontaviranomaiset, puolustusvoimat, tutkivat journalistit ja tietoturva-ammattilaiset. Myös pahantahtoiset toimijat, kuten verkkorikolliset tai valtiolliset uhkatoimijat hyödyntävät avointen tietolähteiden tiedustelun tekniikoita. Tavoitteena on löytää tietoa organisaatiosta tai yksityishenkilöistä, jota voidaan hyödyntää esimerkiksi kyberhyökkäyksissä, tai joka voisi aiheuttaa uhkaa turvallisuudelle. Tätä tietoa voidaan hyödyntää esimerkiksi kyberhyökkäyksissä, hyökkäysten ennaltaehkäisyyn, rikostutkinnassa tai liiketoiminnan strategian tukena. (SANS Institute 2023; Fruhlinger, Sharma & Breeden 2023)

Avointen tietolähteiden tiedustelu on monella tapaa peilikuva operatiivisesta turvallisuudesta (Operational Security, OPSEC), eli prosessista, jolla organisaatiot suojaavat tietoa, jota ei saa päästä julkiseen tietoon. Työntekijän julkaisema kuva sosiaalisessa mediassa voi työntekijän huomaamatta sisältää arkaluontoista tietoa organisaatiosta, tai julkiseen verkkoon yhteydessä olevat IoT-laitteet voivat paljastaa uhkatoimijaa hyödyttävää tietoa organisaation käyttämistä teknologioista. Avointen tietolähteiden tiedustelulla voidaan proaktiivisesti tunnistaa mahdolliset uhkat ja ehdottaa toimenpiteitä niiden pienentämiseksi. (Fruhlinger ym. 2023; Rose 2020)

Euroopan Unionin rahoittama DYNAMO-projekti pyrkii tuottamaan alustan, jonka avulla kriittisten sektoreiden, kuten energia, terveys ja kuljetus, infrastruktuuriin kohdistuvia kyberuhkia voidaan pienentää. Se tarjoaa tukea kaikissa kybersietokyvyn vaiheissa – valmistautuminen, ehkäisy, suojaus, reagointi ja palautuminen – yhdellä alustalla, joka on räätälöity kriittisen infrastruktuurin tarpeisiin. Alustan tavoitteena on parantaa organisaatioiden sekä yhteiskunnan turvallisuutta. (DYNAMO 2023; Packham 2022)

Yksi osa alustaa on uhkatiedon (Cyber Threat Intelligence, CTI) tuottaminen vaarantumisindikaattoreiden (Indicator of Compromise, IOC) avulla, kehittäen organisaation tieto- ja kyberhavainnointikyvykkyyttä (DYNAMO 2023). Vaarantumisindikaattorit ovat vihjeitä tai havaintoja uhkaavasta toiminnasta, kuten esimerkiksi mustalistatut IP-osoitteet, haitalliseksi todetun tiedoston jäännösarvo, tai kalastelusivuksi luokitellun verkkosivuston URL-osoite (Gaucheler 2023). Avointen tietolähteiden tiedustelu on keskeinen osa vaarantumisindikaattoreiden tuottamisessa ja sillä saadaan syvennettyä tietoa mahdollisista uhkista.

2 Opinnäytetyön tutkimusmenetelmä

Opinnäytetyön tutkimusmenetelmänä käytettiin integroivaa kirjallisuuskatsausta, jossa tarkastellaan kriittisesti ja monipuolisesti eri lähteitä, ja muodostetaan niistä laaja kuva tutkittavasta aiheesta.

Työn aikana testattiin useita avointen tietolähteiden tiedusteluun tarkoitettuja työkaluja, joita alan ammattilaiset ovat suositelleet artikkeleissa, blogikirjoituksissa ja tutkimuksissa. Testausten tuloksena saatiin esimerkkejä tiedustelussa syntyvistä tietotyypeistä ja uhkatiedon keräämisen mahdollisuuksista.

Työkaluja testattiin joko niiden omassa web-käyttöliittymässä, tai virtuaalisessa Kali Linux -ympäristössä. Työkalujen käyttöohjeita löytyi yleensä työkalun omasta dokumentaatiosta. Työssä käytetyt työkalut on listattu lähdeluetteloon.

Työssä integroitiin OSINT-tiedustelusta saatava uhkatieto tutkimukseen tiedon jakamisesta ja siihen liittyvästä lainsäädännöstä ja etiikasta. Lähteinä käytettiin lakiasiakirjoja, direktiivi-ehdotuksia, sekä alan ammattilaisten kirjoittamia tieteellisiä artikkeleita ja tutkimuksia.

Tiedonhaussa käytettiin pääasiassa hakukoneita, kuten Google, Google Scholar ja DuckDuckGo. Esimerkkejä hakukoneissa käytetyistä hakusanoista ovat mm. ”OSINT”, ”OSINT in threat intelligence”, ”OSINT regulations”, ”OSINT ethics”, ”Threat intelligence sharing”, ”Open-Source intelligence legislation”. Erityisesti OSINT-tiedustelun tekniikoista ja hyödyistä käsittelevässä opinnäytetyön osiossa käytettiin tutkimuskirjallisuuden lisäksi tutkimusaineistoa, eli alan asiantuntijoiden blogeista ja artikkeleista löytyvää yhtenäistä tietoa ja näkemystä. Lähteitä analysoitiin objektiivisesti sisällönanalyysi-menetelmällä, ja työssä pyrittiin käyttämään mahdollisimman tuoreita lähteitä.

3 Avointen tietolähteiden tiedustelu

Tiedustelu on kansallisen tai kansainvälisen uhkatiedon keräämistä. Tiedustelu voi varoittaa mahdollisista uhkista tai mahdollisuuksista, tuottaa tietoa henkilöistä tai sitä voidaan hyödyntää vastatiedustelussa (counterintelligence). (Office of the Director of National Intelligence 2023)

Avointen tietolähteiden tiedustelu (Open Source Intelligence, OSINT) tarkoittaa tiedustelua, jota tuotetaan keräämällä, arvioimalla ja analysoimalla julkisesti saatavilla olevaa tietoa. Avointen tietolähteiden tiedustelua tekevät, kuten tietoturva-ammattilaiset, pahantahtoiset hakkerit tai valtiolliset tiedustelupalvelut, hyödyntävät edistyksellisiä tekniikoita etsiäkseen

suuresta tietomäärästä juuri sitä tietoa, joka auttaa heitä saavuttamaan tavoitteensa. (SANS Institute 2023; Fruhlinger ym. 2023)

3.1 Käyttötavat

Avointen tietolähteiden tiedustelua käytetään tyypillisesti seuraavilla tavoilla:

- Turvallisuus ja tiedustelu: Avointen tietolähteiden tiedustelulla voidaan kerätä tietoa mahdollisista uhkista turvallisuudelle, kuten terrorismista tai kyberhyökkäyksistä. Sen avulla voidaan myös kerätä tiedustelutietoa ulkomaisista hallituksista, yrityksistä tai yksityishenkilöistä.
- Liiketoiminta ja markkinatutkimus: Avointen tietolähteiden tiedustelulla voidaan kerätä tietoa kilpailijoista, alan trendeistä ja kuluttajakäyttäytymisestä. Tätä tietoa voidaan hyödyntää liiketoiminnan strategiassa ja päätöksenteon tukena.
- Tutkiva journalismi: Avointen tietolähteiden tiedustelulla voidaan kerätä tietoa mm. politiikasta, liiketoiminnasta ja rikollisuudesta.
- Akateeminen tutkimus: Tutkijat voivat käyttää avointen tietolähteiden tiedustelua kerätäkseen tietoa mm. trendeistä, mielipiteistä ja taloudellisista indikaattoreista. (SANS Institute 2023)

Avointen tietolähteiden tiedustelua hyödyntävät esimerkiksi hallitukset, lainvalvontaviranomaiset, puolustusvoimat, tutkivat journalistit, ihmisoikeustutkijat, yksityisetsivät, asianajotoimistot, yritysten tietoturva- ja kybervalvontatiimit ja penetraatiotestaajat (SANS Institute 2023). Sitä käytetään myös yksityiselämässä esimerkiksi ihmisten etsimiseen (Maor 2022).

3.2 Avoimen tiedustelutiedon tietolähteet

Avointa tiedustelutietoa kerätään monista eri tietolähteistä, kuten kirjastoista, uutisista, artikkeleista, nettisivuilta, sosiaalisesta mediasta, blogeista, hakukoneista (mukaan lukien erikoishakukoneet, esim. Shodan), syvästä verkosta, pimeästä verkosta, kuvista, videoista, lähdekoodeista tai metatiedoista. (Kaspersky 2023; SANS Institute 2023)

Tietotyypit voidaan jakaa jäsenneltyihin ja jäsentämättömiin tietoihin. Jäsennellyt tietolähteet tarjoavat jäsenneltyä dataa tarkasti määritellyssä muodossa ja sitä on mahdollista käsitellä koneellisesti. Esimerkkejä jäsennellyistä tiedoista ovat IP-osoitteiden mustat listat tai CVE-tiedostot. Jäsenneltyä tietoa voidaan kerätä esimerkiksi verkkosivujen haravoinnilla. (Ferreira 2018, 9)

Jäsentämättömät tietolähteet tarjoavat jäsentämätöntä dataa, jossa sisältö pääasiassa tekstimuodossa. Vaikka näiden käsittely vaatii enemmän työtä, jäsentämättömät tiedot sisältävät usein runsaasti hyödyllistä tietoa. Esimerkkejä jäsentämättömistä tiedoista ovat uutiset, kir-

joitukset viestipalvelu X:ssä (ent. Twitter) tai pimeän verkon keskustelupalstat. Jäsentämätöntä tietoa voidaan kerätä esimerkiksi verkkosivujen haravoinnilla, luonnollisen kielen käsittelytyökaluilla (Natural Language Processing, NLP) tai koneoppimismalleilla. (Ferreira 2018, 9)

3.3 Tiedustelutekniikat

Avointen tietolähteiden tiedustelua voidaan tehdä passiivisesti, semi-passiivisesti tai aktiivisesti. (SANS Institute 2023)

Passiivisessa menettelytavassa ei olla vuorovaikutuksessa kohdejärjestelmien kanssa, eikä siinä kommunikoida tai olla tekemisissä ihmisten kanssa. Passiivisessa tiedustelussa voidaan esimerkiksi skannata julkisia verkkosivuja, hakea tietoa avoimista sovellusrajapinnoista tai hakea tietoa pimeästä verkosta. (SANS Institute 2023)

Semi-passiivisella tiedustelulla viitataan esimerkiksi verkkoskannaukseen, eli liikenteen ohjaamiseen kohdepalvelimeen tietojen saamiseksi. Skannausliikenteen on oltava samanlaista kuin normaali Internet-liikenne havaitsemisen välttämiseksi. (Imperva 2023)

Aktiivinen tiedustelu tarkoittaa, että ollaan jollain tavalla tekemisissä tai vuorovaikutuksessa kohteen kanssa, eli esimerkiksi lisätään kohde ystäväksi sosiaalisessa mediassa, lähetetään kohteelle viesti, rekisteröidytään nettisivuille lataamaan rekisteröityneille tarkoitettua materiaalia, tai skannataan avoimia portteja ja ohjelmistohaavoittuvuuksia. (Kaspersky 2023; SANS Institute 2023)

4 Kyberuhkatietoa avointen tietolähteiden tiedustelulla

Avointen tietolähteiden tiedustelun avulla organisaatiot voivat varautua niihin kohdistuviin kyberuhkiin räätälöimällä vastatoimensa potentiaalisten kyberuhkatoimijoiden taktisiin, operatiivisiin ja strategisiin menettelytapoihin. (Slinde 2023, 2)

Organisaatiot voivat siirtyä reaktiivisesta puolustamisesta proaktiiviseen toimintaan suojelemalla itseään kyberhyökkäyksiltä ja havaitsemalla poikkeavuuksia avointen tietolähteiden tiedustelun avulla. (Slinde 2023, 19)

Hyökkääjää hyödyttäviä tietoja voivat olla esimerkiksi avoimet portit, haavoittuva ja päivittämätön ohjelmisto, huonosti määritelty pilvitallennustila, tunnukset ohjelmistokoodissa ja tietojärjestelmätiedot kuten laitteiden nimet, IP-osoitteet ja kokoonpanot. Organisaation ulkopuoliset verkkosivustot, kuten sosiaalinen media, sisältävät valtavia määriä tietoa, erityisesti työntekijöistä. Tietoa työntekijöistä voidaan hyödyntää esim. kohdennetussa tietojenkalasteluhyökkäyksessä tai salasanojen murtamisessa. Myös toimittajien ja kumppanien jakamat

tiedot voivat olla merkityksellisiä hyökkäyksen suunnittelussa. (Yadav, Kumar & Singh 2023; Imperva 2023)

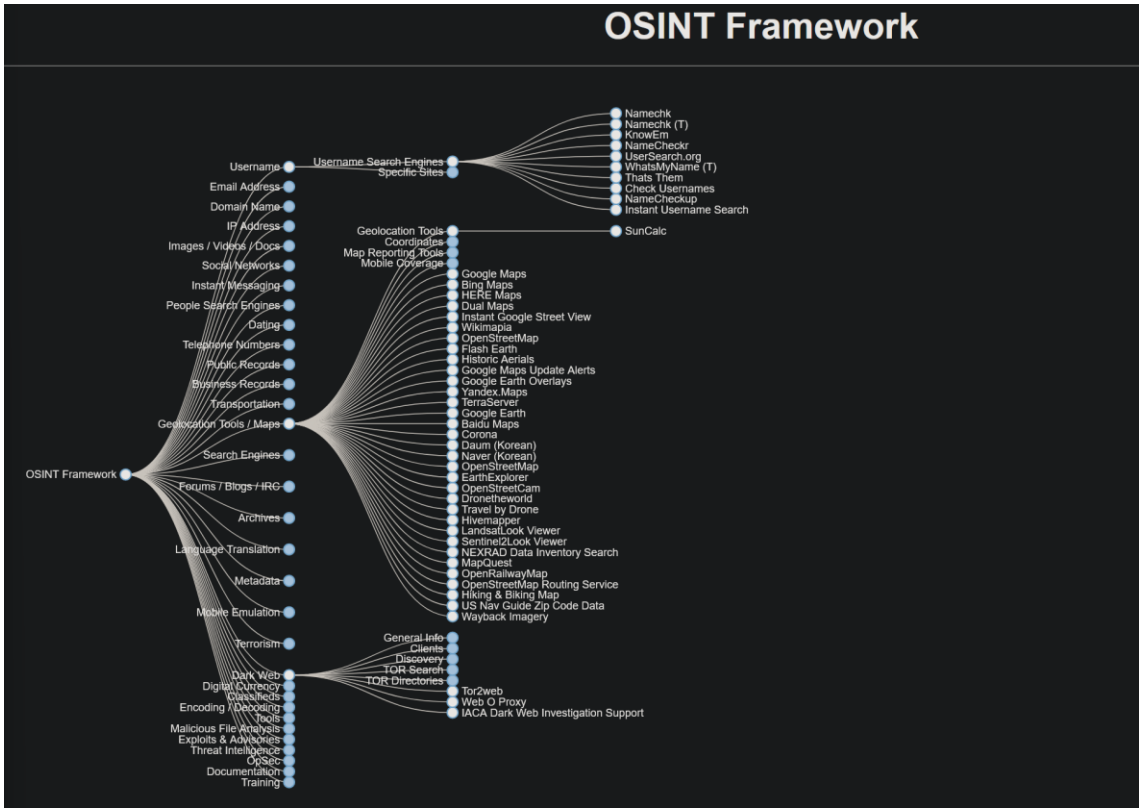
5 Tekniikat ja työkalut

Avointen tietolähteiden tiedusteluun sopivia työkaluja on valtava määrä. Niitä tulee jatkuvasti lisää, samalla kun osa poistuu käytöstä. Tähän opinnäytetyöhön on valikoitu käytetyimpiä työkaluja OSINT-tutkinnassa, keskittyen lähinnä ilmaisiin, avoimen lähdekoodin työkaluihin, jotka voidaan ohjelmointirajapinnan avulla yhdistää esimerkiksi keskitettyyn uhkien tiedustelualustaan (kuten MISP tai Cortex) jatkoanalysointia ja toimenpiteitä varten. Maksullisilla lisensseillä usein saadaan enemmän ominaisuuksia käyttöön tai vapautettua rajoituksia kyselyjen määrästä.

Pelkkiin työkaluihin ei voi luottaa, vaan avointen tietolähteiden tiedusteluun tarvitaan ammattilainen, joka osaa etsiä löydetystä datasta merkittävät asiat. Täyttä automaatiota tarjoavat tuotteet aliarvioivat modernien organisaatioiden tietoverkkojen monimutkaisuuden ja tiedon määrän, ja täysin niihin luottaminen voi johtaa vääriin johtopäätöksiin. (Micallef 2021)

OSINT Framework

OSINT Framework on kattava OSINT-työkalujen viitekehys, josta löytää sopivan työkalun helposti tilanteeseen kuin tilanteeseen. Viitekehys listaa ilmaisia työkaluja tai resursseja, joista on hyötyä avointen tietolähteiden tiedustelussa. Osa työkaluista saattaa vaatia rekisteröitymisen tai maksun, jotta saa kaikki ominaisuudet käyttöön. Kuvassa 1 on näkymä OSINT Frameworkista. (OSINT Framework 2023)



Kuva 1: OSINT Framework

5.1 Tekniikat

Hakuoperaattorit

Edistyneiden hakuoperaattoreiden käyttö hakukoneissa (dorking, hakukonehakkerointi), on tekniikka, jota hyödyntämällä Googlen kaltaisista hakukoneista voi löytää esimerkiksi vuotaneita dokumentteja tai tietoturvahaavoittuvuuksia. (Byron 2023)

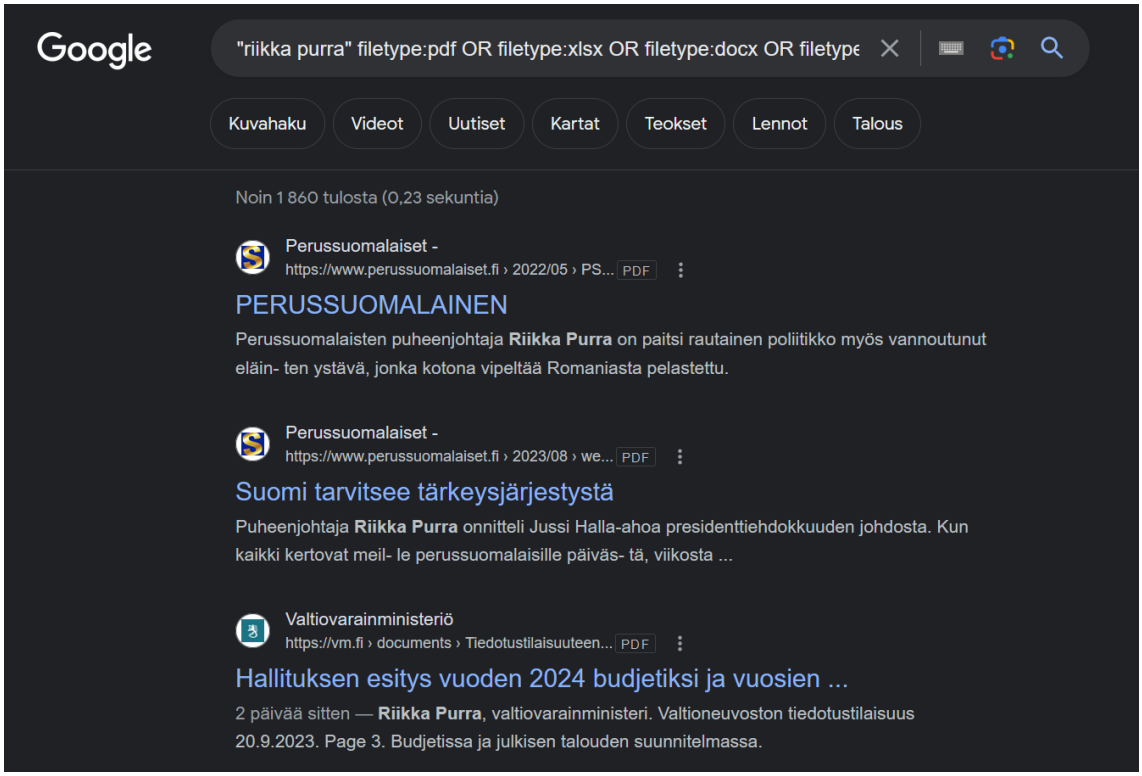
Esimerkkejä usein käytetyistä hakuoperaattoreista ovat:

- Filetype: etsii tiettyjä tiedostotyyppejä (esimerkiksi filetype:(jpg OR png))
- Ext: etsii tiettyjä tiedostotyyppejä tiedostotunnisteen perusteella (esimerkiksi ext:log)
- Intitle/intext/inurl: etsii annettua tietoa sivuston otsikosta, sisällöstä tai osoitteesta (esimerkiksi inurl:admin)
- Lainausmerkeillä voi etsiä tarkkaan määritettyä tekstijonoa (esimerkiksi ”John J. Doe”)
- Before ja after: rajaa hakutuloksia ajan mukaan (esimerkiksi after:2015-10-25 AND before:2016-10-25)
- Site: etsii tuloksia vain määritellyiltä sivuilta (esimerkiksi site:Wikipedia.org)

- Cache: palauttaa viimeksi Googlen välimuistiin tallennetun version verkkosivusta (esimerkiksi cache:petteriorpo.fi)
- -: jättää tuloksista määritellyt tulokset pois (esimerkiksi -site:Wikipedia.org)
- *: antaa hakukoneen täydentää tähden sisällön (esimerkiksi sähköpostiosoitetta hakiessa voi käyttää [käyttäjätunnus]*com) (Byron 2023; Imperva 2023; Maltego 2021)

Seuraavassa esimerkissä käytetään hakuparametreja kaikkien PDF-, Excel-, Word- tai PowerPoint -tiedostojen etsimiseen, joissa esiintyy nimi ”Riikka Purra”:

"Riikka Purra" filetype:pdf OR filetype:xlsx OR filetype:docx OR filetype:PPT



Kuva 2: Esimerkki hakuparametrien käytöstä

Säännölliset lausekkeet

Säännöllisiä lausekkeita (Regular expressions, regex) voidaan hyödyntää määriteltyjen tietojen haravoinnissa jäsentelemätöntä tietoa käsiteltäessä. Tällä tekniikalla voi esimerkiksi etsiä suurenkin tekstin joukosta IP-osoitteita, tiedostojen jäännösarvoja, URL-osoitteita tai Bitcoin-osoitteita. Tekniikan käyttäminen on tehokasta riippumatta tekstin määrästä ja siihen löytyy kattavasti ohjeita. Tähän löytyy myös aputyökaluja, kuten regex101, jolla voi kokeilla lausekkeen toimivuutta ennen käyttöä. (Gaucheler 2023; regex101 2023)

Alla olevassa taulukossa on esimerkkejä erilaisista säännöllisistä lausekkeista, joita voi käyttää haavoittuvuustunnusten (CVE), jäännösarvojen, verkko-osoitteiden, IP-osoitteiden ja Bitcoin-osoitteiden hakemiseen.

Taulukko 1: Esimerkkejä säännöllisistä lausekkeista (mukaillen Gaucheler 2023)

Haavoittuvuustunnus (CVE)	CVE-\d{4}-\d*
Jäännösarvot (MD5, SHA1, SHA256)	[A-Fa-f0-9]{64}|[a-fA-F0-9]{40}|[a-fA-F0-9]{32}
URL	(?:h..ps?|f.p)://(?:www(?:.|[.])|(?!www))[a-zA-Z0-9]+a-zA-Z0-9[^\s]{2,}|www(?:.|[.])[a-zA-Z0-9]+a-zA-Z0-9[^\s]{2,}|https?://(?:www(?:.|[.])|(?!www))[a-zA-Z0-9]+(?:.|[.])[^\s]{2,}|www(?:.|[.])[a-zA-Z0-9]+(?:.|[.])[^\s]{2,}
IPv4	((?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?\.){3}(?:(?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)))
IPv6	[0-9a-fA-F]{1,4}(?::[0-9a-fA-F]{1,4}){7}|(?:(?:[0-9A-Fa-f]{1,4}(?::[0-9A-Fa-f]{1,4}){0,5})?)::(?:(?:[0-9A-Fa-f]{1,4}(?::[0-9A-Fa-f]{1,4}){0,5})?)
Bitcoin-osoite	(?:[13]|bc1)[a-zA-HJ-NP-Z0-9]{26,62}

Seuraava esimerkki käyttää Pythonin sisäänrakennettua *re* -moduulia etsiäkseen tekstin joukosta IPv4-osoitteita:

```
import re

txt = "Osoitteeni on 192.168.1.1. Mikä sinun osoitteesi on?"

regex_ipv4 = "((?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?\.){3}(?:(?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)))"

print(re.findall(regex_ipv4, txt))
```

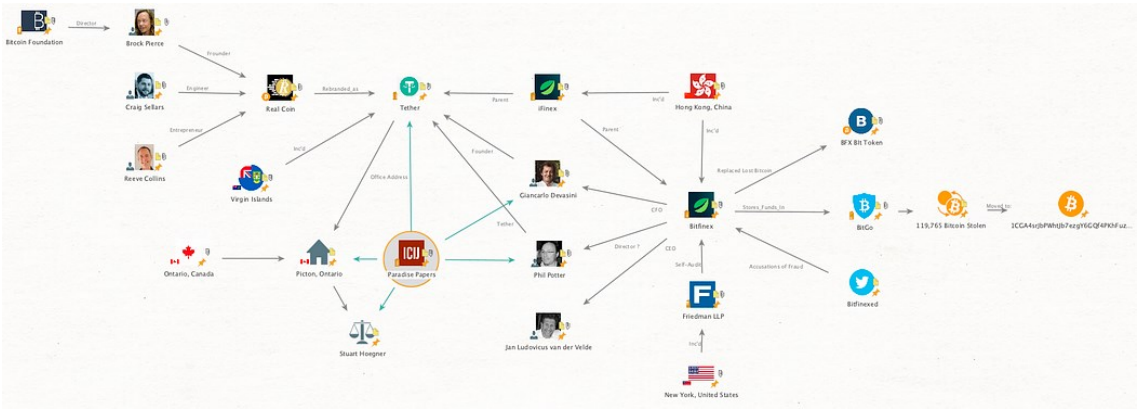
5.2 Monikäyttöiset työkalut

Maltego

Maltego ei ole avoimen lähdekoodin ohjelmisto, ja sen ilmainen kokeiluversio on hyvin rajoitettu, mutta se on yksi suosituimmista avointen tietolähteiden tiedusteluun käytetyistä työkaluista, joten se ansaitsee maininnan myös tässä yhteydessä.

Maltego on Javalla pyörivä ohjelmisto, joka toimii Windowsilla, Macilla ja Linuxilla. Maltegossa on valmiina yli 70 ohjelmointirajapintayhteyttä erilaisiin tietolähteisiin, ja siihen voi yhdistää myös omia tietolähteitä. Maltego erikoistuu suhteiden ja yhteyksien löytämiseen ihmisten, yritysten, verkkotunnusten ja muiden avointen tietojen välillä, sekä niiden visualisoimiseen helppolukuisiksi kaavioiksi ja kuvioiksi. (Fruhlinger ym. 2023; Maltego 2022)

Kuvassa 3 on Maltegolla luotu yhteyskartta Bitfinex-hyökkäyksestä, jossa näkyy hyökkäykseen liitetyt henkilöt, yritykset ja paikat, sekä niiden yhteydet toisiinsa. (Mader 2022)



Kuva 3: Maltegolla luotu yhteyskartta Bitfinex-hyökkäyksestä (Mader 2022)

TheHarvester

TheHarvester on avoimeen lähdekoodiin perustuva avointen tietolähteiden tiedustelutyökalu, jolla voi etsiä mm. sähköpostiosoitteita, nimiä, aliverkkotunnuksia, IP-osoitteita ja verkko-osoitteita. Työkalu on suunniteltu etsimään avointa tiedustelutietoa organisaatiosta tai verkkotunnuksesta. Tietolähteinä käytetään mm. suosittuja hakukoneita, kuten Bing ja DuckDuckGo, mutta myös vähemmän tunnettuja tietolähteitä, kuten cerspotter ja DNSdumpster. (Fruhlinger ym. 2023; Martorella 2023)

Kuva 4: theHarvester-työkalun ohjeet

Työkalua on helppo käyttää. Kuvassa 4 näkyy ohjeet ja mahdolliset lisävalinnat työkalun käyttöön. Seuraava komento etsii kaikki verkkotunnukseen ”laurea.fi” liittyvät autonomisten järjestelmien numerotunnukset (ASN), IP-osoitteet, sähköpostiosoitteet ja aliverkkotunnukset. Komennossa käytettävä valinta ”-b all” käyttää kaikkia mahdollisia tietolähteitä, mutta koska työkaluun ei ole syötetty API-avaimia, käytössä on vain sellaiset lähteet, jotka eivät vaadi API-avainta. Komento hyödyntää vain passiivisia tiedonkeruumenetelmiä. Komentoa voisi vielä tehostaa ottamalla käyttöön esimerkiksi valinnan –c (DNS brute force), mutta silloin mentäisiin eettisesti ja lainsäädännöllisesti harmaalle alueelle, koska silloin olisimme aktiivisesti vuorovaikutuksessa verkkotunnuksen kanssa.

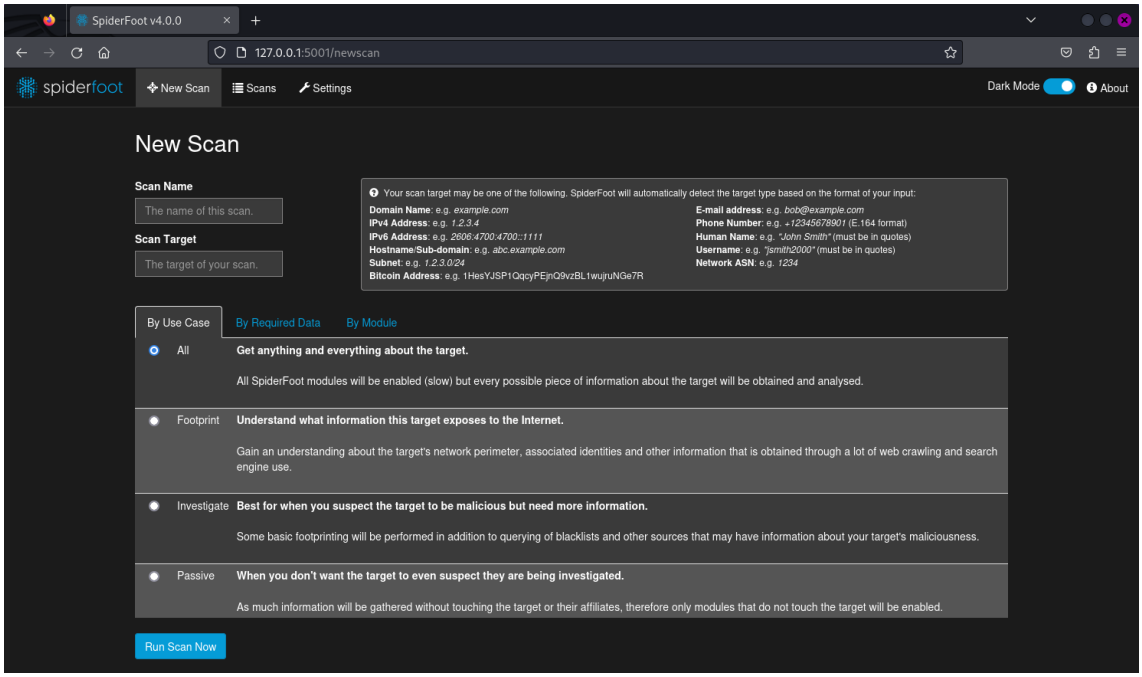
theHarvester -d laurea.fi -b all

TheHarvester löysi viisi ASN-tunnusta, 77 IP-osoitetta, 61 sähköpostiosoitetta ja 414 aliverkkotunnusta. Kyseessä on suuri määrä hyödyllistä tietoa mahdolliselle uhkatoimijalle, ja organisaatioiden on hyvä tietää, mitä kaikkea tietoa yksinkertaisella hakukomennolla voi löytyä.

SpiderFoot

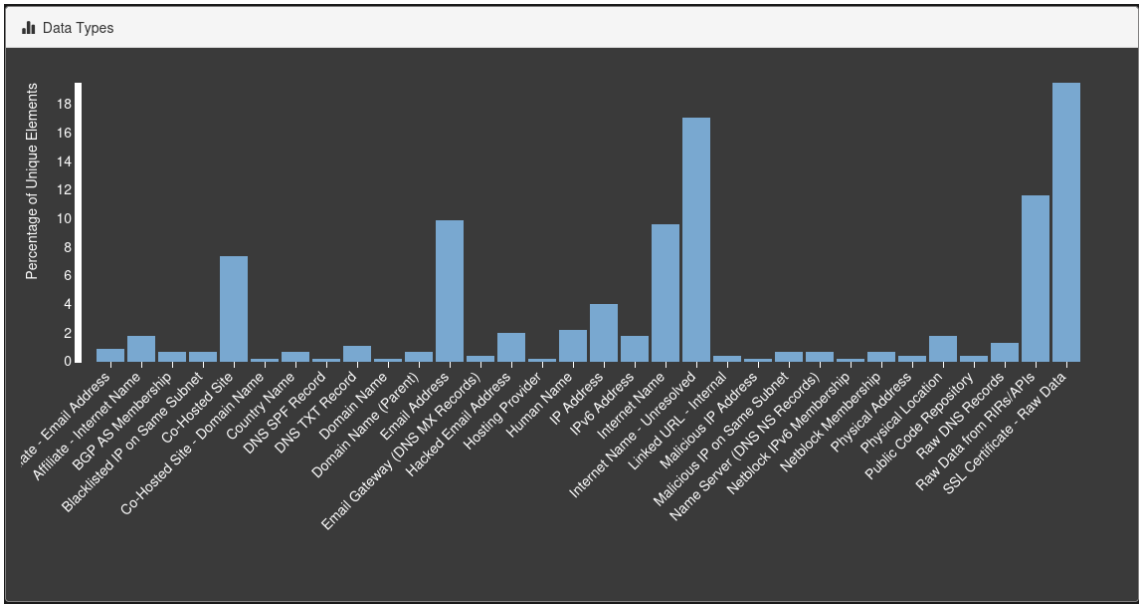
Spiderfoot on myös avoimen lähdekoodin työkalu, jolla voidaan etsiä ja kerätä IP-osoitteita, verkkotunnuksia, aliverkkotunnuksia, ASN-tunnuksia, sähköpostiosoitteita, puhelinnumeroita, nimiä, käyttäjätunnuksia ja Bitcoin-osoitteita. Työkalussa on yli 200 moduulia, joiden avulla voidaan etsiä esimerkiksi omasta organisaatiosta julkiseksi päätyneitä tietoja, kuten vuotaneita kirjautumistietoja. (Fruhlinger ym. 2023; Micallef 2023)

Kuvassa 5 näkyy SpiderFootin käyttöliittymä asentamisen jälkeen. Käytössä on valmiita asetuksia eri tilanteisiin. Myös tässä kannattaa muistaa, että kohteesta riippuen täyden skannauksen tekeminen ei välttämättä ole laillista. Passiivisella menetelmällä usein saadaan jo kattavasti tietoa, joten etsintä kannattaa aloittaa siitä.



Kuva 5: SpiderFoot käyttöliittymä

Tekemällä passiivisen skannauksen satunnaisesta suomalaisesta organisaatiosta, tietoa organisaation käyttämistä teknologioista löytyy kattavasti. Myös työntekijöiden nimiä ja sähköpostiosoitteita löytyi useita kymmeniä ja jopa viittauksia mahdollisiin haavoittuvuuksiin (kuva 6). Kyseinen skannaus löysi esimerkiksi 11 työsähköpostiosoitetta, joista on salasana vuotanut (Hacked Email Address). Onkohan kaikki salasanat vaihdettu? Kuten kuvasta 6 näkee, erilaisia tietotyyppejä on paljon ja kyseisen organisaation olisi hyvä käydä läpi, onko kaikki tieto varmasti tarkoitettu julkiseksi.

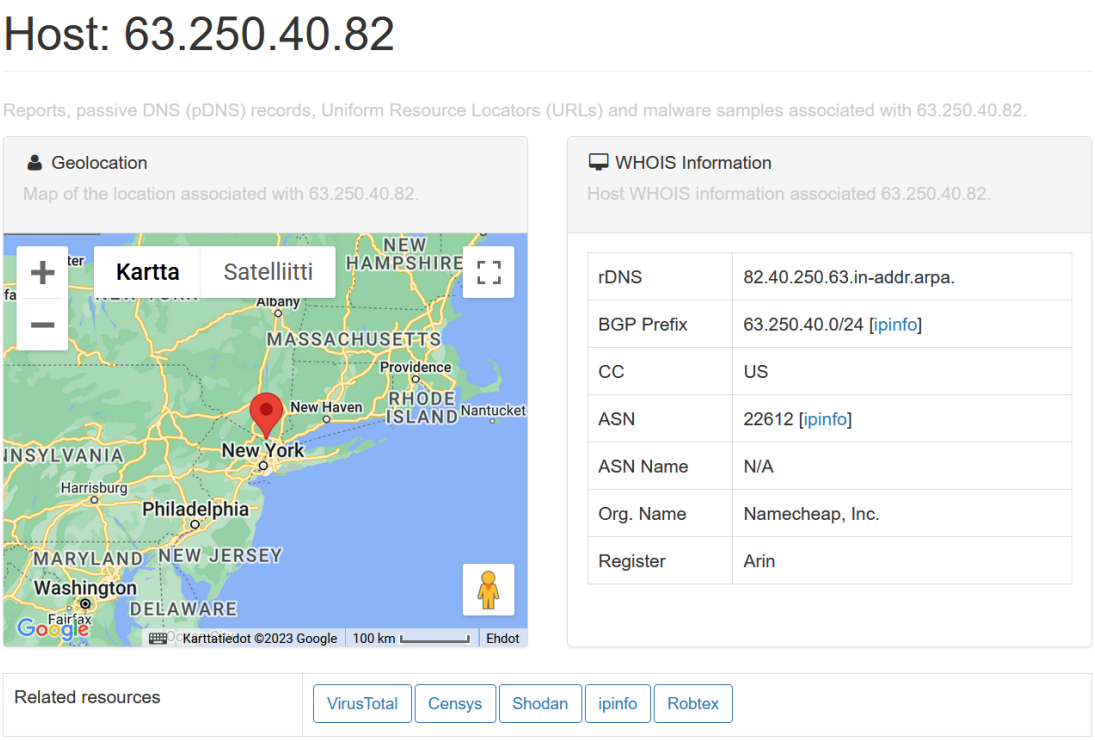


Kuva 6: SpiderFoot yhteenveto löydöksistä

ThreatMiner

ThreatMiner hyödyntää useita eri tietolähteitä (mm. VirusTotal ja Alienvault OTX) etsiäkseen vaarantumisindikaattoreita mm. verkkotunnusten, IP-osoitteiden, sähköpostiosoitteiden ja tiedostonimien perusteella. Se toimii omalla web-käyttöliittymällään ja se on yhdistettävissä ohjelmointipinnalla esimerkiksi MISP tai Splunk -alustoihin. (ThreatMiner 2023)

Kuvassa 7 näkyy syötetyn verkkoisännän hakutulokset. ThreatMiner hakee verkosta olevasta laitteesta mm. sijaintitiedot, WHOIS-tiedot, osoitteeseen linkitetyt uhkatiedot, historialliset tiedot DNS-nimenselvitysten perusteella, sekä isäntään liittyvät URL-osoitteet.



Kuva 7: ThreatMiner hakutulokset

5.3 Sosiaalinen media

Sherlock

Sherlock on Pythonilla ohjelmoitu avoimen lähdekoodin työkalu, joka hakee syötettyä käyttäjätunnusta sosiaalisen median palveluista. (Sherlock Project 2023)

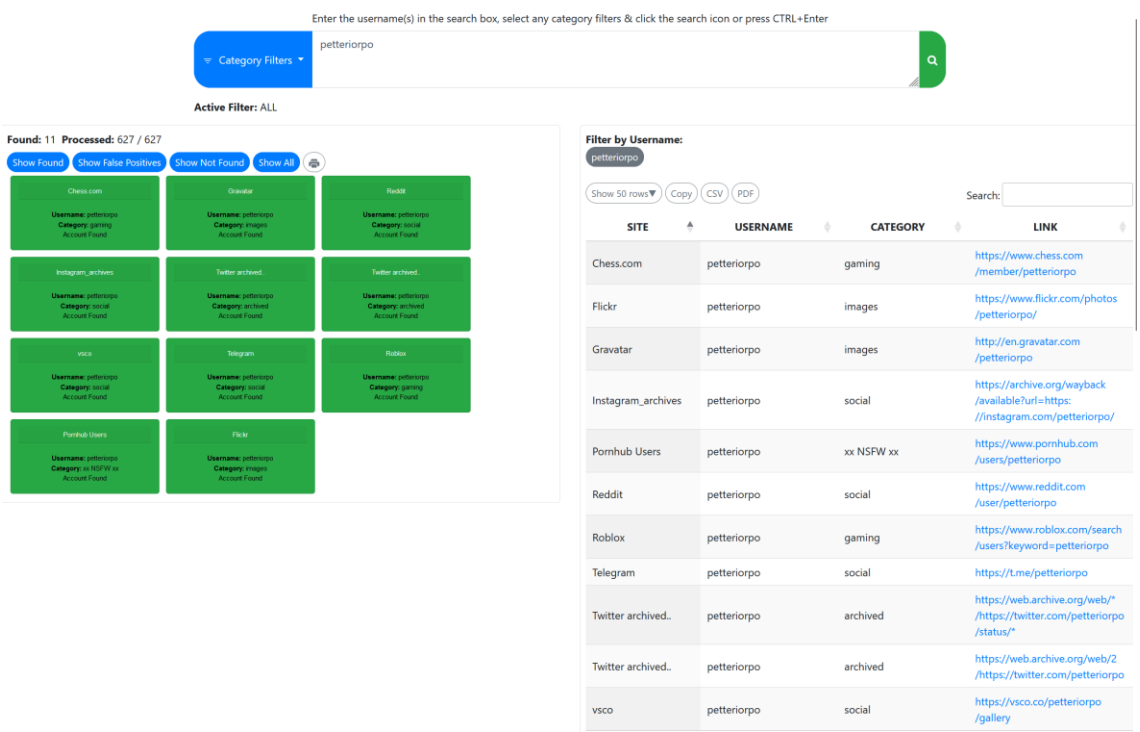
Kuvassa 8 haetaan käyttäjätunnusta ”petteriorpo”, johon Sherlock löytää 15 tulosta eri sosiaalisen median palveluista.

```
┌──(od㉿kali)-[~/sherlock]
└─$ python3 sherlock petteriorpo
[*] Checking username petteriorpo on:

[+] Chess: https://www.chess.com/member/petteriorpo
[+] Fiverr: https://www.fiverr.com/petteriorpo
[+] Flickr: https://www.flickr.com/people/petteriorpo
[+] G2G: https://www.g2g.com/petteriorpo
[+] Gravatar: http://en.gravatar.com/petteriorpo
[+] Lolchess: https://lolchess.gg/profile/na/petteriorpo
[+] Reddit: https://www.reddit.com/user/petteriorpo
[+] Roblox: https://www.roblox.com/user.aspx?username=petteriorpo
[+] RuneScape: https://apps.runescape.com/runemetrics/app/overview/player/petteriorpo
[+] Scribd: https://www.scribd.com/petteriorpo
[+] Telegram: https://t.me/petteriorpo
[+] Twitch: https://www.twitch.tv/petteriorpo
[+] Twitter: https://twitter.com/petteriorpo
[+] VSCO: https://vsco.co/petteriorpo
[+] metacritic: https://www.metacritic.com/user/petteriorpo

[*] Search completed with 15 results
```

Kuva 8: Sherlock hakutuloksia

WhatsMyName

WhatsMyName hakee myös käyttäjätunnuksella yhteyksiä eri sosiaalisen median palveluihin, mutta lisäksi siinä on web-käyttöliittymä. Kuvassa 9 näkyy hakutulokset käyttäjätunnukselle ”petteriorpo”, joita löytyi 11 kappaletta.



Kuva 9: WhatsMyName hakutuloksia

Cree.py

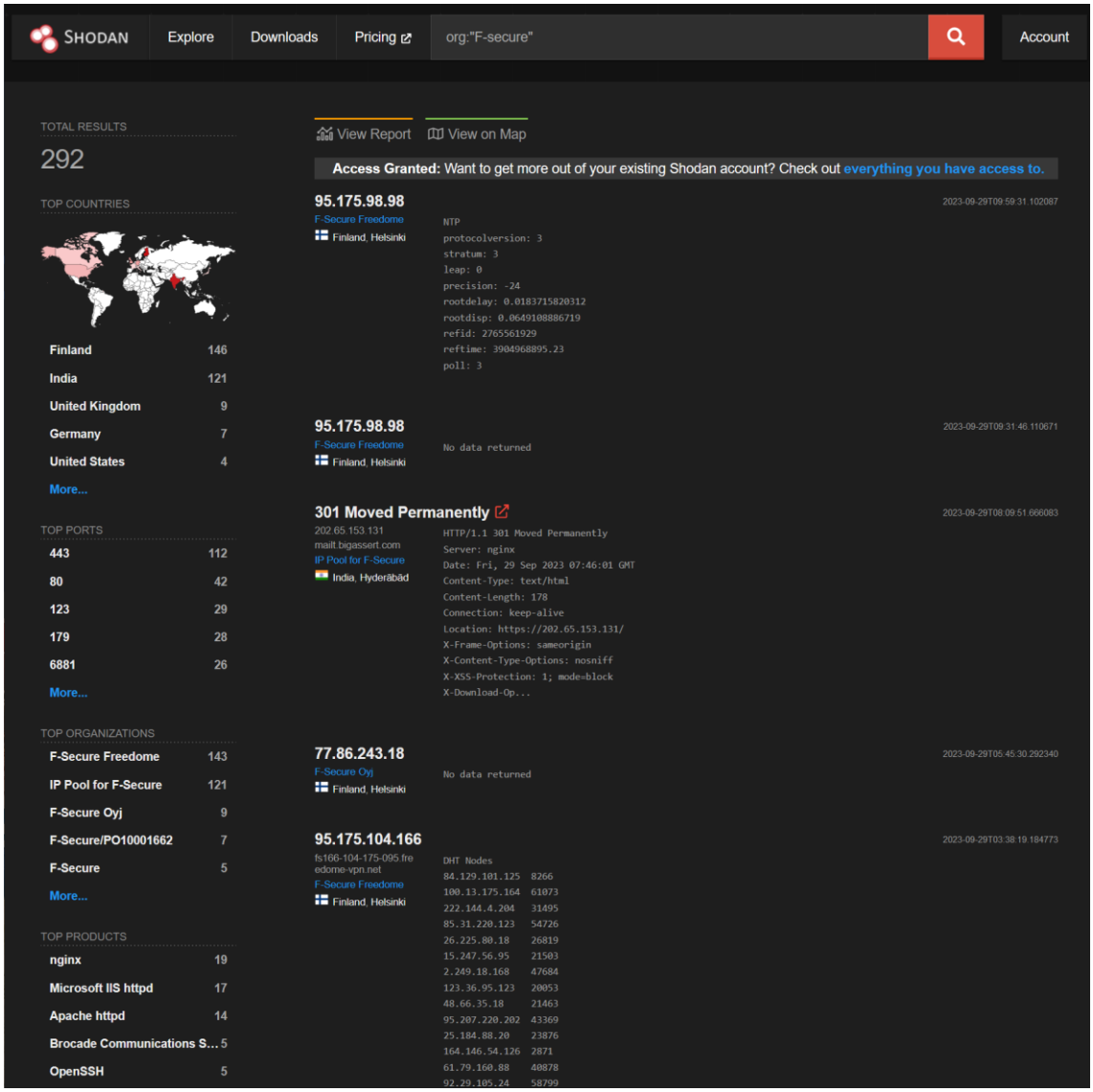
Creepy on Pythonilla ohjelmoitu sijaintitiedon hakutyökalu, joka kerää sijaintitietoja sosiaalisesta mediasta ja näyttää ne kartalla. Sen avulla voidaan katsoa mistä sijainnista sosiaalisen median julkaisut ovat tehty. Työkalua ei enää ylläpidetä, mutta siihen löytyy asennustiedostot ja lähdekoodi sen verkkosivulta. (Cree.py 2023)

5.4 Hakukoneet

Shodan

Shodanilla voi etsiä verkkoon kytkettyjä laitteita (mukaan lukien IoT-laitteita), avoimia portteja ja haavoittuvuuksia. Organisaatio voi Shodanin avulla valvoa omia verkkoon liitettyjä laitteita, niiden käyttöpaikkoja sekä mahdollisia haavoittuvuuksia laitteissa. Shodania voi käyttää rajoitetusti ilmaisella käyttäjätunnuksella, mutta kaikkiin ominaisuuksiin pääsy vaatii kuukausimaksullisen lisenssin. (Fruhlinger ym. 2023; Shodan 2023)

Kuvassa 10 näkyy hakutuloksia F-Securen omistamille laitteille. Tuloksia löytyi yhteensä 292, joista suurin osa sijaitsee Suomessa ja Intiassa. Tuloksia voi rajata käyttämällä Shodanin suodattimia, joista löytyy kattava opas Shodanin verkkosivulta.



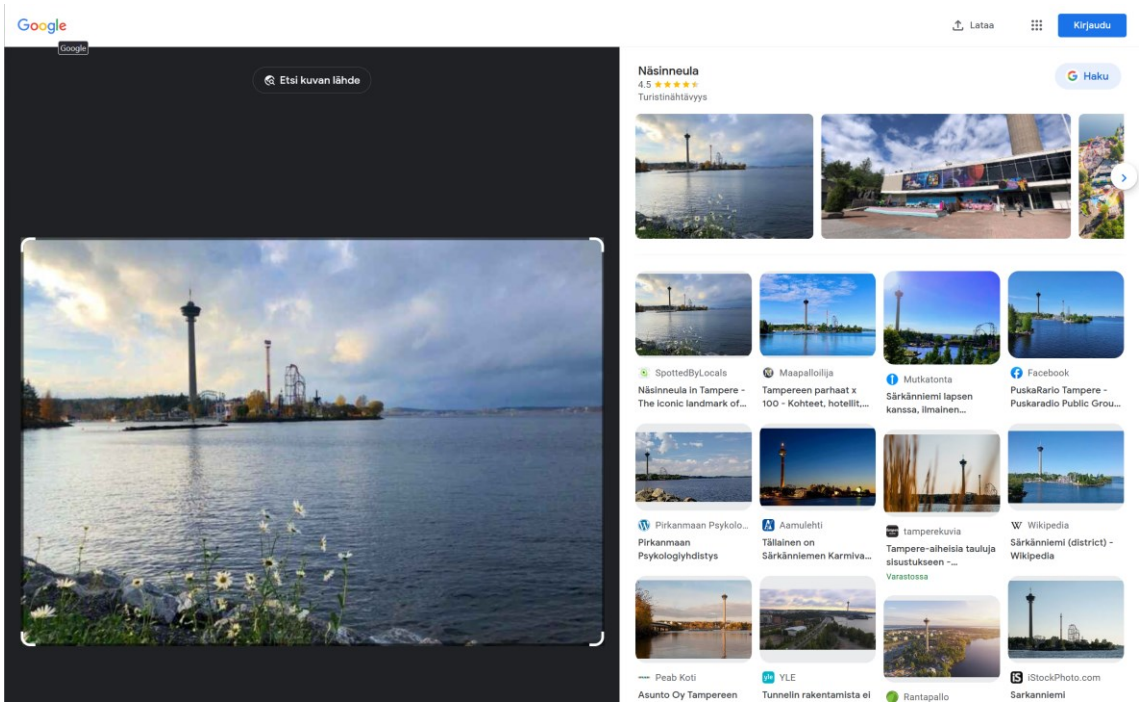
Kuva 10: Shodan hakutuloksia

TinEye

TinEye on tarkoitettu käänteiseen kuvahakuun, eli hakuun syötetään kuva ja TinEye etsii verkosta vastaavia kuvia. Sen avulla voi esimerkiksi moderoida verkkoon julkaistua sisältöä ja havaita brändiin liittyviä väärinkäytöksiä. Sillä voi myös selvittää, minne kaikkialle vuotanut kuva on julkaistu. (Bisson 2021; TinEye 2023)

Google Kuvahaku

Myös Googlella on hakukone käänteiseen kuvahakuun, joka on erityisen hyvä tunnistamaan esimerkiksi sijainteja tai nähtävyyksiä. Kuvassa 11 Google Kuvahaku tunnistaa lähdekuvasta nähtävyyden (Näsinneula) ja tarjoaa hakutuloksia siihen liittyen.

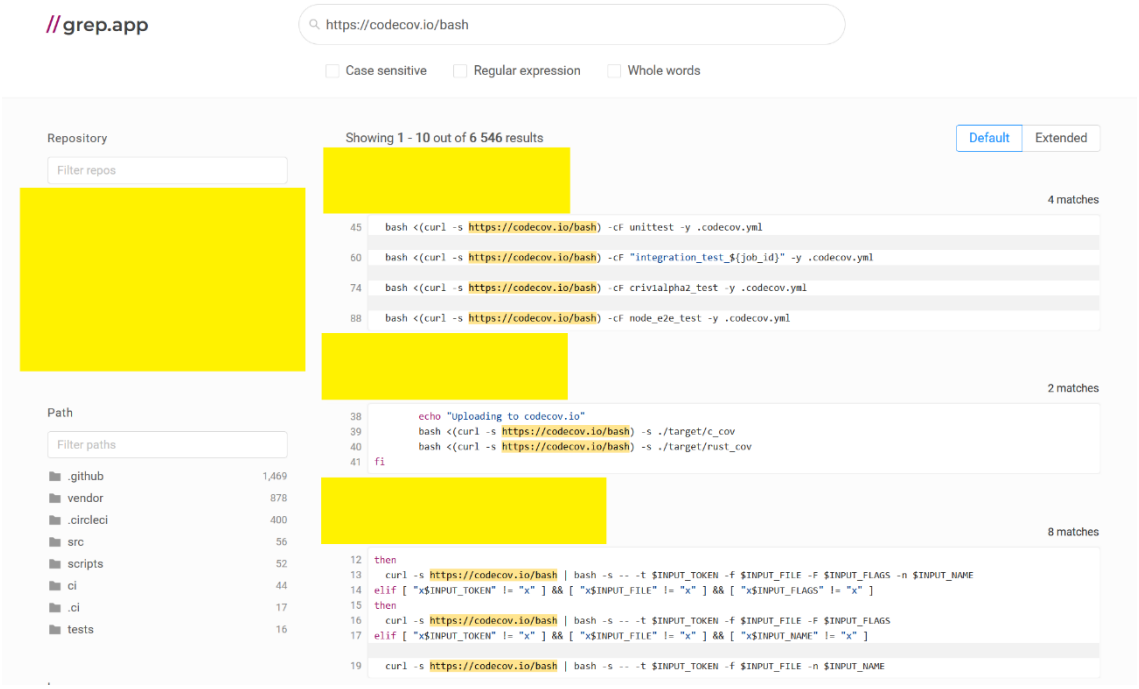


Kuva 11: Google Kuvahaun hakutuloksia

Grep.app

Grep.app on hakukone Git-arkistojen hakemiseen. Sillä voi esimerkiksi etsiä haavoittuvaa ohjelmistokoodia tai haittaohjelmia GitHubista, GitLabista tai BitBucketista. (Fruhlinger ym. 2023)

Kuvassa 12 näkyy Git-arkistoja, joissa on käytetty mm. SolarWinds-hyökkäyksessä hyödynnettyä Codecov Bash Uploaderia.



Kuva 12: Grep.app hakutuloksia

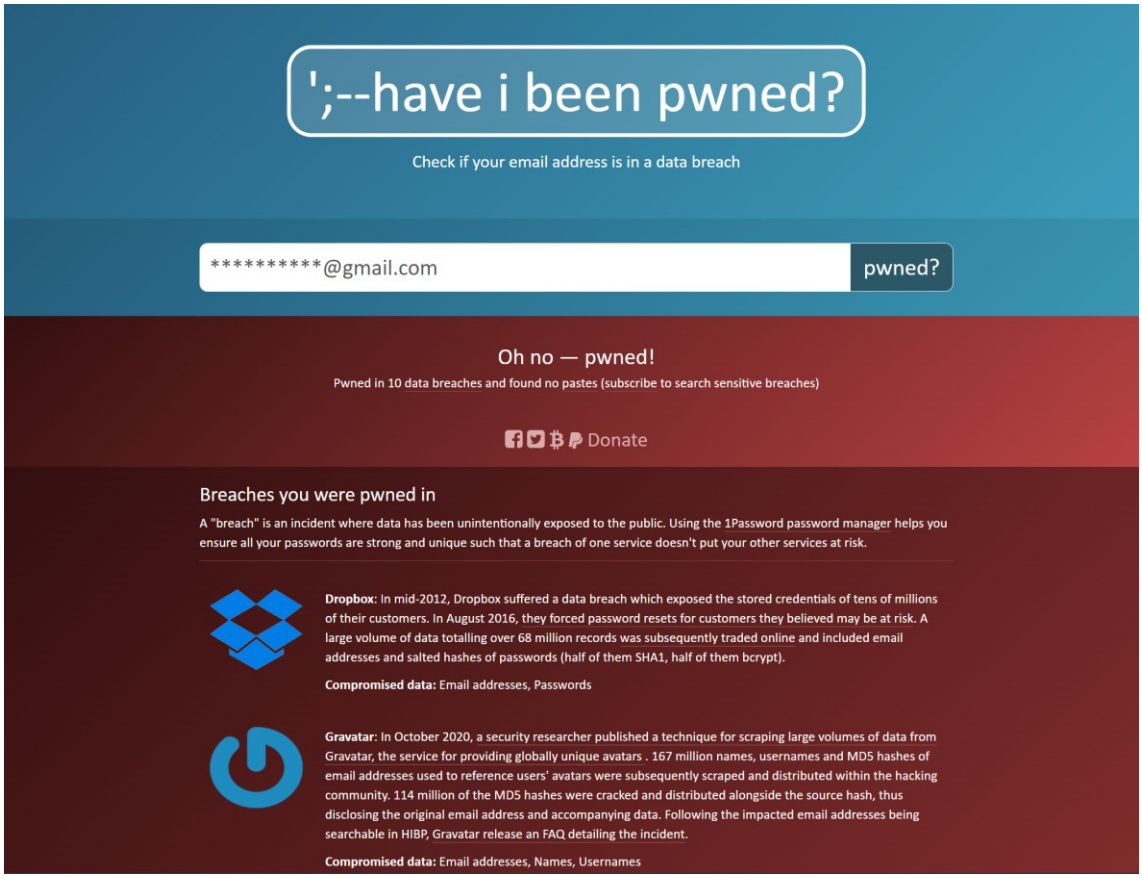
Searchcode

Searchcode hakee hyödyllistä tietoa ohjelmistojen lähdekoodeista. Sillä voi hakea lähdekoodeista esimerkiksi käyttäjätunnuksia, sähköpostiosoitteita tai tietoturvahaavoittuvuuksia (kuten ”eval $_GET”). (Searchcode 2023)

Have I Been Pwned

Have I Been Pwned on hakukone vuotaneiden sähköpostitunnusten hakemiseen. Rekisteröityneet käyttäjät voivat myös tilata ilmoituksia, mikäli heidän sähköpostiosoitteensa havaitaan hyökkäyksen yhteydessä. (Have I Been Pwned 2023)

Kuvan 13 tuloksissa näkyy palvelut, joista kyseinen sähköpostiosoite on vuotanut ja lisätietoa vuotaneista tiedoista.



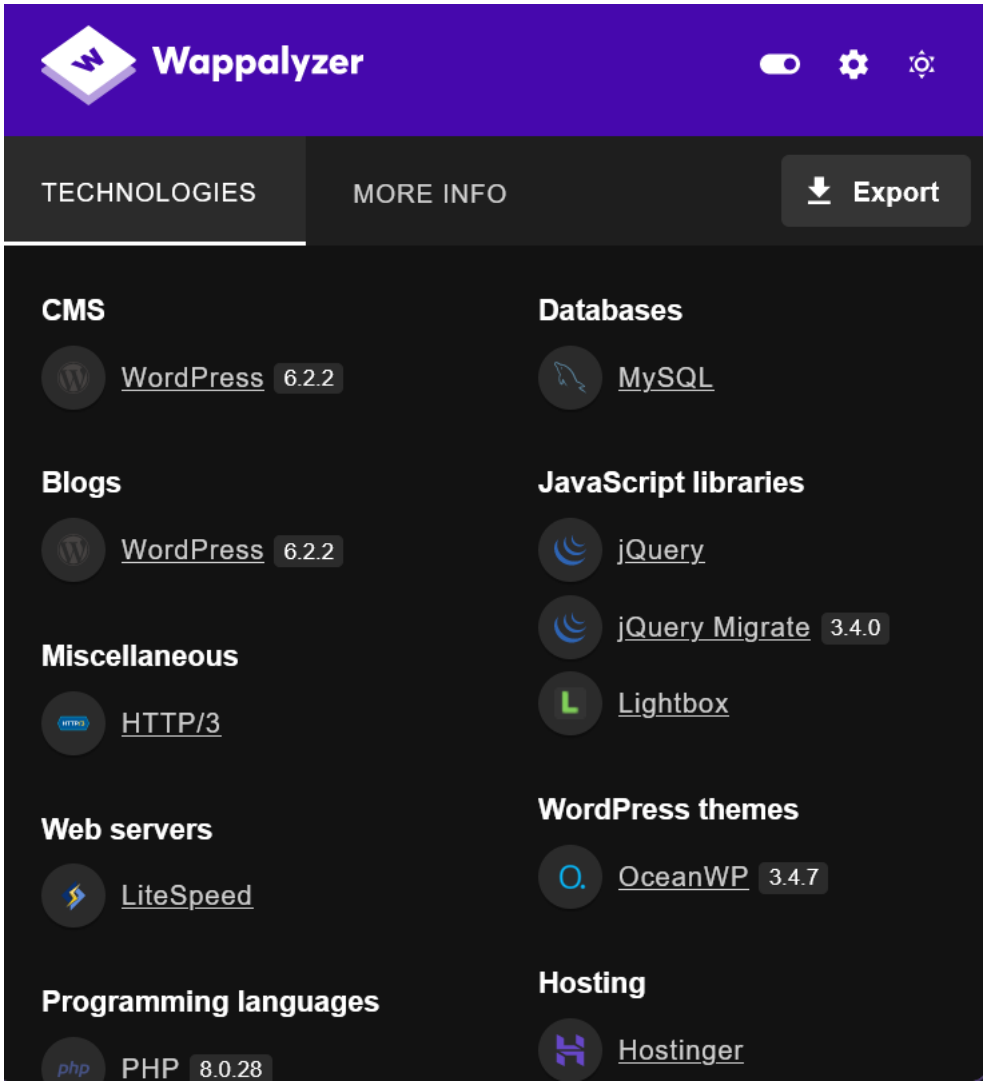
Kuva 13: Have I Been Pwned hakutuloksia

5.5 Verkkosivut

Wappalyzer

Wappalyzer on selainliitännäinen, joka näyttää vierailtavan sivun käytössä olevat teknologiat. Työkalulla voi esimerkiksi tarkistaa, että omien sivujen päivitykset ovat ajan tasalla eikä eri teknologioiden versioissa ole haavoittuvuuksia. Maksullisella lisenssillä Wappalyzer näyttää myös löydetyt yhteystiedot sekä tietoa yrityksestä. (Wappalyzer 2023)

Kuvassa 14 Wappalyzer näyttää yksinkertaisen WordPress-sivuston käytössä olevat teknologiat ja versiot.



Kuva 14: Wappalyzer näkymä

PhishTank

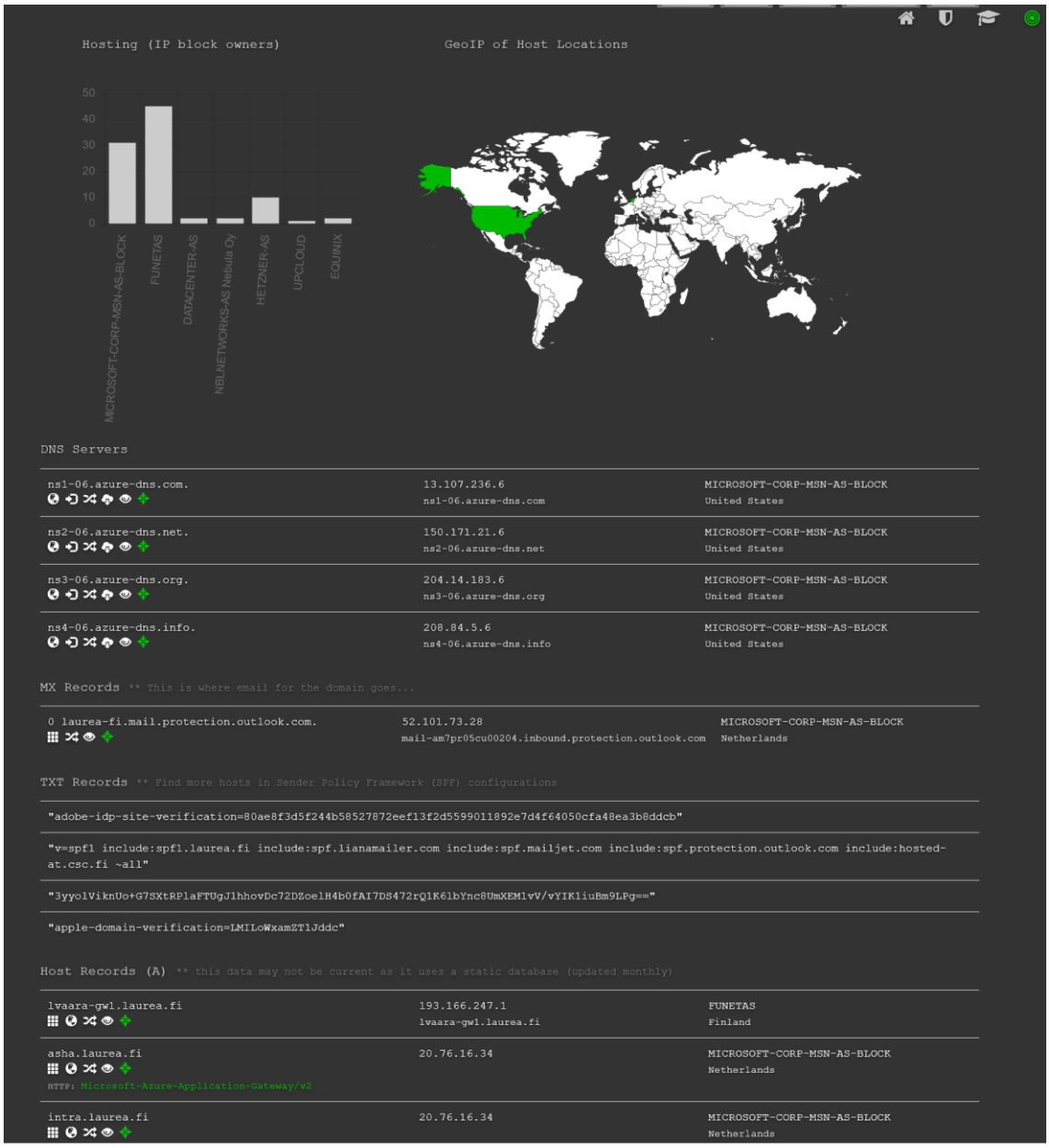
PhishTank on ilmainen verkkosivu, missä kuka vaan voi jakaa, vahvistaa ja seurata kalastelutietoja. Saatavilla on myös ohjelmointirajapinta kehittäjille. Työkalua voi hyödyntää esimerkiksi kalastelusivujen tunnistamiseen. (Authentic8 2023)

5.6 Verkkotunnukset

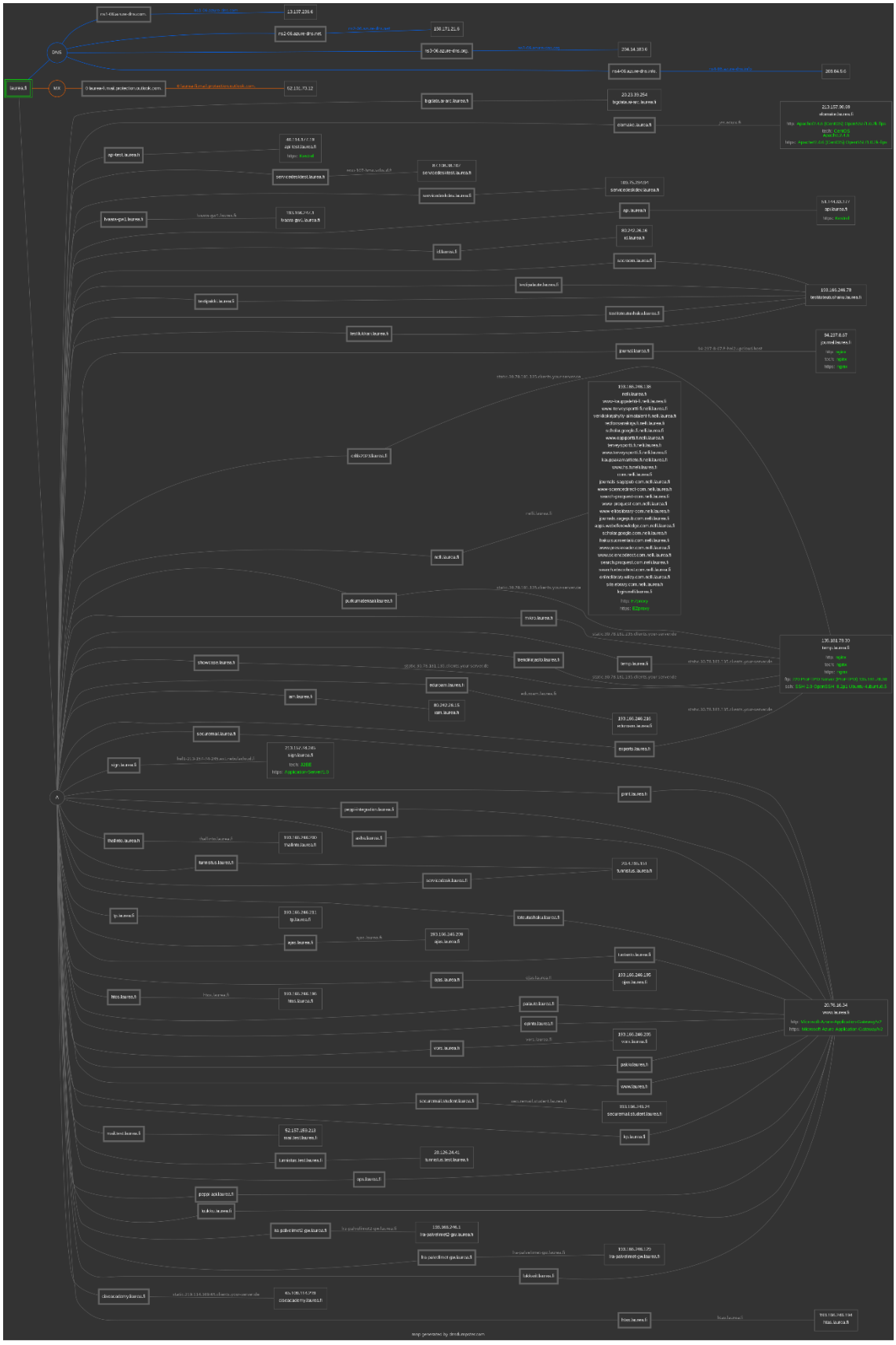
DNSdumpster

DNSdumpster on ilmainen työkalu verkkotunnusten tutkintaan. Se etsii syötettyyn verkkotunnukseen liittyvät aliverkkotunnukset. DNSdumpster ei ole tyypillisen DNS-skannerin tapaan aktiivisessa vuorovaikutuksessa kyseisen verkkotunnuksen kanssa, vaan hakee avoimista tietolähteistä aliverkkotunnuksia ja verkkoisäntiä. (DNSdumpster 2023)

Kuvassa 15 näkyy osa tuloksista haulla ”laurea.fi”. Tuloksissa näkyy verkkoisäntien sijainnit, DNS-palvelimet, MX-tietueet ja aliverkkotunnukset. Lisäksi kuvassa 16 DNSdumpster on hahmotellut verkkokartan kyseisestä verkkotunnuksesta.



Kuva 15: DNSdumpster tuloksia



Kuva 16: DNSdumpsterin verkkokartta

SecurityTrails

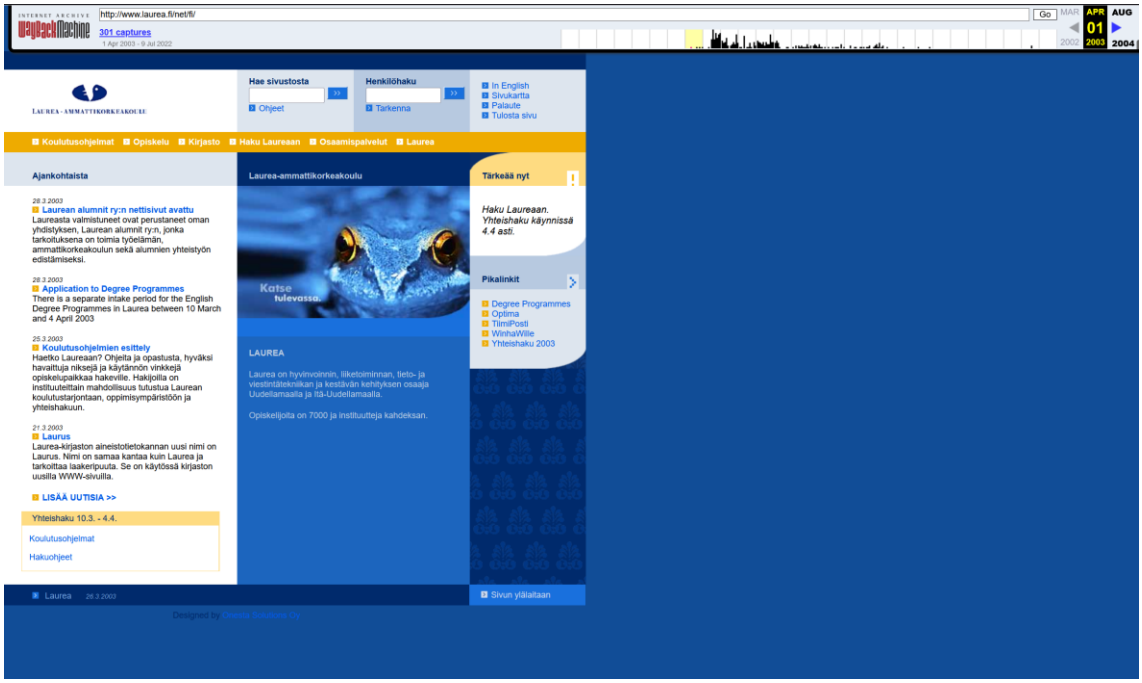
SecurityTrails on DNSdumpsterin tapaan työkalu verkkotunnusten tutkimiseen. SecurityTrails tarjoaa ilmaisversion lisäksi maksullista lisenssiä historiatiedon etsimiseen ja ohjelmointirajapinnan käyttöönottoon.

5.7 Historia, tietovuodot ja pimeä verkko

Wayback Machine

Historiallisen tiedon etsiminen voi osoittautua avointen tietolähteiden tiedustelussa joskus erittäin hyödylliseksi. Wayback Machine tarjoaa näkymän menneisyyteen. Se arkistoi verkkosivuja säännöllisesti ja sitä voi hyödyntää esimerkiksi tutkimalla, miltä jokin verkkosivu on näyttänyt tiettynä ajankohtana. (Wayback Machine 2023)

Kuvassa 17 on näkymä laurea.fi -verkkosivusta vuodelta 2003.



Kuva 17: Laurea.fi -verkkosivu vuonna 2003

Intelligence X

Intelligence X arkistoi Wayback Machinen tapaan verkkosivuja etsittäväksi, mutta sen lisäksi se arkistoi mm. verkkohyökkäyksissä vuotaneita tietokantoja, ja voi osoittautua arvokkaaksi työkaluksi tutkijoille, journalisteille ja tietoturva-ammattilaisille. Työkalua voi kokeilla ilmaiseksi, mutta kaikkien ominaisuuksien hyödyntämiseen tarvitaan maksullinen lisenssi. Myös ohjelmointirajapinta on saatavilla maksaville asiakkaille. (Intelligence X 2023)

Kuvassa 18 näkyy hakutuloksia satunnaiselle Bitcoin-osoitteelle. Huomionarvoista on, että Intelligence X arkistoi tietoja myös Tor-verkosta eli pimeästä verkosta.

Kuva 18: Intelligence X hakutuloksia

Ahmia

Ahmia on suomalainen avoimen lähdekoodin Tor-verkon, eli pimeän verkon hakukone. Haun voi tehdä tavallisellakin selaimella, mutta tulosten avaamiseen tarvitaan Tor-selain. Pimeässä verkossa on paljon tietoa mm. tietovuodoista ja laittomista verkkokaupoista, joten sen indeksointi voi osoittautua tehokkaaksi avointen tietolähteiden tiedustelukeinoksi monissakin käyttötilanteissa. (Kumar 2022; Nurmi 2023)

Dark.fail

Dark.fail pitää listaa suurimmista pimeän verkon sivustoista, kertoo niiden tilasta ja varmistaa listaamiensa linkkien oikeellisuuden PGP-avaimella. Se toimii hyvänä lähtökohtana lähteä etsimään tietoa tai palveluita pimeästä verkosta. (Authentic8 2023)

Torch

Torch on Ahmian tapaan toinen pimeän verkon hakukone, mutta sillä eroavaisuudella, että Torch ei sensuroi hakutuloksia. Tutkijat voivat löytää monenlaista tietoa ja liittyä keskustelupalstoille saadakseen lisätietoja nykyisistä haittaohjelmista, myytävistä varastetuista tiedoista tai ryhmistä, jotka saattavat suunnitella kyberhyökkäystä. (Authentic8 2023)

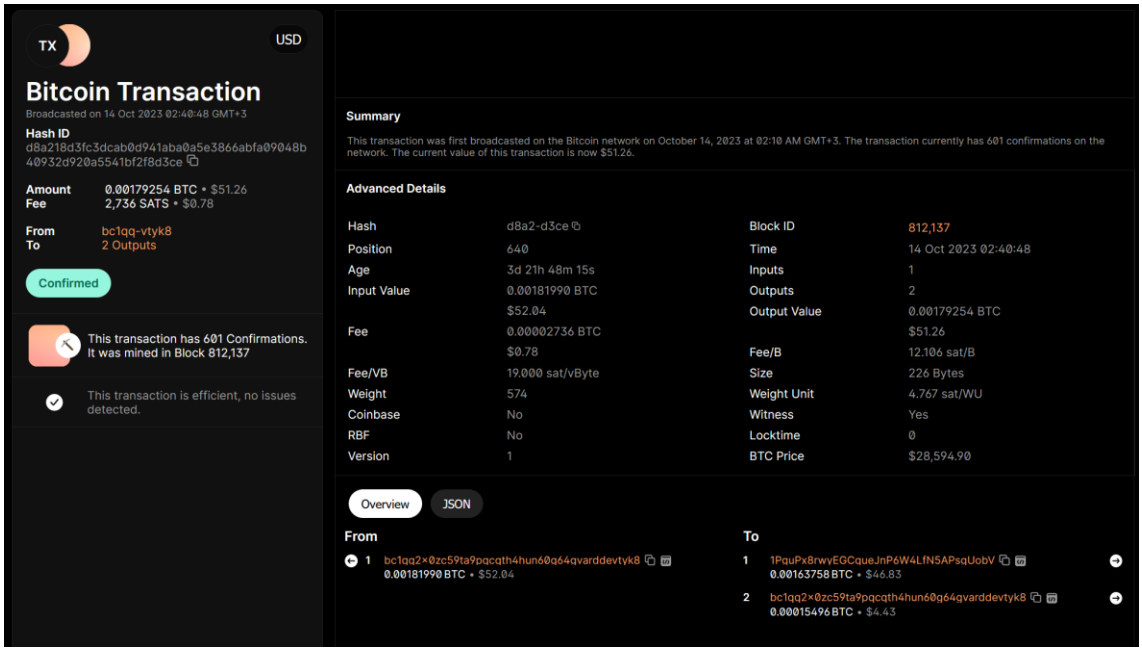
5.8 Kryptovaluutat

Kryptovaluuttojen jäljittäminen voi olla erityisesti rikostutkinnassa merkittävä keino päästä rikollisten jäljille. Rikolliset hyödyntävät kryptovaluuttoja niiden pseudonimettömyyden ja hajautetun luonteen vuoksi. Käyttäjänimien ja tilinumeroiden sijaan käyttäjät tunnistetaan julkisten avainten tiivistearvojen avulla. Kryptovaluutat tarjoavat myös läpinäkyvyyttä, koska kaikki tapahtumat kirjataan julkisesti saatavilla olevaan lohkoketjuun. Kryptovaluuttaosoitteet johdetaan yksityisistä avaimista ja ovat julkisesti saatavilla. Lisäksi jokaisella transaktiolla on yksilöllinen tunniste. (Adams 2022)

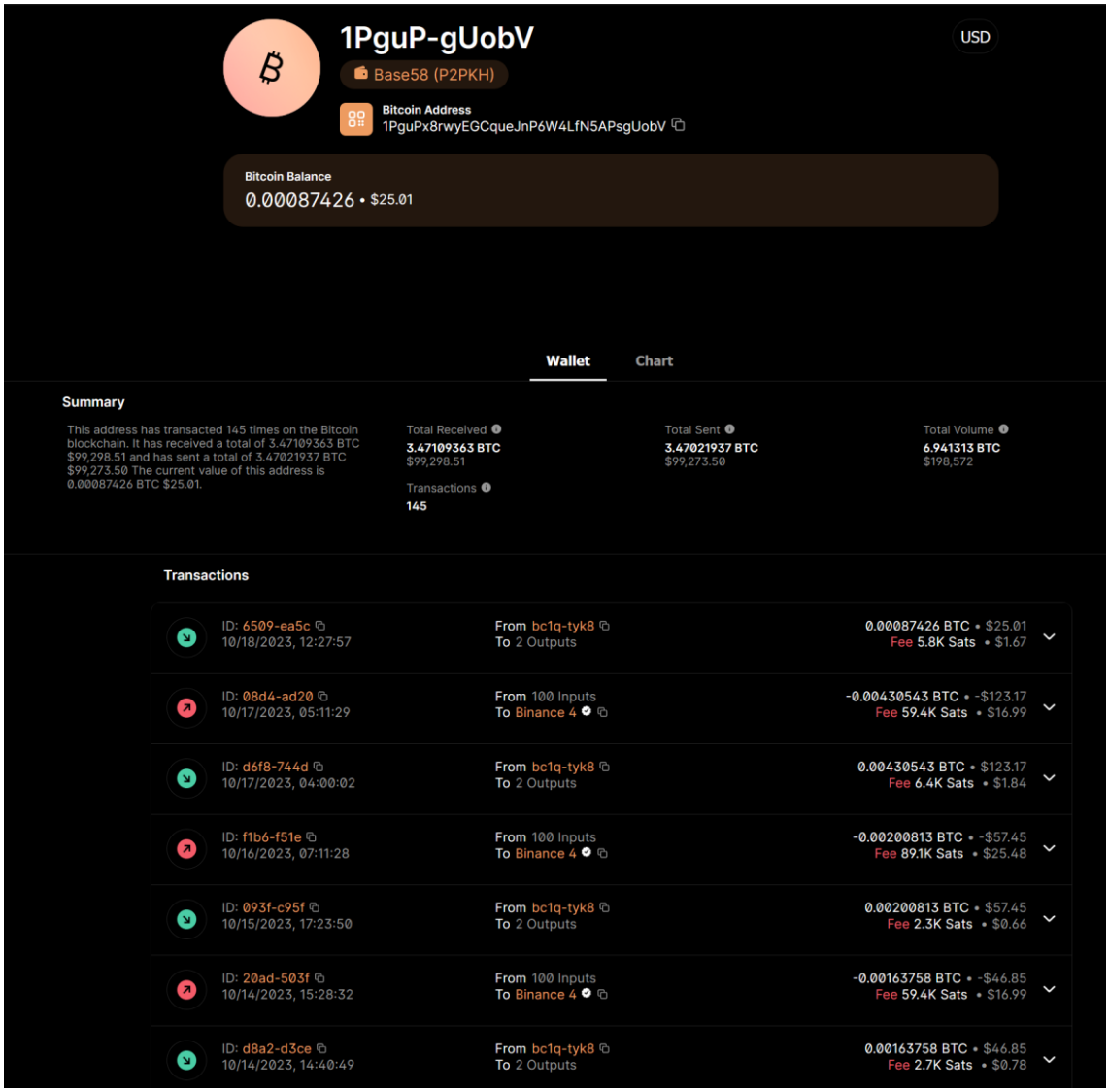
Siitä huolimatta, että esimerkiksi Bitcoin-transaktiot ovat vaikeammin jäljitettävissä kuin tavalliset tilisiirrot, ne eivät kuitenkaan takaa täydellistä suojaa jäljittämistä ja lopulta käyttäjän tunnistamista vastaan. Erilaisten johtolankojen perusteella kryptovaluutan julkinen yksilöllisen lompakon osoite voidaan yhdistää tiettyyn henkilöön. Luovuus ja koneoppimisalgoritmit voivat tehdä jäljittämisestä helppoakin. Lisäksi tulee ottaa huomioon, että Tunne asiakkaasi -määräykset vaativat kryptovaluuttojen välityspalvelut tunnistamaan käyttäjänsä. Vaikka tieto ei ole julkista, voivat lainvalvontaviranomaiset pyytää henkilötiedot etsintäluvan kanssa. On olemassa myös kryptovaluuttoja, jotka väittävät tarjoavansa täyden anonymiteetin, mutta niiden käyttö on toistaiseksi hyvin vähäistä, eikä niiden todellisesta anonymiteetistä ole vielä kattavaa näyttöä. (O'Sullivan 2023)

Blockchain Explorer

Blockchain.com tarjoaa web-käyttöliittymän Bitcoin-lompakoiden osoitteiden ja transaktioiden hakemiseen. Blockchain Explorer näyttää kaikki transaktiot, lompakkoon liittyvät rahansiirrot, transaktioiden ajankohdat, siirrettävän rahan määrän, transaktioon liittyvien lompakoiden osoitteet ja lompakoiden tämänhetkisen saldon. Jokaisen transaktion ketjua voi seurata, kuten mitä rahalle tapahtui seuraavaksi tai mistä raha tuli alkujaan. Kuvassa 19 näkyy tietoja satunnaisesta Bitcoin-transaktiosta Bitcoin Explorerissa. Kuvassa 20 näkyy satunnaisen Bitcoin-lompakon tietoja, kuten sen tämänhetkinen saldo, sekä saapuneet ja lähteneet siirrot. (Adams 2022; Blockchain Explorer 2023)



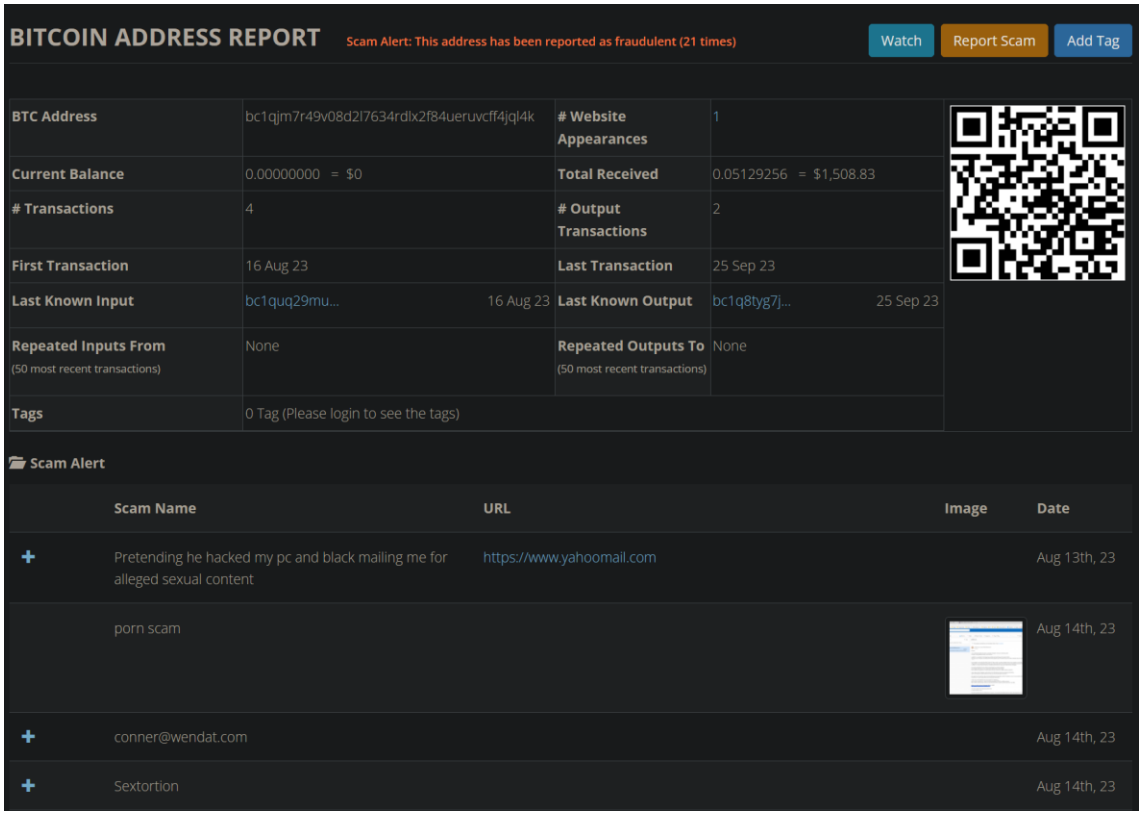
Kuva 19: Satunnainen transaktio Blockchain Explorerissa



Kuva 20: Satunnaisen Bitcoin-lompakon tietoja Blockchain Explorerissa

Bitcoin Who's Who

Bitcoin Who's Who on hakukone, jota hyödyntämällä voidaan yhdistää Bitcoin-osoitteita niiden omistajiin tai heidän yhteistyökumppaneihinsa. Se voi tarjota näkemyksiä osoitteiden omistajista ja siitä, onko osoitteita linkitetty tunnettuihin huijauksiin. Sivustolla on tietokanta osoitteista, joiden on ilmoitettu osallistuneen huijaukseen, sekä tietoja esimerkiksi osoitteista, jotka esiintyvät sosiaalisessa mediassa. Sivuston tavoitteena on auttaa vahvistamaan Bitcoin-osoitteen omistaja ja välttämään Bitcoin-huijaukset tai petokset. Bitcoin Who's Who tarjoaa ohjelmointirajapinnan tietojen hakemiseen rekisteröityneille käyttäjille, ja se on erinomainen lisä vaarantumisindikaattoreiden työkaluarsenaaliin avointen tietolähteiden tiedustelussa. Kuvassa 21 näkyy huijauksissa käytetyn Bitcoin-osoitteen tietoja. (BitcoinWhosWho 2023)



Kuva 21: Bitcoin Who's Who raportti huijauksissa käytetyssä Bitcoin-osoitteesta

5.9 Tiedostot

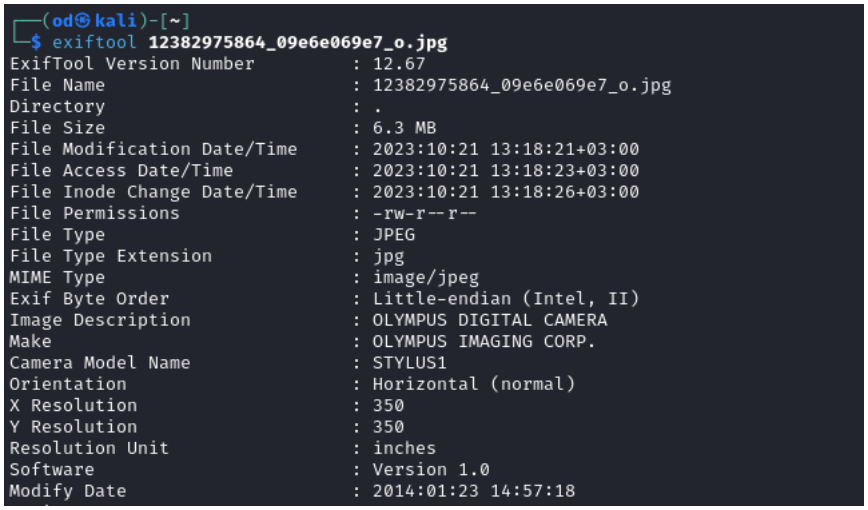
Metagoofil ja Exiftool

Avoimen lähdekoodin Metagoofil tehtiin alkuperäisesti metatietojen hakemiseen julkisista PDF- ja Microsoft Office -tiedostoista. Se teki Google-haun etsien määriteltyjä tiedostoja, latasi ne paikalliselle levylle ja analysoi tiedostojen metadataa käyttäen mm. Hachoir ja PdfMiner -kirjastoja. Työkalu palautti metatietoja tiedostosta, kuten tiedostoihin liitettyjä käyttäjätunnuksia, nimiä, palvelinten nimiä ja tiedostorakenteita. Metatietojen analysointi on kuitenkin myöhemmin otettu pois ja metatietojen haku ladatuista tiedostoista tapahtuu erillisillä työkaluilla, kuten Exiftool -työkalulla. Näiden tietojen saaminen voi olla potentiaaliselle uhkatoimijalle erittäin hyödyllistä esimerkiksi kalasteluhyökkäyksessä tai salasanojen arvuuttelussa. Avointen tietolähteiden tiedustelua tekevä organisaatio voi suojata kyseiset tiedot ennen kuin uhkatoimijat pääsevät niitä hyödyntämään. (Metagoofil 2023)

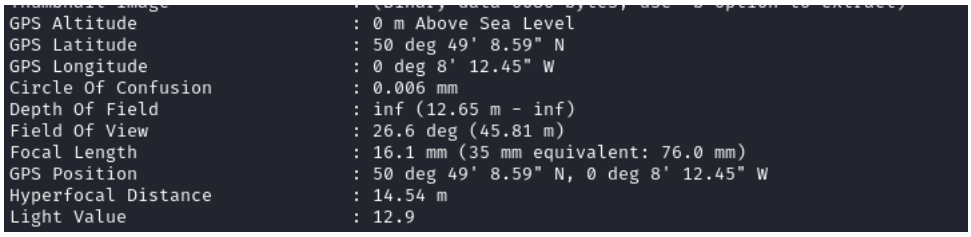
Kuvassa 22 haetaan .pdf, .doc, .xls ja .ppt -tiedostoja annetusta verkkotunnuksesta. Tuloksia tuli noin 30. Seuraavaksi tiedostoja voidaan analysoida esimerkiksi Exiftool-työkalulla. Exiftool-työkalua hyödyntäen satunnaisesta ladatusta tiedostosta löytyy tietoja mm. tekijän nimestä, teko- ja muokkauspäivämääristä sekä käytetystä työkalusta.

Kuva 22: Metagoofil-työkalun käyttöä

Exiftoolia voi käyttää myös kuvien analysointiin. Kuvissa 23 ja 24 on tehty analysointi satunnaiseen järjestelmäkameralla otettuun kuvaan. Tiedoista löytyy mm. kameran malli, sarjanumero, kuvausasetukset ja GPS-sijainti kuvan ottohetkellä.



Kuva 23: Exiftool -tuloksia, osa 1



Kuva 24: Exiftool -tuloksia, osa 2

VirusTotal

VirusTotalia voi käyttää haitallisten tiedostojen tai verkkotunnusten tunnistamiseen. Käyttöliittymään ladataan tiedosto, jonka jälkeen VirusTotal hyödyntää kymmeniä eri lähteitä tunnistaakseen tiedostoon liitettyjä raportoituja haitallisuuksia. Samalla VirusTotal näyttää tiedoston jäännösarvot (mm. MD5, SHA-256) ja metatiedot. VirusTotaliin on saatavilla myös ohjelmointirajapinta. (VirusTotal 2023)

6 Tiedon hyödyntäminen kyberuhkilta puolustautumiseen

Kuten työkalujen käyttöesimerkeissä huomattiin, avointen tietolähteiden tiedustelussa löytyy tietoa mm. kohdeorganisaation työntekijöistä, taloudesta, organisaatioon liitetyistä IP-osoitteista, sen käyttämistä teknologioista, haavoittuvuuksista, henkilöiden tai organisaatioiden suhteista muihin henkilöihin tai organisaatioihin ja niiden fyysisistä sijainneista. Näitä tietoja yhdistämällä uhkatoimija voi suunnitella tehokkaan ja toimivan hyökkäyksen. Tiedoista voi myös aiheutua esimerkiksi mainehaittaa, jos ne päätyvät vääriin käsiin.

Organisaation on hyvä tunnistaa, minkälaisia uhkia siihen kohdistuu, mitkä voisivat olla potentiaalisen uhkatoimijan motiivit tehdä kyberhyökkäys organisaatioon, ja minkälaisen tiedon he kokevat uhkaksi, mikäli se päätyisi julkiseksi. Kun uhkapinta-ala on tunnistettu, organisaatio voi avointen tietolähteiden tiedustelun avulla selvittää, mitä vaarallista tietoa heistä on saatavilla. Potentiaalisesti vaarallisen tiedon löytymisen jälkeen on oltava prosessi, jolla huolehditaan jatkotoimenpiteistä riskin pienentämiseksi. (Hiremath 2023; Rose 2020)

Avointen tietolähteiden tiedustelu ei ole kertaluontoinen projekti, vaan sitä tulee tehdä säännöllisesti. Teknologiasta on tässä apua ja erilaisia skannauksia kannattaakin automatisoida, jotta vaaralliset tiedot huomataan ennen uhkatoimijaa ja niihin ehditään reagoida ennen kuin on myöhäistä. (Hiremath 2023; Makrushin 2022)

Yhteistyö kyberturvallisuusyhteisön sisällä on tärkeää. OSINT-löydösten ja uhkatiedon jakaminen alan kollegoiden kanssa helpottaa yhteistä valmiutta muuttuvia uhkia vastaan. Tarkoitukseen soveltuvat alustat, kuten ISAC:it (Information Sharing and Analysis Center), ovat tässä olennaisia. (Hiremath 2023; ENISA 2023)

7 Säännökset, lainsäädäntö ja eettisyys

Monet nykyaikaiset organisaatiot kohtaavat dilemman avointen tietolähteiden tiedustelusta syntyvän datan käytön kanssa. Tiedon valtavan määrän käyttämättä jättäminen voi altistaa heidät riskeille, jotka olisi voitu helposti tunnistaa. Kuitenkin sen käyttäminen voi johtaa eettisiin huolenaiheisiin ja syytöksiin vastuuttomasta datan käytöstä. Kerätty tiedustelutieto sisältää hyvin usein henkilötietoja, jotka voidaan määritellä henkilökohtaiseksi tunnistetiedoksi (PII). Useimmat organisaatiot kuuluvat yleisen tietosuoja-asetuksen (GDPR) tai muiden tietosuojasäännösten piiriin. Teknologian käyttö, joka mahdollistaa tehokkaamman datan käsittelyn, voi myös herättää huolenaiheita massiivisen datan keräämisen eettisyydestä. Tiedustelutieto saattaa paljastaa myös rikollisia aikomuksia, jolloin tietojen paljastamiselle voi olla erityisiä lakivaatimuksia. Esimerkiksi Isossa Britanniassa voi rikosepäillyn henkilön tietojen julkaisemisesta saada sakkoja ja vankeustuomion. (Bisson 2021; Brown 2021)

Eettinen datan keruu on kuitenkin mahdollista OSINT-tutkimuksissa, erityisesti jos oikeat periaatteet on otettu käyttöön. Luonteensa vuoksi OSINT käyttää avointa, julkisesti saatavilla olevaa tietoa. Lainsäädäntö (GDPR) kuitenkin osoittaa, että julkinen tieto ei ole automaattisesti eettisesti hyväksyttävää, ellei sen keräämiseen ja säilyttämiseen sovelleta tarkkoja sääntöjä. OSINT-tutkijat kohtaavat osittain samoja haasteita kuin journalistit, kuten eettinen kohtelu lähteitä kohtaan, yksityisyys ja puolueellisuus. (Brown 2021)

Avointen tietolähteiden tiedustelutieto perustuu julkisesti saatavilla olevaan tietoon, mutta joskus kyseisen tiedon käyttämisellä voi olla vaikutusta ihmisten elämään. Tiedustelutiedon keräämisessä tulee ottaa huomioon tutkinnan kohde ja tiedon tarpeen laajuus, mutta myös tiedon eettinen vaikutus. Tiedon kerääminen tulee rajata niin, että tutkinnan tavoitteet voidaan saavuttaa vahingoittamatta kenenkään oikeuksia ja yksityisyydensuojaa. Tiedon keräämisen automatisointi teknologiaa hyödyntämällä voi usein johtaa epäeettiseen tai jopa laittomaan tiedonkeruuseen. Keskeinen osa eettistä tiedustelua on varmistaa, että tiedonkeruu on ihmisten hallinnassa ja kaikki tiedusteluun osallistuvat ymmärtävät tiedustelun eettiset ja oikeudelliset rajoitukset välttääkseen tietosuojaongelmia ja muita eettisiä huolenaiheita. (Imperva 2023)

Tiedonkeruun kohdentaminen

Suurien datamäärien kerääminen ei ole suotavaa niin sääntelyn kuin yleisönkin kannalta. Määrätiedoton tiedonkeruu voi altistaa organisaatiot mainehaitalle ja sakkomaksuille tietosuojasäännösten noudattamatta jättämisestä. Näiden riskien välttämiseksi on varmistettava, että tiedonkeruu on mahdollisimman kohdennettua. (Brown 2021)

Selkeiden parametrien asettaminen kunkin tutkimuksen alussa on yksi tapa toteuttaa kohdentamista. Tiedustelusykli on viitekehys tiedon keräämiseen ja tulkintaan. Tiedustelusyklin ensimmäistä vaihetta pidetään yleensä "suunnitteluna ja ohjauksena”, ja siinä kehotetaan määrittämään, mitä kysymyksiä tiedustelussa yritetään selvittää, mitä tietoa niiden selvittämiseen saatetaan tarvita ja mistä tämä tieto voisi olla peräisin. Viitekehystä käyttämällä voidaan aloittaa tutkimus selkeällä tavoitteella ja prosessilla. Tämä mahdollistaa tiedonkeruun kohdentamisen tarkasti ja välttää tarpeetonta tiedonkeruuta. (Brown 2021)

Teknologia voi myös olla suurena apuna eettisen datan keruun varmistamisessa. OSINT-työkalut ohjaavat usein käyttäjää määrittämään tarkemmat hakuparametrit kuin manuaalisesti olisi mahdollista. (Brown 2021)

Päätöksenteko ihmisten käsissä

Viimeaikaisissa ehdotuksissa Euroopan tekoälyasetukseen korostuvat huoli tekoälyn tekemästä automaattisesta päätöksenteosta, eli päätöksenteosta ilman ihmisen osallistumista (Tekoälyä

koskeva sääntelykehysehdotus 2023). OSINT-tutkimuksissa voi tulla ihmisellekin erittäin vaikeita päätöksiä vastaan. Esimerkiksi pimeästä verkosta laittomasti hankitun tietokannan ostaminen voi edistää omaa, muilta osin eettistä tutkimusta, jolla voisi olla hyviä vaikutuksia yhteiskunnalle, mutta ostamalla kyseistä tietoa rahoittaa rikollista toimintaa (Hanham 2022, 56). Useimmissa tapauksissa ihmisten tekemä, harkitusti arvioitu päätöksenteko, on eettisempää kuin tekoälyratkaisujen tekemä automaattisen päätöksenteko (Brown 2021).

Samaan aikaan teknologia on olennainen väline työnkulun nopeuttamiseen. Yksi hyvä tapa on automatisoida tiedon kerääminen useista lähteistä, mutta jättää keskeiset päätökset ihmisen tehtäväksi. Teknologiaa käytetään auttamaan saamaan tarvittu tieto nopeammin, mutta ihmistaidot ja tieto varmistavat, että jokainen päätös on mahdollisimman eettinen. (Brown 2021)

Toiminnan dokumentointi

Toiminnan dokumentointi kunkin tutkimusvaiheen aikana mahdollistaa eettisyyden varmistamisen tiedonkeruussa todistamalla, että tuloksiin perustuvat päätökset tehtiin eettisesti tallentamalla kaikki prosessit ja tunnistetut todisteet. Dokumentoinnilla voi myös todistaa tehneensä perusteellisen tutkimuksen tarvittaessa esimerkiksi lainvalvontaviranomaisen pyynnöstä. Esimerkiksi kriittisen infrastruktuurin organisaatioiden voi olla tarpeen osoittaa sääntelyelimille, että he ovat ottaneet proaktiivisen lähestymistavan riskien tutkimiseen. (Brown 2021)

Kaikkien tutkimusvaiheiden dokumentointi voi kuitenkin olla erittäin aikaa vievää. Teknologia auttaa myös tässä: automatisoimalla tutkimusvaiheiden tallennuksen, ihminen voi keskittyä asioihin, missä hän tuottaa enemmän arvoa. (Brown 2021)

7.1 Uhkatiedon jakamista koskeva lainsäädäntö

Uhkatiedon luonteen vuoksi sen käsittelyyn ja jakamiseen liittyy useita vaatimuksia ja sääntöjä. Tärkeimpinä huomioon on otettava yleinen tietosuoja-asetus (GDPR), NIS2-direktiivi ja valmisteilla oleva tekoälyasetus. Näiden lisäksi alasta riippuen voi astua voimaan muita alakohtaisia säännöksiä tai lakeja (esim. HIPAA terveydenhuollossa).

7.1.1 Yleinen tietosuoja-asetus

Yleinen tietosuoja-asetus (GDPR) pyrkii vähentämään oikeudellista epävarmuutta ja rajoittamaan tulkintoja asettamalla selkeät säännöt ja ehdot henkilötietojen käsittelylle ja jakamiselle sekä luonnollisten henkilöiden suojelulle henkilötietojen käsittelyn yhteydessä (yleinen tietosuoja-asetus 2016 Art. 1). Organisaatioiden on varmistettava, että he käsittelevät vain välttämättömän määrän henkilötietoja tarkoituksensa saavuttamiseksi (yleinen tietosuoja-

asetus 2016 Art. 5). GDPR määrittelee tarkat roolit ja velvollisuudet tietojen käsittelijöille ja tietojen hallinnoijille (yleinen tietosuoja-asetus 2016 Art. 4).

Yleisen tietosuoja-asetuksen mukaan henkilötiedoilla tarkoitetaan kaikkia tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyviä tietoja, kuten nimi, henkilötunnus, sijaintitieto, verkkotunnistetieto tai yksi tai useampi hänelle tunnusomainen fyysinen, fysiologinen, geneettinen, psyykkinen, taloudellinen, kulttuurillinen tai sosiaalinen tekijä (yleinen tietosuoja-asetus 2016 Art 4). On tärkeää huomata, että IP-osoitteita voidaan pitää henkilötietoina, koska ne voivat mahdollistaa henkilön tunnistamisen (MISP 2018).

Tietoturvatiimeille suunnatussa MISP-käytössä jakaminen perustuu ensisijaisesti uhkien ja haitallisten toimijoiden tietojen vaihtoon eikä henkilötietojen jakamiseen. Nämä tiedot kuitenkin yleensä sisältävät henkilötietoa, jonka takia tiedon jakamista tekevien tahojen olisi oltava mahdollisimman hyvin tietoisia, mitä tarkoitusta varten tietoa käsitellään ja mikä on asianomaisten osapuolten rooli tietojen käsittelyssä. (MISP 2018)

GDPR:n mukaisesti tietoturvatiimit voivat käsitellä ja jakaa tietoa, kun se on linjassa tarkoituksen ja tarpeen periaatteiden kanssa, ja kun se on välttämätöntä verkon ja tietojen tuvallisuuden varmistamiseksi. (Yleinen tietosuoja-asetus 2016 Art. 6)

”On asianomaisen rekisterinpitäjän oikeutetun edun mukaista rajoittaa henkilötietojen käsittely siihen, mikä on ehdottoman välttämätöntä ja oikeasuhteista, jotta viranomaiset, tietoturvaloukkauksiin ja niiden ennaltaehkäisyyn keskittyvät CERT-ryhmät (Computer Emergency Response Teams), tietoturvaloukkauksiin reagoivat ja niitä tutkivat CSIRT-toimijat (Computer Security Incident Response Teams), sähköisten viestintäverkkojen ja -palvelujen tarjoajat sekä turvallisuusteknologian ja -palvelujen tarjoajat voivat varmistaa verkko- ja tietoturvallisuuden eli verkon tai tietojärjestelmän kyvyn suojautua tietyllä suojatasolla onnettomuuksilta tai laittomilta taikka ilkivaltaisilta toimilta, jotka vaarantavat tallennettujen tai siirrettävien henkilötietojen saatavuuden, aitouden, eheyden ja luottamuksellisuuden ja niihin liittyvien, verkoissa ja tietojärjestelmissä tarjottujen tai välitettävien palvelujen turvallisuuden. Tähän voisi kuulua esimerkiksi luvattoman sähköisiin viestintäverkkoihin pääsyn ja vahingollisen koodin jakamisen ehkäiseminen sekä palvelunestohyökkäysten ja tietokoneille ja sähköisille viestintäjärjestelmille koituvien vahinkojen estäminen.” (Yleinen tietosuoja-asetus 2016 selonteko 49)

Vaikka GDPR antaa mahdollisuuden henkilötietojen käsittelyyn turvallisuuden varmistamiseksi, tulee varmistua, että tietojen käsittely on laillista ja perustuu yleiseen etuun (yleinen tietosuoja-asetus 2016 Art. 6). Henkilötietojen käsittelyn on täytettävä GDPR:n kuusi periaatetta (yleinen tietosuoja-asetus 2016 Art. 5). Nämä ovat:

- Lainmukaisuus, kohtuullisuus ja läpinäkyvyys

- Käyttötarkoitussidonnaisuus
- Tietojen minimointi
- Täsmällisyys
- Säilytyksen rajoittaminen
- Eheys ja luottamuksellisuus

Kun tietoja jaetaan MISP:n kautta, useimmissa tapauksissa henkilötietoja ei olla saatu rekisteröidyltä itseltään, vaan tiedot tulevat uhka-analyysien tuloksina (MISP 2018). Tällaisessa tapauksessa GDPR:n artikla 14 laukaisee läpinäkyvyysperiaatteen soveltamisen, edellyttäen, että rekisterinpitäjän identiteetti ja yhteystiedot ilmoitetaan rekisteröidylle (MISP 2018; Yleinen tietosuoja-asetus 2016). Tällaisen tiedon antaminen uhkatoimijoille voi kuitenkin vaarantaa tutkinnan eikä edistää yleistä etua, jolloin artiklan 14 kohdan 5 poikkeustapauksia voidaan soveltaa: ”kyseisten tietojen toimittaminen osoittautuu mahdottomaksi tai vaatisi kohtuutonta vaivaa, erityisesti kun käsittely tapahtuu yleisen edun mukaisia arkistointitarkoituksia tai tieteellisiä ja historiallisia tutkimustarkoituksia taikka tilastollisia tarkoituksia varten siten, että noudatetaan 89 artiklan 1 kohdassa esitettyjä edellytyksiä ja suojatoimia, tai niiltä osin kuin tämän artiklan 1 kohdassa tarkoitettu velvollisuus todennäköisesti estää kyseisten yleisen edun mukaisten arkistointitarkoitusten tai tieteellisten ja historiallisten tutkimustarkoitusten taikka tilastollisten tarkoitusten saavuttamisen tai vaikeuttaa sitä suuresti; tällaisissa tapauksissa rekisterinpitäjän on toteutettava asianmukaiset toimenpiteet rekisteröidyn oikeuksien ja vapauksien sekä oikeutettujen etujen suojaamiseksi, mukaan lukien kyseisten tietojen saattaminen julkisesti saataville” (yleinen tietosuoja-asetus 2016 Art. 14 (5)(b)). Esimerkiksi CSIRT-toimijat voivat saattaa tietoja käsittelystään julkisesti saataville RFC 2350:n ja GDPR:n artiklan 14 (1) ja (2) mukaisesti (MISP 2018).

Useimmissa tapauksissa MISP:iin syötetyt tiedot liittyvät suoraan hyökkäykseen ja ne on jo valikoitu erikseen analysoitavaksi suuresta datamäärästä, jotta ne olisivat hyödyllisiä hyökkäyksen havaitsemiseen ja/tai lieventämiseen. Näissä tapauksissa MISP:n käyttö täyttää tietojen minimoinnin periaatteen ja tarkoituksen rajoittamisen periaatteen. MISP sisältää myös ominaisuuksia, jotka arvioivat vaarantumisindikaattoreiden käyttökelpoisuutta uhkien havaitsemiseen ja/tai lieventämiseen. (MISP 2018)

Säilytysaika voi vaihdella suuresti riippuen luottamusympäristön käyttötarkoituksesta (MISP 2018). GDPR määrää, että henkilötietoja on ”säilytettävä muodossa, josta rekisteröity on tunnistettavissa ainoastaan niin kauan kuin on tarpeen tietojenkäsittelyn tarkoitusten toteuttamista varten; henkilötietoja voidaan säilyttää pidempiä aikoja, jos henkilötietoja käsitellään ainoastaan yleisen edun mukaisia arkistointitarkoituksia taikka tieteellisiä tai historiallisia tutkimustarkoituksia tai tilastollisia tarkoituksia varten 89 artiklan 1 kohdan mukaisesti edellyttäen, että tässä asetuksessa vaaditut asianmukaiset tekniset ja organisatoriset toimen-

piteet on pantu täytäntöön rekisteröidyn oikeuksien ja vapauksien turvaamiseksi” (yleinen tietosuoja-asetus 2016 Art. 5 (1)(e)).

MISP:n tapauksessa henkilötiedot ovat monesti pseudonymisoitu, jolloin rekisteröityä ei voida tunnistaa ilman lisätietoja. MISP:n käyttötapauksiin kuuluu myös yleensä uhkatoimijoiden ja hyökkäysten tutkimusta, ja tietoa voidaan joutua säilyttämään pitkään, jotta esimerkiksi hyökkäyskuvioita voitaisiin tunnistaa ja tuottaa tilastoja. Jos entiteetti ei enää tarvitse tiettyjä attribuutteja MISP:ssä, sille on mahdollista poistaa attribuutit (ja tapahtumat) paikallisesta MISP-instanssistaan. (MISP 2018)

Yhteenvetona voidaan siis todeta, että tietoturvaan liittyvät toimijat voivat jakaa tietoa henkilötietolainsäädännön puitteissa, kun niiden toiminta on suunniteltu palvelemaan yleistä etua ja tietoturvaa, ja kun henkilötietoja käsitellään vain tarpeellisessa laajuudessa ja määrin.

7.1.2 Tekoälyä koskeva sääntelykehysehdotus

EU:n tekoälyasetus on maailman ensimmäinen konkreettinen aloite tekoälyn sääntelemiseksi. Sen tavoitteena on tehdä Euroopasta luotettavan tekoälyn globaali keskus asettamalla yhtenäisiä sääntöjä, jotka ohjaavat tekoälyn kehittämistä, markkinointia ja käyttöä EU:ssa. Tekoälylailla pyritään varmistamaan, että EU:n alueella käytettävät tekoälyjärjestelmät ovat turvallisia ja kunnioittavat perusoikeuksia ja arvoja. Lisäksi sen tavoitteena on edistää investointeja ja innovaatioita tekoälyssä, parantaa hallintoa ja valvontaa sekä kannustaa yhtenäisen EU:n markkinan kehittymistä tekoälyn alalla. (Sathe & Ruloff 2023; Tekoälyä koskeva sääntelykehysehdotus 2023)

Asetus ottaisi käyttöön riskiperusteisen lähestymistavan tekoälyn sääntelyyn, keskittyen sovelluksiin, joilla on suurin potentiaali aiheuttaa ihmisille vahinkoa. Tämä kattaisi tilanteet, joissa tekoälyjärjestelmiä käytetään kriittisen infrastruktuurin, kuten veden tai energian, toiminnassa, oikeusjärjestelmässä sekä julkisten palveluiden ja hallituksen etuuksien saannin määrittämisessä. Teknologian valmistajien olisi suoritettava arviointeja vaatimustenmukaisuudesta ennen tekoälyjärjestelmien käyttöön saattamista, sekä otettava huomioon kuhunkin riskitasoon liittyvät vaatimukset. (Satariano 2023; Tekoälyä koskeva sääntelykehysehdotus 2023)

Uhkatiedon jakamisessa ja luottamusympäristöissä kannattaa jo rakentamisvaiheessa ottaa tulevan asetuksen vaatimukset huomioon, sillä on hyvin todennäköistä, että tekoälyä tullaan siinä hyödyntämään. Ensimmäinen vaihe on tunnistaa kaikki käytössä olevat tekoälymallit, sekä oma rooli niiden kehittäjänä tai hankkijana, ja listata tunnistetut tekoälymallit tietokantaan. (Sathe & Ruloff 2023; Tekoälyä koskeva sääntelykehysehdotus 2023)

Seuraava vaihe on arvioida kyseisten tekoälymallien riskiluokitus. EU:n tekoälyasetus erottaa neljä erilaista riskikategoriaa. Hyväksymättömän riskin mallit ovat kiellettyjä. Korkean riskin mallit ovat sallittuja, mutta niiden on täytettävä useita vaatimuksia ja niiden vaatimustenmukaisuus on arvioitava ennen mallin markkinoille saattamista. Näiden mallien on myös rekisteröidyttävä EU:n perustamaan tietokantaan. Korkean riskin tekoälymallien käyttö edellyttää asianmukaista riskienhallintajärjestelmää, tapahtumien seurantakykyä ja ihmisen valvontaa tai omistajuutta. (Tekoälyä koskeva sääntelykehysehdotus 2023)

Vähäisen riskin mallien osalta (kuten chatbotit) vaaditaan läpinäkyvyyttä, eli käyttäjän on tiedettävä, että hän on vuorovaikutuksessa tekoälyn kanssa. Ehdotuksessa sallitaan vähäriskisen tekoälyn, kuten tekoälyä hyödyntävien videopelien tai roskapostisuodattimien vapaa käyttö. (Tekoälyä koskeva sääntelykehysehdotus 2023)

Uhkatiedustelussa tulee ottaa huomioon, että esimerkiksi biometrisiä tunnistusjärjestelmiä pidetään suuririskisinä ja niihin sovelletaan tiukkoja vaatimuksia (Satariano 2023). Voi olla, että sosiaalisesta mediasta biometristen tietojen haravoiminen olisi kiellettyä (Sathe & Ruloff 2023). Myös ennakoivat työkalut, joilla pyritään ennustamaan, kuka tulee tekemään rikoksia, olisivat kiellettyjä (Sathe & Ruloff 2023). Poikkeuksia saattaa esiintyä esimerkiksi silloin, kun ne ovat tarpeen terrorismin uhan estämiseksi tai rikollisen tai epäillyn havaitsemiseksi, paikantamiseksi, tunnistamiseksi tai syytteeseen asettamiseksi (Satariano 2023). Tällainen käyttö edellyttää lainkäyttöelimen tai muun riippumattoman elimen lupaa ja asianmukaisia määräaikoja, maantieteellistä kattavuutta ja haun kohteena olevia tietokantoja (Tekoälyä koskeva sääntelykehysehdotus 2023).

Komissio kuitenkin ehdottaa toimenpiteitä, joilla edistetään luottamusta tekoälyyn, kuten tekoälyä hyödyntävät turvallisuusoperaatiokeskukset (Tekoälyä koskeva sääntelykehysehdotus 2023). Voisi siis odottaa, että tekoälyn hyödyntäminen yhteisen hyvän ja turvallisuuden parantamiseksi olisi jopa kannustettavaa.

7.1.3 NIS2-direktiivi

NIS2-direktiivi, joka on päivitys alkuperäiseen EU:n NIS (Network and Information Security) -direktiiviin, määrittelee vaatimukset kyberturvallisuuden yleisen tason parantamiseksi EU:ssa. Avointen lähteiden tiedustelua ja uhkatietoja jakavien organisaatioiden tulee ottaa huomioon NIS2:n asettamat vaatimukset. Uusia velvoitteita tulee noudattaa viimeistään lokakuusta 2024 alkaen. (NIS2-direktiivi 2023)

NIS2 korostaa asianmukaisten ja suhteellisten teknisten ja organisatoristen toimien toteuttamisen tärkeyttä verkko- ja tietojärjestelmien turvallisuuteen kohdistuvien riskien hallitsemiseksi. Direktiivi edellyttää mm. johdon vastuunkantamista, merkittävien kyberturvallisuustapahtumien nopeaa raportointia (24 tunnin kuluessa tapahtuman havaitsemisesta), asianmu-

kaista turvallisuuspolitiikkaa, häiriönhallintaa, jatkuvuudenhallintaa, riskienhallintaa, toimitusketjun turvallisuuden varmistamista, jäsenvaltioiden toteuttamia kansallisia kyberturvallisuusstrategioita, säännöllistä turvatoimien arviointia sekä tietoisuuden ja koulutuksen lisäämistä. (Aleksiev, Oberschelp de Meneses & Young 2023; NIS2-direktiivi 2023)

NIS2 kannustaa tietojen jakamiseen ja yhteistyöhön toimijoiden välillä samassa sektorissa, eri sektoreiden välillä sekä relevanttien kansallisten ja EU:n viranomaisten kanssa (Aleksiev ym. 2023; NIS2-direktiivi 2023). Tiedon jakamista varten ja säännösten noudattamisen helpottamiseksi uhkatiedon jakamiseen tarkoitetut alustat, kuten MISP tai ISAC:it (Information Sharing and Analysis Centre) tulevat hyödylliseksi.

8 Uhkatiedon jakaminen luottamusympäristössä

DYNAMO-projektissa tuotettava alusta tarjoaa tukea kaikkiin kybersietokyvyn vaiheisiin. Yksi osa alustaa on kyvykkyys uhkatiedon tallentamiseen ja jakamiseen DYNAMO-sidosryhmien kesken, kiinnittäen erityistä huomiota jaetun tiedon suojaamiseen. Tätä tarkoitusta varten projektissa kehitetään luottamusympäristöä DYNAMO-alustan käyttäjille. Kehitetty moduuli perustuu avoimen lähdekoodin ratkaisuihin, kuten MISP, hyödyntäen sen edistyneitä tallennus- ja jakamistoimintoja, mutta mahdollistaen myös integraation muiden avoimen lähdekoodin työkalujen, kuten Cortexin ja TheHiven kanssa. (DYNAMO 2023)

Uhkatoimijat erikoistuvat kohdistamaan hyökkäyksensä tiettyihin sektoreihin, kuten hallituksiin tai pankkitoimintaan. Kun he löytävät tekniikan, joka toimii yhtä organisaatiota vastaan, yleinen seuraava vaihe on yrittää samoja tekniikoita samankaltaisia organisaatioita vastaan. Siksi luottamusympäristössä jaettu uhkatieto hyödyttää kaikkia osapuolia. Uhkatiedon jakaminen vähentää todennäköisyyttä sille, että samankaltaiset organisaatiot joutuvat samojen uhkien uhreiksi. (Postolovski 2023)

8.1 MISP, TheHive ja Cortex

MISP Open Source Threat Intelligence and Sharing Platform on uhkatiedon jakamiseen tarkoitettu, avoimen lähdekoodin alusta. Sen tavoitteena on auttaa ennaltaehkäisemään ja tunnistamaan kohdennettuja kyberhyökkäyksiä esimerkiksi vaarantumisindikaattoreiden avulla. MISP antaa uhkatiedoille yhtenäisen rakenteen ja yhdistää automaattisesti samankaltaista tietoa, tehden tiedon suuresta määrästä huolimatta sen tallentamisesta ja jakamisesta helpompaa organisaatioiden kesken, jotka todennäköisesti kohtaavat samankaltaisia uhkia. MISPissä on REST API -ohjelmointirajapinta toiminnallisuuksien ja ulkoisten tietolähteiden lisäämistä varten. (CIRCL 2023; Postolovski 2023)

TheHive on avoimen lähdekoodin Security Incident Response -alusta, eli tietoturvahäiriöiden hallintaan tarkoitettu alusta. Sen voi synkronoida MISP-instanssien kanssa tutkinnan tekemiseksi MISP-tapahtumista. Tutkimustulokset voi viedä ja jakaa MISP-tapahtumana luottamusympäristön kesken. (TheHive 2022)

Cortex on TheHive Projectin tekemä ohjelmisto vaarantumisindikaattoreiden analysointiin, joka on tarkoitettu nimenomaan MISPin ja TheHiven kanssa käytettäväksi. (Cortex 2023)

8.2 Luottamusympäristössä jaettavat tiedot

Uhkatiedon jakamisessa täytyy löytää tasapaino läpinäkyvyyden, yksityisyyden ja luottamuksellisuuden välillä. Organisaation on hyvä jakaa uhkatietoa, joka auttaa muita organisaatioita luottamusympäristössä ymmärtämään uhkat paremmin. Toisaalta tulisi välttää tunnistettavan tai sensitiivisen tiedon jakamista. Tällaista tietoa ovat esimerkiksi yritysten nimet, sisäiset IP-osoitteet, henkilöstön nimet, asiakastunnisteet tai liiketoimintaan liittyvät tiedot. Jaettava tieto sisältää siis tiedustelutietoa vain uhkista, eikä omasta liiketoiminnasta, infrastruktuurista, työntekijöistä tai asiakkaista. Jaettavien uhkatietojen tarkistaminen on tärkeää yksityisyyden ja luottamuksellisuuden ylläpitämiseksi. Riskien vähentämiseksi ja luottamuksellisuuden ja tarkkuuden varmistamiseksi tiedon jakamisessa voidaan käyttää hyväksyntäketjua ennen tiedon jakamista muiden organisaatioiden kanssa. (Postolovski 2023)

9 Yhteenveto, jatkotutkimusaiheet ja pohdinta

Avointen tietolähteiden tiedustelu (OSINT) on merkittävä väline organisaatioiden kyberturvallisuuden ylläpitämisessä. Sen avulla voidaan proaktiivisesti tunnistaa ja ennakoida mahdollisia uhkia sekä kehittää toimenpiteitä näiden uhkien pienentämiseksi. Erityisesti tämä koskee sellaisen tiedon löytämistä, joka on julkisesti saatavilla ja jota uhkatoimijat voisivat hyödyntää organisaatiota vastaan. Tämä voi sisältää esimerkiksi työntekijöihin liittyvää tietoa, jota voitaisiin käyttää hyväksi esimerkiksi tietojenkalastelussa tai salasanojen murtamisessa.

OSINT:n harjoittaminen vaatii asiantuntijuutta. Ammattilaisen on kyettävä tunnistamaan merkittävä data laajasta informaatiomäärästä ja arvioitava sen relevanssi organisaation turvallisuuden kannalta. Eettiset ja lainsäädännölliset näkökohdat ovat keskeisiä tiedustelutiedon keräämisessä. On välttämätöntä, että tiedonkeruu rajataan niin, ettei se loukkaa kenenkään yksityisyyttä tai oikeuksia. Teknologian käyttö tiedonkeruussa voi johtaa epäeettisiin tai laittomiin käytäntöihin, ellei ihmisten hallinta ja ymmärrys tiedustelun rajoituksista ole etusijalla.

Tiedusteluprosessiin on hyvä sisällyttää selkeä viitekehys, jossa tutkimuksen tavoitteet ja prosessit on määritelty. Teknologian rooli tiedonkeruussa on merkittävä nopeuttaessaan tie-

donhankintaa, mutta ihmisarvioinnin merkitys korostuu eettisissä päätöksissä. Toiminnan dokumentointi on keskeistä eettisyyden varmistamisessa. Automaattiset järjestelmät voivat tallentaa tutkimusvaiheita, jolloin ihmiset voivat keskittyä analyysiin ja arvonluontiin.

Uhkatiedon jakamisessa käytetään alustoja kuten MISP, jonka tavoitteena on ennaltaehkäistä ja tunnistaa kohdennettuja kyberhyökkäyksiä sekä jakaa uhkatietoa toimijoiden välillä. Uhkatiedon jakamisessa on tärkeää varmistaa, että jaettavat tiedot eivät loukkaa yksityisyyttä tai luottamuksellisuutta. Tietoturvaan liittyvät toimijat voivat jakaa tietoa henkilötietolainsäädännön puitteissa, kun toiminta palvelee yleistä etua ja henkilötietoja käsitellään vain tarpeellisessa laajuudessa. Lisäksi on otettava huomioon esimerkiksi NIS2-direktiivin ja tekoälyasetuksen asettamat vaatimukset.

OSINT tarjoaa organisaatioille tärkeän työkalun kyberturvallisuuden ylläpitämiseen, mutta sen tehokas ja eettinen käyttö edellyttää asiantuntemusta, selkeitä prosesseja ja teknologian sekä ihmistaidon tasapainoista yhdistämistä.

OSINT-toiminnan eettisyys ja tekoälyn hyödyntäminen ovat aiheita, joista riittäisi pohdittavaa erillisiin jatkotutkimuksiin. Nämä kaksi asiaa voisi hyvin myös yhdistää yhdeksi tutkimuskokonaisuudeksi.

Lähteet

Adams, Steve. 2022. Cryptocurrency and NFT OSINT Investigations Tips & Techniques. Viitattu 14.10.2023. https://www.skopenow.com/resource-center/cryptocurrency-and-nft-osint-investigations-tips-techniques

Aleksiev, Aleksander, Oberschelp de Meneses, Anna & Young, Mark. 2023. New EU Cyber Law "NIS2" Enters Into Force. Inside Privacy. Viitattu 2.11.2023. https://www.insideprivacy.com/cybersecurity-2/new-eu-cyber-law-nis2-enters-into-force/

Authentic8. 2023. 21 OSINT research tools for threat intelligence. Viitattu 11.9.2023. https://authentic8.com/resources/21-osint-research-tools-threat-intelligence

Bisson, David. 2021. 10 Open-Source Intelligence Tools (That Actually Work With Your Existing Security Software). Security Intelligence. Viitattu 11.9.2023. https://securityintelligence.com/articles/10-open-source-intelligence-tools-existing-security-software/

Brown, Charles. 2021. How to conduct ethical OSINT Investigations. Blackdot Solutions Videris. Viitattu 22.10.2023. https://blackdotsolutions.com/blog/ethics-in-data-collection/

Byron, Aubrey. 2023. OSINT need-to-knows: Intro to advanced search and Google dorking. authentic8. Viitattu 22.9.2023. https://authentic8.com/blog/osint-need-knows-intro-advanced-search-and-google-dorking

CIRCL. 2023. CIRCL » MISP - Open Source Threat Intelligence Platform. Viitattu 24.10.2023. https://www.circl.lu/services/misp-malware-information-sharing-platform/

Cree.py. 2023. Creepy. Viitattu 29.9.2023. https://www.geocreepy.com/

DYNAMO. 2023. Horizon Dynamo EU. Viitattu 21.9.2023. https://horizon-dynamo.eu/about/

ENISA. 2023. Information Sharing and Analysis Centers (ISACs). Topic. ENISA. Viitattu 23.10.2023. https://www.enisa.europa.eu/topics/national-cyber-security-strategies/information-sharing

Ferreira, Pedro. 2018. Techniques and tools for OSINT-based threat analysis. DiSIEM - Diversity-enhancements for SIEMs; .

Fruhlinger, Josh, Sharma, Ax & Breeden, John. 2023. 15 top open-source intelligence tools. CSO Online. Viitattu 11.9.2023. https://www.csoonline.com/article/567859/what-is-osint-top-open-source-intelligence-tools.html

Gaucheler, Mathieu. 2023. Advanced IOCs Collection with OSINT and Threat Intelligence Feeds. Maltego. Viitattu 11.9.2023. https://www.maltego.com/blog/advanced-iocs-collection-with-osint-and-threat-intelligence-feeds/

Hanham, Melissa. 2022. Setting Your Moral Compass: A Workbook for Applied Ethics in OSINT. Ethical Journalism Network. Viitattu 22.10.2023. https://ethicaljournalismnetwork.org/setting-your-moral-compass-a-workbook-for-applied-ethics-in-osint

Hiremath, Omkar. 2023. Protecting Your Organization With Open-source Intelligence (OSINT). Software Secured. Viitattu 23.10.2023. https://www.softwaresecured.com/protecting-your-organization-with-open-source-intelligence-osint/

Imperva. 2023. Open-Source Intelligence (OSINT). Viitattu 5.9.2023. https://www.imperva.com/learn/application-security/open-source-intelligence-osint/

Kaspersky. 2023. OSINT (Open-Source Intelligence). Viitattu 5.9.2023. https://encyclopedia.kaspersky.com/glossary/osint/

Kumar, Dinesh. 2022. 10 OSINT Tools Hackers Need to Know About. Medium. Viitattu 1.10.2023. https://0xtmux.medium.com/10-osint-tools-hackers-need-to-know-about-9cbb3519ea47

Mader, Jeffrey. 2022. Maltego Basics: Building a Network Diagram. Medium. Viitattu 28.9.2023. https://osintteam.blog/maltego-basics-building-a-network-diagram-5d7ee843e423

Makrushin, Denis. 2022. Attack Surface Monitoring using Open-Source Intelligence. Medium. Viitattu 11.9.2023. https://infosecwriteups.com/attack-surface-monitoring-using-open-source-intelligence-90415e863e93

Maltego. 2021. Useful Google Dorks for Open Source Intelligence Investigations. Viitattu 22.9.2023. https://www.maltego.com/blog/using-google-dorks-to-support-your-open-source-intelligence-investigations/

Maltego. 2022. Writing Custom Maltego Integrations. Viitattu 28.9.2023. https://www.maltego.com/blog/writing-custom-maltego-integrations/

Maor, Etay. 2022. How and Why to Apply OSINT to Protect the Enterprise. Dark Reading. Viitattu 23.10.2023. https://www.darkreading.com/attacks-breaches/how-and-why-to-apply-osint-to-protect-the-enterprise

Martorella, Christian. 2023. theHarvester. Python. https://github.com/laramies/theHarvester

Micallef, Steve. 2021. Attack Surface Management. You're (probably) doing it wrong. Medium. Viitattu 11.9.2023. https://medium.com/@micallst/attack-surface-management-youre-probably-doing-it-wrong-608719da1cab

Micallef, Steve. 2023. SpiderFoot. Python. https://github.com/smicallef/spiderfoot

MISP. 2018. Information sharing and cooperation enabled by GDPR. MISP. Viitattu 26.10.2023. https://www.misp-project.org/compliance/GDPR/

Nurmi, Juha. 2023. Ahmia — Search Tor Hidden Services. Viitattu 1.10.2023. https://ahmia.fi/

Office of the Director of National Intelligence. 2023. What is Intelligence? Viitattu 21.9.2023. https://www.dni.gov/index.php/what-we-do/what-is-intelligence

O'Sullivan, Fergus. 2023. Is Crypto Anonymous in 2023 & Is Bitcoin More Traceable Than Cash? Cloudwards. Viitattu 14.10.2023. https://www.cloudwards.net/is-crypto-anonymous/

Packham, Karen. 2022. RHEA to Contribute to Develop Cyber Threat Intelligence Solution for European DYNAMO Project. RHEA Group. Viitattu 11.9.2023. https://www.rheagroup.com/rhea-to-contribute-to-develop-cyber-threat-intelligence-solution-for-european-dynamo-project/

Postolovski, Tash. 2023. What is MISP? The Ultimate Introduction. Cosive. Viitattu 26.10.2023. https://www.cosive.com/blog/what-is-misp-the-ultimate-introduction

regex101. 2023. regex101: build, test, and debug regex. regex101. Viitattu 22.9.2023. https://regex101.com/

Rose, Zoë. 2020. OSINT – Using Threat Intelligence to Secure Your Organisation. Tripwire. Viitattu 11.9.2023. https://www.tripwire.com/state-of-security/osint-using-threat-intelligence-secure-organisation

SANS Institute. 2023. What is Open-Source Intelligence? Viitattu 28.8.2023. https://www.sans.org/blog/what-is-open-source-intelligence/

Satariano, Adam. 2023. Europeans Take a Major Step Toward Regulating A.I. The New York Times. 14.6.2023, , osa Technology. https://www.nytimes.com/2023/06/14/technology/europe-ai-regulation.html

Sathe, Madan & Ruloff, Karl. 2023. The EU AI Act: What it means for your business. EY. Viitattu 2.11.2023. https://www.ey.com/en_ch/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business

Searchcode. 2023. searchcode source code search engine. Viitattu 29.9.2023. https://searchcode.com/

Shodan. 2023. Shodan. Shodan. Viitattu 29.9.2023. https://www.shodan.io

Slinde, Johanna Sofie. 2023. Unveiling the Potential of Open-Source Intelligence (OSINT) for Enhanced Cybersecurity Posture

ThreatMiner. 2023. About ThreatMiner | ThreatMiner.org. Viitattu 29.9.2023. https://www.threatminer.org/about.php

Yadav, Ashok, Kumar, Atul & Singh, Vrijendra. 2023. Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security. Artificial Intelligence Review. https://doi.org/10.1007/s10462-023-10454-y. DOI: 10.1007/s10462-023-10454-y

Yleinen tietosuoja-asetus. 2016. Luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta annettu Euroopan parlamentin ja neuvoston asetus (EU) 2016/679 (yleinen tietosuoja-asetus). OJ L. vsk. 119; . http://data.europa.eu/eli/reg/2016/679/oj/fin

TheHive. 2022. Viitattu 26.10.2023. https://github.com/TheHive-Project/TheHive

BitcoinWhosWho. 2023. Viitattu 21.10.2023. https://www.bitcoinwhoswho.com/

Blockchain Explorer. 2023. Viitattu 18.10.2023. https://www.blockchain.com/explorer

Cortex. 2023. Scala. TheHive Project. Viitattu 26.10.2023. https://github.com/TheHive-Project/Cortex

DNSdumpster. 2023. Viitattu 30.9.2023. https://dnsdumpster.com/

Have I Been Pwned. 2023. Viitattu 29.9.2023. https://haveibeenpwned.com/

Intelligence X. 2023. Viitattu 30.9.2023. https://intelx.io/

Metagoofil. 2023. Python. Viitattu 21.10.2023. https://github.com/opsdisk/metagoofil

NIS2-direktiivi. 2023. Euroopan komissio. Viitattu 2.11.2023. https://digital-strategy.ec.europa.eu/fi/policies/nis2-directive

OSINT Framework. 2023. Viitattu 22.10.2023. https://osintframework.com/

Sherlock Project. 2023. Python. Sherlock. Viitattu 29.9.2023. https://github.com/sherlock-project/sherlock

Tekoälyä koskeva sääntelykehysehdotus. 2023. Euroopan komissio. Viitattu 2.11.2023. https://digital-strategy.ec.europa.eu/fi/policies/regulatory-framework-ai

TinEye. 2023. Viitattu 29.9.2023. https://tineye.com/

VirusTotal. 2023. Viitattu 21.10.2023. https://www.virustotal.com/gui/home/upload

Wappalyzer. 2023. Viitattu 30.9.2023. https://www.wappalyzer.com/

Wayback Machine. 2023. Viitattu 23.11.2023. https://archive.org/web/

Kuvat

Kuva 1: OSINT Framework 11
Kuva 2: Esimerkki hakuparametrien käytöstä 12
Kuva 3: Maltegolla luotu yhteyskartta Bitfinex-hyökkäyksestä (Mader 2022) 14
Kuva 4: theHarvester-työkalun ohjeet 15
Kuva 5: SpiderFoot käyttöliittymä 16
Kuva 6: SpiderFoot yhteenveto löydöksistä 17
Kuva 7: ThreatMiner hakutulokset 18
Kuva 8: Sherlock hakutuloksia 19
Kuva 9: WhatsMyName hakutuloksia 19
Kuva 10: Shodan hakutuloksia 21
Kuva 11: Google Kuvahaun hakutuloksia 22
Kuva 12: Grep.app hakutuloksia 23
Kuva 13: Have I Been Pwned hakutuloksia 24
Kuva 14: Wappalyzer näkymä 25
Kuva 15: DNSdumpster tuloksia 26
Kuva 16: DNSdumpsterin verkkokartta 27
Kuva 17: Laurea.fi -verkkosivu vuonna 2003 28
Kuva 18: Intelligence X hakutuloksia 29
Kuva 19: Satunnainen transaktio Blockchain Explorerissa 31
Kuva 20: Satunnaisen Bitcoin-lompakon tietoja Blockchain Explorerissa 32
Kuva 21: Bitcoin Who's Who raportti huijauksissa käytetyssä Bitcoin-osoitteesta 33
Kuva 22: Metagoofil-työkalun käyttöä 34
Kuva 23: Exiftool -tuloksia, osa 1 34
Kuva 24: Exiftool -tuloksia, osa 2 34

Taulukot

Taulukko 1: Esimerkkejä säännöllisistä lausekkeista 13