OSINT-tiedustelun automatisointi pimeässä verkossa ja CAM-materiaali tutkinnan haasteet

Iiro Lahti

2021 Laurea

Laurea-ammattikorkeakoulu

OSINT-tiedustelun automatisointi pimeässä verkossa ja CAM-materiaali tutkinnan haasteet

Iiro Lahti

Tietojenkäsittelyn koulutusohjelma
Opinnäytetyö
Marraskuu, 2021

Laurea-ammattikorkeakoulu Tiivistelmä
Tietojenkäsittely
Tietojenkäsittely (AMK)

Iiro Lahti

OSINT-tiedustelun automatisointi pimeässä verkossa ja CAM-materiaali tutkinnan haasteet

Vuosi 2021 Sivumäärä 25

Tässä opinnäytetyössä esitellään avoimen lähteen tiedustelua (OSINT). Työssä esitellään myös erilaisia automaatiotyökaluja, joita voidaan käyttää OSINT-tiedustelun suorittamiseen. Opinnäytetyö on tehty osana AiBOLOTE -projektin valmisteluprosessia, johon Laurean ammattikorkeakoulun on tarkoitus osallistua. Hankkeen tarkoituksena on tutkia OSINT-tiedustelun keinoja pimeässä verkossa. Sen avulla voidaan esimerkiksi tutkia Child Abuse Material (CAM-materiaalia).

Opinnäytetyön tutkimus on toteutettu teoreettisen tutkimuksen keinoin. Työn keskeinen tuotos on selvitys, miten OSINT-tiedustelua voidaan automatisoida ja mitä haasteita automatisoitu tiedustelu mahdollisesti sisältää. Pohdin myös niitä haasteita, joita esiintyy CAM-materiaalin tutkinnassa pimeässä verkossa.

Työssä kuvataan niitä haasteita, joita syntyy CAM-materiaalin tutkimisessa pimeässä verkossa ja tarkastellaan myös tapaa, jolla tämä tutkimus voitaisiin suorittaa OSINT:n avulla altistamatta tutkijoita itse CAM-materiaalin sisällölle. Menetelmänä voisi olla keskittää tiedustelua jo tunnettuihin linkkeihin ja sivustoihin.

Asiasanat: OSINT, Avoimen lähteen tiedustelu, pimeäverkko, Internetin tasot, CAM-materiaali

Laurea University of Applied Sciences
Degree Programme in Business Information Technology
Bachelor's Thesis

Iiro Lahti

Automating OSINT intelligence and child pornography investigation challenges

Year 2021 Pages 25

Abstract

This thesis project introduces open-source intelligence (OSINT) and the various automation tools that can be used to utilize OSINT. The thesis has been done as a part of the preparation process for the AiBOLOTE project, in which Laurea University of Applied Sciences is to participate. The aim of the project is to study the uses of OSINT in the dark web, which can be used, for example, to study Child Abuse Material (CAM).

The thesis has been carried out utilizing extensive theoretical research. The key output of the work is an explanation of how OSINT can be automated and what challenges automated intelligence possibly entails. An additional output is a reflection on the challenges that arise in the study of CAM in the dark web.

The work also describes the challenges that arise in the study of CAM in the dark web. The work reviews the way in which this study could be conducted using OSINT without exposing researchers to the content of the CAM itself. The method could be to focus the inquiry on already known links and sites.

Keywords: OSINT, Open-source intelligence, Dark web, Internet layer, CAM-material

Sisällys

1 Johdanto

Tämä opinnäytetyö on tehty osana AiBOLOTE hankkeen valmisteluprosessia. Hankkeen tarkoituksena on tutkia avointen lähteiden tiedustelun (OSINT) keinoja, joilla esimerkiksi Child Abuse Materialin (CAM-materiaali), eli lapsipornotutkintaa voidaan suorittaa. Opinnäytetyöhön pohjautuva tutkimus on osa ICCWS-konferenssia.

Opinnäytetyössä kerrotaan aluksi mitä OSINT tarkoittaa ja miten se toimii, sekä millaisia hyötyjä ja haasteita OSINT-tiedustelussa on. Työssä esitellään muutama yleisin OSINT-tutkimukseen hyödynnetty työkalu. Tämän jälkeen kuvataan internetin tasot ja keskitytään pimeän verkon toimintaan. Näiden jälkeen paneudutaan enemmän siihen, miten automatisoida OSINT-tiedustelua ja miten haasteellista CAM-materiaalin tutkinta on pimeässä verkossa. Viidennessä kappaleessa paneudutaan tutkimustuloksiin ja työkaluihin, joita hyödyntämällä OSINT-tiedustelu onnistuu. Työssä käsitellään myös vaiheita, miten aloittaa OSINT-tiedustelu pimeässä verkossa. Lopussa on pohdintaa OSINT-tiedustelun tulevaisuudesta, tutkimukseni luotettavuutta sekä mahdollisia jatkotutkimuskohteita.

Tämä tutkimus on suoritettu teoreettisen tutkimuksen keinoin. Opinnäytetyössä tutkitaan teoriaan pohjautuen OSINT-tiedustelua ja eri työkaluja, joita OSINT-tiedustelussa voidaan käyttää. Lisäksi selvitetään, miten voitaisiin automatisoida rikostutkintaa ja etenkin CAM-materiaalin tutkintaa. Tutkinta on haastavaa, sillä tutkija ei saisi törmätä vahingossa CAM-materiaaliin. Tässä työssä pohditaan, miten tämä voitaisiin välttää ja silti suorittaa tutkintaa mahdollisimman tehokkaasti.

Opinnäytetyön aluksi luvussa kaksi kerron, mikä on OSINT ja mitä se tarkoittaa. Tämän jälkeen käyn läpi OSINT-tiedustelun historiaa ja modernia avointein lähteiden tiedustelua. Modernin tiedustelun jälkeen pohditaan OSINT-tiedustelun eettisyyttä. Luvun lopussa kuvataan internetin tasot ja OSINT-tiedustelun työkalujen määritelmät. Kolmannessa luvussa puolestaan syvennytään tarkemmin pimeään verkkoon: miten se toimii ja miten pimeässä verkossa on turvallista toimia. Neljännessä luvussa siirrytään itse tutkimukseen ja selvitetään automatisoinnin mahdollisuuksia pimeän verkon tiedustelussa. Lisäksi pohditaan, kuinka pimeässä verkossa voidaan tutkia CAM-materiaalia ja mitä haasteita tähän sisältyy. Luvussa viisi syvennytään automatisointiin, millaisia hyötyjä siitä on ja sisältyykö siihen jotain haasteita. Lisäksi paneudutaan automatisointityökaluihin. Luvun lopussa käydään seikkaperäisesti läpi vaiheet, miten aloittaa OSINT-tiedustelu pimeässä verkossa. Opinnäytetyön viimeinen luku sisältää pohdintaa edellä olleista aiheista ja lisäksi paneudutaan itse tutkimuksen tuloksiin. Näin muodostuu selkeä ja looginen kokonaisuus, jossa käydään läpi OSINT-tiedustelua ja sen tuomia hyötyjä ja haasteita CAM-materiaalin tutkinnassa.

2 Mikä on OSINT?

OSINT (Open Source Intelligence) tarkoittaa avoimien lähteiden tiedustelu. OSINT sisältää tiedon hankinnan, koostamisen ja analysoinnin sellaisista lähteistä, joihin on vapaa tai kaupallisesti saatavilla oleva pääsy. (Borges 2021.)

Avainsana OSINT-konseptin takana on tieto. Tarkemmin sanottuna tämä tarkoittaa tietoja, jotka ovat saatavissa ilmaiseksi. Sillä ei ole merkitystä, onko se sanomalehdissä, blogeissa, verkkosivuilla, twiiteissa, sosiaalisen median korteissa, kuvissa, podcasteissa tai videoissa, kunhan se on julkista, ilmaista ja laillista. (Borges 2021.)

Avointen lähteiden tiedustelua suorittavat käytännössä kaikki. Yritykset ja yksityiset henkilöt hyödyntävät OSINT-tiedustelua, mutta he eivät ole ainoita. OSINT-tiedustelu on hyvin yleisesti käytössä myös viranomais- ja lainvalvontaorganisaatioissa.

Monet arviot osoittavat, että 90 prosenttia tiedustelupalvelujen saamasta hyödyllisestä tiedosta tulee julkisista lähteistä, eli toisin sanoen OSINT-lähteistä. Sosiaalisen median sivustot avaavat lukuisia mahdollisuuksia tutkimuksiin, koska yhdessä paikassa on valtava määrä hyödyllistä tietoa. Voit esimerkiksi saada paljon henkilökohtaisia tietoja kaikista ihmisistä maailmanlaajuisesti tarkistamalla heidän Facebook-sivunsa. (Hassan & Hijazi 2018 xix.)

OSINT sisältää kaikki julkisesti saatavilla olevat tietolähteet. Nämä tiedot löytyvät joko verkosta tai verkon ulkopuolelta. Verkosta saatavat tiedot ovat foorumit, blogit, sosiaaliset verkostoitumissivustot, videonjakosivustot (kuten YouTube.com), wikit, rekisteröityjen whois-tietueet verkkotunnukset, metatiedot ja digitaaliset tiedostot. Tietolähteisiin kuuluvat myös dark webin resurssit, maantieteelliset sijaintitiedot, IP-osoitteet, hakukoneet ja ylipäätään kaikki mitä löytyy verkosta. Verkon ulkopuolelta, eli niin sanottuja perinteisiä lähteitä ovat joukkotiedotusvälineet (televisio, radio, sanomalehdet, kirjat, lehdet), akateemiset julkaisut, väitöskirjat, konferenssijulkaisut, yritysprofiilit, vuosikertomukset, yritysuutiset, työntekijäprofiilit ja yhteenvetotiedot. Tietoa antavat myös valokuvat ja videot, mukaan lukien metatiedot, paikkatiedot, kuten kartat ja kaupalliset kuvatuotteet.

2.1 OSINT tiedustelussa

Yhdysvaltain puolustusministeriö (DoD) määrittelee OSINT:n seuraavasti: ”Avoimen lähdekoodin tiedustelu (OSINT) on tiedustelu, joka tuotetaan julkisesti saatavilla olevasta tiedosta ja jota kerätään, hyödynnetään ja levitetään oikeaan aikaan asianmukaiseen yleisöön tätä tarkoitusta varten. tietyn tiedusteluvaatimuksen täyttämiseksi. "

Vuonna 2001 julkaistun Naton avoimen lähdekoodin tiedustelukäsikirjan V1.2 mukaan avointa tietoa ja tiedustelutietoja on neljä luokkaa.

* Avoimen lähdekoodin tiedot (OSD): Nämä ovat yleisiä tietoja, jotka tulevat ensisijaisesta lähteestä. Esimerkkejä ovat satelliittikuvat, puhelutiedot ja metatiedot, datajoukot, kyselytiedot, valokuvat sekä ääni- tai videotallenteet, jotka ovat tallentaneet tapahtuman.

*Avoimen lähdekoodin tiedot (OSINF): Nämä ovat yleisiä tietoja, jotka on ensin suodatettu jonkin tietyn kriteerin tai tarpeen täyttämiseksi. Näitä tietoja voidaan kutsua myös toissijaiseksi lähteeksi. Esimerkkejä ovat kirjat tietystä aiheesta, artikkelit, väitöskirjat, taideteokset ja haastattelut.

*Avoimen lähdekoodin älykkyys (OSINT): Tämä sisältää kaiken tiedon, joka on löydetty, suodatettu ja määritetty vastaamaan tiettyä tarvetta tai tarkoitusta. Tätä tietoa voidaan käyttää suoraan missä tahansa tiedustelutilanteessa. OSINT voidaan määritellä pähkinänkuoressa avoimen lähdekoodin materiaalinkäsittelyn tulokseksi.

*Vahvistettu OSINT (OSINT-V): Tämä on OSINT erittäin varmalla tavalla: tiedot olisi vahvistettava käyttämällä muuta kuin OSINT-lähdettä tai erittäin arvostetusta OSINT-lähteestä. Tämä on välttämätöntä, koska jotkut ulkopuoliset vastustajat voivat levittää epätarkkoja OSINT-tietoja tarkoituksenaan johtaa harhaan OSINT-analyysejä. Hyvä esimerkki tästä on, kun televisioasema lähettää suorana lähetyksenä presidentin saapumisen toiseen maahan. Tällainen tieto on OSINT, mutta sillä on suuri varmuus.

2.2 OSINT:n Historia

Toisen maailmansodan aikana Yhdysvaltain tiedustelupalvelu The Office of Strategic Services (OSS), joka oli CIA:n edeltäjä, alkoi hyödyntää avointen lähteiden tiedustelua Japanin iskettyä Pearl Harbouriin. Ison-Britannian tiedustelupalvelu Special Operations Executive (SOE) oli jo hyödyntänyt sodan aikana OSINT-tiedustelua. (Colquhoun 2016.)

OSS katsoi kuolinilmoituksia Saksan alueellisissa sanomalehdissä etsien uutisia tärkeistä natseista. Kuvia uusista taistelulaivoista, pommikraattereista ja lentokoneista koottiin huolellisesti, ja yhdessä arvioituna ne antoivat OSS:lle mahdollisuuden arvioida vihollistensa tilaa. On hämmästyttävää, kuinka OSS:n toiminta on samanlaista kuin nykypäivän OSINT-tutkimukset, vaikkakin ilman tietokoneita. OSS:n ja SOE:n perusteella voidaan väittää, että avoimen lähdekoodin älykkyyden juuret ulottuvat lähes vuosisadan taakse. (Colquhoun 2016.)

Toisen maailmansodan jälkeen avointen lähteiden tiedustelua hyödynnettiin luonnollisesti kylmän sodan aikana. Datan määrä, jota voitiin hyödyntää OSINT-tiedustelussa ei missään nimessä vähentynyt varsinaisen sodan jälkeen ja siirryttäessä kylmään sotaan. (Colquhoun 2016.)

Kylmän sodan päättymisen jälkeen globaaleista yhteiskunnista on tullut avoimempia, ja Internetin vallankumous ja sen laaja käyttö ovat muuttaneet maailman pieneksi kyläksi. Internet-

verkon vapauttaminen miljardeille ihmisille ympäri maailmaa kommunikoimaan ja vaihtamaan digitaalista dataa on siirtänyt koko maailman nykyiseen tietokauteen. Tämä muutos digitaaliseen aikakauteen toi valtavia etuja yhteiskunnallemme. Muutoksen nopeus ja laajuus ovat aiheuttaneet myös erilaisia riskejä. Verkkorikolliset, terroristiryhmät, sortavat järjestelmät ja kaikenlaiset haitalliset toimijat käyttävät Internetiä tehokkaasti rikostensa suorittamiseen. Juniper Research ennusti, että tietoverkkorikollisuus maksaisi yrityksille yli 2 biljoonaa dollaria vuoteen 2019 mennessä (Hassan & Hijazi 2018, 1.)

2.3 Moderni avointen lähteiden tiedustelu

Tiedustelun tieteenaloja on kuusi, joista OSINT on luonnollisesti yksi. Nämä tieteenalat tai ehkä paremmin tiedusteluhaarat ovat:

SIGINT - Signal Intelligence eli signaalitiedustelu

IMINT - Imagery Intelligence eli kuvatiedustelu

MASINT - Measurement and Signature Intelligence eli mittaus- ja tunnusmerkkitiedustelu

HUMINT – Human Intelligence eli henkilötiedustelu

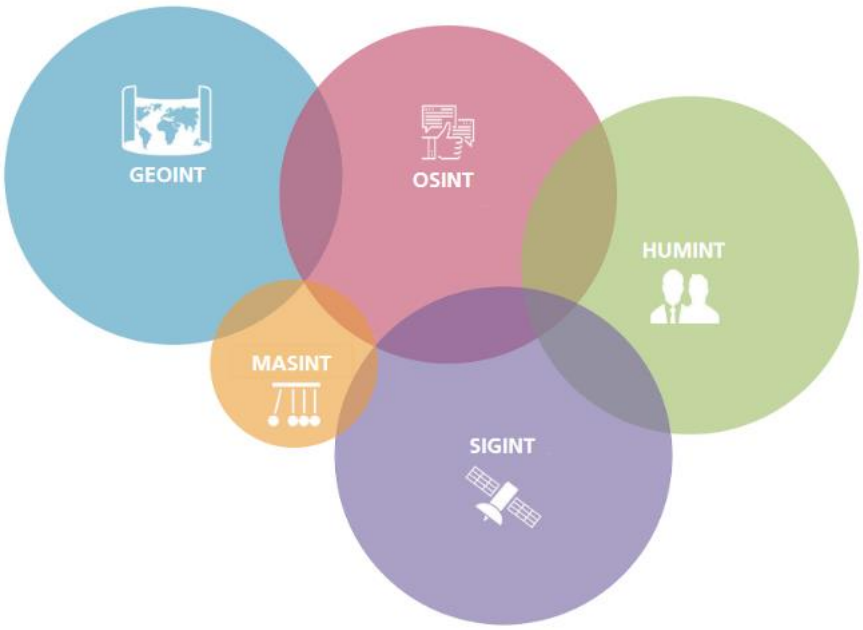
OSINT – Open Source Intelligence eli avoimen lähdetiedon tiedustelu

GEOINT - Geospatial Intelligence eli sijaintitiedustelu

(Office of the Director of National Intelligence 2021.)

OSINT-tiedustelu on alkanut muistuttamaan yhä enemmän muita tiedusteluluokkia teknologisen kehityksen vuoksi. GEOINT on aikaisemmin palvellut oikeastaan vain valtiollisia toimijoita, mutta nykyään kaupalliset satelliittikuvantamiset ovat tulleet julkisesti saataville. Sosiaalisessa mediassa suoritettava OSINT-tiedustelulla voidaan profiloida kohdehenkilöitä ja saada näin tietoja kohteen yhteyksistä, toimintatavoista sekä vapaa-ajasta. Tätä voidaan verrata HUMINT-tiedusteluun tai tämä voi olla ensimmäinen askel, kun suunnitellaan HUMINT-tiedustelu operaatiota. OSINT-tiedusteluoperaation tuoma valtava datamäärä alkaa puolestaan muistuttamaan SIGINT-tiedustelua. (Williams & Blum 2018, 7–8.)

OSINT-tiedustelun rooli ja merkitys on siis kasvanut viime vuosien aikana merkittävästi. OSINT-tiedustelun tuomat suuret tietomäärät vaikuttavat jokaiseen tiedustelualaan. Samaan aikaan tiedustelualat ovat alkaneet lähestymään toisiaan ja rajat hämärtyä. On huomioitava, että OSINT-tiedustelu kuitenkin eroaa muista tiedustelualoista, sillä se hyödyntää ainoastaan laillisesti saatavilla olevia lähteitä rikkomatta tekijänoikeuksia tai yksityisyyttä koskevia lakeja. Seuraava kuva esittää hyvin kuinka keskeisessä asemassa OSINT-tiedutelu nykyään on. OSINT-linkittyy erittäin vahvasti neljään muuhun tiedustelun haaraan.



Kuvio 1: OSINT suhteessa muut tiedustelumuotoihin. (Williams & Blum 2018.)

2.4 Avointen lähteiden tiedustelun eettisyys

Avointen lähteiden tiedustelu sisältää tiettyjä eettisiä ja laillisia haasteita. Sitä voidaan käyttää luonnollisesti hyvää ja pahaan. Mahdollista on myös, että tietoja kerättäisiin vain tietyn päämäärän tai lopputuloksen saamiseksi. (Hassan & Hijazi 2018, 17–18.)

Joku voi hankkii OSINT-lähteitä laittomin keinoin perustellakseen laillisen tutkimuksen. Miten oikeusjärjestelmän pitäisi käsitellä sitä? Toinen ongelma on, kun OSINT-lähteitä rajataan tai valitaan käyttäjän tarpeen mukaan. Käyttäjät voisivat tehokkaasti hylätä tärkeät lähteet tarkoituksellisesti tietyn tuloksen aikaansaamiseksi. (Hassan & Hijazi 2018, 17–18.)

Toinen huolenaihe on se, että tietynlaisia piilotettuja julkisia tietoja kerätään ja julkistetaan laajasti osana skandaalia. Tavallinen Internetin käyttäjä ei voi tarkastella paljoakaan julkista tietoa, ja sen hankkiminen vaatii erityisiä tekniikoita ja menetelmiä. Mikä on seuraus tällai-

sista asioista? Mitä vaikutuksia joihinkin ryhmiin tai yksilöihin tulee, kun heistä paljastetaan tällaisia tietoja? Mitkä ovat moraaliset seuraukset? (Hassan & Hijazi 2018, 17–18.)

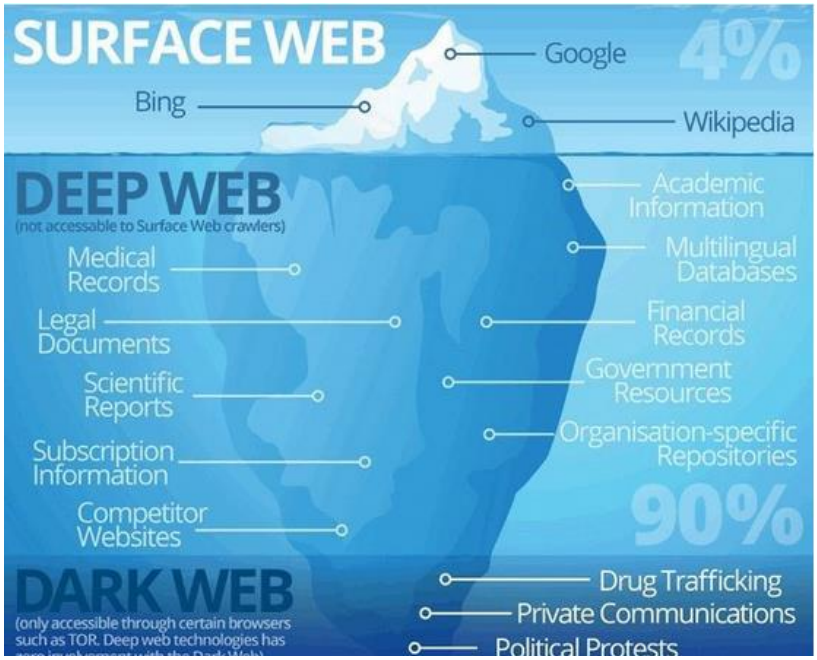
OSINT-tiedustelua suorittavilla tahoilla tulisi olla eettinen ohjeistus henkilöstölleen. Italian Confidential Intelligencetin Osint Centerillä on yksitoistakohtainen eettinen ohjeistus henkilöstölleen. (Osint-ltd 2021.)

Turvallisuusalalla on suuri vastuu ja lakisääteisten määräysten lisäksi tärkeätä on myös yksilöiden kunnioittaminen. Osint Centerin henkilöstön korkean ammattitaidon varmistamiseksi henkilöstölle annetaan tiukka ja läpinäkyvä eettinen ohjeistus, jota on noudatettava. (Osint-ltd 2021.)

2.5 Internetin tasot

Tässä kappaleessa viitataan lähdeaineistoon lähdekirjallisuuden sijaan. Syy tähän on selkeys. Digital on kasannut yhteen mielestäni hyvin ja selkeästi saman teoriatiedon, kuin mitä lähdekirjallisuuskin.

Internetistä saatavat tiedot ovat avointen lähteiden tiedustelun kannalta elintärkeitä. Nämä tiedot mahdollistavat OSINT:n tehokkaan käytön, kun verkosta saadaan hankittua valtavasti tietoa valitusta kohteesta. Internet koostuu valtaisan kehityksensä vuoksi kolmesta eri tasosta, jotka ovat World Wide Web eli verkkopinta, Deep Web, eli syvästä verkosta ja Dark Web, eli pimeästä verkosta. OSINT-tiedustelussa pyrkimyksenä on hyödyntää kaikkien näiden tasojen mahdollistamaa avointen lähteiden kokonaisuutta. Perinteisessä verkossa tapahtuvassa OSINT-tiedustelussa tiedon keruu kohdistetaan verkkopintaan. Verkkopinta, eli WWW koostuu kuitenkin vain noin viidestä prosentista kaikesta verkkoliikenteestä. Tämä viisi prosenttia siis sisältää kaikki julkiset sivustot, kuten Googlen, Facebookin ja Amazonin sekä lähtökohtaisesti kaikki sivustot, jotka on indeksoitu hakukoneiden järjestelmiin. (Digital 2021.)



Kuvio 2: Verkon jakauma. (Slide to doc 2021)

Deep Web eli syvä verkko tarkoittaa puolestaan tarkoittaa verkkosivustoja, joihin ei päästä tavanomaisilla hakukoneilla, kuten Googlella tai Bingillä. Näitä deep webin sivustoja ei löydy perinteisten hakukoneiden hakutuloksista, sillä niitä ei ole näiden hakukoneiden indeksointijärjestelmissä. Näille deep webin sivuille pääsee vain tarkan URL-osoitteen kautta. Deep webin tarkoituksena on siis periaatteessa pitää halutut verkkosivut poissa normaalien käyttäjien näköpiiristä. Deep webiä hyödyntävät suuryritykset, koulut, järjestöt sekä valtiolliset ja kunnalliset toimijat. Näiden toimijoiden tarkoituksena on pitää sivustonsa vain oman henkilöstönsä käytössä. Sivut ovat usein suojattu salasanalla, sillä ne sisältävät salassa pidettävää tietoa. (Digital 2021.)

Tasoista syvimpänä on dark web eli pimeä verkko. Dark web on osa deep webiä, eli myöskään näitä dark webin sivustoja ei ole indeksoitu hakukoneisiin. Dark webin ja deep webin välisenä erona on vain se, ettei dark webin sivustoihin pääse kuin hyödyntäen salattua reititystekniikkaa. Tällaista tekniikkaa hyödyntäviä selaimia ovat muun muassa Tor, I2P ja Freenet. Nämä selaimet salaavat käyttäjän tietoliikenteen ja sijainnin. Suosituin dark web selaimista on Tor. Tämän verkon toiminta on täysin anonyymiä. (Ozkaya & Islam 2019, 47.)

Tor nimi tulee sanoista The Onion Router. Tämä nimi kuvaan hyvin myös verkon toimintaa. Verkon tietoliikenne on kerroksellinen, kuten sipuli. Kerroksellisessa suojauksessa käyttäjän tietoliikenne salataan ja reititetään useiden eri yhdyspisteiden kautta. Näitä yhdyspisteitä eli solmuja on noin 7000 kappaletta. Tämän vuoksi mikään yksittäinen yhdyspiste ei vie käyttäjää määränpäähän, joten käyttäjän paikantaminen on käytännössä mahdotonta. (Ozkaya & Islam 2019, 47.)

Suoritettaessa OSINT-operaatiota deep webissä ja dark webissä, tulee käytössä olla tähän tehtävään tarkoitettuja erityisiä keräily-, käsittely- ja analysointityökaluja. Tietojen etsintä deep- ja dark webistä on hyvin haastava ja aikaa vievä operaatio, joka vaatii suuret resurssit. Tietojen etsintä saattaa olla myös haitallista, etenkin dark webissä. Tiedot saattavat sisältää rikollista materiaalia, sekä haitta- tai kalasteluohjelmia. (Ozkaya & Islam 2019, 47.)

2.6 OSINT-työkalun määritelmä

Automaattisten työkalujen avulla online-hakija voi automatisoida hakuprosessin suurimpien hakukoneiden, kuten Googlen, Bingin ja Shodanin avulla. Automaattiset työkalut ovat nopeita ja mahdollistavat jatkuvan testin suurelle määrälle hakulausekkeita. Menetelmä palauttaa kattavammat tulokset, koska hakutyökalu voi rakentaa monimutkaisia hakulausekkeita paremmin kuin ihminen. (Hassan & Hijazi 2018.)

2.6.1 Pintaverkossa ja syvässä verkossa käytetyt työkalut

Pintaverkossa tapahtuva OSINT-tiedustelu on helpoin tapa aloittaa tiedusteluoperaatio. Pintaverkosta kerättävät tiedot ovat helposti saatavilla hakukoneindeksoitujen sivujen vuoksi. Pintaverkko antaa tiedustelijalle hyvin laajan valikoiman työkaluja ja erilaisia tekniikoita tarvittavan tiedon keräämiseen (Bertram 2015, 21, 56.)

Edellisessä kappaleessa on kuvattu useita eri OSINT-tiedustelutyökaluja. Nämä kaikki työkalut sopivat mainiosti pintaverkon tiedusteluun. Pintaverkon tiedusteluun ei ole välttämätöntä käyttää mitään muuta työkalua, kuin vain hakukonetta kuten Googlea. Työkalut tosin helpottavat ja nopeuttavat tiedustelua suuresti.

3 Mikä on pimeä verkko?

1990-luvun lopulla kaksi Yhdysvaltain puolustusministeriön tutkimusorganisaatiota pyrkivät kehittämään anonyymin ja salatun verkon, joka suojaisi Yhdysvaltain vakoojien arkaluontoista viestintää. Tämä salainen verkko ei olisi tavallisten Internet-surffailijoiden tuntema tai käytettävissä. Vaikka alkuperäinen salainen aikomus ei koskaan toteutunut täysin, jotkut tutkijat näkivät tarpeelliseksi perustaa voittoa tavoittelematon järjestö, joka keskittyy ihmisoikeus- ja yksityisyysaktivistien nimettömyyteen. (Kumar & Rosenbach 2019.)

Nykyään Internet määritellään verkkojen verkostoksi tai useiksi toisiinsa yhdistetyiksi tietokoneverkoiksi, jotka tarjoavat viestintä- ja tieto-ominaisuuksia käyttäen standardoituja viestintäprotokollia, kuten liikenteenohjausprotokollaa/Internet -protokollaa (TCP/IP). (Retzkin 2018.)

Pintaverkko tai WWW on Internetissä. Dark Web on olemassa Dark Netissä tai pikemminkin useissa pimeissä verkoissa. On tärkeää huomauttaa, että termit Dark Web ja Dark Net eivät ole sama asia. Dark net oli termi, jota käytettiin 1970-luvulla verkoissa, jotka oli eristetty ARPANET:sta pääasiassa turvallisuustarkoituksiin. Ne on määritetty vastaanottamaan ulkoista dataa, mutta ne oli piilotettu ARPANET-verkkoluetteloista eivätkä vastanneet verkkotiedusteluihin, kuten ping-pyyntöihin. (Retzkin 2018.)

Ajan myötä termiä käytettiin myös peiteverkoissa, jotka ovat pääasiassa verkkoja, jotka käyttävät ohjelmistoja ja laitteistoja useiden kerrosten luomiseen. Nämä kerrokset ajetaan useiden ja erillisten verkkokerrosten päällä tai yhteisen verkon kautta, joihin pääsee vain erityisillä selaimilla tai ohjelmistolla tai joissa niiden IP-osoitteet eivät ole maailmanlaajuisesti reititettävissä. Muutamia esimerkkejä tällaisista peittoverkoista ovat Tor, Invisible Internet Project (I2P) tai FreeNet. (Retzkin 2018.)

Totuus pimeästä verkosta on, että se tarjoaa äärimmäistä yksityisyyttä ja suojaa autoritaaristen hallitusten valvonnalta, helpottaa kasvavaa maanalaista markkinaa, jota rikolliset käyttävät huumeiden, varastettujen henkilöllisyyksien, lapsipornografian ja muiden laittomien tuotteiden ja palvelujen liikenteeseen. Koska jäljittämätön kryptovaluutta on ensisijainen maksuväline pimeässä verkossa, tarvitaan tiivistä yhteistyötä lainvalvontaviranomaisten, rahoituslaitosten ja sääntelyviranomaisten välillä ympäri maailmaa. (Kumar & Rosenbach 2019.)

3.1 Miten pimeä verkko toimii?

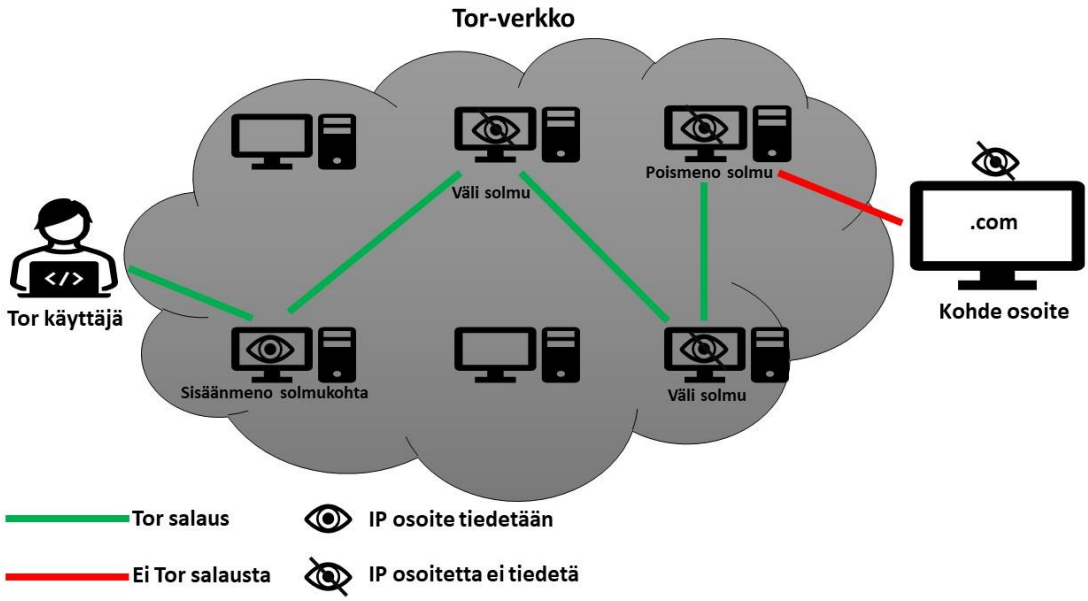
Tässä otsikossa viitataan tutkimusaineistoon. Syy tähän on, että artikkelin kirjoittaja Louis DeNicola on koonnut mielestäni oivallisesti yhteen seikat, jotka ovat löydettävissä myös tutkimuskarjallisuudesta. DeNicola on myös lisännyt esimerkiksi CIA:n osoitteen, joka kuvaa mielestäni hyvin Tor-verkon eroavaisuutta pintaverkosta. Tällaista en ole löytänyt tutkimuskirjallisuudesta. (DeNicola 2021.)

Tor-verkko on suurin ja ehkä tunnetuin pimeän verkon selain. Usein viitataan juuri siihen, kun puhutaan pimeästä verkosta. (DeNicola 2021.)

Päästäksesi pimeään verkkoon voit ladata ja asentaa ilmaisen Tor-selaimen. Voit käyttää sitä myös pinnan ja syvän verkon selaamiseen. Se on täysin laillista, ja Yhdysvaltain hallitus on merkittävä rahoittaja Tor-projektille, joka luo selaimen. (DeNicola 2021.)

Mutta selaimen käyttö on vasta ensimmäinen askel. Toisin kuin pintaverkko, pimeälle verkolle ei ole suuria hakukoneita. Sinun on ehkä löydettävä sipulisivuston osoite itse, eikä nimiä ole helppo muistaa. Esimerkiksi CIA:n sipuliosoite on ciadot-gov4sjwlzihbbgxnqg3xiyrg7so2r2o3lt5wz5ypk4sxyjstad.onion ja voittoa tavoittelematon uutishuone ProPublica's on propub3r6espa33w.onion. (DeNicola 2021.)

Nämä ovat kaksi esimerkkiä laillisista organisaatioista, joilla on sivustoja pimeässä verkossa. Mutta pimeä verkko tunnetaan parhaiten siellä tapahtuvasta laittomasta toiminnasta. (DeNicola 2021.)



Kuvio 3: Miten Tor-verkko toimii?

3.2 Miten liikkua turvallisesti pimeässä verkossa?

Kun aloitetaan tiedustelutoimet pimeässä verkossa, on hyvä ymmärtää tähän sisältyvät riskit ja kuinka tiedustelua voidaan suorittaa turvaten omaa tietoliikennettä ja identiteettiä. Pimeässä verkossa toimiminen on turvallista, mikäli pysytään poissa epäilyttäviltä sivuilta. Tämä tosin on erityisen haastavaa, jos tarkoituksena on tutkia ja tiedustelle rikollista toimintaa.

Terbium Labin mukaan noin 55 % pimeän verkon sivuistoista on laillisia, joten näillä sivustoilla tehtävä tiedustelu on lähtökohtaisesti varsin turvallista. Samaan aikaan on huomattava, että noin 45 % pimeän verkon sivuista sisältävät laitonta sisältöä. Tiedustelu pimeässä verkossa voi siis olla varsin riskialtista toimintaa.

Erdal Ozkaya ja Rafiqul Islam kuvaavat kirjassaan haavoittuvuuksia ja heikkouksia, joita hyödyntämällä henkilön identiteetti voidaan paljastaa pimeässä verkossa. Useat eri tavat vaativat kuitenkin suuria resursseja, jotta pystytään suorittamaan henkilöiden identiteettien paljastamisen. Tämä tarkoittaa käytännössä valtiollisen tason resursseja, joskin joillain suuryrityksillä voi olla mahdollisuus suorittaa henkilötiedustelua pimeässä verkossa. Erilaisia tapoja suorittaa tiedustelua ovat: Website fingerprinting, Eavesdropping, Traffic Analysis, Exit Node Block, Browser Vulnerabilities ja The Bad apple attack.

4 Tutkimus

Tässä luvussa käsittelen aluksi tutkimusmenetelmiäni ja strategiaani tutkimukseni suorittamisessa. Tämä jälkeen tutkin, miten automatisoida OSINT-tiedustelua pimeässä verkossa. Tarkastelen myös CAM-materiaalin tutkintaa pimeässä verkossa. Käytän lähdeaineistonani aluksi Jyväskylän yliopiston aineistoa tutkimusmenetelmän määrittelystä. Tämän jälkeen paneudun teoriaan Inside The Dark Web ja Automating Open Source Intelligence: Algorithms for OSINT-

kirjojen avulla. Nämä kirjat ovat erittäin informatiivisia ja selkeitä sekä käsittelevät erinomaisesti aihetta. CAM-materiaalia käsittelevä otsikko on kirjoitettu varsin pitkälti vain tutkimusaineiston pohjalta, sillä CAM-materiaali aiheesta löytyy varsin heikosti tieteellistä ja luotettavaa kirjallisuutta. Sen vuoksi tässä osiossa olen keskittynyt haastatteluihin. Näiden haastatteluiden luotettavuuden olen pyrkinyt varmistamaan siten, että haastattelun sisältö on ollut löydettävistä myös muista lähteistä ja että viitattavat lähteet työskentelevät yleisesti luotettuina pidetyille tahoille, kuten Suomen tietotoimistolle.

4.1 Tutkimusmenetelmä ja -strategia

Toteutus on usein tutkimuksen työläin vaihe, jossa kerätään tutkimuksen empiirinen aineisto ja analysoidaan se suunnitelman mukaisesti. Usein aineiston kerääminen ja analysoiminen tuottaa tutkimukseen sellaisia ajatuksia ja näkökulmia, joita ei vielä tutkimussuunnitelman laatimisvaiheessa osattu ottaa huomioon. (Jyväskylän yliopisto 2015.)

Teoreettisessa tutkimuksessa ei havainnoida tutkimuskohteita välittömästi, vaan kohteesta pyritään hahmottamaan käsitteellisiä malleja, selityksiä ja rakenteita aiemman tutkimuskirjallisuuden pohjalta. (Jyväskylän yliopisto 2015.)

Tutkimukseni on suoritettu empiirisen tutkimuksen keinoin. Tutkimuksessani laajasti kerätty tutkimuskirjallisuus ja -aineisto ovat keskeisessä roolissa, kun käsittelen avointein lähteiden tiedustelua ja pimeän verkon rakennetta. Lisäksi pohdintaani ohjaa vahvasti keräämiäni ja käsittelemäni lähdeaineisto. Miksi tutkimusta ei suoritettu sitten esimerkiksi laadullisella tutkimusmenetelmällä? Kiinnostusta herättää myös, miten ihmiset pärjäävät todellisissa olosuhteissa. Laadulliset tutkimukset voivat kiinnittää huomiota näiden asetusten kontekstuaaliseen rikkauteen, tutkimuksesi avulla voidaan tutkia monenlaisten ihmisten jokapäiväistä elämää ja mitä he ajattelevat monissa eri olosuhteissa. (Yin 2015, 3.) Tällainen tutkimustapa ei olisi ollut mahdollinen tässä tutkimustyössä. CAM-materiaalin tutkintaa OSINT-tiedustelun keinoin ei ole mahdollista suorittaa laadullisin keinoin.

Strategianani tutkimuksen suorittamisessa oli kerätä lähdeaineistoa mahdollisimman laajalti ja eri tyylisistä lähteistä. Tutkimuksen lähtökohtani on ollut tarkastella OSINT-tiedustelua ja pimeää verkkoa viranomaisten näkökulmasta. Tämä on toki vaatinut aluksi sen, että avaan OSINT:a ja internetin rakennetta laajemmin.

4.2 Miten olisi mahdollista automatisoida OSINT-tiedustelua pimeässä verkossa?

Datan kerääminen pimeästä verkosta ei ole yksinkertaista. Monet organisaatiot epäonnistuvat siinä. (Ozkaya & Islam 2019, 220.)

Avointen lähteiden tiedustelua on siis hyvin haastavaa automatisoida. Automatisoidessa tulisi voida rajata erittäin tarkasti se tieto mitä haluttaisiin saada. Avointen lähteiden tiedustelussa kerätään suuri määrä tietoa erinäisistä lähteistä, joten vaarana on tiedon liiallinen määrä. Tämän tiedon analysointiin tarvittaisiin osaava henkilöstö.

Tietojen keräämisen laajuus on suurta ja laaja-alaista. Nopeus, määrä ja monipuolisuus ovat niin suuria, että OSINT saattaa luoda Big Data-ongelman. Työkalut, joilla käsitellään tietoja käsitteleviä työkaluja, kuten Maltego ja Recon-Ng, ovat tulossa yhä suositummiksi ja yleisemmiksi. Nämä lähestymistavat vaativat kuitenkin edelleen asetuksia ja tietyn määrän osaamista. Nämä vaaditut asetukset sisältävät myös tietyn määrän toimintoja, joita ei voida automatisoida tai ainakin se olisi vaikeaa. Tietojen hakeminen ja jossain määrin väärien positiivisten rajoittaminen voidaan automatisoida. (Layton & Watters 2015, 3–4.)

Avointen lähteiden tiedustelua on kuitenkin mahdollista automatisoida ainakin jossakin määrin. Parhaiten tiedustelun automatisointi onnistuu tiedon keruun osalta. Analysointiin automatisointi on haastavampaa ja johtopäätöksiin sekä jatkotoimien tekemiseen automatisointi ei välttämättä sovellu.

Automaatio edellyttää työkaluja. Ensimmäinen työkalu on indeksointirobotit. Koska tiedon lähteitä on monia, tiedonkeruuprosessi vie aikaa, jos käytetään vain manuaalisia tekniikoita. Verkkoindeksointi alkaa URL-luettelolla eli siemenluettelolla. Kun indeksointirobotti käy näissä, se tunnistaa kaikki sivun hyperlinkit ja lisää uudet linkit URL-osoitteiden luetteloon. Luettelon URL -osoitteita käytetään rekursiivisesti käytäntöjen mukaisesti. Jos indeksointirobotti arkistoi verkkosivustoja, se kopioi ja tallentaa tiedot sellaisinaan. Sitten arkistot tallennetaan, jotta niitä voidaan katsella, lukea ja navigoida sellaisina kuin ne olivat live-verkossa. (Layton & Watters 2015, 9.)

Toinen työkalusarja on sovellusliittymät. Monilla sivustoilla on sovellusliittymät, jotka palauttavat tulokset JSON-muodossa. Sovellusliittymät edellyttävät ilmaista tai maksullista access token-palvelua. Sovellusliittymän JSON-lähtö voidaan tuoda ja käsitellä Pythonilla. Tämä, sekä mahdollisuus tehdä sovellusliittymän eräpyynnöt voivat helpottaa tietojen etsimistä ja keräämistä. (Layton & Watters 2015, 9.)

4.3 Miten CAM-materiaalin tutkinta toteutetaan pimeässä verkossa ja mitkä ovat sen tuomat haasteet?

CAM-materiaalin tekijöille ja kuluttajille Suomi on oikea turvasatama, sillä viranomaisten keinot toiminnan paljastamiseksi ovat olleet alkeellisia. Kun muualla Euroopassa poliisilla on käytössä luokitteluun ja jopa kasvojen ja paikkojen tunnistukseen käytettäviä automaattisia sovelluksia, Suomessa työkaluna ovat olleet poliisin silmät ja jokaisen poliisilaitoksen oma järjestelmä. Samaa materiaalia on saatettu samaan aikaan tutkia Lounais- ja Itä-Suomessa. Tieto ei liiku ja työtä tehdään päällekkäin. (STT 2019.)

CAM-materiaalin tutkiminen on erittäin haasteellista ja tarkoin säänneltyä toimintaa. Se on tutkijalle erittäin rankkaa. ”Näistä hommista on päädytty työkyvyttömäksikin” sanoi lapsipornoa tutkiva rikosylikomisario Sari Sarana Ilta Sanomien haastattelussa vuonna 2019. Tutkittaessa pimeän verkon rikollisuutta on varmistuttava, ettei esimerkiksi huumerikoksia tutkiva viranomainen päädy käsittelemään CAM-materiaalia edes vahingossa. Jos kuitenkin näin kävisi, ei tätä materiaalia saa katsoa, tutkia tai säilyttää. Materiaaliin saa perehtyä vain sellainen viranomainen, joka suorittaa nimenomaan CAM-materiaalin tutkintaa.

CAM-materiaalin tutkinnassa maanrajat ylittävä viranomaistoiminta on hyvin suuressa roolissa. Materiaalia saatetaan tuottaa yhdessä maassa, katsoja on toisesta maasta ja uhri sekä tekijä puolestaan ovat kolmannesta maasta. Lisäksi tutkimusta suorittava viranomainen saattaa olla vielä neljännestä maasta. Tämän vuoksi hyvä tiedonvaihto ja Europolin sekä Interpolin hyödyntäminen on erityisen tärkeätä. Suomessa on muutama vuosi sitten otettu käyttöön CAM-tietokanta, joka parantaa CAM-materiaali tutkintaa ja helpottaa yhteistyötä toisten valtioiden kanssa.

Yksityinen sektori sopisi hyvin viranomaisten avuksi. IOSI-yhteisö voisi olla hyvä apu OSINT-tiedustelussa. IOSI kertoo itsestään verkkosivuillaan seuraavasti: ”IOSI on yhteiskuntaan keskittyvä organisaatio, joka muokkaa turvallisuutta ja älykkyyttä ja on sitoutunut edistämään ja parantamaan kansainvälistä turvallisuutta. IOSI toimii osittain konsultoiden, osittain laboratoriona ja osittain ajatushautomona. Käytännön ratkaisuja kehitetään nykyisiin ja uusiin turvallisuusuhkiin. Näillä keinoilla autetaan julkista ja yksityistä sektoria, sekä kansalaisyhteiskuntaa edistämään tehokkaasti julkista turvallisuutta, demokratiaa ja ihmisoikeuksia.”

IOSI-hanke ja sen jäsenet pyrkivät OSINT:n avulla pääsemään käsiksi tietoon lasten seksuaalisesta hyväksikäytöstä, joka voi auttaa lainvalvontaviranomaisia tunnistamaan väärinkäytöksen tekijät ja löytämään uhrit. IOSI:lla on yhteys yksittäisiin OSINT-asiantuntijoihin ympäri maailmaa. OSINT:n käyttö lasten seksuaalista hyväksikäyttöä koskevissa tapauksissa voi lisätä ja auttaa lisäämään löydettyjen ja pelastettujen uhrien määrää ja lyhentämään siihen kuluvaa aikaa. (Iosi 2020.)

Haasteena on jälleen se, että IOSI:n OSINT-asiantuntijat eivät saa käsitellä CAM-materiaalia. Toisaalta heidän tietojaan OSINT-tiedustelusta voidaan hyödyntää viranomaisten toiminnassa.

CAM-materiaalin tutkintaa voidaan suorittaa pimeässä verkossa toki siten, että tutkija ei itse mene CAM-materiaalia sisältävälle sivustolle. Tosin tässä tilanteessa tulee olla tiedossa sivusto tai linkki, jota kautta päästään kiellettyä materiaalia sisältävälle sivustolle. Tällaisessa tilanteessa OSINT-tiedustelu voisi olla hyvä keino suorittaa tutkintaa. Näin tiedustelu kohdentuisi käyttäjään, joka käyttää kiellettyä sivustoa tai linkkiä, eikä tutkijan itse tarvitse nähdä CAM-materiaalia. Näin toimii muun muassa Yhdysvaltain Department of Homeland Security.

Yhdysvaltain Department of Homeland Security on tunnistanut pimeän verkon käyttäjiä sen jälkeen, kun he ovat ladanneet tiedostoja tiedostonjakopalveluiden kautta. Department of Homeland Security sai useiden epäiltyjen IP-osoitteet, jotka vierailivat Tor-verkossa isännöidyillä lapsipornosivustoilla. Tutkijat seurasivat kaikkia käyttäjiä, jotka käyttivät linkkejä saadakseen arkiston, joka sisälsi pimeässä verkossa ylläpidetyn CAM-materiaalin. (Information security newspaper 2021.)

5 Automatisointi

Tässä luvussa käsittelen aluksi avointen lähteiden tiedustelun hyötyjä ja haasteita. Tämän jälkeen käsittelen automatisointityökaluja ja tarkastelen syvemmin paria automatisointityökalua, joita yhdessä käyttämällä voitaisiin saada kattava näkemys tiedusteltavasta kohteesta. Lisäksi käsittelen OSINT-tiedusteluun liittyviä riskejä ja haasteita, sekä seikkoja ja vaiheita, joiden avulla voidaan aloittaa OSINT-tiedustelu.

5.1 Avointen lähteiden tiedustelun hyödyt ja haasteet

On selvää, että OSINT:lla kuten kaikilla tiedusteluhaaroilla ja tavoilla on omat haasteensa ja rajoitteensa. Avointen lähteiden tiedustelulla on vahva potentiaali, mutta sen sisältämiä rajoitteita ei kuitenkaan tule unohtaa.

Jos halutaan pysyä nimettömänä käytettäessä internetiä, Tor on vähintään yhtä hyvä kuin paraskaan VPN. On kuitenkin muistettava, että Tor ei ole VPN. Se on välityspalvelin, joka suojaa vain sen kautta kulkevaa liikennettä. Yksin Tor ei voi siis taata turvallisuutta ja yksityisyyttä verkossa. On ymmärrettävä parhaat käytännöt ja käyttövinkit, jotta varmistetaan maksimaalinen turvallisuus ja hyödyt. Nämä ovat: Älä käytä henkilökohtaisia tietojasi Pidä järjestelmäsi ajan tasalla. Älä käytä Toria Google -hauissa. Poista Java, JavaScript ja Flash käytöstä. Älä torrent tai käytä P2P -verkostoitumista. Poista säännöllisesti evästeet ja muut tiedot sekä älä käytä HTTP-sivustoja. (James 2018.)

5.1.1 Hyödyt

Yksi OSINT:n käytön suurimmista eduista on hinta. OSINT on paljon halvempi verrattuna perinteisiin tiedonkeruutyökaluihin. Kustannusetujen lisäksi OSINT:lla on monia etuja tiedonsaannissa ja jakamisessa. Tiedot voidaan jakaa laillisesti ja helposti kenelle tahansa ja avoimet lähteet ovat aina saatavilla sekä jatkuvasti ajan tasalla mistä tahansa aiheesta. Julkisista lähteistä kerätty tieto on loistava voimavara kansallisen turvallisuuden tiedusteluun, ja sitä voidaan käyttää tukemaan pitkän aikavälin strategioiden luomista erilaisiin liiketoimintatavoitteisiin. (Expert.ai 2021.)

5.1.2 Haasteet

OSINT:lla on myös omat haasteensa. Yksi OSINT:n suurimmista ongelmista on mahdollinen tiedon ylikuormitus. Oikean tiedon suodattaminen suuresta datamäärästä voi olla vaikeaa. Itse asiassa ilman arvokkaita OSINT-työkaluja oikean tiedon löytäminen ja etsiminen voi olla aikaa vievää toimintaa. OSINT ei myöskään ole käyttövalmis. Se vaatii ihmisiltä paljon analyyttistä työtä, jotta voidaan erottaa kelvolliset, tarkistetut tiedot väärästä, harhaanjohtavasta tai yksinkertaisesti epätarkasta uutisesta ja tiedosta. (Expert.ai 2021.)

5.2 OSINT-työkalut

Avoimen lähteen tiedusteluun on useita erilaisia työkaluja. Geekflare.com on listannut sivuillaan kahdeksan käytetyintä OSINT työkalua. Työkalut ovat seuraavat: Shodan, Spyse, Google Dorks, Maltego, The Harvester, Recon-Ng, SpiderFoot, ja Creepy. Nämä samat työkalut ovat myös muiden sivustojen perusteella hyvin suosittuja ja käytettyjä.

- **Shodan** tarjoaa tuloksia, jotka ovat järkeviä ja liittyvät turvallisuusammattilaisiin. Se sisältää pääasiassa tietoja verkkoon liitettävistä resursseista. Laitteet voivat vaihdella kannettavista tietokoneista, liikennevaloista, tietokoneista ja monista muista IoT-laitteista. Tämä avoimen lähdekoodin työkalu auttaa lähinnä turva-analyytikkoa tunnistamaan kohteen ja testaamaan sen eri haavoittuvuuksia, salasanoja, palveluja ja portteja.
- **Spyse** on kyberturvallisuushakukone, joka hakee teknisiä tietoja, joita hakkerit käyttävät yleisesti tietoverkkotutkimuksessa. Spyse tarjoaa laajaa tietoa kohteen tutkimiseen eri tulopisteiden kautta. Käyttäjä voi aloittaa yhdellä verkkotunnuksella ja laajentaa tutkintasädettä tarkistamalla erityyppisiä kohteisiin liittyviä tietoja, kuten haavoittuvuuksia, IP-osoitteita, ASN-osoitteita, DNS-tietueita, saman IP-osoitteen verkkotunnuksia, saman MX/NS-verkkotunnuksia.
- **Google Dorks** antaa tehokkaita tuloksia erinomaisella suorituskyvyllä. Tämä kyselypohjainen avoimen lähdekoodin työkalu on pääasiassa kehitetty ja luotu auttamaan käyttäjiä kohdistamaan hakemistoon tai hakutuloksiin oikein ja tehokkaasti.

- **Maltego** on suunnitellut ja kehittänyt Paterva. Se on yksi Kali Linuxin sisäänrakennetuista työkaluista. Tätä avoimen lähdekoodin älykkyystyökalua käytetään pääasiassa merkittävän etsinnän suorittamiseen eri kohteita vastaan useiden sisäänrakennettujen muunnosten avulla. Tämän työkalun käyttäminen edellyttää rekisteröitymistä. Rekisteröinti on maksutonta ja käyttäjän tulee rekisteröityä Paterva-sivustolle. Kun rekisteröinti on suoritettu, käyttäjät voivat käyttää tätä työkalua luodakseen ja kehittääkseen tehokkaita digitaalisia jalanjälkiä tietystä kohteesta Internetissä.
- **The Harvester** on työkalu, joka löytää erilaisista julkisista tiedostoista sähköpostit, aliverkkotunnukset, IP-osoitteet sekä vastaavat erilaisista julkisista tiedoista.
- **Recon-Ng**-työkalun koko teho on täysin modulaarisessa lähestymistavassa. Recon-Ng:ssä on useita sisäänrakennettuja moduuleja, joita käytetään kohdistamaan pääasiassa samalla, kun kerätään tietoja käyttäjän tarpeiden mukaan.
- **SpiderFoot** on avoimen lähdekoodin tiedustelutyökalu, joka on saatavana Linuxille ja Windowsille. Se on kehitetty Python-kielellä ja toimii käytännössä kaikilla alustoilla. Sen integroiminen on helppoa ja tehokasta vuorovaikutteisen graafisen käyttöliittymän komentoriviliittymän takia. Se on automaattisesti mahdollistanut sen, että voimme käyttää yli 100 OSINT-lähteen kyselyjä sähköpostien, nimien, IP-osoitteiden, verkkotunnusten tiedottamiseen. Se kerää laajan valikoiman tietoja kohteesta mm. verkkolohkoja, sähköpostiviestejä, verkkopalvelimia.
- **Creepy** on avoimen lähdekoodin maantieteellisen sijainnin älykkyystyökalu. Se kerää tietoja paikkatiedoista käyttämällä erilaisia sosiaalisen verkostoitumisen alustoja, jotka on jo julkaistu muualla. Creepy esittelee raportit kartalla käyttämällä hakusuodatinta tarkan sijainnin ja päivämäärän perusteella. Nämä raportit ovat saatavana CSV- tai KML-muodossa lisäanalysoitavaksi viemistä varten.

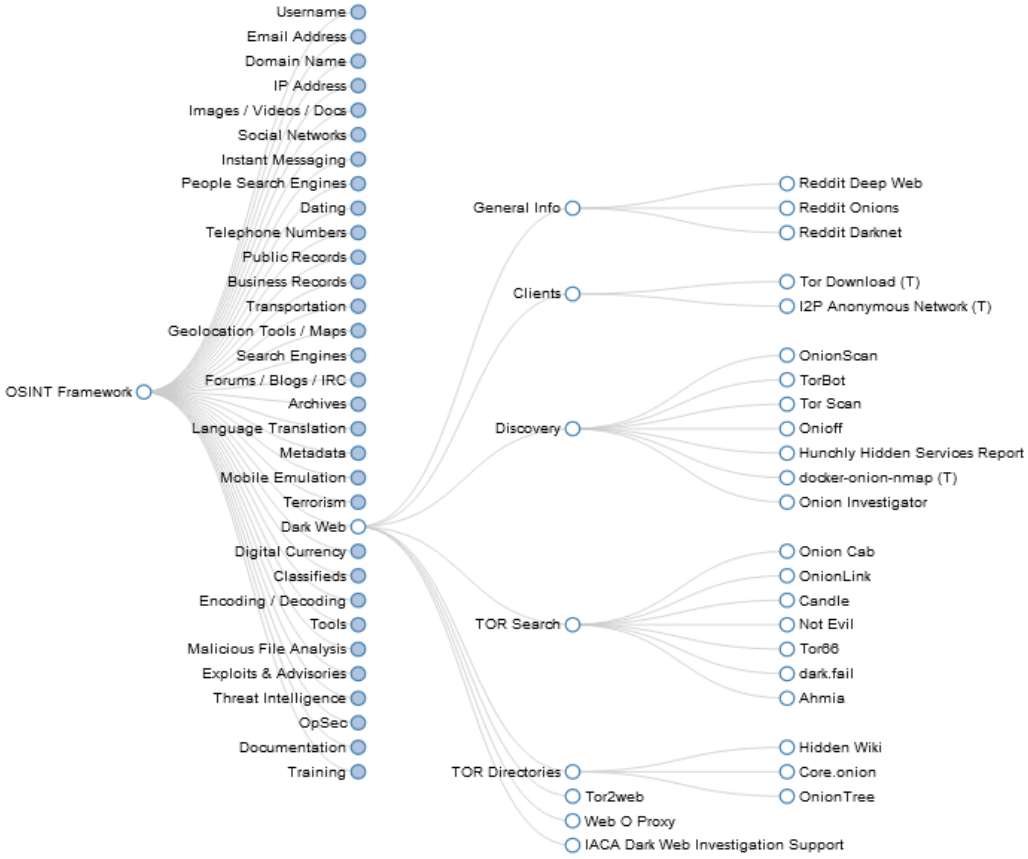
5.3 Avointen lähteiden tiedustelu pimeässä verkossa ja käytetyt työkalut

Pimeän verkon tiedusteluun on olemassa hieman eri tarkoituksiin useita eri työkaluja. Osa työkaluista on pimeän verkon hakukoneita. Github-sivustolla on lueteltu erilaisia työkaluja pimeän verkon tiedusteluun. Näitä ovat muun muassa Katana, DarkSearch ja Ahmia Search Engine. Osa työkaluista taas hankkii onion-linkkejä. Näitä ovat esimerkiksi Hunchly ja Tor66 Fresh Onions. Toiset työkalut puolestaan skannaavat näitä samoja linkkejä. Tällaisia ovat Onionscan ja Onion-nmap. Osa taas indeksoi dataa pimeästä verkosta. TorBot ja OnionIngestor ovat tällaisia työkaluja.

5.4 Millä työkalulla automatisointi onnistuisi?

Pimeän verkon tiedustelu on varsin pitkälti manuaalista työtä. Tiedustelua voi toki suorittaa pimeässä verkossa ja tähän on olemassa ihan hyviä työkalujakin. Varsinaista automaattista tiedustelua ei pimeässä verkossa voida suorittaa. Tiedustelua toki voidaan tehdä hyödyntäen

vaikkapa pimeän verkon hakukoneita, mutta tällöin tiedustelun pitäisi olla varsin hyvin kohdennettua ja tulisi olla tiedossa kohde, jota tiedustella. Työkalua, jolle voitaisiin antaa tehtäväksi etsiä esimerkiksi kaikki asekauppa, huume tai CAM-materiaalisivustot ei ole tiettävästi olemassa. Syykin tälle on selvä. Tällaisen täysin automaattisen työkalun luominen olisi erityisen haastavaa, sillä pimeässä verkossa toimitaan useilla kielillä. Lisäksi pimeän verkon henkilöiden etsintään ei ole saatavilla täsmällisiä työkaluja. Näin ollen tiedon etsintä ja analysoiminen olisi erittäin haastavaa suorittaa automaattisesti. Seuraava kuva havainnollistaa, kuinka haastavaa täysin automaattisen OSINT-tiedustelutyökalun luominen olisi.



Kuvio 4: OSINT runko. (OSINT framework 2021.)

5.4.1 Maltego

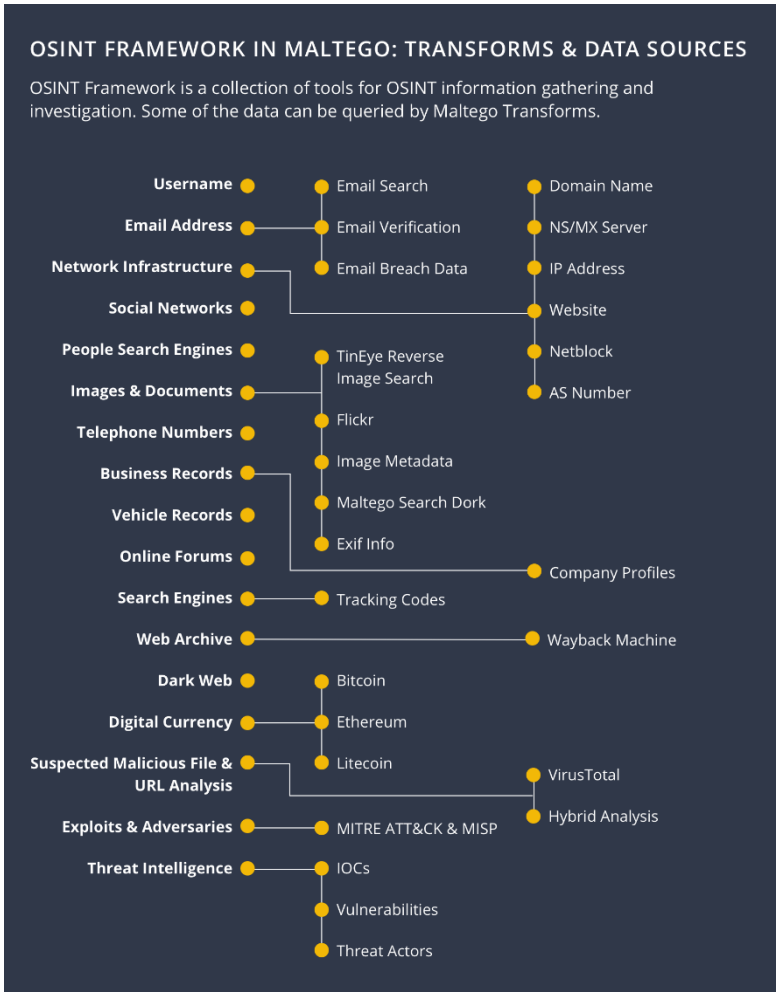
Maltego on Patervan kehittämä työkalu, joka on yleisesti käytössä kyberturvallisuusalalla. Maltego on hyvä työkalu kaivamaan tietoja syvältä kohteesta. Se käy jatkuvasti läpi internetin tietovirtaa, kunnes se löytää etsimänsä. Mikäli halutaan etsiä esimerkiksi varastettua dataa verkosta hän voi vain antaa tietyn lauseen, kuten ”varastettu data XYZ pankista.” Kun ohjelma löytää tätä vastaavan tiedon antaa se tästä ilmoituksen. Näin kuvaavat Erdal Ozkaya ja Rafiqul Islam teoksessaan Inside the dark web.

Moniin OSINT-tietolähteisiin integroitu Maltego antaa käyttäjille mahdollisuuden käyttää näitä tietoja ja kartoittaa datasuhteet kuvallisesti. (Maltego 2020.)

Maltego tarjoaa useita muunnoksia eri OSINT-tietointegraatioista. Nämä muunnokset hakevat dataa sekä aktiivisesti että passiivisesti aina infrastruktuuritiedoista uhkatietoihin, kiinnostaviin henkilöihin liittyviin tietoihin ja kryptovaluuttatoimintaan. Tämä ei kuitenkaan tarkoita sitä, että Maltegon käyttäjät olisivat vaarassa paljastaa IP-osoitteensa tai identiteettinsä suorittaessaan aktiivisia OSINT-kyselyitä. Koska kaikki Transform-kyselyt suoritetaan oletuksena Maltegon julkisen palvelimen kautta, vaikka tutkittavana oleva kohde huomaa nämä datapyyntöpyynnöt. Tällöin he näkisivät vain, että Maltego kyselee tällaisia tietoja. (Maltego 2020.)

Kun aloitetaan suorittamaan OSINT-tiedustelua, termi OSINT Framework tulee hyvin tutuksi. OSINT Framework tarkoittaa kokoelmaa työkaluja OSINT-tietojen keräämiseen ja tutkimuksiin.

Vaikka osa tiedoista isännöidään kolmannen osapuolen sivustoilla tai valtion tietokannoissa, osa tietolähteistä on integroitu Maltegoon ja Maltego Transformsiin ja niitä voi hakea. (Maltego 2020.)



Kuvio 5: OSINT framework Maltegossa. (Maltego 2020.)

OSINT kattaa monenlaisia tietoja, mikä tarkoittaa, että on myös lukuisia työkaluja ja palveluita, joita voidaan käyttää tietojen keräämiseen. Pelkästään Maltegossa käyttäjät voivat hakea kaikenlaisia tietoja Shodan-, WHOIS-, TinEye-, The Wayback Machine-, VirusTotal-, ATT & CK- ja MISP-, Pipl-, Orbis- ja muiden tietojen integroinnin ansiosta. Tämä tekee Maltegosta tehokkaan linkkien analysointityökalun eri alojen tutkijoille, jotka voivat hyödyntää Maltegoa ja sen tietointegraatioita kaikenlaisten tutkimusten suorittamiseen esimerkiksi verkon jalanjälki, kyberturvallisuustutkimus, uhka-analyysit, POI-tutkimus, petostutkimus, IoT-haavoittuvuusanalyysi. (Maltego 2020.)

Pimeän verkon rikollisuuta tutkittaessa Maltego voisi olla hyvä työkalu, etenkin rahoituksen tiedustelussa. Pimeässä verkossa rahoja siirretään yleensä Bitcoininen avulla, joten tämän valuutan seurantaan Maltego sopisi hyvin.

Toisaalta Maltego on hyvä työkalu, kun pyritään selvittämään trolliarmeijoiden toimintaa ja sidoksia. Trolliarmeijat ovat tulleet osaksi poliittista vaikuttamista etenkin vuoden 2016 Yhdysvaltojen presidentin vaalien jälkeen.

Internet-lähteiden välisten suhteiden tutkiminen voi auttaa meitä ymmärtämään, miten väärät tiedot leviävät ja kuka hyötyy niistä. Graafisena linkkien analysointityökaluna Maltego voi auttaa tutkijoita visualisoimaan ja ymmärtämään nämä verkot. (Maltego 2020.)

5.4.2 Recon-Ng

Recon-Ng on toinen varsin hyödyllinen avointen lähteiden tiedustelun työkalu. Se toimii Linux käyttöjärjestelmässä ja tulee Kali Linuxin mukana. Recon-Ng on tehty monesta eri moduulista. Recon-Ng toimii työtilan kautta, jolloin toiminta suhde tiettyyn kohteeseen luodaan yhden työtilan sisällä. Näin kuvaavat Erdal Ozkaya ja Rafiqul Islam kirjassaan Inside the dark web.

Ozkaya ja Islman mukaan Recon-Ng toimii pääosin hyödyntäen URL:ia, minne sen pitäisi hakea sisältöä analysoitavaksi. Sen vuoksi, kun käyttäjä luo työtilan tulisi tarjota verkkotunnuksia, joissa on epäilyttävää sisältöä. Erilaiset moduulit voivat ottaa talteen enemmän informaatiota verkkotunnuksesta. Työkalu voi jopa käyttää hakukoneita kuten Bing löytääkseen informaatiota kohteena olevasta verkkotunnuksesta.

5.5 Liittyykö tähän riskejä?

Kaikkeen tiedusteluun ja tiedon keräämiseen liittyy luonnollisesti riskejä. OSINT-tiedustelu ei ole poikkeus. OSINT-tiedustelun riskejä kuvataan Ntrepidcorp verkkosivulla. Niitä ovat mm. identiteetin paljastuminen, vastahyökkäykset online-vastustajilta ja virheellisen tiedon kokoaminen. Oikealla OSINT-koulutuksella voidaan tutkimuksia tehdessä välttää nämä haasteet.

Hassan ja Hijazi nostavat puolestaan esille seuraavat kolme haastetta avointen lähteiden tiedustelussa.

- Datamäärä: OSINT:n kerääminen tuottaa valtavan määrän dataa, joka on analysoitava, jotta sitä voidaan hyödyntää. Tätä tarkoitusta varten on olemassa monia automatisoituja työkaluja, ja monet hallitukset ja suuryritykset ovat kehittäneet omat tekoälytyökalunsa ja -tekniikkansa hankitun datan suodattamiseksi. Valtava tietomäärä on kuitenkin edelleen haaste OSINT-keräilijälle.

- Lähteiden luotettavuus: OSINT-lähteet, kun niitä käytetään tiedustelutietojen yhteydessä, on tarkistettava perusteellisesti luokiteltujen lähteiden avulla, ennen kuin niihin voidaan luottaa. Monet hallitukset lähettävät virheellistä tietoa OSINT-keräysprosessin harhaanjohtamiseksi.

- Ihmistyö: Pelkkää tietomäärää pidetään OSINT-kokoelman suurimpana haasteena. Ihmisten on tarkasteltava automatisoitujen työkalujen tuotosta tietääkseen, ovatko kerätyt tiedot luotettavia. Heidän on myös mahdollisesti vertailtava niitä joihinkin turvaluokiteltuihin tietoihin sen luotettavuuden ja tarkoituksenmukaisuuden varmistamiseksi. Tämä kuluttaa valtavasti aikaa ja arvokkaita henkilöresursseja.

Avointen lähteiden tiedusteluun siis liittyy riskejä ja haasteita. Nämä riskit eivät kuitenkaan ole kovinkaan suuria ja riskien haitallisuus on myös varsin matala. Automatisoidessa OSINT-tiedustelua tulee kiinnittää huomiota henkilöstön määrään ja lähteiden luotettavuuteen, samoihin seikkoihin kuin muussakin tiedustelussa.

5.6 Kun aloitat pimeän verkon tiedustelun

Yhdistyneiden kansakuntien terrorismin vastainen yksikkö kehottaa määrittämään työpaikan ja aseman asetukset, kun tiedustelu pimeässä verkossa aloitetaan. Myös seuraavat seikat on hyvä huomioida ennen tiedustelun aloittamista pimeässä verkossa: käytä Tor -verkkoa puhtaalla laiteella, käytä Tor -verkkoa virtuaalikoneen sisällä, valitse syöttöverkon solmu huolellisesti ja huomioi myös tietojen tallennusratkaisu sekä asetukset.

Lisäksi ennen kuin avataan Tor-selain, tulee kaikki muut käynnissä olevat ohjelmistot sulkea ja poistaa kaikki selaimen laajennukset käytöstä. Lisäksi tulee luoda "Uusi identiteetti" tai "Uusi Tor -piiri" aina, kun käytetään uutta .onion-linkkiä. Myöskään mitään sisältöä ei tule ladata, ellei se ole välttämätöntä. Myös valetilejä tulee käyttää.

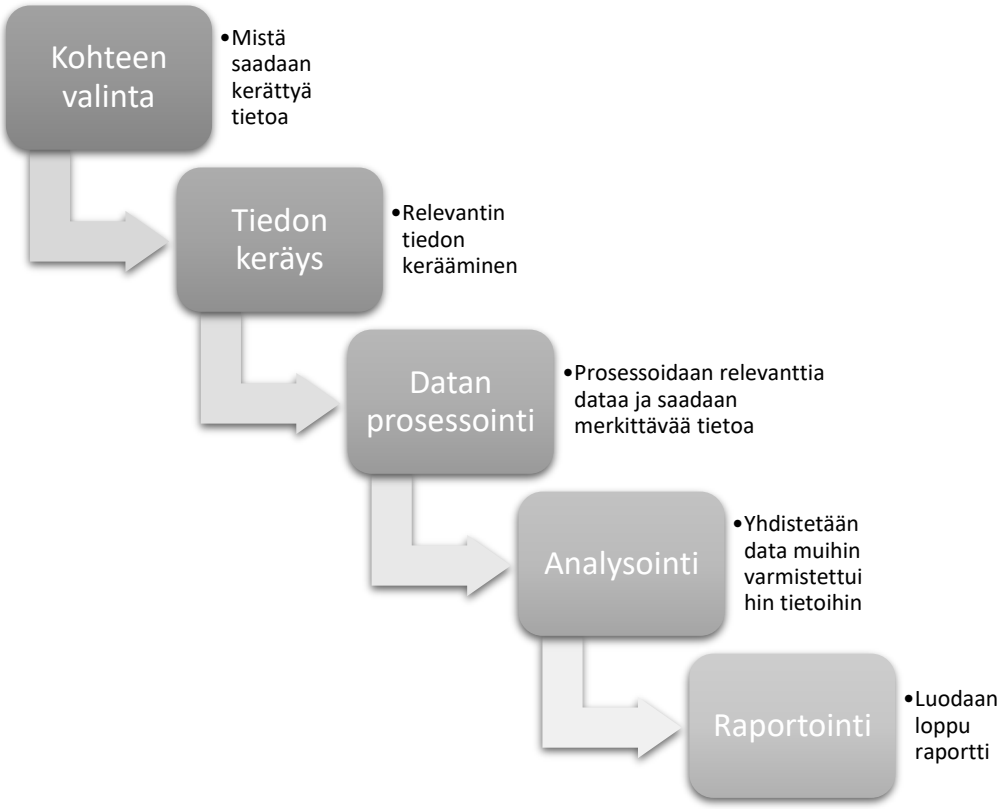
Tor -verkon suosio on ollut melko tasainen viime vuosina. Tor on ensisijainen kohde turvallisuuspalveluille, joilla on kiire tunnistaa ja hyödyntämään haavoittuvuuksia selaimissa. Tästä

syystä turvallisuuden ylläpitäminen Tor:n käytön aikana on nyt tärkeämpää kuin koskaan. (Benetis 2020.)

Kun siis aloitetaan tiedustelu OSINT:a hyödyntäen tulee alkuun ensinnäkin valita tiedusteltava kohde huolella. Halutaanko kerätä tietoja tietystä kohdehenkilöstä vai esimerkiksi jokin sivuston toiminnasta ja tietoliikenteestä. Ennen OSINT-tiedustelua on mahdollista hyödyntää muita perinteisempiä tiedustelun keinoja, kuten henkilötiedustelua.

Kun kohde on valittu, aloitetaan tiedustelu. Tiedusteluoperaatio olisi hyvä aloittaa puhtailla laitteilla, etenkin jos tiedustelua suoritetaan Tor-verkossa. Tor-verkossa suoritettavassa tiedustelussa on huomioitava se, että sivustot toimivat useilla kielillä, joten tiedustelun alussa tämä on otettava huomioon. Tiedustelussa olisi hyvä käyttää useampaa työkalua, jotka ovat erikoistuneet tietynlaisen tiedon keräämiseen. Tällaisia työkaluja voisi olla esimerkiksi Maltego ja Recon-Ng.

Nämä työkalut keräävät valtavan määrän dataa, joten tiedon analysointiin tulee varata aikaa ja resursseja. Tiedon analysoinnin jälkeen on mahdollista jatkaa tiedusteluoperaatiota toisilla tiedustelumuodoilla sekä jatkaa ja tarkentaa OSINT-tiedustelua.



Kuvio 6: OSINT-prosessi.

6 Pohdinta

Tässä opinnäytetyön viimeisessä luvussa käsittelen aluksi OSINT-tiedustelun nykyisyyttä ja tulevaisuutta. Pohdin tutkimukseni luotettavuutta ja millaisia riskejä saattaa tutkimukseni luotettavuuteen liittyä. Lopuksi pohdin jatkotutkimuskohteita: mitä seikkoja ja aiheita olisi kiinnostavaa tutkia seuraavaksi?

Tämän tutkimustyöni aikana on myös tullut esiin sellaisia näkökulmia ja seikkoja, joita en työn alkuvaiheessa osannut ajatella. Yksi tällainen seikka on esimerkiksi OSINT-tiedustelun keräämä valtava datamäärä ja se millaiset vaikutukset sillä on OSINT-tiedustelun automatisointi mahdollisuuksiin. Tämän havainnon jälkeen pyrin keräämään tietoa siitä, millaisilla työkaluilla tätä valtaisaa datamäärää olisi helpointa käsitellä. Tämä muutti hieman tutkimuksen painopistettä ja näkökulmaa.

6.1 Tutkimustulokset

Tutkimusongelma viittaa yleensä johonkin vaikeuteen, jonka tutkija kokee joko teoreettisen tai käytännön tilanteen yhteydessä ja haluaa löytää ratkaisun siihen. (Kothari 2004, 24.)

Tutkimusongelmina tässä opinnäytetyössä ovat olleet miten olisi mahdollista automatisoida OSINT-tiedustelua pimeässä verkossa ja miten CAM-materiaalin tutkinta luonnistuu pimeässä verkossa ja sen tuomat haasteet.

Tutkimuksen perusteella OSINT-tiedustelun automatisointi on mahdollista ainakin tiettyyn pisteeseen asti. Automatisoinnista huolimatta ei voida unohtaa tiedustelua suorittavan henkilöstön panosta, sillä automatisoitu OSINT-tiedustelu tuottaa suuren määrän dataa, joka tulee analysoida. Analysointiin tarvitaan henkilöstöä, joka voi ja osaa yhdistää saadun tiedustelutiedon luotettavaan ja varmistettuun dataan. Tämä luotettava ja varmistettu data on erittäin tärkeää etenkin silloin, kun tiedustelun kohteena on CAM-materiaalisivustot. Tarkkuutta tarvitaan, jos OSINT-tiedustelua suoritetaan tiedettyyn tai epäiltyyn CAM-materiaali sivustoon tai linkkiin, joka johtaa tällaiselle sivustolle.

OSINT-tiedustelua on siis mahdollista automatisoida. Automatisoidusta tiedustelusta on erittäin suuri apu, kun suoritetaan tiedustelua CAM-materiaalin parissa. Tällainen automaatio vähentää riskiä siihen, että tutkija joutuisi näkemään tai tallentamaan CAM-materiaalia. Automatisoitu tiedustelu myös vähentää rikostutkijan kuormitusta. Samalla automatisoitu OSINT-tiedustelu lisää rikollisten kiinnijäämisriskiä, etenkin kun OSINT yhdistetään muihin perinteisiin tiedustelutapoihin.

6.2 OSINT: missä olemme nyt ja mihin menemme?

OSINT ei rajoitu vain tiedustelupalveluihin, lainvalvontaviranomaisiin ja sotilasvirastoihin. OSINT:sta on tullut olennainen osa hallitusten, liikeyritysten, YK:n järjestöjen, kansalaisjärjestöjen, korkeakoulujen, tiedotusvälineiden ja kansalaisyhteiskuntien, kuten kansalaisjärjestöjen ja ammattiliittojen, päätöksentekoprosessia. Nykyään yritykset käyttävät OSINT:a sisäisten vuotojen tutkimiseen, kilpailijoiden tietojen keräämiseen ja ulkomaanmarkkinoiden suuntausten ennustamiseen. OSINT:a käyttävät myös black hat-hakkerit ja rikollisjärjestöt tutkiakseen tietoja, joita voitaisiin käyttää hyökkäykseen kohteen kimppuun. (Hassan & Hijazi 2018, 341–343.)

Informaatioaika on johtanut suureen määrään mahdollisia tiedustelulähteitä, ja se muokkaa OSINT-keräyksen tulevaisuutta. Tiedustelualueella ennustetaan, että käytäntö kerätä verkkotietoja terrorismin torjumiseksi ja rikollisuuden ratkaisemiseksi lisääntyy. Lisäksi OSINT tarjoaa edelleen halvan menetelmän tiedon hankkimiseksi kaikista yhteisöistä ympäri maailmaa. Esimerkiksi monet tutkimukset osoittavat, että länsimaiset turvallisuuspalvelut ennustivat viimeaikaisia mielenosoituksia arabimaissa analysoidessaan arabialaisten käyttäjien käyttäytymistä sosiaalisilla alustoilla tuolloin. (Hassan & Hijazi 2018, 341–343.)

Millainen on avointen lähteiden tiedustelu ja mihin suuntaan se voisi mennä? Voisiko tekoäly ja koneoppiminen olla tulevaisuudessa yhä tärkeämmässä osassa OSINT-tiedustelua?

Jos tekoäly on OSINT:n tulevaisuus, kuinka konenäkö, oppiminen, luonnollisen kielen käsittely (NLP), itsenäiset koneet ja robotiikka voivat auttaa OSINT:n kehitystä? Nämä kysymykset ovat ajankohtaisia, koska tekoäly voi olla täydellinen liittolainen OSINT-prosessien tehostamisessa, kun kyse on kyberturvallisuudesta, sotilaallisista tarkoituksista, kodista tai jopa terveydestä tai kyseen ollessa tiedustelusta, tiedonkeruusta, analysoinnista ja suurten tietomäärien suodattamisesta. (Borges 2021.)

Hallitukset ja tiedustelupalvelut käyttävät jo tekoälyä sosiaalisen keräyksen edistämiseen. Erityisesti sotilasvoimat luottavat, että tekoäly auttaa heitä menestymään taistelussa terrorismia, tietohyökkäyksiä, väärennettyä propagandaa ja kansallista turvallisuutta vastaan. (Borges 2021.)

Millaisia seurauksia ihmisten yksityisyyteen voisi olla, jos viranomaistasoa olevat OSINT-tiedustelutyökalut olisivat kaikkien käytössä? Etenkin mitä riskejä voisi olla, jos rikolliset saavat yhä tehokkaammat työkalut käyttöönsä?

Toisaalta siviilimaailmassa yritykset hyödyntävät myös OSINT-tiedustelua kaupallisten intressiensä takaamiseksi jo nyt. Tällainen OSINT-tiedustelun hyödyntäminen on positiivinen ilmiö ja kehityksen suunta.

Yritykset ovat halukkaampia kehittämään omia OSINT-kykyjään saadakseen kilpailuetua ja turvatakseen investointinsa jatkuvasti muuttuvassa maailmassa. Suuret organisaatiot pyrkivät saamaan omat OSINT-tiiminsä, kun taas kaupalliset OSINT-palveluntarjoajat tarjoavat edelleen palvelujaan pienille ja keskisuurille yrityksille, joilla ei ole varaa itsenäiseen OSINT-keräysosastoon. (Hassan & Hijazi 2018, 341–343.)

6.3 Pohdintaa tutkimuksen luotettavuudesta

Kuinka luotettavana tätä tutkimusta voidaan pitää? Toki tutkimuksen tulosta tulee tarkastella kriittisesti. Olen kuitenkin pyrkinyt käyttämään tiedossani olevaa lähdeaineistoa hyvin laajalti. Lähdeaineistoa on kerätty yrityksiltä, yhteisöiltä, julkisilta toimijoilta ja perinteisestä kirjallisuudesta. Näitä yhdistelemällä koen, että kokonaisuus on varsin luotettava. Lähdeaineistoa kerätessäni olen pyrkinyt varmistamaan selittämäni faktan myös toisesta lähteestä ja näin lisäämään tutkimuksen luotettavuutta. Toisaalta tutkimuksen objektiivisuus voi herättää kysymyksiä, kun tutkimusta on suorittanut vain yksi henkilö. Vaikka lähteiden luotettavuutta on pyritty tarkastelemaan kriittisesti ja selvittämään sama fakta myös toisesta lähteestä, voidaanko silti olla täysin varmoja, että käsitelty asia ja faktana tarjottu asia olisi absoluuttinen totuus. Tästäkin huolimatta koen, että tutkimukseni validi ja reliabiliteetti ovat hyvällä tasolla.

6.4 Jatkotutkimuskohteet

Avointen lähteiden tiedustelu (OSINT) on siis hyvin yleinen ja käytetty tiedustelutapa. OSINT-tiedustelun merkitys on viime vuosien aikana vain lisääntynyt, kun yhä suurempi osa ihmisten elämästä ja toiminnasta on siirtynyt verkkoon. OSINT-tiedustelun merkitys on kasvanut myös siksi, että yhä suurempi osa rikollisuudesta on siirtynyt verkkoon.

Esimerkiksi huumekauppaa hoidetaan nykyään pimeässä verkossa. Lainvalvojien tulee keskittyä tulevaisuudessa yhä enemmän OSINT-tiedusteluun unohtamatta kuitenkaan muita tiedustelumuotoja. Muut perinteisemmät tiedustelun muodot tukevat ja ohjaavat OSINT-tiedustelua. OSINT-tiedustelun heikkoutena voidaan nähdä sen keräämä valtava datamäärä. Tästä kerätystä datamäärästä voi olla hyvin haasteellista ja työlästä löytää oikeita ja keskeisiä tietoja. Tämän tiedon ja tiedustelun rajaamiseen on hyvä hyödyntää perinteisempiä tiedustelun muotoja. Henkilötiedustelun ja signaalitiedustelun avulla saadaan jo merkittävästi rajattua OSINT-tiedustelua. Toisaalta OSINT-tiedustelu saattaa puolestaan ohjata esimerkiksi henkilötiedustelua haluttuun ja oikeaan suuntaan. Suoritettaessa vaikka CAM-materiaaliin liittyvää tiedustelua pimeässä verkossa, voidaan saada sellaista tietoa, jonka perusteella voidaan aloittaa tai kohdentaa henkilötiedustelua paremmin epäiltyyn.

Verkossa suoritettava OSINT-tiedustelu tuottaa siis valtavasti dataa. Tämä valtaisa datamäärä luokin suurimmat haasteet ja riskit OSINT-tiedustelulle. Kuinka kohdentaa tiedustelua juuri

oikein? Valtavasta datamäärästä tiedon hakeminen on kuin etsisi neulaa heinäsuovasta. Lisäksi riskinä voi olla tarkoituksella OSINT-tiedustelulle syötetty väärä tieto. Miten tämän tiedon voi erottaa oikeasta?

Automaatio on merkittävässä osassa nykypäivän OSINT-tiedustelua. Ilman tarkkoja parametrejä automattiset työkalut saattavat kuitenkin lisätä analysoitavaa datan määrää. Automaatiotyökalut ovat kuitenkin hyvin tarpeellisia ja jopa välttämättömiä, sillä verkossa suoritettavaa OSINT-tiedustelua on käytännössä mahdotonta suorittaa käsin varsinkaan kovin tehokkaasti. Automaatiossa on haasteita. Ei pelkästään suuren kerätyn datamäärän vuoksi, vaan myös siksi, että tehokas tiedustelu pimeässä verkossa on haastavaa. Tämä haaste muodostuu pimeän verkon rakenteesta ja käyttötarkoituksesta. Pimeässä verkossa on erittäin suuri määrä eri kielillä toteutettuja sivustoja. Tiedustelun kohdentaminen on siis varsin haastavaa, yhdellä englanninkielisessä hakuehdolla ei voi löytää läheskään kaikkea tarpeellista dataa.

Jatkossa olisi mielenkiintoista tutkia OSINT-tiedusteluun suunniteltuja työkaluja, joita viranomaiset käyttävät. Nämä työkalut ovat oletettavasti paljon hienostuneempia ja kehittyneempiä kuin yleisesti käytössä olevat ja mainostetut OSINT-työkalut. Lisäksi olisi kiinnostavaa nähdä tai tehdä tutkimusta pimeässä verkossa olevista kielistä ja siitä millainen jakauma kielten välillä pimeässä verkossa on. Tämä voisi antaa näkemystä siitä, minkä valtion rikollisuus on keskittynyt eniten verkkoon. Toisaalta tässä olisi huomioitava se, että vain noin 45 % pimeän verkon sivuista sisältää laitonta materiaalia.

Lähteet

Tutkimuskirjallisuus

Painetut

Ozkaya, E & Islam, R. 2019. Viitattu 9.9.2021. Inside the dark web. CRC press Taylor & Francis Group.

Sähköiset

Benetis, V. 2020. Viitattu 3.9.2021. https://www.itu.int/en/ITU-D/Cybersecurity/Documents/CyberDrill-2020/How%20to%20conduct%20effective%20OSINT%20investigation%20online.pdf

Colquhoun, C. 2016. Viitattu 3.9.2021. https://www.bellingcat.com/resources/articles/2016/07/14/a-brief-history-of-open-source-intelligence/

Hassan, N & Hijazi, R. 2018. Viitattu 15.9.2021. Open Source Intelligence Methods and Tools: A Practical Guide to Online Intelligence. E-kirja. Apress L. P.

Jyväskylän yliopisto. 2015. Viitattu 16.9.2021. https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/tutkimusstrategiat/teoreettinen-tutkimus

Jyväskylän yliopisto. 2015. Viitattu 16.9.2021. https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/tutkimusprosessi/tutkimuksen-toteuttaminen

Kothari, C. 2004. Viitattu 21.9.2021. *Research Methodology: Methods and Techniques*. E-kirja. New Age International Ltd.

Kumar, A & Rosenbach, E. 2019. Viitattu 31.8.2021.https://www.imf.org/external/pubs/ft/fandd/2019/09/the-truth-about-the-dark-web-kumar.htm

Layton, R & Paul A. Watters. 2015. Viitattu 15.9.2021. *Automating Open Source Intelligence: Algorithms for OSINT*. E-kirja. Elsevier Science & Technology Books.

Office of the Director of National Intelligence. 2021. Viitattu 5.9.2021. https://www.dni.gov/index.php/what-we-do/what-is-intelligence

Retzkin, S. 2018. Viitattu 8.9.2021. *Hands-On Dark Web Analysis: Learn What Goes on in the Dark Web, and How to Work with It*, E-kirja. PAIKKA. Packt Publishing, Limited.

Williams, H & Blum, I. 2018. Defining Second Generation Open Source Intelligence (OSINT). RAND CORPORATION. Viitattu 10.9.2021. https://www.rand.org/content/dam/rand/pubs/research_reports/RR1900/RR1964/RAND_RR1964.pdf

Wilson, E & Gollnick, C. 2017. Separating fact from fiction the truth about the dark web. Viitattu 13.9.2021. https://www.cyberdefensemagazine.com/annual-editions/RSA-2017/mobile/index.html#p=42

Yin, R. 2015. Viitattu 21.9.2021. *Qualitative Research from Start to Finish. E-kirja.* Guilford Publications.

Tutkimusaineisto

Sähköiset

Borges, E. 2021. Viitattu 3.9.2021. https://securitytrails.com/blog/what-is-osint-how-can-i-make-use-of-it

DeNicola, L. 2021. Viitattu 4.9.2021. https://www.experian.com/blogs/ask-experian/what-is-the-dark-web/

Digital. 2021. Viitattu 1.9.2021. https://digital.com/online-privacy/deep-dark-web/

Github. 2021. Viitattu 8.9.2021. https://github.com/apurvsinghgautam/dark-web-osint-tools

Information security newspaper. 2021. Viitattu 17.9.2021. https://www.securitynewspaper.com/2017/06/01/dark-web-users-child-porn-website-tracked-visiting-file-sharing-site/

Iosi. 2020. Viitattu 16.9.2021. https://www.iosi.global/child-sexual-abuse/

James, L. 2018. Viitattu 8.9.2021. https://www.makeuseof.com/tag/tor-browser-safety-tips/

Maltego. 2020. Viitattu 7.9.2021. https://www.maltego.com/blog/what-is-open-source-intelligence-and-how-to-conduct-osint-investigations/

Viljakainen, M. 2019. Viitattu 16.9.2021. https://www.is.fi/kotimaa/art-2000006069404.html

Ntrepidcorp. 2021. Viitattu 10.9.2021. https://ntrepidcorp.com/case-studies/osint-investigations-the-benefits-and-risks/

OSINT framework. 2021. Viitattu 2.9.2021. https://osintframework.com/

Osint-ltd. 2021. Viitattu 8.9.2021. https://www.osint-ltd.com/en/code-of-ethics/

Slide to doc. 2021. Viitattu 2.9.2021. https://slidetodoc.com/web-shell-attacks-dr-bhawana-rudra-national-institute/

STT. 2019. Viitattu 16.9.2021. https://www.sss.fi/2019/01/lapsipornon-vaihto-vertaisverkossa-kertoo-siita-etta-myos-suomessa-tuotetaan-materiaalia-ja-jopa-live-striimeja-hyvaksikaytoista/

Techjournalist. 2020. Viitattu 14.9.2021 https://techjournalism.medium.com/modern-espionages-lesson-for-open-source-investigations-4d596c243217

Kuviot