SAVONIA

OPINNÄYTETYÖ - **AMMATTIKORKEAKOULUTUTKINTO**
TEKNIIKAN JA LIIKENTEEN ALA

AVOINTEN LÄHTEIDEN TIEDUSTELU JA HEN-KILÖPROFILOINTI

TEKIJÄ/T: Eemeli Tolppanen

SAVONIA-AMMATTIKORKEAKOULU

OPINNÄYTETYÖ

Tiivistelmä

Koulutusala Tekniikan ja liikenteen ala	
Koulutusohjelma/Tutkinto-ohjelma Sähkötekniikan koulutusohjelma	
Työn tekijä(t) Eemeli Tolppanen	
Työn nimi Avointen lähteiden tiedustelu ja henkilöprofilointi	
Päiväys 29.05.2020	Sivumäärä/Liitteet 70 / 1
Ohjaaja(t) Arto Toppinen	
Toimeksiantaja/Yhteistyökumppani(t) Nixu Oyj / Antti Niemelä	

Tiivistelmä

Sosiaalinen media on tuonut meille reaaliaikaisen tiedon välityksen sekä valtavien tietomäärien virtauksen maailmanlaajuisesti. Sosiaalista mediaa ja muita viestintäkanavia käyttävät yksityiset henkilöt ja yritykset. Kaikki henkilökohtainen data, mitä ihmiset jakavat sähköisiin palveluihin on julkisesti kaikkien saatavilla. Avoimiin lähteisiin perustuvalla tiedustelulla OSINT:illa (Open Source Intelligence) pystytään profiloimaan verkosta löytyvällä datalla henkilöiden ja yritysten haavoittuvuuksia. Tämä on ensimmäinen aste henkilön- tai organisaation tietoturvahyökkäyksessä. Näiden seurauksena OSINT on noussut hyvin merkittävään roolin nyky-yhteiskunnassa.

Opinnäytetyön tavoitteena oli tehdä kehittämistyö tilaajayritys Nixu Oyj:lle. Kehittämistyössä pyrittiin kehittämään ja automatisoimaan henkilöprofilointia, sekä perehtymään avointen lähteiden tiedustelussa käytettyihin toimintatapoihin, ominaisuuksiin ja työkaluihin. Tilaajayritys hyödyntää avointen lähteiden tiedustelua osana testauspalvelua. Avointen lähteiden tiedustelua hyödynnetään muun muassa murtotestauksessa ja Red Teaming-kampanjoissa, joissa simuloidaan hyökkäystilanne kohteeseen. Näissä pyritään löytämään haavoittuvuusreittejä, eli hyökkäysvektoreita muun muassa ihmisiin, organisaatioihin, prosesseihin, alihankkijoihin ja teknologiaan.

Työssä perehdyttiin avointen lähteiden tiedustelun teoreettiseen puoleen analysoiden, kuinka se on ajan saatossa kehittynyt nykypäivän moderniksi tiedusteluluokaksi. Teoriaosuudessa pohdittiin myös mitä merkittäviä haasteita ja eettisyyskysymyksiä avointen lähteiden tiedustelu pitää sisällään. Opinnäytetyössä tarkasteltiin myös tiedustelussa käytettävien tiedon keräys- ja analysointimenetelmiä, pohtien tietojen keräyksen merkitystä eri Internetin tasoilla. Opinnäytetyön tutkimuskysymyksessä tarkasteltiin, kuinka henkilöiden profiloiminen suoritetaan ja voidaanko tiedon etsintää ja analysointia automatisoida ja kehittää jollain tasolla.

Lopputuloksena luotiin kattava kokonaisuus avointen lähteiden tiedustelusta ja siinä hyödynnetyistä työkaluista. Suoritetussa henkilöprofiloinnissa perehdyttiin tutkimuskysymykseen ja saatuja tuloksia on havainnollistettu anonyymisti. Tulokset koostuivat henkilöprofiloimisessa hyödynnetyistä käytännöistä, tiedon etsintätyökaluista, sekä havainnollistettiin tiedon etsinnällisesti merkittävimpiä löydöksiä.

Avainsanat

OSINT, avointen lähteiden tiedustelu, tiedustelu, henkilöiden profiloiminen

SAVONIA UNIVERISTY OF APPLIED SCIENCES

THESIS

Abstract

Field of Study Technology, Communication and Transport			
Degree Programme Degree Programme in Electrical Engineering			
Author(s) Eemeli Tolppanen			
Title of Thesis Open source intelligence and personal profiling			
Date	29 May 2020	Pages/Appendices	70 / 1
Client Organisation /Partners Nixu Plc			

Abstract

The purpose of this thesis was to do a development project for the client company Nixu Plc. The aim of this development was to improve and automate personal profiling as well as to get acquainted with methods and features of the open source intelligence (OSINT). The thesis followed the implementation plan which consists of examining, executing and analysing the procedure.

The thesis explored the theoretical aspects of open source intelligence, analyzing how it has evolved over time into today's modern intelligence class. The theoretical part also reflects what significant challenges and ethical issues open source intelligence encounters. The methods of gathering and analyzing information in this kind of reconnaissance were studied considering the importance of data collection at different levels of the Internet. The research question of the thesis considered how profiling of individuals is performed and whether information gathering, and analysis can be automated and improved at some level.

As a result of this thesis, an extensive description of open source intelligence and use of selected data gathering and analyzing tools was created. The results of the research question focusing on personal profiling were illustrated anonymously. The results were obtained, and they consisted of the practices used to profile individuals, information gathering tools and the most significant findings in terms of information gathering.

Keywords

OSINT, open source intelligence, reconnaissance, profiling of individuals

ESIPUHE

Opinnäytetyö on toteutettu tilaajayritykselle Nixu Oyj:lle Espoossa kevään aikana vuonna 2020. Haluan kiittää Nixu Oyj:tä mahdollisuudesta toteuttaa opinnäytetyö heille. Haluan kiittää erityisesti työnohjaajaa Antti Niemelää ja esimiestäni Tommi Vihermaata määrätietoisesta ja opettavaisesta ohjaamisesta. Lopuksi haluan kiittää puolisoani annetusta tuesta ja kannustuksesta opinnäytetyön aikana, joka mahdollisti opinnäytetyön etenemisen hyvinä ja vaikeina aikoina.

Espoossa 29.05.2020

Eemeli Tolppanen

SISÄLTÖ

1 JOHDANTO

Sosiaalinen media ja muut Internetin tuomat palvelut ovat helpottaneet ja lisänneet ihmisten tiedon jakamista, mutta suurella osalla sosiaalisen median käyttäjistä ei ole tietoa tietojen jakamisen haittapuolista. Ihmiset jakavat arviolta 682 miljoonaa Twiittia, 67 miljoonaa Instagram julkaisua ja jopa 4,3 miljardia Facebook julkaisua päivittäin (Schultz 2019). Ymmärtävätkö sosiaalisen median käyttäjät, mitä heidän julkaisemillaan tiedoilla voidaan tehdä ja miten paljon informaatiota heistä voi kerätä?

Digitalisaation kehityksen seurauksena Internetin käyttö on kasvanut räjähdysmäisesti. Tämän vuoden ensimmäisellä kvartaalilla globaali digitaalinen väestö, eli henkilöt, jotka ovat yhteydessä Internettiin on saavuttanut 4,54 miljardin käyttäjän rajan. Määrän arvioidaan kasvavan vuoteen 2022 mennessä kuuteen miljardiin käyttäjään. (Statista 2020; Morgan 2019.) Kyseisestä väestöstä 3,8 miljardia ihmistä käyttää jonkinlaista sosiaalisenmedian alustaa, jolla pystytään viestimään reaaliajassa maailmanlaajuisesti. Alustoilta, kuten Facebook, YouTube, Instagram, Snapchat ja Twitter, ihmiset pystyvät jakamaan informaatiota julkisesti ja ilmaiseksi. (Datareportal s. a.) Tämän vuoksi sosiaalisen median suosio kasvaa koko ajan.

Avointen lähteiden tiedustelulla tarkoitetaan tiedon keräystä julkisista lähteistä, sekä niiden analysointia ja hyödyntämistä tiedustelutarkoituksiin (Oikarinen 2020). Avointen lähteiden tiedustelulla kerättyjä tietoja voidaan hyödyntää muun muassa taloudellisissa tutkimuksissa, rikollisen toiminnan ehkäisemisessä, liiketoimintakilpailijoiden analysoinnissa ja henkilöiden ja yrityksien tiedustelutietojen hankkimisessa (Hassan ja Hijazi 2018, xix). Avointen lähteiden tiedustelun avulla voidaan löytää esimerkiksi henkilötietoja, puhelinnumeroita tai muita merkittäviä tietoja, joita kohde on jakanut julkisiin lähteisiin. Saadut tiedot voivat toimia muun muassa pohjatietona sosiaalisessa manipuloinnissa. (Cherkasets 2019.)

Sosiaalinen manipulointi on yksi yleisimmistä nettipetoksista, jonka avulla manipuloidaan käyttäjää luovuttamaan luottamuksellisia tietoja, kuten esimerkiksi käyttäjätunnuksia, salasanoja, pankkitietoja ja henkilöllisyystietoja (Webroot s. a.). Avointen lähteiden tiedustelulla saatujen pohjatietojen avulla luodaan luottamus kohteeseen, joka mahdollistaa käyttäjän luovuttamaan arkaluontoisia tietoja. Maailman laajuisesti vuodesta 2013 alkaen on varastettu arviolta noin 10 miljoonaa identiteettiä (Sobers 2020). Eikä Suomikaan ole suojassa tältä haitalliselta toiminnalta. Suomalaista yli puoli miljoonaa henkilöä on joutunut identiteettivarkauden yrityksen kohteeksi menneen kahdentoistakuukauden aikana. Näistä noin 45 000 suomalaisen identiteetti varastettiin onnistuneesti. Varkauksista on seurannut uhreille merkittäviä taloudellisia haittoja. (mySafety s. a.)

Avointen lähteiden tiedustelua on hyödynnetty jo ensimmäisestä maailmansodasta asti, jolloin tiedustelu pohjautui konkreettisiin lähteisiin, kuten sanomalehtiin ja radiolähetyksiin (Colquhoun 2016). Ajan kuluessa teknologian kehityksen seurauksena saatavilla olevien avointen tietojen määrä on kasvanut ja tietojen keräys metodit kehittyneet. Tietojen keräyksessä hyödynnetään nykypäivänä automatisoituja ohjelmistoja, eli työkaluja, jotka on ohjelmoitu etsimään tiettyjen kriteerien avulla mahdollisimman paljon tietoa kohteesta automatisoidusti. Työkalut pystyvät keräämään valtavasti tietoa eri lähteistä muodostaen suuren tietomäärän, joka täytyy analysoida luotettavaksi. (Passi 2018). Automaation tuoman nopeuden ja kattavien tietomäärien seurauksena on arvioitu, että 90 prosenttia tiedustelupalveluiden hankkimasta hyödyllisistä tiedoista pohjautuu avoimista julkisista lähteistä kerättyyn tietoon (Hassan ja Hijazi 2018, xix).

Opinnäytetyön tavoitteena on tuoda esiin avointen lähteiden tiedustelun teoreettinen puoli ja pyrkiä sen pohjalta kehittämään ja automatisoimaan henkilöprofilointia. Teoria osuudessa käsitellään avointen lähteiden tiedustelun merkittävyyttä nykypäivänä sekä kuinka se on mullistunut ajan saatossa yhdeksi merkittävimmäksi tiedustelu luokaksi. Opinnäytetyössä havainnollistetaan avointen lähteiden käyttäjiä ja heidän motiivejansa hyödyntää OSINT-tiedustelua, sekä mitä merkittäviä haasteita ja eettisyys kysymyksiä avointen lähteiden tiedustelu pitää sisällään. Opinnäytetyössä esitellään valittuja avointen lähteiden tiedustelussa käytettäviä käyttöjärjestelmiä sekä työkaluja eri Internetin tasoilla. Käsiteltävien aiheiden perusteella perehdytään tutkimuskysymykseen, kuinka automatisoida ja kehittää henkilöprofilointia avointen lähteiden tiedustelun näkökulmasta. Opinnäytetyössä käsiteltävien aiheiden perusteella luodaan yhteenveto, joka tiivistää opinnäytetyön yhdeksi kokonaisuudeksi.

2 AVOINTEN LÄHTEIDEN TIEDUSTELUN MÄÄRITELMÄ JA HISTORIA

OSINT (Open Source Intelligence), eli avointen lähteiden tiedustelulla tarkoitetaan julkisista lähteistä kerättyä tietoa. OSINT nimen osa 'OS', kuvaa avointa lähdettä, jolla tarkoitetaan tietolähdettä. Tällaisia tietolähteitä on esimerkiksi sanomalehdet, sosiaalisen median julkaisut, blogit tai kuvat, kunhan ne ovat vain julkisia, avoimia ja laillisia. Kerättyjä tietoja voidaan hyödyntää tietyn päämäärän saavuttamiseen, olkoon se henkilönprofiloinnin tietojen saanti, verkkoturvallisuuden kartoittaminen tai yrityksen taloudellisten suuntautumisten selvittäminen. (SecurityTrails Team 2018 a.)

> *"Avointen lähteiden tiedustelu (OSINT) on tiedustelua, jonka tiedot saadaan julkisista saatavilla olevista lähteistä ja kerätään, hyödynnetään ja levitetään oikeaan aikaan sopivalle yleisölle tietyn tiedustelutarpeen toteuttamiseksi"*
> (U.S. Government Printing Office 2006, Osa. 931.).

Ensimmäiset konkreettiset havainnot avointen lähteiden tiedustelusta johtaa juurensa jo 1800-luvun loppupuolelle vuoteen 1889, jolloin perustettiin Preussin Saksan sotilaallinen tiedusteluosa IIIb. Ensimmäisen maailmansodan aikana tiedusteluosa IIIb kehittyi hybridi turvallisuusjärjestöksi, joka hoiti tiedustelua ja vastatoimintaa, lehdistö- ja postin sensuuria, sekä liittoutuneiden ja puolueettomien maiden valvontaa ja sotapropagandaa. Osaston toimintatavat perustuivat pääosin avointen lähteiden tiedusteluun. Sodan aikana ihmisläheinen tiedustelu HUMINT uudelleen suuntautui klassisesta vakoilusta, sotavankien organisoituun kuulusteluun ja teknologian kehityksen panostamiseen. Kehitys tapahtui etenkin ilmakuvien ja signaalitiedustelun osalta. Tämän seurauksena avointen lähteiden tiedustelun rooli heikkeni ja jäi muiden tiedusteluluokkien varjoon. (Pöhlmann 2017.)

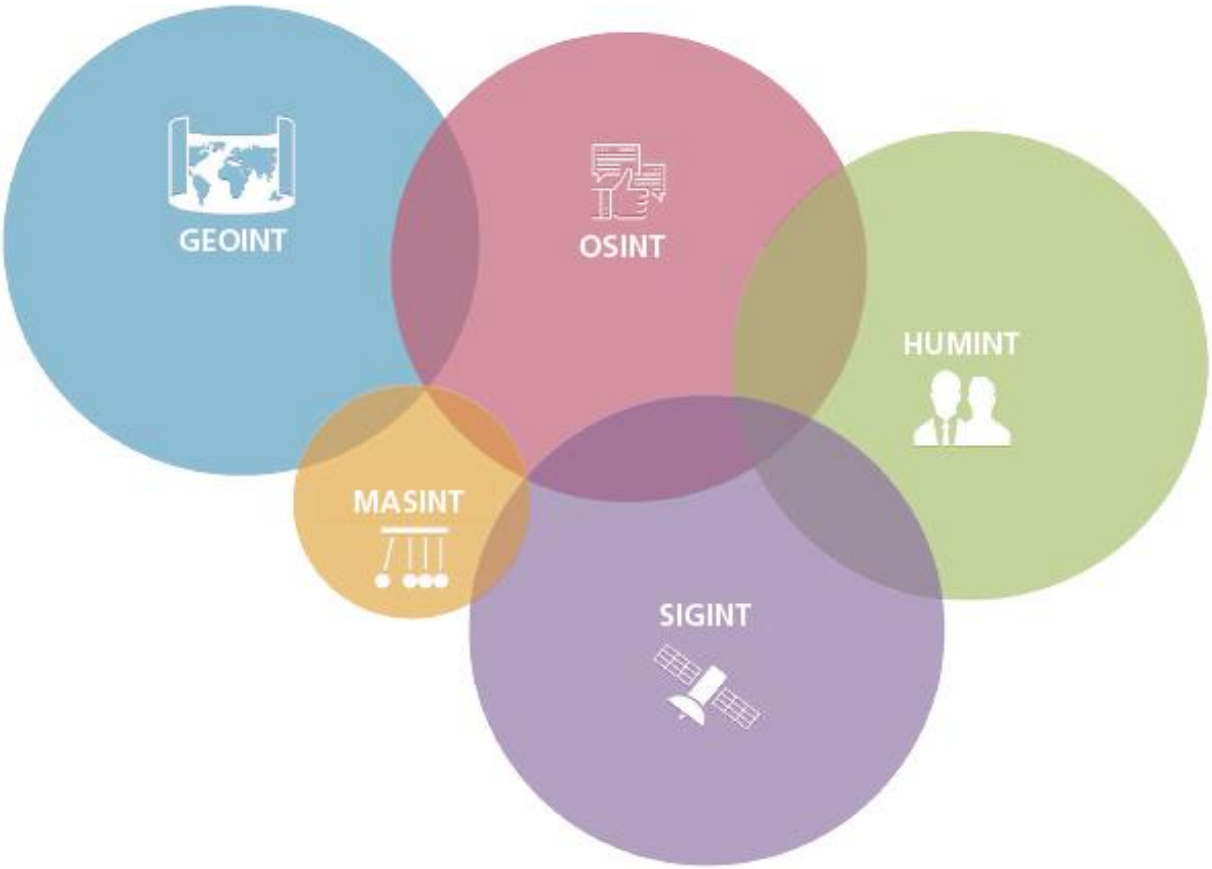
Avointen lähteiden tiedustelu heräsi uudelleen toisen maailmansodan aikana, jolloin William Donavan, johti tiedustelupalvelua Yhdysvalloissa. Pearl Harbor iskun jälkeen erilaisen tiedustelun tarve havaittiin tarpeelliseksi. Tämän seurauksena Donavanin osasto nimitettiin OSS:äksi (Office of Strategic Services) CIA:n edeltäjäksi. Tähän järjestöön kuului haara, joka oli omistettu avointen lähteiden tiedustelulle. Tämä osasto keräsi huolellisesti tietoa vihollisista saatavilla olevista avoimista lähteistä. Lähteinä toimivat muun muassa sanomalehdet, radiolähetykset, valokuvat ja artikkelit. Näiden tietojen perusteella pystyttiin muodostamaan tärkeitä tietoja vihollisista ja heidän suuntautumisistaan. (Colquhoun 2016.)

Ajan kuluessa avointen lähteiden tiedustelun merkittävyyttä alettiin taas kyseenalaistamaan tiedustelunpiirissä, sillä sen tuomat tiedustelu tulokset eivät olleet niin huomattavia ihmisläheiseen tiedusteluun verrattuna. Tämän seurauksena OSINT hukkui taas muiden tiedustelu haarojen alle, kunnes vuonna 2009 Iranissa puhkesi "Vihreä vallankumous", jonka agendana oli protestoida Iranin hallitusta vastaan. Protestoijat käyttivät hyväkseen juuri julkaistua 3G-teknologiaa ja älypuhelimia, jakaakseen omaa agendaansa ja koordinoidakseen omaa toimintaansa. Tämä oli ensimmäinen kerta historiassa, jolloin yksilöt pystyivät jakamaan julkisesti reaaliajassa sisältöä maan sisäisistä poliittisista suuntautumisistaan. Kaikki tieto ja sisältö jaettiin muutamiin käytettävissä oleviin sovelluksiin muun muassa Twitteriin ja Facebookiin, joiden sisältö oli maailmanlaajuisesti kaikkien saatavilla ja valmiina analysoitavaksi. Tämä on yksi merkittävimmistä tapahtumista avointen lähteiden tiedustelun historiassa. Tämä merkittävä tapahtuma noteerattiin myös maailmanlaajuisesti, jonka seurauksena moderni avointen lähteiden tiedustelu syntyi. Hallitukset, lehdistöt ja jopa rikollisorganisaatiot ymmärsivät avointen lähteiden merkittävyyden nyky-yhteiskunnassa ja ovat sen myötä käyttäneet sen tarjoamia tietoja hyväkseen. (Colquhoun 2016.)

Tämän myötä OSINT on noussut vertaiseksi muiden tiedusteluluokkien kanssa. OSINT:in lisäksi tiedusteluluokkiin kuuluu myös IMINT (Imagery Intelligence; kuvatiedustelu), MASINT (Measurement and Signature Intelligence; mittaus- ja tunnusmerkkitiedustelu), SIGINT (Signal Intelligence; signaalitiedustelu), HUMINT (Human Intelligence; henkilötiedustelu) ja GEOINT (Geospatial Intelligence; sijaintitiedustelu). (ODNI s. a.; Kuva 1.)

Teknologian kehityksen myötä OSINT-tiedustelu on alkanut muistuttamaan entistä enemmän muita tiedusteluluokkia. Esimerkiksi GEOINT:in tuoma satelliitti kuvantaminen on palvellut vain pääsääntöisesti armeijaa ja valtiollisia toimijoita, mutta kaupallisten satelliittien tulemisen jälkeen satelliittikuvien hyödyntäminen on tullut myös mahdolliseksi julkisille tahoille. Sosiaalisen median jaettavien tietojen avulla pystytään profiloimaan henkilöitä ja tämän kautta samaan tietoa heidän yhteyksistään ja toimintatavoistaan. Tätä voidaan verrata HUMINT tiedusteluun, jossa kerätään samankaltaista tietoa, mutta ihmislähteistä. Puolestaan profiloimisen tuomat valtavat tietomäärät alkavat muistuttamaan SIGINT-tiedustelua, eli elektronisen tiedon keräämistä ja analysointia. Tietojen analysoinnilla pyritään seulomaan saaduista tiedoista kaikista tärkeimmät asian haarat luotettavamman lopputuloksen takaamiseksi. (Williams, Blum 2018, 7-8.)

Tiedustelualat ovat siis pikkuhiljaa alkaneet lähentymään toisiaan, ja alojen väliset rajat ovat alkaneet hämärtyä. OSINT on tässä merkittävimmässä roolissa, sillä sen tuomat suuret tietomäärät ujuttautuvat jokaiseen edellä mainittujen tiedustelujen piiriin. Vaikka tiedustelualat ovat alkaneet sekoittua, pohjimmiltaan OSINT-tiedustelu eroaa muista tiedustelualoista sen tavasta hyödyntää vain laillisesti saatavilla olevia lähteitä rikkomatta tekijäoikeuksia tai yksityisyyttä koskevia lakeja.



Kuva 1. Tiedustelualojen hämärtyneet rajat (Williams, Blum 2018, 9.)

2.1 Moderni avointen lähteiden tiedustelu

Nykypäivänä avointen lähteiden tiedustelun tiedon saanti perustuu verkosta löytyvään tietoon. Kaikki julkiset tiedot kuuluvat OSINT-tiedustelun piirin, olivatpa ne sitten kirjoja tai raportteja julkisessa kirjastossa, artikkeleita sanomalehdessä tai lausuntoja lehdistötilaisuudessa (SentinelOne 2019). Näin ollen kuka vain pystyy etsimään tietoa pääsääntöisesti kenestä vain, kun kohde, henkilö tai yritys on julkaissut tietojaan julkisiin lähteisiin.

Avointen lähteiden tiedustelun kohteen profiilin rakentaminen aloitetaan kuvan mukaisesti (Kuva 2). Lähteiden määrittämisellä tarkoitetaan tietolähteiden keräämistä, joista voidaan löytää tietoa kohteesta. Tämä määrittää myös hyvin pitkälti kuinka laaja suoritettava tiedustelu tulee olemaan. Kohteena voi toimia yritys, organisaatio tai henkilö, johonka tiedustelu kohdistetaan. (z3roTrust 2018.) Yritykseen ja organisaatioon kohdistuvassa tiedustelussa pyritään keräämään tietoa muun muassa kohteen taloustiedoista, tietoverkon infrastruktuurista, henkilöstöstä, sekä tehdyistä testauspöytäkirjoista, joita voi esimerkiksi olla tietoturvatestaus raportit (Kinzie 2020). Henkilöiden profilointia, voidaan myös tehdä yrityksiin ja organisaatioihin kohdistuneessa tiedustelussa, jos se on vain määritelty käsiteltäväksi. Tämän kaltaisessa tiedustelussa kartoitetaan muun muassa kohteen henkilökohtaisiatietoja, elintapoja, koulutusta, ammatteja, taloudellista tilannetta, vapaa-ajan toimintaa ja verkkotunnusta (Cherkasets 2019). Tämän jälkeen aloitetaan tiedonkeräys määritellyistä lähteistä ja keräyksen aikana löydettyjen lähteiden perusteella. Kerätyt tiedot prosessoidaan, eli seulotaan saaduista tiedoista kaikki hyödylliset tiedot, jonka jälkeen tiedot on helpompi analysoida ja raportoida luotettavaksi sekä hyödylliseksi. (z3roTrust 2018.)

KUVA 2. Avointen lähteiden tiedustelun prosessi.

Avointen lähteiden tiedustelu sisältää kaikki julkisesti saatavilla olevat tietolähteet, joita voidaan etsiä verkosta tai konkreettisista lähteistä. Kuten Hassan ja Hijazi (2018, 5) kuvaavat kirjassaan, kuinka lähteet voidaan jakaa viiteen eri ryhmään:

1. Perinteiset joukkotiedotusvälineet televisio, radio, uutislehdet, kirjat ja lehdet.
2. Kuvat ja videot, jotka sisältävät metadataa.
3. Internet, joka sisältää foorumit, blogit, sosiaalisen verkostoitumisen sivustot, videoiden jakelu sivustot, wikit, Whois-rekisteröidyt verkkotunnukset, metatiedot ja digitaaliset tiedostot, pimeät verkkoresurssit, paikkatiedot, IP-osoitteet, hakukoneet ja pääasiallisesti kaikki mitä verkosta löytyy.
4. Erikoislehdet, akateemiset julkaisut, väitöskirjat, konferenssijulkaisut, yritysprofiilit, vuosikertomukset, yritysuutiset, työntekijäprofiilit ja tiivistelmät.
5. Maantieteelliset tiedot kartat ja kaupalliset kuvalliset tuotteet

Avointen lähteiden tiedustelussa käytettävät lähteet voidaan kerätä kolmella eri tavalla: passiivisella-, puolipassiivisella ja aktiivisella tiedonkeruulla. Käytetty keräysprosessi on riippuvainen kohteen merkityksestä, haluttavasta tiedosta ja sen määrästä. Keräysmenetelmät pyrkivät kuvaamaan kuinka tiedon kerääminen jättää jäljen tietojen etsinnästä. Jos etsintä on hyvinkin laajaa, sivustot pitävät tällaista toimintaa haitallisena tai epäilyttävänä ja estävät sen. (The PTES Team s. a.; Hassan ja Hijazi 2018, 14.)

Passiivinen tiedonkeräys perustuu tietojen keräämiseen julkisesti saatavilla olevista resursseista, ilman että lähetetään suorasti tai epäsuorasti tietoliikennettä kohdepalvelimelle. Tämän tyyppinen tiedonkeruu on täysin salaista,

eikä kohde tiedä, että hänestä kerätään tietoa. Tämä on eniten käytetty tiedonkeräys menetelmä OSINT-tietojen keräyksessä, jota tulisi useimmissa tapauksissa käyttää. Ainoana huonona puolena on saatavilla olevien tietojen rajoittuminen vain arkistotietoihin. Ne ovat usein vanhentuneita ja suojaamattomia tiedostoja, jotka on jätetty serverille ja kohteen verkkosivulla olevaan hakemistoon. (Hassan ja Hijazi 2018, 14.)

Tiedon keräystä voidaan tehdä myös puolipassiivisella tavalla, jossa tietoliikenne ohjataan suoraan kohdepalvelimelle, mutta tiedon keräys pyritään naamioimaan mahdollisimman tavanomaiseksi tietoliikenteeksi. Etsinässä käydään perusteellisesti läpi kohteen verkosta löytyvät tiedot havaitsemattomalla tavalla, jottei kohde havaitse toimintaa haitalliseksi tai epäilyttäväksi. Tämän kaltaista tiedon keräystä pidetään pääosin nimettömänä tiedon etsintänä, sillä tiedustelulla ei tehdä niin radikaaleja tietojen etsintöjä. Kohde voi kuitenkin tarkistaa palvelimelle tehdyt yhteydenotot tai verkkolaitetiedot, josta voidaan saada tietoa tapahtuneesta tiedustelusta, mutta tiedustelijaa ei kuitenkaan pystytä tällä tavalla jäljittämään. (Hassan ja Hijazi 2018, 14.)

Puolestaan aktiivisella tiedonkeräyksellä, ollaan suoraan yhteydessä järjestelmän kanssa, josta tiedon keräys tapahtuu. Tämän seurauksena tietoliikenne näyttää epäilyttävältä ja haitalliselta käyttäytymiseltä. Aktiivinen tiedonkeräys jättää selviä jälkiä tunkeilijan havaitsemisjärjestelmään, eli IDS:ssään (Intrusion Detection System) ja tunkeutumisen estämisen järjestelmään IPS:ssään (Intrusion Prevention System). IDS:llä tunnistetaan tietoverkkoon suuntautuvat hyökkäysyritykset ja IPS:n avulla estetään tietoverkkoon tunkeutuminen ja myös haitallisten datapakettien lähettäminen. (Petters 2020; Hassan ja Hijazi 2018, 15.) Tiedusteluprosessissa käytetään edistyneitä tekniikoita. Niiden avulla saadaan tietoa muun muassa kohteen IT-infrastruktuurista, kuten avoimista porteista, haavoittuvuuksista ja web-palvelinsovelluksista. Tiedon keräyksessä voidaan tehdä myös sosiaalisen manipuloinnin hyökkäyksiä, jonka avulla voidaan saada käyttäjän paljastamaan tai antamaan pääsy salattuihin tiedostoihin. (Hassan ja Hijazi 2018, 15.)

Tiedonkeräyksen jälkeen seuraa tietojen analysointivaihe. NATO Open Source Intelligence Handbook (2001, 10-11) kirjassa havainnollistetaan, kuinka avoimista lähteistä kerätyt tiedot voidaan jakaa neljään eri kategoriaan.

1. Avointen lähteiden data, Open Source data (OSD): Ensisijaisesta lähteestä peräisin olevaa yleistä tietoa. Tietoina voi muun muassa olla satelliittikuvat, puhelutiedot ja metatiedot, tietojoukot, tutkimustiedot, valokuvat ja ääni- tai videotallenteet, jotka ovat tallentaneet tietynlaisen hetken. Tietoja ei olla käsitelty eikä editoitu millään tavalla, saadut tiedot ovat siis perusmuodossa.
2. Avointen lähteiden tiedot, Open Source information (OSINF): Yleiset tiedot, jotka ovat ensin suodatettu ja analysoitu tietyn kriteerin tai tarpeen täyttämiseksi. Näitä tietoja voidaan kutsua myös toissijaisiksi lähteiksi. Esimerkkejä ovat kirjat, artikkelit, väitöskirjat, taideteokset ja haastattelut.
3. Avointen lähteiden tiedustelu, Open Source intelligence (OSINT): Sisältää kaikki tiedot, jotka on löydetty, suodatettu ja analysoitu vastaamaan tiettyä tarvetta tai tarkoitusta. Tätä tietoa voidaan käyttää suoraan missä tahansa tiedusteluyhteydessä. OSINT voidaan määritellä pähkinänkuoressa avointen lähteiden prosessoinnin tuotoksena.
4. Vahvistetut OSINT tiedot, Validated OSINT (OSINT-V): Kerätyt tiedot on varmennettava ja vahvistettava erittäin korkean varmuusasteen omaavasta luotettavasta lähteestä. Tällaisiin turvaluokiteltuihin tiedustelulähteisiin pääsee vain valtuutetut henkilöt. Lähteiden varmennuksen avulla pystytään varmistamaan lähteiden virheettömyys, sillä tiedot saattavat olla epätarkkoja ja tahallisesti levitettyjä ja ne voivat pyrkiä hankaloittamaan OSINT-analyysin luotettavuutta.

2.2 Ketkä hyödyntävät avointen lähteiden tiedustelua?

Pääasiallisesti avointen lähteiden tiedustelua hyödyntävät kaikki Internettiä käyttävät ihmiset, sillä avoimet lähteet ovat kaikkien saatavilla. Avointen lähteiden tiedustelua on esimerkiksi tiedon etsintää uudesta autosta ostovaiheessa. Jos uppoudutaan syvemmälle avoimiin lähteisiin käyttämällä tiedonetsintätyökaluja, niin näiden työkalujen avulla voidaan helpottaa ja nopeuttaa tiedonetsintää. Näitä työkaluja käyttävät yleensä hallitukset, erilaiset järjestöt, lainvalvontaviranomaiset, tietoturva-asiantuntijat ja rikollisjärjestöt. Kaikilla näillä toimijoilla on omat motiivinsa ja hyötynäkökulmansa, käyttäessään OSINT-lähteitä. Pääasiallisesti tiedustelua käytettään potentiaalisten heikkouksien tunnistamiseen yrityksistä ja henkilöistä. (THE RECORDER FUTURE TEAM 2019; Huff 2018.)

Hallitukset ja niiden armeijaosastot ovat OSINT-lähteiden suurimpia hyödyntäjiä, sillä heillä on tarvittavat resurssit pystyäkseen keräämään ja analysoimaan valtavia tietomääriä, sekä ylipäätään suoriutumaan mittavista tiedusteluista. He tarvitsevat OSINT-lähteitä ylläpitääkseen kansallista turvallisuutta. OSINT-lähteitä hyödynnetään terrorismin torjuntaan ja seurantaan. OSINT-lähteiden avulla voidaan toimittaa johtoasemassa oleville tarvittavia tietoja kotimaisen- ja ulkomaisen politiikan vaikuttavista tekijöistä, sekä ulkomaisten tiedotusvälineiden tietoja. Tiedustelupalvelut pyrkivät myös hyödyntämään avointen lähteiden tietoja peilaten niitä salaisiin lähteisiin kuten OSINT-V-lähteille tehdään. Näin he pystyvät mahdollisesti saamaan tarvittavia tietoja henkilöistä ja ennustamaan parhaimmassa tapauksessa tulevaisuutta. (Hassan ja Hijazi 2018, 10.)

Kansainväliset järjestöt, kuten YK ja Punainen Risti, käyttävät OSINT-lähteitä rauhanturvaoperaatioissa ja avustaessaan kriiseissä ja katastrofien sattuessa. He analysoivat myös sosiaalisen median sivustoja ja verkkosovelluksia pyrkiäkseen estämään terroristiryhmien haitalliset toiminnat heidän järjestöjänsä kohtaan. Poliisi ja muut lainvalvontaviranomaiset suojelevat kansalaisia väärinkäytöksiltä, seksuaaliväkivallalta, identiteettivarkauksilta ja muilta rikoksilta, OSINT-lähteistä saatujen tietojen avulla. (Hassan ja Hijazi 2018, 11.)

Viranomaiset pyrkivät seuraamaan sosiaalisen median palstoja ja etsimään haitallisia- ja uhkaavia avainsanoja, sekä profiloimaan tietojen pohjalta rikollisia ja heidän sosiaalisia verkostojansa. Näiden tietojen avulla voidaan päästä jäljille mahdollisista rikoksista ja väärinkäytöksistä. Yritykset käyttävät OSINT-lähteitä seuratakseen markkinatilannetta ja mahdollisia nousevia kilpailevia yrityksiä, pyrkien ennustamaan heidän toimintaansa vaikuttavia tekijöitä. Yritykset hyödyntävät myös OSINT-tiedustelun avulla löydettäviä heikkouksia heidän IT-infrastruktuuristansa muodostaen uhkatietostrategiansa, jottei heidän yrityksensä joutuisi tietomurron kohteeksi. (Hassan ja Hijazi 2018, 11–12.)

Tietoturva-asiantuntijat ja verkkorikolliset hyödyntävät OSINT-tiedustelua saadakseen tietoa heidän kohteistansa. Ainoana erona on, että tietoturva-asiantuntijat etsivät näitä haavoittuvuuksia, jotta ne voidaan korjata ennen kuin verkkorikolliset hyödyntävät heikkouksia ja pääsevät käsiksi luottamuksellisiin tietoihin. Kuitenkin molemmat pyrkivät toimimaan samanlaisilla metodeilla. Terroristi- ja rikollisjärjestöt pystyvät hyödyntämään OSINT-lähteitä heidän hyökkäyksiensä suunnittelussa, kuten kohteen tiedustelussa. Näitä OSINT-lähteitä ovat esimerkiksi satelliittikuvat ja sosiaalinen media. Sosiaalisen median ja erilaisten verkkosivujen avulla he pystyvät rekrytoimaan uusia taistelijoita ja rahoittajia, sekä levittämään propagandaansa. Tämän kaltaiset toimijat pyrkivät keräämään tiedustelupalveluiden luottamuksellisia tietoja, hyödyntäen niitä OSINT-lähteinä. Lähteinä voi esimerkiksi olla tiedustelupalveluiden sotilastiedot ja hyökkäyssuunnitelmat. (Hassan ja Hijazi 2018, 12–13.)

2.3 Avointen lähteiden tiedustelun hyödyt ja haasteet

Kaikilla tiedustelualoilla on omat rajoituksensa, eikä OSINT eroa siinä suhteessa. Avointen lähteiden tiedustelulla on vahva potentiaali tiedon keräyksessä, mutta sen tuomia haasteita ei pidä jättää huomiotta. OSINT tuo mahdollisuuksia ja tietoa monelta eri tiedusteluosa-alueelta, kuten aikaisemmin mainittiin. Tämän seurauksena tiedustelun tuomat edut koskevat monia eri aloja nykymaailmassa.

OSINT-lähteiden avulla saadut suuret tietomäärät tuo avointen lähteiden tiedustelulle suuren edun, mutta se on myös samalla sen suurin heikkous. Informaatiota on niin paljon, että sen suodattaminen, analysointi ja luotettavuuden todentaminen tuo suuria haasteita OSINT-tiedustelulle. Tämä on yksi avointen lähteiden tiedustelun keskeinen ongelma: tiedon määrän ylikuormitus. OSINT-tiedustelu on kuitenkin todella halpa tiedonkeruutyökalu, jonka ansiosta sitä pystyvät hyödyntämään kaikki pienyrittäjistä valtiollisiin hallintoelimiin. Resurssit mahdollistavat vain suurempien tietomassojen analysoinnin ja tietojen todenperäisyyden todennettavuuden. Pääasiallisesti kuka vain pystyy hyödyntämään OSINT-lähteitä ja näitä tietoja pystytään jakamaan laillisesti kenelle tahansa ilman pelkoa, että rikkoo tekijänoikeuksia. (Expert System Team 2017; Hassan ja Hijazi 2018, 15-16.). Tietenkin tietylle osalle tietoja on omat rajoituksensa, muun muassa harmaa kirjallisuus on tarkoitettu vain hyvin suppeaan levitykseen. Harmaalla kirjallisuudella tarkoitetaan aineistoa, jota ei olla rekisteröity kansallisbibliografiaan. Kirjallisuuden tietoihin käsiksi pääsemiseen täytyy saada tekijänoikeuksien haltijan lupa tai maksaa siitä tietty summa. (Sluijter 2020.)

Avointen lähteiden tiedustelun etuna on myös sen saatavuus. Käytännössä OSINT-lähteitä voidaan hyödyntää missä tahansa ja lähteet ovat yleensä aina ajan tasalla. Tiedustelun suorittamiseen tarvitaan vain tietoa, osaamista ja työkalut tiedustelun suorittamiseksi ja tietojen analysoimiseksi. Pitää ottaa kuitenkin huomioon, että OSINT-järjestelmän hyödyntäminen vaatii yksityiskohtaista analyysiä ja ymmärrystä sen käytön tuomista vaatimuksista. Useimmat tahot ja valtiot käyttävät OSINT-lähteitä levittääkseen tietolähteiksi harhaanjohtavia tai epätarkkoja tietoja. Tavoitteenaan heillä on pyrkiä luomaan hämmennystä ja häiritä tiedustelevia toimijoita. Näin ollen tietojen analysointi vaiheessa on syytä tarkistaa löydetyt lähteet luokitelluista ja varmennetuista lähteistä, jotta voidaan vakuuttua tietojen totuudenmukaisuudesta (OSINT solutions, Inc 2016; Hassan ja Hijazi 2018, 17.)

Tiedon keräyksen aikana voidaan myös kohdata etsinnällisiä haasteita erityisesti, kun tietoa etsitään syvien- ja pimeiden verkostojen sivuilta. Syvän- ja pimeän verkon sivustoilla tarkoitetaan verkosta löytyviä sivustoja, joita ei olla indeksoitu mihinkään hakukoneisiin ja pimeän verkkoon päästään käsiksi vain hyödyntäen salaista reititystekniikkaa. (Ozkaya ja Islam 2019, 3-48.) Indeksoimattomuuden seurauksena tiedustelijan on tiedettävä tarkka URL-osoite, jotta hänellä on pääsy kyseiselle sivustolle. Kuinka tiedustelija pääsee käsiksi näihin indeksoimattomiin sivustoihin? Tämä on syvän- ja pimeän verkon yksi suurimmista haasteista avointen lähteiden tiedustelun näkökulmasta. Pimeän verkon tiedustelu luo vielä lisää haasteita verkon salaisenreititystekniikan ja nimettömyyden seurauksena. (Bertram 2015, 17-18).

2.4 Avointen lähteiden tiedustelun eettisyys

Avointen lähteiden tiedustelun tärkein pätevyysvaatimus on, että tietojen kerääminen ei edellytä minkäänlaista salaista keräystekniikkaa. Tiedot on hankittava keinoilla, jotka eivät loukkaa lähteiden tekijöiden tekijänoikeuksia ja kaupallisia vaatimuksia (Media Sonar 2020). Tästä huolimatta avointen lähteiden tiedustelu kohtaa eettisiä huolenaiheita tietojen keräämisessä ja analysoimisessa. Erityisesti huomiota herättää onko avointen lähteiden tiedustelu oikein ihmisoikeudellisten, yksityisyyden- ja tietosuojalain perusteella. (Eijkman ja Weggemans 2013.)

Tiedustelun näkökulmasta juridisia kysymyksiä herättää erityisesti se onko tiedustelussa hyödynnetty laittomia- tai jopa varastetuksi luokiteltuja lähteitä todisteina. Onko tiedustelijalla oikeus hyödyntää näitä tietoja vai onko tiedustelija vastuuvelvollinen jättämään tiedot käsittelemättä ihmisoikeuksien ja rikosoikeudellisten syiden perusteella (Hu 2016). Toisaalta jos OSINT-lähteet hankitaan laittomin keinoin pyrkien käyttämään tietoja juridisesti hyvään, niin onko se oikeudellisesti väärin? Puolestaan, kun tietoa kerätään julkisista lähteistä, jotka on piilotettu, onko se moraalisesti oikein julkaista näistä ilmenneitä tietoja. Tietojen julkaisemisella saattaa olla suuret vaikutukset joihinkin ryhmiin ja yksilöihin. Julkaiseminen saattaa mahdollisesti laukaista haitallisia skandaaleja, vaikka tiedot on etsitty julkisista lähteistä hyödyntäen vain erityisiä tekniikoita ja tietojen analysointia. Tiedot ovat olleet kaikkien saatavilla, jos on vain osannut etsiä niitä. (Hassan ja Hijazi 2018, 17.)

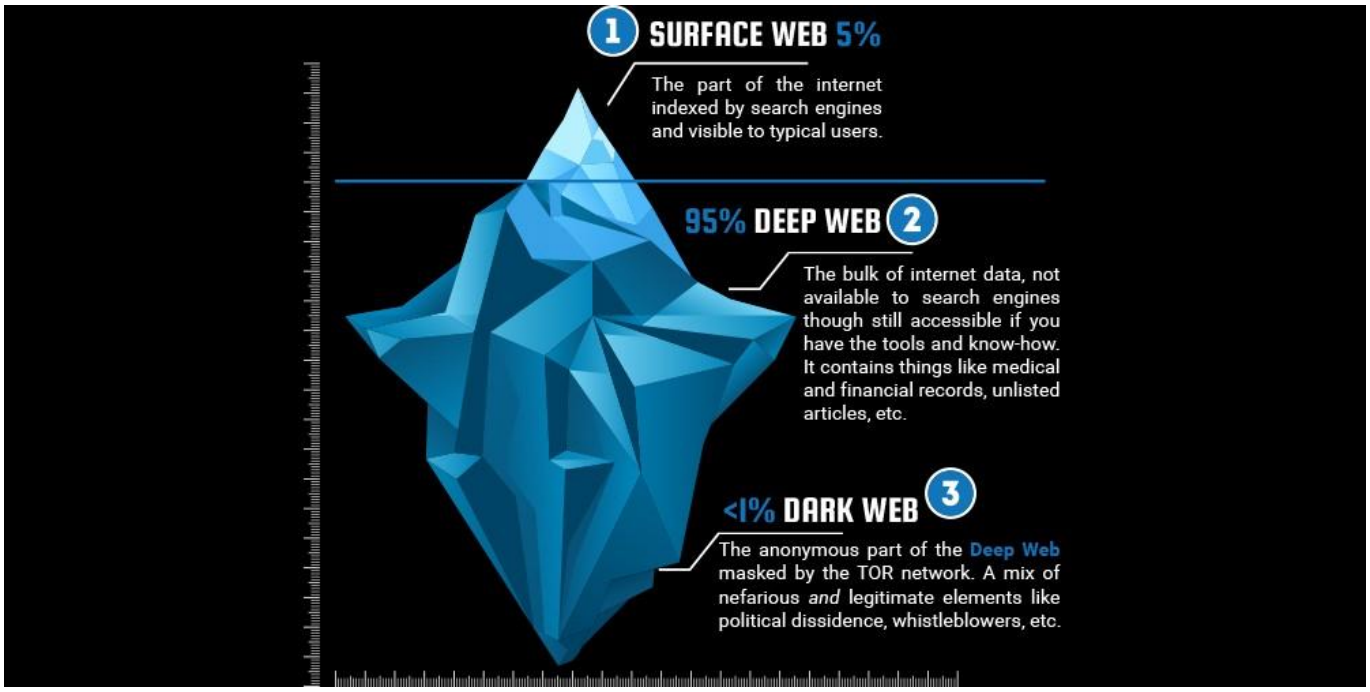
Täytyy pitää myös mielessä tiedon etsinnässä käytettyjen automatisoitujen työkalujen seurauksena ilmenevät ohjelmistovirheet, jotka saattavat johtaa tietojen lopputuloksen vääristymiseen. Tämä johtaa kysymykseen voidaanko täysin luottaa ohjelmistojen tiedon keräykseen ja analysointiin samalla tavalla, kuin ihmisen tekemään manuaaliseen työhön? Toisaalta täytyy olla myös tietoinen tiedustelijan mahdollisesta alttiudesta pyrkiä valitsemaan tai jopa väärentää kerättyjä lähteitä tietyn tuloksen saavuttamiseksi. (Hassan ja Hijazi 2018, 18.)

Suuret yritykset, jotka käsittelevät valtavasti verkon käyttäjien henkilötietoja, kuten Google ja Facebook, hyödyntävät näitä tietoja kaupallista tiedustelua varten. Tällaisia tietotyyppejä on esimerkiksi henkilökohtaiset tiedot, kuten nimi, syntymäpaikka, sosiaaliturvatunnus ja perhetiedot. Tekniset tiedot, kuten käyttöjärjestelmä, selaimen versio, IP-osoite, laitteen sijainti ja kaikki tiedot, jotka käyttäjä mahdollistaa löydettäväksi. Yritykset perustelevat tiedon keräyksen hankkimalla vain nimettömiä tietoja, mutta muita lähteitä hyödyntäen voidaan muodostaa kokonaisvaltainen kuva käyttäjästä. (Hassan ja Hijazi 2018, 18.)

Voidaan pohtia, kuinka tämän kaltaiseen tiedonkeräykseen pitäisi suhtautua. Onko moraalisesti oikein, että suuret yritykset saavat hyödyntää käyttäjien henkilötietoja kaupallisiin tarkoituksiin? Käyttäjien on usein myös hyväksyttävä käyttöehdot tietojen luovuttamiseen, tai muuten käyttäjä ei saa oikeutta käyttää sovellusta. Onko oikein käydä kauppaa sovelluksen käytön ja henkilötietojen välillä? Suuret yrityksetkin kohtaavat tietomurtoja, kuten vuonna 2019 Zynga niminen yritys. Zynga on luonut Farmville mobiilipelin, johon voidaan kirjautua myös Facebook tunnuksella. Tietomurto hyökkäyksen seurauksena 218 miljoonan käyttäjätunnukset ja salasanat varastettiin. Tämä määrä pitää sisällään Facebook:in ja Zunga:n käyttäjätietoja. (Swinhoe 2020.) Tämän seurauksena käyttäjän täytyy olla tietoinen mihin sovelluksiin henkilökohtaisia tietojaan luovuttaa.

3 INTERNETIN TASOT

Internetistä löytyvät tiedot ovat avointen lähteiden tiedustelun kultasuoni, jonka avulla saadaan ammennettua valtavasti tietoa kohteesta. Internetin valtavan kehityksen seurauksena verkko koostuu kolmesta eri tasosta: maailmanlaajuisesta verkosta (World Wide Web), eli verkonpinnasta, syvästä verkosta (Deep Web) ja pimeästä verkosta (Dark Web). Avointen lähteiden tiedustelussa pyritään hyödyntämään kaikkien näiden tasojen mahdollistamaa avointen lähteiden kokonaisuutta. Perinteisen verkossa tapahtuvan avointen lähteiden tiedustelu kohdistetaan verkonpintaan, jonka sisältö koostuu vain noin viidestä prosentista kaikesta verkossa tapahtuvasta liikenteestä (Kuva 3). Tämä sisältää kaikki perinteiset julkiset sivustot kuten esimerkiksi Google, Facebook ja Amazon sekä pääsääntöisesti kaikki sivustot, jotka on indeksoitu hakukoneiden järjestelmään.



Kuva 3. Internetin tasot (z3roTrust 2019.)

Puolestaan kun sukelletaan ”syvään verkkoon”, sillä tarkoitetaan mitä tahansa verkkosivustoa, johon ei voida päästä tavanomaisen hakukoneiden, kuten Googlen tai Yahoon avulla. Sivustot eivät löydy perinteisten hakutulosten joukosta sillä sivustoja ei olla syötetty perinteisten hakukoneiden indeksointijärjestelmään. Järjestelmän avulla hakukonepalvelimet järjestelevät syötetyt indeksit algoritmien joukkoon, joiden avulla hakuja tehdessä haettu tieto osataan indeksoida haettuun asiayhteyteen. Tämän takia näille sivustoille päästään vain syöttämällä tarkka URL-osoite tai mahdollisen verkkolinkin kautta. Syvän verkon tarkoitus on käytännössä pitää halutut verkkosivustot poissa normaalien käyttäjien näköpiiristä. Tätä hyödyntävät suuret yritykset, koulutusjärjestelmät ja järjestöt, jotka haluavat tiettyjen sivustojen olevan vain heidän henkilökuntansa käytössä. Usein miten nämä sivustot ovat suojattu salasanalla, sillä sivustot sisältävät salassa pidettäviä tietoja yrityksistä ja järjestöistä. (Sheils 2020.)

Tasojen pohjalla sijaitsee pimeä verkko, joka on osa syvää verkkoa, eli sivustoja, joita ei ole indeksoitu hakukoneisiin. Ainoana erona on, että pimeään verkkoon päästään käsiksi vain hyödyntäen salattua reititystekniikkaa. Tekniikkaa hyödyntää muun muassa Tor-, I2P- ja Freenet selain. Näiden avulla käyttäjän tietoliikenne ja sijainti on salattu. Suosituin salattu reititystekniikka on Tor, eli The Onion Router: verkosto koostuu täysin anonyymistä toimin-

nasta. Sen tietoliikenne on sipulireititetty, jolla tarkoitetaan kerroksellista salausta. Tällaisessa suojauksessa käyttäjän luoma tietoliikenne salataan ja ohjataan useisiin eri yhdistyspisteisiin, eli solmuihin, joita on noin 7000 kappaletta. Tämän seurauksena mikään yksittäinen solmu ei voi assosioida käyttäjää määränpäähänsä, jolloin käyttäjän paikantaminen on todella hankalaa. (Vacca 2013; Ozkaya ja Islam 2019, 48.) Pimeän verkon alkuperäinen päämäärä oli mahdollistaa ihmisten sananvapaus, sellaisissa maissa, jotka rajoittavat ihmisten sananvapautta ja yksityisyyden suojaa. Kuitenkin verkon tuoman nimettömyyden seurauksena pimeää verkkoa on alettu hyödyntämään rikolliseen toimintaa muun muassa huumeiden, aseiden, lapsi pornon, haittaohjelmien ja varastettujen omaisuuksien myymiseen. (Ozkaya ja Islam 2019, 34-48.)

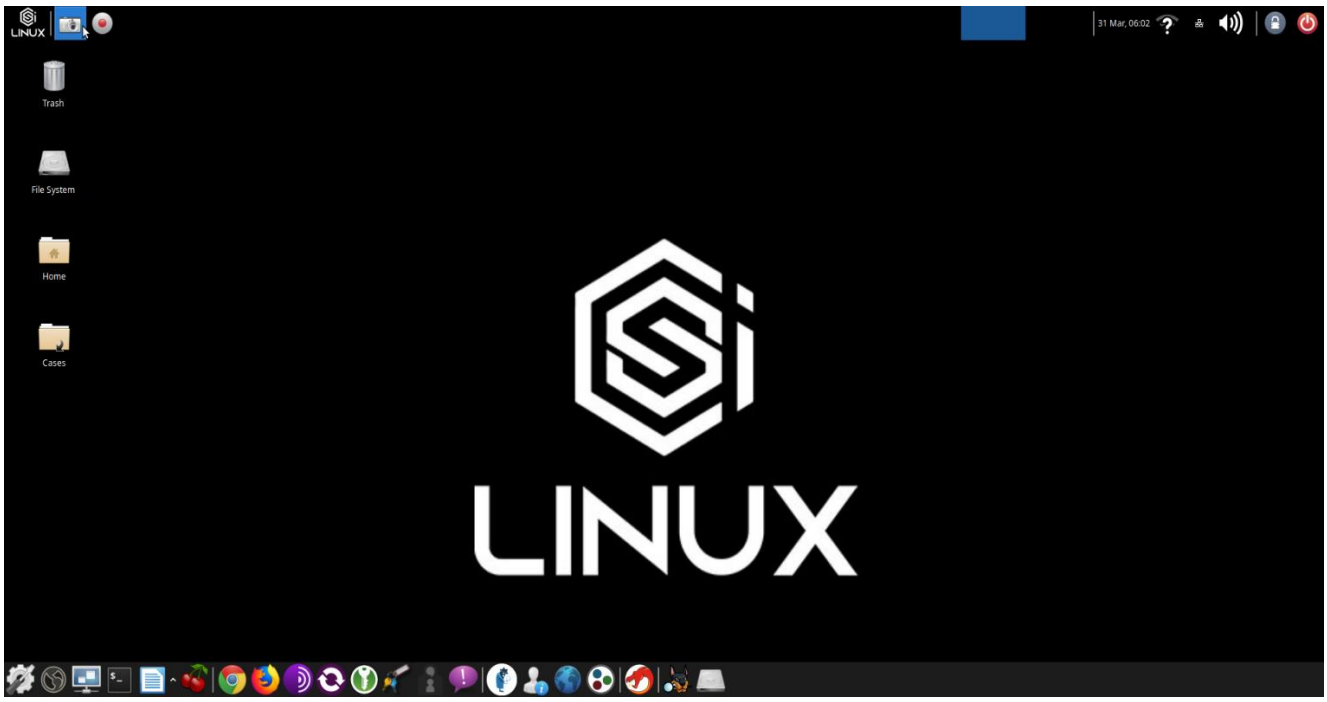
Näiden syvän- ja pimeän verkkojen avointen lähteiden tiedustelu vaatii erikoistuneiden keräys-, käsittely- ja analysointityökalujen käyttöä. Niiden avulla pystytään keräämään halutuilta sivustoilta tietoa, jota analysoimalla voidaan saada tietoa etsitystä kohteesta. Tietojen etsintä syvästä- ja pimeästä verkosta on erittäin hankalaa ja aikaa vievää. Erityisesti pimeän verkon tiedon etsintä voi olla myös haitallista, sillä kerätty tieto saattaa sisältää traumaattista sisältöä sekä haitta- ja kalasteluohjelmia. (z3roTrust 2019.)

4 KÄYTETTY KÄYTTÖJÄRJESTELMÄ

Avointen lähteiden tiedusteluun on saatavilla monia eri käyttöjärjestelmiä. Käytetyimpiä ovat muun muassa Kali Linux, Buscador, Hurrot ja CSI Linux Investigator. (PentestIT 2020). Suurimmaksi osaksi kaikki järjestelmät perustuvat Linux käyttöjärjestelmään, joista on jakautunut kaksi eri linjaa Debian ja Ubuntu. Nämä Linux-pohjaiset käyttöjärjestelmät, ovat hyvin pitkälti samanlaisia, mutta tiettyjä eroavaisuuksia niissä on. Debiania pidetään enemmän asiantuntijoiden käyttöjärjestelmänä, toisin kuin Ubuntua, joka soveltuu hyvin aloittelijoille. Tämä johtuu Ubuntun helppokäyttöisyydestä, koska se sisältää valmiiksi asennettuja uusia ominaisuuksia ja ohjelmistoja. Debiania pidetään vakaampana alustana, sillä se ei sisällä ladattaessa ylimääräisiä ohjelmistoja, jotka saattavat tuoda mahdollisia virhetilanteita tai arvaamattomia kaatumisia. (Congleton 2018.)

Opinnäytetyössä perehdytään Ubuntu pohjaiseen CSI Linux Investigator käyttöjärjestelmään, jonka on suunnitellut Computer Forensics. Käyttöjärjestelmä on avoimen lähdekoodin sovellus, jonka seurauksena kuka vain pystyy lataamaan ja hyödyntämään sitä. Tämän käyttöjärjestelmän tarkoituksena on tuoda käyttäjälleen monikäyttöinen ympäristö OSINT-tiedustelussa käytettävien työkalujen avulla. CSI Linux, pystytään käynnistämään virtuaalikoneen kautta. Käyttöjärjestelmä sisältää kolme erilaista ympäristöä: CSI Linux Analyst, CSI Linux Gateway ja CSI Linux SIEM. Analyst virtuaalikone on käytetyin työasema, joka sisältää kaikki tarvittavat työkalut tiedustelua varten (Kuva 5). Gateway puolestaan lähettää kaiken tietoliikenteen suojatun Tor verkon kautta, jonka avulla pyritään suojelemaan käyttäjän nimettömyyttä. SIEM virtuaalikonetta käytetään tietomurtojen tunnistamisen havaitsemiseen. (Information Warfare Center s. a. a; Information Warfare Center 2020 s. a. b.)

CSI Linux Investigator, pystytään käynnistämään virtuaalikoneen käynnistysalustan avulla (Kuva 4). Käytettynä alustana toimi Oraclen kehittämä VirtualBox-alusta, jolla pystytään luomaan virtuaalikäyttöjärjestelmiä. Virtuaalikäyttöjärjestelmät toimivat myös virtuaalikovalevyinä. (Oracle s. a.). Virtuaalikoneiden hyvänä puolena on mahdollisuus käyttää monia eri käyttöjärjestelmiä samalla tietokoneella. Virtuaalikovalevyt luovat turvallisen ympäristön ohjelmistojen suorittamiseen. Sillä jos virtuaalikäyttöjärjestelmä altistuu haittaohjelmalle, niin käyttöjärjestelmä voidaan tuhota ja poistaa kyseinen ongelma ilman, että käytettävä tietokone altistuisi minkäänlaiselle haitalliselle toiminnalle. (Totounji 2017.)



Kuva 4. CSI Linux Investigator käyttöympäristö

Open Source Tools:

* ABE (Android Backup Extractor)
* ADB
* AFLogical OSE
* AlienVault-OTX Python-SDK
* Amass
* Autopsy GUI
* Bleachbit
* Catfish Search
* Cewl
* CherryTree
* ClamAV (Antivirus)
* CS-QuickTunnel
* DC3DD
* DCFLDD
* Docker
* EXIFTool
* EyeWitness
* FastCrackZip
* FBI (Facebook Information)
* FFMpeg
* FileCompare (Forensics-Colorize)
* File Roller
* Foremost
* FreePlane (Mind Mapping)
* Ghidra
* GIMP
* GNUPG
* GoBuster
* GPA
* Guymager
* Hashcat
* HTTrack
* (x)Hydra
* Infoga
* Instaloader
* iPhone Backup Decoder and Analyzer 2
* KeePassXC
* LibMobileDevice
* LibreOffice
* LittleBrother
* MagicRescue
* Maltego CE
* Medusa
* Metagoofil

Open Source Tools (Continued):

* NMap
* MISP
* OnionShare
* OphCrack
* OSINTFramework
* OSINT-Search
* OutGuess
* PDFMeta
* Pidgin (Off-The-Recod plugin)
* PhotoRec
* qTox
* Recon-NG
* RecordMyDesktop
* RecoverDM (Bad Sector Recovery)
* RevoverJPEG
* RecoverMOV
* Radare2
* Rootkit Hunter
* Scalpel
* Sherlock
* Skiptracer
* Slack
* Spiderfoot
* StegHide
* StegSnow
* StegoSuite
* Sublist3r
* theHarvester
* TInfoleak
* Tor Browser
* Transmission Bittorrent
* UFW (Firewall)
* Vinetto
* YouTube_DL
* VLC Player
* Volatility 3
* Whonix Gateway (Until CSI Gateway is released)
* Wireshark
* Zulucrypt

Kuva 5. CSI Linux Investigator sisältämät työkalut (Information Warfare Center 2020 s. a. c.)

5 KÄYTETYT TYÖKALUT JA NIIDEN MÄÄRITELMÄT

Avointen lähteiden tiedustelu pitää sisällään useita eri työkaluja. Tässä työssä perehdytään valittujen työkalujen käyttöön eri Internetin tasoilla. Valitut työkalut ovat avointen lähdekoodin sovelluksia, jotka ovat tällä hetkellä käytetyimpiä työkaluja avointen lähteiden tiedustelun piirissä (Naini 2019). Haasteita työkalujen käytölle on tuonut verkkosivustojen suojausmekanismien kehittyminen, jonka tarkoituksena on estää tiedustelu tyyppinen tietojen etsintä ja kerääminen. Onneksi myös työkalut kehittyvät ja tällä hetkellä niiden saatavuus on todella laaja. Työkalujen käytön tarkoituksena on helpottaa ja nopeuttaa tietojen keräystä, prosessointia ja analysointia. Työkalujen avulla pystytään suorittamaan kyselyitä kohdennettuun tarkoitukseen tai sitten suorittamaan laaja mittaisia kyselyitä, joilla saadaan hyvin paljon tietoa kohteesta. (Hassan ja Hijazi 2018, 15, 152.)

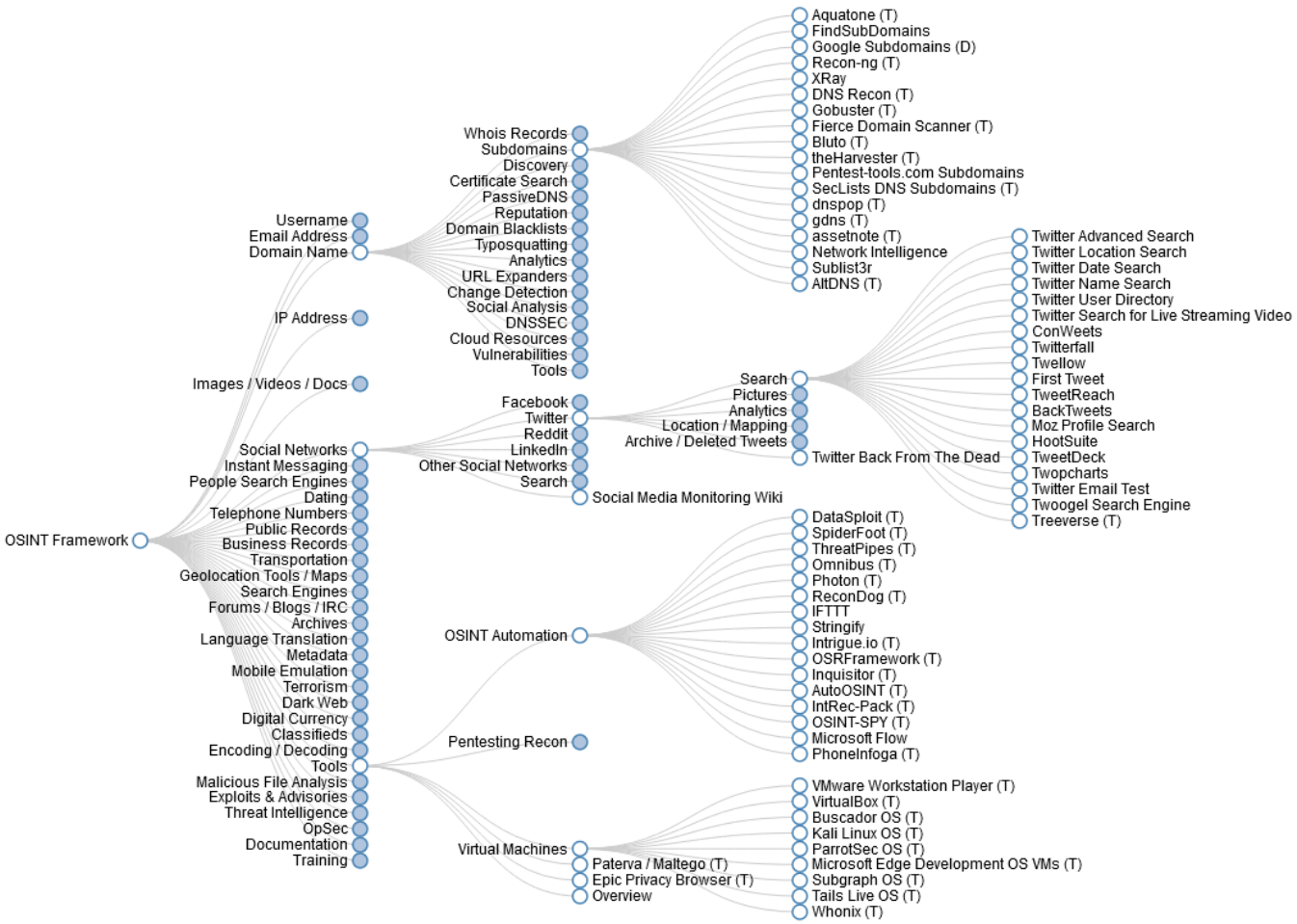
Suurten tietomäärien analysointi ja luotettavuuden todentaminen on mahdollista osalla valituista työkaluista. Kuitenkin hallitukset, valtiontason toimijat ja suuret yritykset, ovat luoneet itselleen tehokkaita ja automatisoituja työkaluja kehittyneillä tiedon suodatus- ja todennusmenetelmillä. Niillä pystytään kattavampaan ja luotettavampaan tietojen etsintään. Nämä työkalut ovat luotu pelkästään tällaisten toimijoiden käyttöön, jonka seurauksena tällaisten työkalujen käyttöä on rajoitettu. (Hassan ja Hijazi 2018, 16.) Avoimen lähdekoodin työkaluilla pystytään keräämään myös merkittäviä tietoja kohteesta. Tiedon keräystä pystytään tehostamaan hyödyntämällä useita avoimen lähdekoodin työkaluja tiedon etsinnässä. Tämän avulla saadaan laajennettua kohteesta kerätyn tiedonmäärää. (Naini 2019.)

Avointen lähteiden tiedustelun suuren tietosisällön seurauksena avointen lähdekoodin työkalut kohdistavat tiedon keräämisen tietylle osa-alueelle. Tämän seurauksena työkaluja on saatavilla todella paljon, eikä tiedon etsintään ole olemassa suoraviivaista tekniikkaa eikä mekanismia. Käytettyjen työkalujen käyttöä täytyy aina pohtia tehtäväkohtaisesti, riippuen määritellystä tehtävänannosta ja kohteesta. Ennen tiedustelua on syytä pohtia kysymyksiä: Mitä etsitään? Mikä on tiedustelun päämäärä? Kuka tai mikä on kohteena? Kuinka tiedustelu suoritetaan? (SecurityTrails Team 2018 a; SecurityTrails Team 2018 b.)

Valittujen työkalujen esittelyissä pyritään antamaan kattava kokonaisuus käsiteltävistä työkaluista ja niiden tuomasta tietosisällöstä. Tiedon etsintätyökalut on jaoteltu, etsimään tietoa verkon eri tasoista. Joitakin työkaluja on myös mahdollista käyttää eri pintojen välillä. Työkalujen esittelyssä ei suoriteta mittavia tiedon keräyksiä, vaan pääpainona esittelyssä on tuoda esiin: kuinka työkaluja käytetään, miten tiedon etsintä tapahtuu ja millaisia tietoa etsinnällä voidaan saada.

5.1 Pinta- ja syvässä verkossa käytetyt työkalut

Verkon pinnalle kohdistuva tiedustelu on usein miten helpoin tapa aloittaa OSINT-tiedustelu, sillä kerättävät tiedot ovat helposti saatavilla hakukoneiden indeksoitujen sivustojen ansiosta. Pinta verkko tarjoaa tiedustelijalle laajan valikoiman työkaluja ja tekniikoita tiedon etsinnän soveltamiseksi. (Bertram 2015, 21, 56.) Tietoa voidaan kerätä monella eri tekniikalla: hakukoneiden kyselyiden perusteella ja verkkosivujen tietosisällön läpikäymisellä, sosiaalisen median tiedon etsinnällä ja tarkkailulla (Akhgar, Bayerl ja Sampson 2016, 124). Alla olevasta kuvasta (Kuva 6) huomataan, kuinka suuri avointen lähteiden tiedustelun rakenne on. Rakennetta hyödyntämällä saadaan laaja käsitys erilaisista tavoista tehdä etsintöjä tiettyyn aihepiiriin. Työkaluilla saadaan informaatiota muun muassa: sosiaalisesta mediasta, julkisista- ja yritys pöytäkirjoista, puhelinnumeroista, verkkotunnuksista, IP-osoitteista ja sähköpostiosoitteista (Kuva 6). Tiedon keräyksen aikana saadut tiedot voivat sisältää informaatiota syvän verkon piilotetuista sivustoista, jonka seurauksena päästään myös niihin käsiksi. Kuten aikaisemmin mainittiin, tämä on yksi syvän verkon suurimmista haasteista avointen lähteiden tiedustelun näkökulmasta (Bertram 2015, 17.)



Kuva 6. OSINT rakenne (Nordine s. a.)

5.1.1 Maltego

Maltego on Etelä-Afrikkalaisen yrityksen Patervan suunnittelema avointen lähteiden tiedustelussa käytetty linkki-analyysityökalu, jolla pystytään yhdistämään kerättyjen tietojen väliset suhteet selkeään solmupohjaiseen kaavioon. Maltego pohjautuu Java ohjelmointikieleen ja se on tällä hetkellä yksi käytetyimmistä työkaluista OSINT-lähteiden piirissä. Tämän seurauksena se on asennettu valmiiksi Kali Linux käyttöjärjestelmään. Työkalua voidaan käyttää verkosta löytyvien lähteiden välisten suhteiden tutkimiseen. Tutkimus tapahtuu automatisoidun kyselyprosessin avulla. Käyttäjä voi luoda yhteyksiä muun muassa verkkotunnuksen, nimen, käyttäjänimen, organisaation ja puhelinnumeroiden avulla. Pääasiallisesti työkalu jäsentelee suuren määrän tietoa eri lähteistä ja muodostaa sen pohjalta solmupohjaisen kaavion tietojen analysointia varten. (Maltego s. a. a; Wondersmith_rae 2019.)

Maltegosta on tehty neljä erilaista versiota, jotka on esitetty alla olevassa taulukossa (Taulukko 1). Jokainen versio on mahdollista ladata Windows-, Linux- tai Mac käyttöjärjestelmälle. Maltegon asennus onnistuu helposti ja asennusohje löytyy Maltegon kotisivulta (Maltego s. a. c). Asennuksen jälkeen täytyy vain rekisteröityä, joka mahdollistaa ilmaisversioiden käytön. Tässä työssä käytetty versio on Maltego CE.

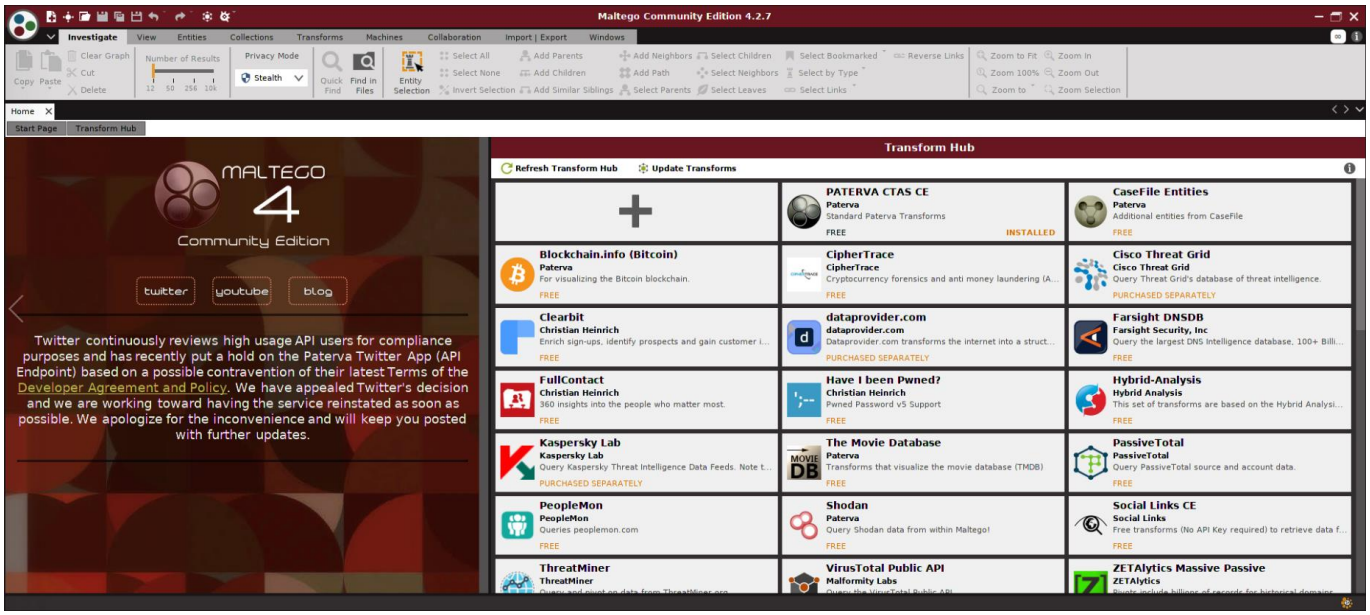
TAULUKKO 1. Maltego neljä eri versiota (Maltego s. a. b.)

Versio	Sisältää	Ilmainen / Maksullinen
Casefile	Ainoastaan linkkien tutkimiseen offline-tilassa	Ilmainen
Maltego CE	Sisältää rajallisen määrän sovellusohjelmointirajapintojen hyödyntämis- vaihtoehtoja	Ilmainen
Maltego Classic	Sisältää kaikki sovellusohjelmointirajapintojen hyödyntämis- vaihtoehdot	Maksullinen – 899€
Maltego XL	Sisältää kaikki edellä mainitut ominaisuudet ja paranneltu kyky analysoida ja suoriutua suurista tietomääristä.	Maksullinen – 1799€

Maltegon käynnistämisen jälkeen avautuu Maltegon aloitusruutu (Kuva 7). Kuvasta nähdään, että Maltego sisältää monia eri ”Transform hubeja”, joiden avulla haetaan annetun syötteen perusteella liittyviä tietoja. Löydetyt tiedot esitetään ”kokonaisuuksina” solmupohjaisessa kaaviossa. Useat hubit vaativat kuitenkin API-avaimen, eli käyttäjäkohtaisen tunnisteavaimen, jotta lisättyjä hubeja voidaan hyödyntää Maltegon kyselyissä. On huomattava kuitenkin, että ilmaiset versiot eivät sisällä kaikkia hubeja, vaan ne ovat saatavilla vain maksullisissa versioissa.

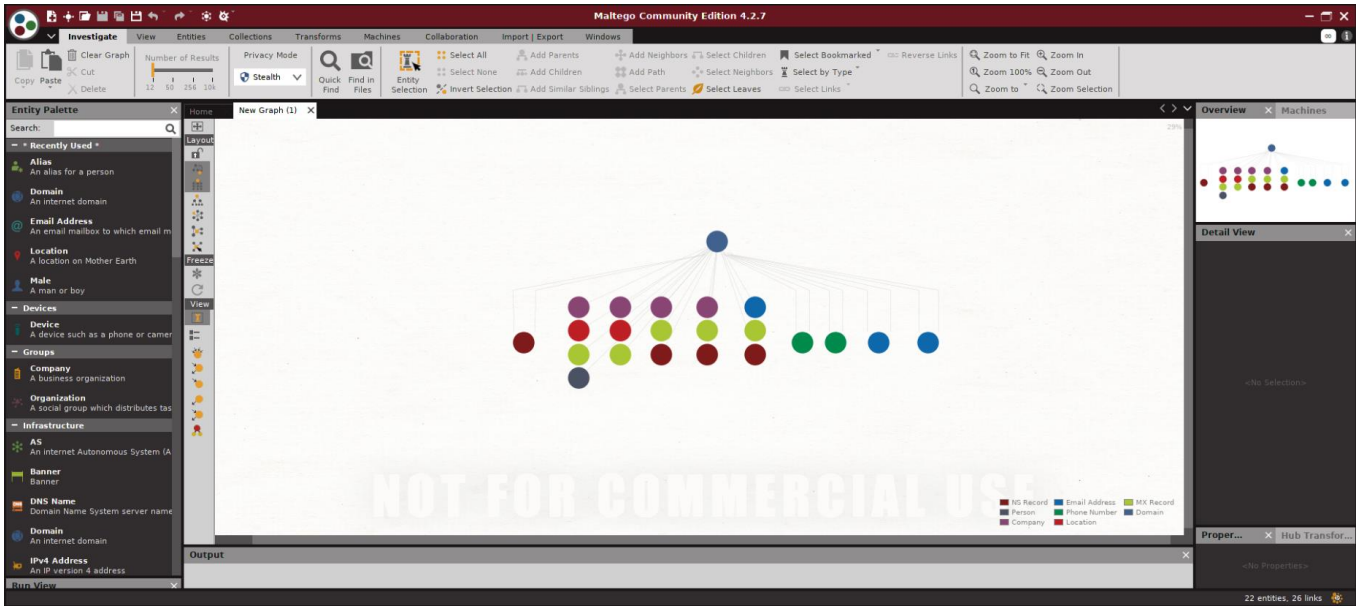
Uuden graafisen työn aloitus tapahtuu klikkaamalla vasemmasta yläreunasta löytyvää ”Create a new graph” kuvaketta, jonka jälkeen avautuu tyhjä ikkuna. Tähän ikkunaan voidaan lisätä vasemmasta reunasta haluttu kokonaisuus, eli ”entity”. Kokonaisuutena voi toimia muun muassa verkkotunnus, sähköpostiosoite, puhelinnumero, twiittaus tai käyttäjätunnus. Esimerkiksi jos kokonaisuudeksi valitaan verkkotunnus, voidaan kirjoittaa kohteen verkkotunnus osoitteen syöte kenttään. Klikkaamalla oikealla hiirellä verkkotunnusta, avautuu ”Run Transforms” ikkuna. Ikkunasta voidaan valita suoritettava hakuprotokola tai klikkaamalla ”All Transforms” voidaan suorittaa laaja haku, jossa Maltego etsii kaikki kohteesta löytyvät linkittyvyydet ja esittää ne solmupohjaiseen kaavioon (Kuva 8). On otettava myös huomioon, että Maltego CE versio ei anna suorittaa kuin 50 hakua kerralla, jonka seurauksena, jos kohteesta löytyy enemmän kuin 50 linkittyvyyttä, haut on tehtävä pienimmissä osissa. Tietomäärien

kasvaessa suuriksi graafista mallinnusta voidaan muokata haluamaansa muotoon vasemmasta ”Layout” paneelista (Kuva 9). Tämä voi helpottaa tietojen analysointia ja tietojen välisten suhteiden ymmärtämistä.



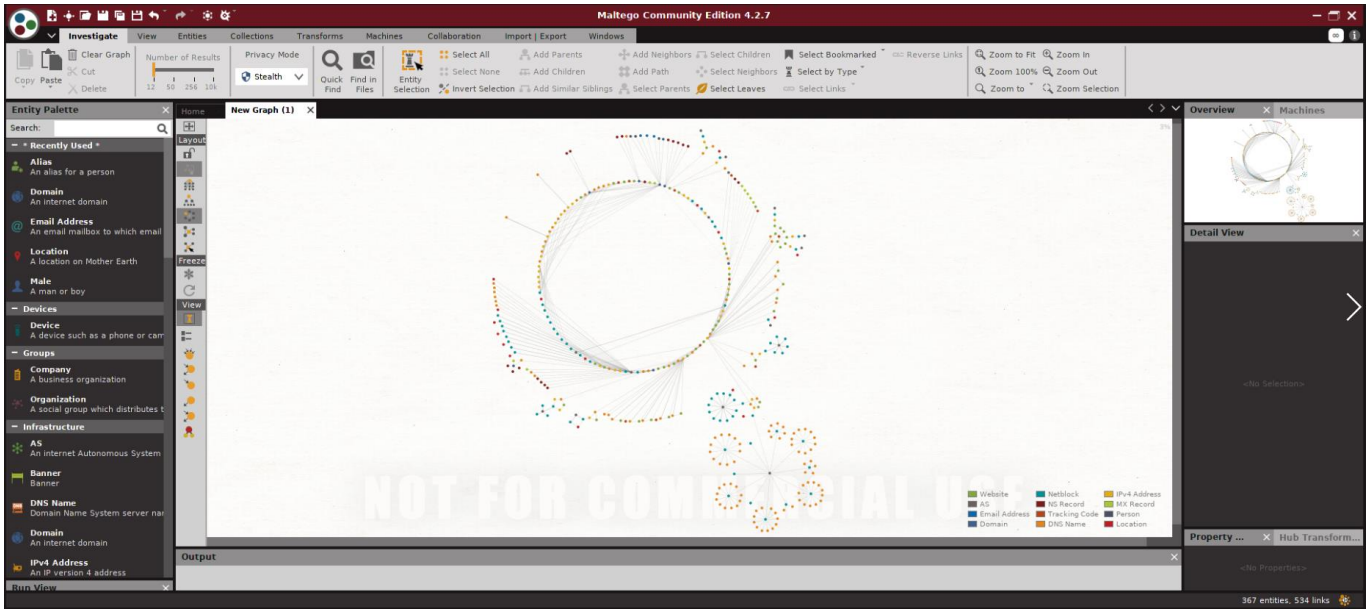
Kuva 7. Maltego CE aloitusruutu

Tulokset voivat sisältää uusia verkkosivuja, henkilöitä, sijainteja, verkkotunnuksia, IP-osoitteita, sähköpostiosoitteita, AS numeroita , sivustojen seurantakoodeja, DNS nimiä, NS- ja MX kirjanpitoja. AS numeroilla tarkoitetaan verkko-operaattorin ylläpitämiä IT-verkkojen ryhmiä, joilla on vain yksi selkeästi määritelty reitityskäytäntö (APNIC s. a.). Sivustojen seurantakoodien ”Tracking Codes” avulla kerätään tietoa verkkosivuston käyttäjästä (RYTE WIKI s. a.). Maltego etsii sivuston seurantakoodin avulla sivustoja, jotka hyödyntävät samankaltaista seurantakoodia. DNS nimien, avulla käännetään sivustojen IP-osoitteet ihmisystävällisenpään muotoon (CLOUDFLARE s. a. a). Netblocks eli IP-osoitteiden välityspalikka, joka reitittää tietyn IP-osoite ryhmän osoitteet. NS kirjanpidot, joka siirtää aliverkkotunnuksen nimipalvelimien joukkoon (Dnsimple s. a.). MX kirjapidot, joka määrittää postipalvelimen, joka vastaa sähköpostiviestien hyväksymisestä verkkotunnuksen puolesta (CLOUDFLARE s. a. b).



Kuva 8. Domain haku

Maltegon vahvuutena on selkeä käyttöliittymä ja sen tapa selvittää automaattisesti suuria tietomääriä ja niiden välisiä suhteita useista eri avoimista lähteistä ja esittää ne selkeässä graafisessa mallinnuksessa. Tämä auttaa tiedonetsijää analysoimaan tietoja, joiden avulla voidaan tehdä merkittäviä löytöjä. (Wondersmith_rae 2019.) Automatisoitujen kyselyprosessien yhdellä verkkotunnuksella voidaan saada valtavasti tietoa IT-verkkoinfrastruktuurista. Löydettyjen tietojen pohjalta saatujen tietojen määrä saattaa kasvaa todella suureksi, kuten huomataan kuvasta (Kuva 9). Ainoana Maltegon heikkoutena on sen rajoittuneisuus mahdollistaa käyttäjien kehittää ohjelmistoa oman tarpeen mukaan. Tämän seurauksena käyttäjä ei pysty vaikuttamaan, kuinka haut suoritetaan ohjelmallisesti, vaan käyttäjän täytyy hyväksyä Maltegon tavat suorittaa etsintöjä.

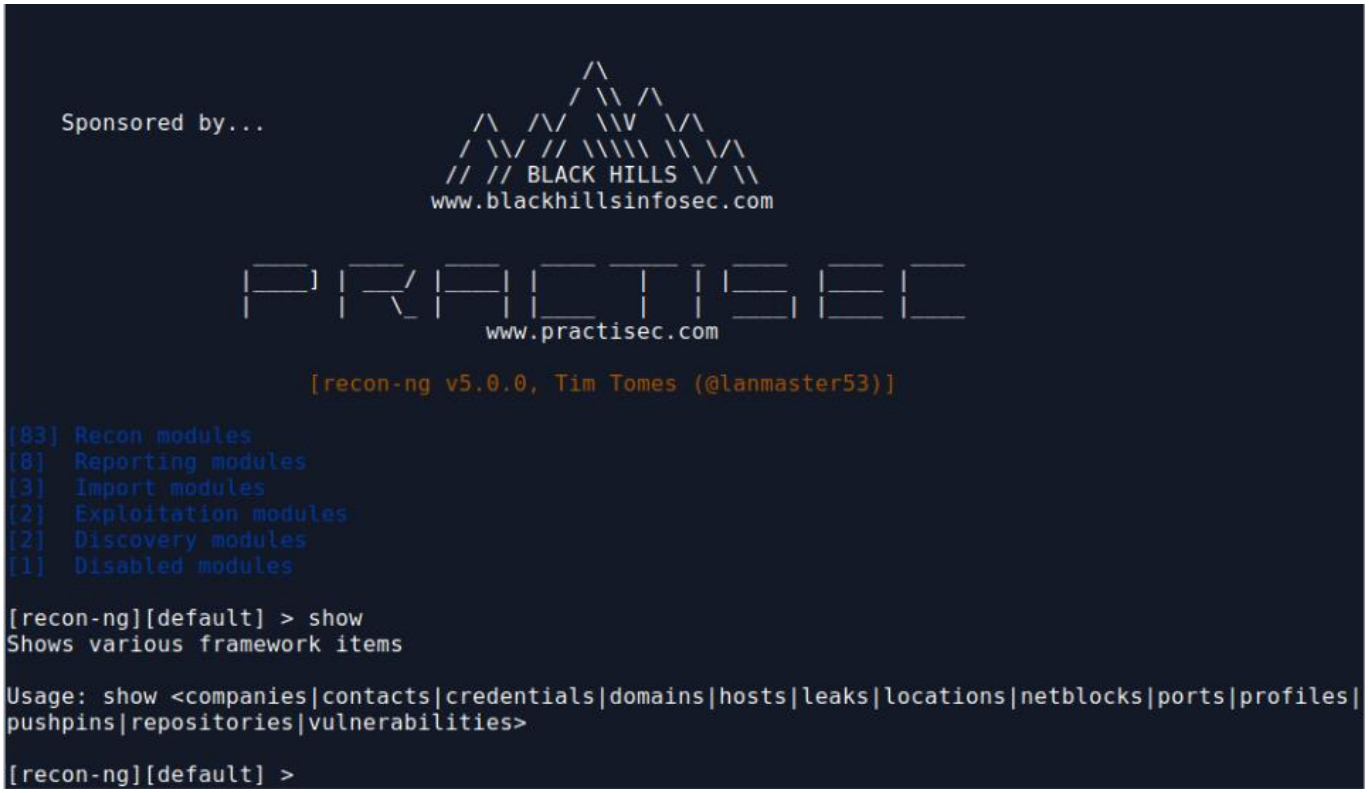


Kuva 9. Laajennetun haun tulokset

5.1.2 Regon-ng

Recon-ng on suunnitellut Tim Tomes (@LaNMaSteR53). Työkalu on tehokas laajaan verkkopohjaiseen avointen lähteiden tiedusteluun, joka toimii vain Linux käyttöjärjestelmällä. Tämän takia Recon-ng on tietoturva-asiantuntijoiden ja verkkorikollisten yleisessä käytössä. Recon-ng on komentokehote ohjelmisto (Kuva 10). Se sisältää itsenäisiä moduuleita, oman tietokannan ja selkeät ja helppokäyttöiset komennot halutun tiedon etsinnän suorittamiseen. Recon-ng avulla voidaan etsiä muun muassa verkkotunnuksia, jotka sisältävät tietyn merkkijonon. Sillä voidaan myös skannata verkkosivun mahdolliset haavoittuvuudet. Käytännössä Recon-ng avulla voidaan suorittaa saatavilla olevien moduulien sallimissa rajoissa mitä tahansa tiedusteluita. Käytettävä moduuli määrittelee tiedustelun laajuuden ja tavan, sekä onko tiedustelu passiivista-, puolipassiivista- vai aktiivista tiedonkeräystä. (Tomes 2020; GeekWire 2020.)

Ohjelmistokehys on koodattu Python ohjelmointikielellä ja itsenäiset moduulit mahdollistavat ulkopuolisten liittymisen mukaan ohjelmiston kehittämiseen, mutta Tim Tomes on tällä hetkellä pääsijainen ylläpitäjä. Recon-ng komentojen suorittaminen tapahtuu komentosyöte riville (Kuva 10). Saatavilla olevat komennot saadaan näkyviin kirjottaen komennon ”help” avulla. Näillä komennoilla voidaan muun muassa luoda uusia tietokantoja (Workspaces), lisätä kohteeksi haluttu verkkotunnus (Domain), etsiä tietokantaan lisättyjä tietoja (Show), etsiä uusia moduuleita ja ladata niitä (Marketplace). Tietyt moduulit, esimerkiksi Twitter ja LinkedIn, kuitenkin vaativat API-avaimen, joka saattaa vaatia rekisteröitymisen moduulin käyttämälle ohjelmistopohjalle.



Kuva 10. Recon-ng käyttöliittymä

Ala puolella olevien kuvien avulla havainnollistetaan Recon-ng peruskomentoja ja yleisesti, kuinka käyttöliittymä toimii. Perusteiden jälkeen perehdytään, miten ja mihin moduuleita voidaan käyttää.

Tiedon etsintä aloitetaan luomalla tietokanta (Kuva 11). Tietokannan avulla tehdyt tiedustelun tulokset liitetään tehdyn tietokannan sisältöön, joka helpottaa tietojen käsittelyä jälkeen päin. Tietokanta lista sisältää kaikki luodut

tietokannat ja tietokannan voi valita ”workspaces select <tietokannan nimi>” komennon avulla. Tietokantoihin voidaan luoda verkkotunnuksia, joihin halutut kyselyt suoritetaan (Kuva 12). Syötetty verkkotunnus näkyy verkkotunnus listassa, joka on tietokanta kohtainen.

```
[recon-ng][default] > workspaces create Test
[recon-ng][Test] > workspaces list

  +------------+
  | Workspaces |
  +------------+
  | Test       |
  | default    |
  +------------+
```

Kuva 11. Tietokannan luonti

```
[recon-ng][Test] > db insert domains
domain (TEXT): Test.fi
[recon-ng][Test] > show domains

  +-------------------------------+
  | rowid |  domain |    module   |
  +-------------------------------+
  | 1     | Test.fi | user_defined |
  +-------------------------------+
```

Kuva 12. Verkkotunnuksen lisäys tietokantaan

Käytettävien moduulien API-avainten lista nähdään kuvasta (Kuva 14). Listassa on kaikki ladatut moduulit, jotka vaativat etsinnöissä API-avaimen. Avaimen lisäys tietylle moduulille tapahtuu helposti komennolla ”keys add <Moduulin nimi> <API-avain>”, kuten huomataan kuvasta (Kuva 13; Kuva 15).

```
[recon-ng][Test] > keys add shodan_api 12345
```

Kuva 13. Esimerkki API-avaimen lisäyksestä.

```
[recon-ng][Test] > keys list

  +--------------------------+
  |       Name       | Value |
  +--------------------------+
  | binaryedge_api   |       |
  | bing_api         |       |
  | builtwith_api    |       |
  | censysio_id      |       |
  | censysio_secret  |       |
  | flickr_api       |       |
  | fullcontact_api  |       |
  | github_api       |       |
  | google_api       |       |
  | hashes_api       |       |
  | hibp_api         |       |
  | ipinfodb_api     |       |
  | ipstack_api      |       |
  | namechk_api      |       |
  | pwnedlist_api    |       |
  | pwnedlist_iv     |       |
  | pwnedlist_secret |       |
  | shodan_api       |       |
  | twitter_api      |       |
  | twitter_secret   |       |
  | virustotal_api   |       |
  +--------------------------+
```

Kuva 14. Tyhjä API-avain lista

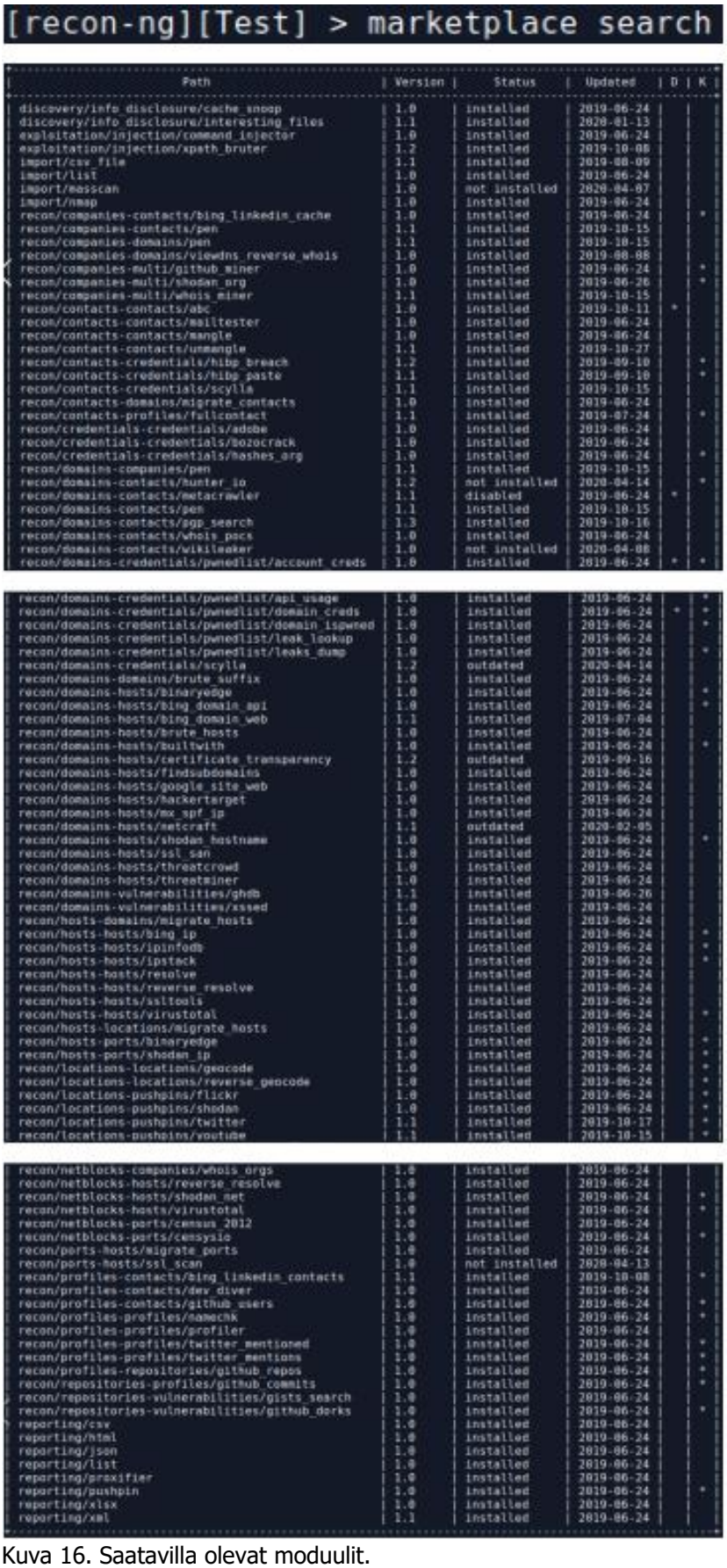
```
[recon-ng][Test] > keys list

  +--------------------------+
  |       Name       | Value |
  +--------------------------+
  | binaryedge_api   |       |
  | bing_api         |       |
  | builtwith_api    |       |
  | censysio_id      |       |
  | censysio_secret  |       |
  | flickr_api       |       |
  | fullcontact_api  |       |
  | github_api       |       |
  | google_api       |       |
  | hashes_api       |       |
  | hibp_api         |       |
  | ipinfodb_api     |       |
  | ipstack_api      |       |
  | namechk_api      |       |
  | pwnedlist_api    |       |
  | pwnedlist_iv     |       |
  | pwnedlist_secret |       |
  | shodan_api       | 12345 |
  | twitter_api      |       |
  | twitter_secret   |       |
  | virustotal_api   |       |
  +--------------------------+
```

Kuva 15. Listaan lisätty API-avain.

Kaikki saatavilla olevat moduulit löytyvät marketplace lataus sivulta (Kuva 16). Moduulin lataus tapahtuu ”marketplace install <Moduulin polku>” komennolla. Kaikki ladatut moduulit löytyvät modules search sivulta (Kuva

17) ja moduulin käyttöönotto tapahtuu ”modules load <moduulin nimi>” komennolla. On hyvä myös osata liikkua takaisin valitun moduulin valikosta ja se tapahtuu ”back” komennolla, jolla päästään takaisin perustilaan.



Kuva 16. Saatavilla olevat moduulit.

Kuva 17. Ladatut moduulit

Recon-ng sisältää monia eri moduuleita eri osa-alueilta, kuten Discovery, Exploitation, Import, Recon ja Reporting, kuten kuvasta huomataan (Kuva 17). Kaikkien moduuleiden läpi käyminen ei ole mahdollista tässä tutkielmassa. Työssä perehdytään vain osaan käytetyimmistä moduuleista, saaden täten peruskäsityksen kuinka Recon-ng toimii käytännössä. Ala puolella on esitetty käytettävät moduulit:

- recon/domains-contact/whois-pocs
- recon/domains-hosts/bing_domain_web
- recon/domains-hosts/brute_hosts
- recon/hosts-hosts/resolve
- recon/hosts-hosts/reverse_resolve
- discovery/info_disclosure/interesting_files
- recon/profiles-profiles/profiler

Moduulit ovat yksilöllisiä ja toimivat tietyn periaatteen mukaan, mutta niiden käyttäminen ja arvojen muuttaminen toimivat samalla tavalla. Esimerkiksi moduuleiden ”Options” välilehden tietoja voidaan muuttaa ”options set <option nimi> <uusi arvo>” komennolla, kuten nähdään esimerkistä (Kuva 18). Moduulien kerätyt tiedot kerätään taulukoihin, jotka on esitetty ”show” komennon avulla (Kuva 19). Halutun taulukon tiedot esitetään komennolla ”show <taulukko>”, kuten kuvassa (Kuva 20).

```
[recon-ng][Test][whois_pocs] > options set SOURCE Test.fi
SOURCE => Test.fi
```
Kuva 18. Moduulin Options tietojen muuttaminen

```
Usage: show <companies|contacts|credentials|domains|hosts|leaks|locations|netblocks|ports|profiles|pushpins|repositories|vulnerabilities>
```
Kuva 19. Taulukot, jotka sisältävät kerättyjä tietoja

```
[recon-ng][Test] > show contacts
```
Kuva 20. Komento, jolla saadaan taulukon tiedot näkyviin

WHOIS POCS MODUULI

ARIN Whois-palvelu on julkinen resurssi, jonka avulla käyttäjä voi hakea tietoja IP-numeroresursseista, organisaatioista, POC-tiedoista, asiakkaista ja muista yhtiöistä (Arin s. a. a). Moduulin whois_pocs avulla kerätään pelkästään yhteyspiste (POC, Points of Contact) tietoa, eli henkilöä tai tiettyä osaa ARIN tietokannasta. POC-tiedot määritetään yhteystiedoilla, mukaan lukien henkilön nimi, sähköpostiosoite, postiosoite ja puhelinnumero (Arin s. a. b). Moduulin valinnan jälkeen voidaan syöttää ”info” komentokenttään, joka antaa tiedot käytettävästä moduulista (Kuva 21).

```
[recon-ng][Test.fi] > modules load whois_pocs
[recon-ng][Test.fi][whois_pocs] > info

      Name: Whois POC Harvester
    Author: Tim Tomes (@lanmaster53)
   Version: 1.0

Description:
  Uses the ARIN Whois RWS to harvest POC data from whois queries for the given domain. Updates the
  'contacts' table with the results.

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  default        yes       source of input (see 'show info' for details)

Source Options:
  default        SELECT DISTINCT domain FROM domains WHERE domain IS NOT NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs
```

Kuva 21. Moduulin sisältävä informaatio.

Moduulin tiedonkeräyksen perehtymisen jälkeen voidaan aloittaa moduulin tietojenkeräys ”run” komennolla (Kuva 22). Komennon jälkeen moduuli alkaa keräämään POC-tietoja. Tämän kaltainen tiedustelu hyödyntää pelkästään julkista resurssia, jonka seurauksena tiedustelu on passiivista tiedonkeräystä.

```
[recon-ng][Test][whois_pocs] > run
```

Kuva 22. Whois_pocs moduulin suoritus komento.

BING DOMAIN WEB MODUULI

Bing_domain_web moduulin avulla voidaan kerätä hosteja, jotka sisältävät syötetyn verkkotunnuksen Bing.com hakukoneen avulla (Kuva 23). Moduuli kerää kaikki hostit ”site” etsintäparametrin avulla ja etsinnän jälkeen esittää ne hosts taulukossa.

```
[recon-ng][Test.fi] > modules load bing_domain_web
[recon-ng][Test.fi][bing_domain_web] > info

      Name: Bing Hostname Enumerator
    Author: Tim Tomes (@lanmaster53)
   Version: 1.1

Description:
  Harvests hosts from Bing.com by using the 'site' search operator. Updates the 'hosts' table with the
  results.

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  default        yes       source of input (see 'show info' for details)

Source Options:
  default        SELECT DISTINCT domain FROM domains WHERE domain IS NOT NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs
```

Kuva 23. Bing_domain_web moduulin sisältävä informaatio.

Etsintä voidaan suorittaa samalla tavalla, kun aikaisemmin ”run” komennolla (Kuva 24). Komennon jälkeen moduuli alkaa keräämään hosteja. Etsinnän jälkeen tulokset voidaan käydä tarkistamassa ”show hosts” komennolla.

```
[recon-ng][Test][bing_domain_web] > run
```

Kuva 24. Bing_domain_web moduulin suoritus komento.

BRUTE_HOST MODUULI

Brute_host moduulin toimintaperiaate perustuu suurien määrien tietoliikenteen lähettämiseen DNS-serverille pyrkien arvaamaan serverin salasana hyödyntäen salasanalistaa, kuten nähdään kuvasta (Kuva 25). Brute force hyökkäyksen avulla pakotetaan DNS-serveri luovuttamaan sen käsittelemät IP-osoitteet hyökkääjän tietoisuuteen. Tämän kaltainen tiedustelu määritellään aktiiviseksi tiedonkeruuksi. DNS-serverin ylläpitäjä pitää tämän kaltaista tiedustelua epäilyttävänä-, haitallisena- ja jopa rikollisena toimintana. Sen seurauksena täytyy olla hyvin tietoinen mihinkä tämän kaltaista tiedustelua hyödyntää.

```
[recon-ng][Test.fi] > modules load brute_hosts
[recon-ng][Test.fi][brute_hosts] > info

      Name: DNS Hostname Brute Forcer
    Author: Tim Tomes (@lanmaster53)
   Version: 1.0

Description:
  Brute forces host names using DNS. Updates the 'hosts' table with the results.

Options:
  Name      Current Value                          Required  Description
  --------  -------------                          --------  -----------
  SOURCE    default                                yes       source of input (see 'show info' for details)
  WORDLIST  /home/csi/.recon-ng/data/hostnames.txt  yes       path to hostname wordlist

Source Options:
  default        SELECT DISTINCT domain FROM domains WHERE domain IS NOT NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs
```

Kuva 25. Brute_host moduulin sisältävä informaatio.

Komennon suorittaminen tapahtuu, kuten aikaisemminkin ”run” komennolla, jonka jälkeen brute_host moduuli, alkaa hyökkäämään kohteen DNS-serveriin (Kuva 26).

```
[recon-ng][Test.fi][brute_hosts] > run
```

Kuva 26. Brute_host moduulin suoritus komento.

RESOLVE MODUULI

Resolve moduuli selvittää IP-osoitteiden avulla niiden host-nimet (Kuva 27). Tämän etsinnän perusteella saadaan tietoa siitä, kenelle sivusto kuuluu ja mitä mahdollisesti sivusto saattaa sisältää. Tämän moduulin suorittaminen vaatii, että ”hosts” taulukko sisältää IP-osoitteita, jotta voidaan selvittää IP-osoitteiden host-nimet.

```
[recon-ng][Test.fi] > modules load recon/hosts-hosts/resolve
[recon-ng][Test.fi][resolve] > info

      Name: Hostname Resolver
    Author: Tim Tomes (@lanmaster53)
   Version: 1.0

Description:
  Resolves the IP address for a host. Updates the 'hosts' table with the results.

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  default        yes       source of input (see 'show info' for details)

Source Options:
  default        SELECT DISTINCT host FROM hosts WHERE host IS NOT NULL AND ip_address IS NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs

Comments:
  * Note: Nameserver must be in IP form.
```

Kuva 27. Resolve moduulin sisältävä informaatio.

Komennon suorittaminen tapahtuu ”run” komennolla, jonka jälkeen moduuli alkaa kääntämään IP-osoitteita host-nimiksi (Kuva 28). Saadut host-nimet esitetään ”hosts” taulukossa.

```
[recon-ng][Test.fi][resolve] > run
```

Kuva 28. Resolve moduulin suoritus komento.

REVERSE RESOLVE MODUULI

Reverse_resolve moduuli toimiin toisinpäin kuin reverse moduuli, joka käsiteltiin yläpuolella. Moduuli selvittää IP-osoitteet host-nimen perusteella (Kuva 29). Moduulin suorittaminen vaatii, että ”hosts” taulukko sisältää host-nimiä, jotta voidaan selvittää host-nimien IP-osoitteet.

```
[recon-ng][Test.fi] > modules load recon/hosts-hosts/reverse_resolve
[recon-ng][Test.fi][reverse_resolve] > info

      Name: Reverse Resolver
    Author: John Babio (@3vi1john), @vulp1n3, and Tim Tomes (@lanmaster53)
   Version: 1.0

Description:
  Conducts a reverse lookup for each IP address to resolve the hostname. Updates the 'hosts' table
  with the results.

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  default        yes       source of input (see 'show info' for details)

Source Options:
  default        SELECT DISTINCT ip_address FROM hosts WHERE ip_address IS NOT NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs
```

Kuva 29. Reverse_resolve moduulin sisältävä informaatio.

Moduuli suoritetaan ”run” komennolla, jonka jälkeen moduuli alkaa kääntämään host-nimiä IP-osoitteiksi (Kuva 30). Saadut IP-osoitteet esitetään ”hosts” taulukossa.

```
[recon-ng][Test.fi][reverse_resolve] > run
```

Kuva 30. Reverse_resolve moduulin suoritus komento.

INTERESTING_FILES MODUULI

Interesting_file moduuli etsii perusasetuksilla syötetyn host-nimen perustella portin 80 kautta sivuston hakemistossa sisältäviä teksti-, loki- ja status tietoja. Tiedostojen avulla voidaan saada tietoa sivuston haavoittuvuuksista ja rakenteesta (Kuva 31).

```
[recon-ng][Test] > modules load interesting_files
[recon-ng][Test][interesting_files] > info

      Name: Interesting File Finder
    Author: Tim Tomes (@lanmaster53), thrapt (thrapt@gmail.com), Jay Turla (@shipcod3), and Mark Jeffery
   Version: 1.1

Description:
  Checks hosts for interesting files in predictable locations.

Options:
  Name      Current Value  Required  Description
  --------  -------------  --------  -----------
  DOWNLOAD  True           yes       download discovered files
  PORT      80             yes       request port
  PROTOCOL  http           yes       request protocol
  SOURCE    default        yes       source of input (see 'show info' for details)

Source Options:
  default        SELECT DISTINCT host FROM hosts WHERE host IS NOT NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs

Comments:
  * Files: robots.txt, sitemap.xml, sitemap.xml.gz, crossdomain.xml, phpinfo.php, test.php, elmah.axd,
  server-status, jmx-console/, admin-console/, web-console/
  * Google Dorks:
    - inurl:robots.txt ext:txt
    - inurl:elmah.axd ext:axd intitle:"Error log for"
    - inurl:server-status "Apache Status"
```

Kuva 31. Interesting_files moduulin sisältävä informaatio.

Moduuli suoritetaan ”run” komennolla, jonka jälkeen moduuli etsii host-nimen avulla sivuston sisältämät hakemiston tiedot (Kuva 32).

```
[recon-ng][Test][interesting_files] > run
```

Kuva 32. Interesting_files moduulin suoritus komento.

PROFILER MODUULI

Profiler moduuli etsii käyttäjätunnusta vastaavaa käyttäjää lukuisilta sivustoilta (Kuva 33). Näin voidaan kartoittaa syötetyn käyttäjätunnuksen verkkopalveluiden käyttöä ja saada paljon hyödyllistä tietoa kohteen verkko käyttäytymisestä.

```
[recon-ng][Test] > modules load profiler
[recon-ng][Test][profiler] > info

      Name: OSINT HUMINT Profile Collector
    Author: Micah Hoffman (@WebBreacher)
   Version: 1.0

Description:
  Takes each username from the profiles table and searches a variety of web sites for those users. The
  list of valid sites comes from the parent project at https://github.com/WebBreacher/WhatsMyName

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  default        yes       source of input (see 'show info' for details)

Source Options:
  default        SELECT DISTINCT username FROM profiles WHERE username IS NOT NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs

Comments:
  * Note: The global timeout option may need to be increased to support slower sites.
  * Warning: Using this module behind a filtering proxy may cause false negatives as some of these
  sites may be blocked.
```

Kuva 33. Profiler moduulin sisältävä informaatio.

Profiler moduuli suoritetaan ”run” komennolla, jonka jälkeen moduuli alkaa käymään läpi sivustoja käyttäjänimen avulla ja kokoaa löydetyt käyttäjätunnus sivustot ”profiles” taulukkoon (Kuva 34).

```
[recon-ng][Test][profiler] > run
```

Kuva 34. Profiler moduulin suoritus komento.

Recon-ng mahdollistaa laajojen tiedusteluiden tekemisen eri moduuleiden avulla, kun käyttäjällä on moduulin ohjelmiston API-avain. Käyttäjälle on myös annettu oikeudet modifioida ja kehittää Recon-ng:n rakennetta oman tarpeensa mukaan. Tämä on yksi merkittävimmistä Recon-ng:n ominaisuuksista. Recon-ng on pääsääntöisesti helppokäyttöinen, jos käyttäjä on tottunut käyttämään komentokehotetta. Ainoana Recon-ng:n huonona puolena voidaan pitää tietojen analysointia ja prosessointia. Kerätyt tiedot on tallennettu perinteisiin taulukkoihin ja ne voidaan esittää myös verkkosivulla ”./recon-web” moduulin avulla tai ladata Excel-tiedostoon. Tiedot ovat staattisessa muodossa ja tietojen analysointi, ja tietojen välisten suhteiden määrittely on tämän seurauksena hieman haastavaa. Toisaalta Recon-ng mahdollistaa moduuleiden kehittämisen sekä muokkaamisen ja se antaa käyttäjälle mahdollisuuden luoda käyttäjäystävällisimpiä tietojen analysointi moduuleita.

5.1.3 Google Dorks

Google Dork:silla tarkoitetaan Googlen hakukoneelle suoritettavia kyselyitä, joiden avulla voidaan löytää verkkoon vahingossa vuodettuja tietoja, jotka eivät ole saatavilla tavanomasilta verkkosivustoilta. Tämän kaltainen passiivinen tietojen keräys perustuu avainsanoihin, joiden avulla saadaan suodatettua hakutulokset hakukoneen indekseistä. Kyselyt voivat tuoda esiin tietoja, joita ei ole tarkoitettu julkiseen jakoon, mutta niitä ei ole myöskään merkittävästi suojattu. Kysely parametrien avulla voidaan saada tietoa käyttäjätunnuksista, salasanoista, sähköposteista, arkaluontoisista dokumenteista, henkilökohtaisista tunnistettavista taloudellisista tiedoista, verkkosivujen haavoittuvuuksista ja avoimista porteista. (SecurityTrails Team 2019).

Etsintäparametrit syötetään Googlen syöte kenttään, kuten perinteiset Google haut. Etsintäparametreina toimivat muun muassa taulukossa esitetyt parametrit (Taulukko 2).

TAULUKKO 2. Google Dorks parametrit (SecurityTrails Team 2019; Alexis 2020.)

Parametri	Määritelmä	Esimerkki
site:	Etsii vain määritellyltä sivustolta tulokset.	site: Google.com
filetype:	Haetun tiedostotyypillä löytyvät tiedot.	filetype: jpg
inurl:	Etsii URL-osoitteita, jotka sisältävät syötteen.	inurl: http
intext:	Etsii sivustoja, jotka sisältävät syötetyn tekstin.	intext: admin
cache:	Etsii välimuistissa olevat sivuston versiot.	cache: www.google.com
allintext:	Sivustot, joissa on käytetty kyseisiä syötettyjä sanoja.	allintext: password list
intitle:	Etsii verkkosivujen otsikoita, kyseisen syötteen perusteella.	intitle: index of
inanchor:	Etsii linkkejä, jotka sisältävät kyseisen syötteen.	inanchor: my resume
link:	Sivustot, jotka sisältävät linkin kyseiselle URL syötteelle.	link: www.google.com
*:	Etsii kaikki sivustot, jotka sisältävät annetun syötteen.	*:8000
|	Looginen operaattori, jolla voidaan etsiä sivustoja, jotka sisältävät molemmat annetut syötteet.	login | password
+	Etsii sivustoja, joissa on enemmän, kuin yksi syöte	username + password
-	Sivustoja, jotka eivät sisällä syötettä.	-error
after:	Verkkosivuja, jotka on julkaistu syötteen jälkeen.	after:2018

Näiden hakuparametrien avulla voidaan etsiä tietoja verkosta haluttujen kriteerien mukaan. Tietojen väärinkäyttö voi tuoda suuria ongelmia yrityksille ja yksityisille henkilöille. Salattujen tietojen löytyminen verkosta perustuu Googlen toimintamalliin käydä läpi verkkoon laitettavan sivuston sisältämät tiedot. Näiden tietojen perusteella

Googlen hakukone indeksoi tiedot hakukoneen kriteereihin, jonka avulla tiedot voidaan etsiä hakukoneella. Hakukone löytää huonosti suunnitellut verkkosivut, esimerkiksi sellaiset, jotka mahdollistavat käyttäjien pääsyn sivuston hakemistoon. Hakemisto sisältää sivustolle tärkeitä kansioita liittyen esimerkiksi käyttäjätunnusten ja salasanojen säilyttämiseen. Näin ollen kuka tahansa voi löytää nämä arkaluontoiset tiedostot, jos osaa vain hyödyntää hakuparametrejä. (Cyber Army 2020.)

Ala puolella on esitetty kyselyitä käyttäen parametreja, joiden avulla voidaan saada merkittäviä tietoja. Tietoja voidaan saada muun muassa tietokannoista, raporteista, sähköposteista, salasanoista ja käyttäjätunnuksista. (Alexis 2020; Kinzie 2019.)

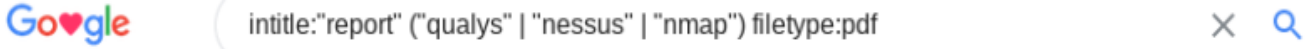
1) Kyselyllä etsitään Excel-tiedostoja, jotka sisältävät sähköpostiosoitteita.

KUVA 35. Google Dorks haku 1

2) Kyselyllä etsitään sivuja, jotka mahdollistavat pääsyn sivun hakemistoon ja sitä kautta wp-config.php tiedostoon. Tämän kaltaiset config.php tiedostot sisältävät pääsytiedot tietokantoihin, joka mahdollistaa pääsyn kaikkiin tietoihin mitä tietokantaan on tallennettu.

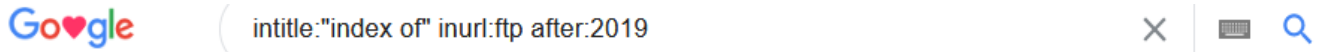
KUVA 36. Google Dorks haku 2

3) Kyselyllä etsitään sivustoja, jotka sisältävät tietoturvaraportteja PDF muodossa. Qualys, Nessus ja Nmap ovat haavoittuvuustestauksessa käytettäviä työkaluja. Sen takia, jos raportti sisältää, jonkun näistä työkaluista se tarkoittaa, että raportti saattaa sisältää tärkeää tietoa kohteena olevan tietoturvallisuudesta.



KUVA 37. Google Dorks haku 3

4) Kyselyllä etsitään tiedonsiirtopalvelimelle (FTP, File Transfer Protocol) julkaistuja tietoja. Tiedot eivät ole julkisia, mutta Googlen hakukoneen indeksoinnin ansiosta tiedostot tulevat julkisiksi.



KUVA 38. Google Dorks haku 4

5) Kyselyllä etsitään salasanoja, jotka on tallennettu Excel-tiedostoon.

KUVA 39. Google Dorks haku 5

6) Kyselyllä etsitään avoimia verkossa olevia web-kameroita, joiden käyttöä ei ole rajoitettu millään tavalla.

KUVA 40. Google Dorks haku 6

Google Dorks on tarkoitettu passiiviseen tiedon keräykseen, johonka OSINT-tiedustelussa yleensä pyritään. Kerätyt tiedot pohjautuvat Google hakukoneen indeksoimiin sivustojen sisältämiin tietoihin. Tietoja etsitään erityisillä hakukonekyselyiden avulla, pyrkien saamaan tietoa kohteen haavoittuvuuksista ja mahdollisista vuodetuista tiedoista. Google Dorks voi olla yllättävän tehokas tietojen keräämisessä, mutta tiedot voivat olla vanhentuneita ja voivat rajoittua vain arkistotietoihin. Tämä on Google Dork:sin kaltaisen tiedon keräyksen haittapuoli.

5.1.4 Shodan

Shodan on hakukone, joka indeksoi lähes kaikki laitteet, jotka ovat yhteydessä Internettiin maailmanlaajuisesti. Shodanin käyttö vaatii vain rekisteröitymisen Shodanin käyttäjäksi, jotta saa oikeuden 10 000 kyselyn suorittamiseen kuukauden aikana. Maksullisissa versioissa voidaan kasvattaa kyselyiden-, skannattujen IP-osoitteiden- ja filttereiden määrää (Taulukko 3). Shodanin hakukone toimii samankaltaisesti kuin Googlen hakukone, toisin vain Shodan ei indeksoi verkkosivuja vaan IPv4- ja IPv6 osoitteita, selvittäen näin verkkoon liitetyt laitteet ja avoimet portit. Liitettyjen laitteistojen ja avoimien porttien tiedot saadaan FTP-bannerin avulla, joka ilmoittaa avoimesti, mitä palvelua se tarjoaa ja kuinka siihen on mahdollista ottaa yhteys. FTP-banneri on pääasiallisesti metadataa, joka sisältää tietoa sijainnista, oletusarvo käyttäjätunnuksesta ja salasanasta, IP osoitteesta, ohjelmiston versiosta, tekijästä ja mallista. (Shodan 2020a).

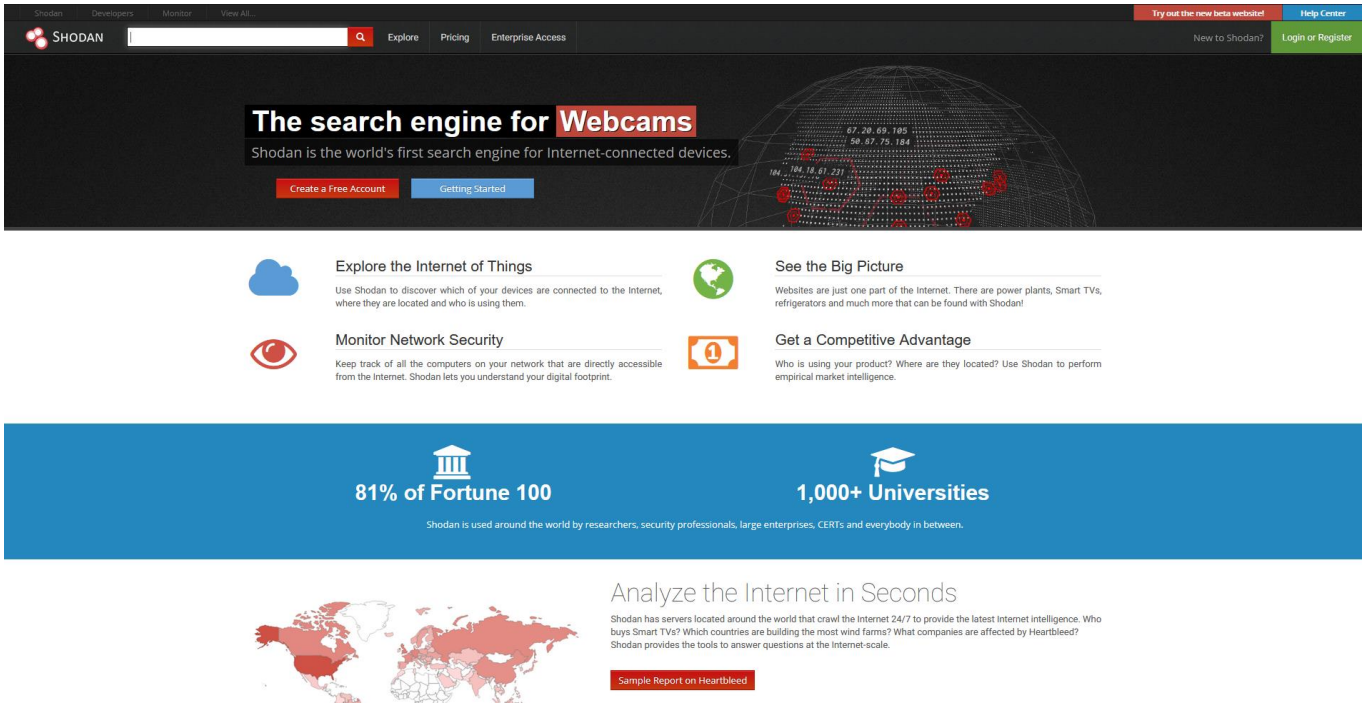
TAULUKKO 3. Shodan versiot (Shodan 2020c.)

Versio	Sisältää		Kuukaudessa
Freelance	- Miljoona kyselyn tulosta kuukaudessa. - Voidaan skannata ja valvoa 5120 IP-osoitetta kuukaudessa. - Mahdollistaa pääsyn useampiin filttereihin. - Perusoikeus streaming-sovellusliittymään	- Sähköpostituki	54,44 €
Small Business	- 20 miljoonaa kyselyn tulosta kuukaudessa. - Voidaan skannata ja valvoa 65 536 IP-osoitetta kuukaudessa. - Mahdollistaa pääsyn useampiin filttereihin. - Mahdollistaa tulosten selailun. - Perusoikeus streaming-sovellusliittymään	- Sähköpostituki - Haavoittuvuushakusuodatin	275,9 €
Corporate	- Rajaton kyselyiden suorittaminen. - Voidaan skannata ja valvoa 300 000 IP-osoitetta kuukaudessa. - Mahdollistaa pääsyn kaikkiin filttereihin. - Mahdollistaa tulosten selailun. - Perusoikeus streaming-sovellusliittymään.	- Premium-tuki. - IP-joukkohakujen suoritus. - Tag-hakusuodatin. - Lisäjäsenyyspäivitykset.	829,54 €

Shodanin tarjoamia tietoja voidaan hyödyntää verkkosivujen- ja IoT-laitteiden etsinnässä, verkkoturvallisuudessa sekä yleisessä tutkimuksessa; mitä laitteistoja nykypäivänä käytetään. Tämän seurauksena Shodan on yksi merkittävimmistä OSINT työkaluista, sillä teknologian kehityksen myötä suurin osa nykypäivän laitteista on yhteydessä verkkoon. Näitä tietoja pyrkivät hyödyntämään tietoturva-asiantuntijat, akateemiset tutkijat ja valtionvirastot. (Wilson 2019.)

Shodanin käyttö perustuu kyselyihin, samankaltaisesti, kuin Google Dorks:issa. Hakukonetta voidaan käyttää, joko Shodanin verkkosivulta (Kuva 41) tai sitten komentokehote-kirjaston avulla, joka pohjautuu Python ohjelmointikieleen. Tässä työssä hyödynnetään Shodanin tarjoamaa verkkosivua kyselyiden suorittamiseen. Kyselyitä voidaan suorittaa bannerin sisältämän tiedon pohjalta käyttämällä filttereitä tietojen kohdentamiseen. Tällä hetkellä Shodanin suosituimpia hakuja ovat Webcam, Cams, Router, FTP, Netgear, SSH ja Ufanet. (Shodan 2020b). Näitä hakuja

voidaan soveltaa filttereiden kanssa, joiden avulla tarkennetaan hakukriteereitä. Hyödynnetyimmät filtterit on listattu taulukkoon (Taulukko 4).



KUVA 41. Shodan aloitussivu

TAULUKKO 4. Shodanin käytetyimmät filtterit (JavierOlmedo 2018.)

Filtterit	Määritelmä	Esimerkki
country:	Kysellyllä voidaan määrittää haku tiettyyn maahan.	country:FI
city:	Kysellyllä voidaan määrittää haku tiettyyn kaupunkiin.	city:Helsinki
port:	Kyselyllä etsitään laitteistoja, jossa on tietty portti avoinna.	port:8080
os:	Kyselyllä voidaan etsiä haluttua laitteistoa, jossa on tietty käyttöjärjestelmä	os:windows xp
before:	Kysely, joka palauttaa tulokset ennen päiväystä.	before:13/04/2020
after:	Kysely, joka palauttaa tulokset päiväyksen jälkeen.	after:01/04/2020

Shodania käytetään suurimmaksi osaksi avoimien porttien etsimiseen, joka on yksi tärkeimmistä avointen lähteiden tiedustelun tiedoista. Avoimet portit voivat mahdollistaa verkkorikollisille pääsyn arkaluontoisiin tietoihin. Tämän seurauksena mitä enemmän organisaatiolla on avoimia portteja, sitä suurempi riski sen on kohdata tietoturvahyökkäys. (GELNAW 2019.) TCP-protokola on merkittävässä roolissa yhteyksien luonnissa porttien välille. TCP (Transmission Control Protocol) on tietoliikenneprotokolla, joka huolehtii tiedonsiirron tietokoneiden välillä, joilla on pääsy Internettiin. TCP toimii Internet Protocol (IP)-sovelluksen kanssa, jonka avulla tiedonsiirto saadaan tapahtumaan haluttuun IP-osoitteeseen. (Rose 2020.) Alapuolelle on listattu tärkeimpiä portteja ja niiden määritelmiä, jotka Shodan kartoittaa etsinnöissään (Taulukko 5).

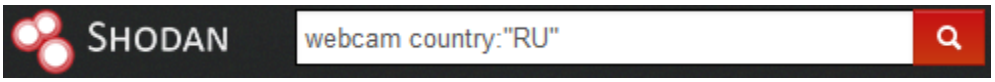
TAULUKKO 5. Avointen porttien määritelmät (ExamCollection s. a.; Wilson 2019; Zeng 2018; Tyson ja JavaWorld 2019.)

Port	Määritelmä
20, 21 = FTP	File Transfer Protocol (FTP) käytetään verkonvälisessä tiedonsiirrossa, joka hyödyntää käyttäjän todennusta ennen tietojen siirtoa. Liitos TCP-protokolaan tapahtuu vasta, kun molemmat portit 20 ja 21 saavat luvan tiedonsiirtoon, joka on todennettu käyttäjätunnuksella ja salasanalla.
22 = SSH	Secure Shell (SSH) porttia käytetään salattuun tietoliikenteeseen, joka muodostaa yhteyden etäpalvelimeen tai hostiin. Portti mahdollistaa myös valtuutettujen henkilöiden kirjautumisen järjestelmiin ja etuoikeuden tietojensiirtoon eri verkkojen välillä.
23 = TELNET	Portin pääasiallinen tehtävä on luoda yhteys palvelimen ja etätietokoneen-, reitittimen- ja kytkimien välille. Porttia ei voida käyttää suojattujen yhteyksien luomiseen.
25 = SMTP	Simple Mail Transfer Protocol (SMTP), tunnetaan sähköpostien välitys protokolana. Varmistaa, että sähköpostiviestit toimitetaan turvallisesti verkon kautta. Sähköpostiviestejä ei kuitenkaan voi ladata tätä kautta, vaan sitä käytetään ainoastaan viestien siirtämiseen.
53 = DNS	Domain Name System (DNS), toimii TCP- ja UDP-protokolien portissa 53. Portti hyödyntää relaatiotietokantoja tietokoneiden tai verkkojen verkkotunnuksien linkittämiseen vastaaviin IP-osoitteisiin. Portti odottaa DHCP pyyntöä datan siirtämiseksi verkon yli. Yhteyden muodostamisen jälkeen palvelin lähettää vyöhyketiedot TCP 53-porttia käyttäen. UDP protokolaa käytetään, kun kysely suoritetaan asiakastietokoneen kautta, mutta jos palvelin ei vastaa viiden sekunnin kuluttua DNS-kysely lähetetään TCP 53-portin kautta.
67, 68 = DHCP	Dynamic Host Configuration Protocol (DHCP), portti jakaa IP-osoitteiseen liittyvät tiedot verkossa oleville laitteille. Tiedot sisältävät reitin aliverkosta ulkoiseen verkkoon, nimipalvelimen (DNS) ja IP-osoitteen. Portin 67 tehtävänä on ottaa vastaan osoitepyyntöjä DHCP:ltä ja lähettää tiedot palvelimelle. Puolestaan portti 68 vastaa kaikkiin DHCP-pyyntöihin ja välittää tiedot eteenpäin.
80 = HTTP	Hypertext Transfer Protocol (HTTP) on yksi maailman käytetyimmistä porteista ja se kuuluu TCP protokolaan. Sen tarkoitus on sallia selaimen muodostaa yhteys Internetin verkkosivulle. Sallimisen jälkeen protokolaa voidaan käyttää tiedonsiirtoon.
8080 = HTTP	Samankaltainen, kuin portti 80. Porttia 8080 käytetään tyypillisesti henkilökohtaisesti ylläpidetyssä verkkopalvelimessa. Portti 8080 on vain oletusarvoinen valinta verkkopalvelimelle.
110 = POP3	Post Office Protocol versio 3 (POP3), portin avulla voidaan ladata sähköpostiviestit SMTP-palvelimelta. Latauksen yhteydessä viestit poistetaan palvelimelta. Verkon ylisiirretyt tiedot ovat tekstimuodossa, eikä niitä ole suojattu millään tavalla.
143 = IMAP	Internet Message Access Protocol (IMAP), portin tarkoitus on hakea sähköposteja etäpalvelimelta tarvitsematta ladata sähköpostia. Sähköpostit ovat virtuaalimuistissa, joka mahdollistaa sähköpostien lataamisen ja lukemisen, käyttäjän todentamisen jälkeen.
993 = IMAP	Samankaltainen portti kuin 143, ainoana erona on vain, että viestintä on salattu Secure Socket Layer (SSL) avulla.
443 = HTTPS	Pohjautuu HTTP protokolaan, ainoana erona on, että siihen on lisätty Secure Socket Layer (SSL) tietoturvaominaisuus. Tämän ansiosta verkkoliikenne salataan ja todennetaan, ennen tiedonsiirtoa.
8443 = HTTPS	Samankaltainen portti kuin 443. 8443 on oletusportti, jota Tomcat hyödyntää salatun viestinnän SSL avaamiseen. Tomcat on alusta, jolla voidaan suorittaa java-verkkosovelluksia.

3389 = RDP	Remote Desktop Protocol (RDP), portin tarkoitus on muodostaa etäyhteys etätietokoneeseen ja antaa täydet oikeudet tähän tietokoneeseen. Portin käyttö vaatii Windows-käyttöjärjestelmän, sekä käyttäjäkohtaisten asetuksien määrittämisen.

Ala puolella on esitetty Shodan kyselyitä, hyödyntäen Shodan verkkopalvelua. Kyselyiden avulla pyritään havainnollistamaan, kuinka Shodan kyselyitä ja filttereitä voidaan käyttää ja mitä tietoa kyselyillä voidaan saada.

1. Kyselyllä etsitään kaikki avoinna olevat web-kamerat venäjän alueelta. Kameroiden avulla voidaan saada tietoa, paikkakunnista, rakennuksista, liikenteestä ja henkilöistä.

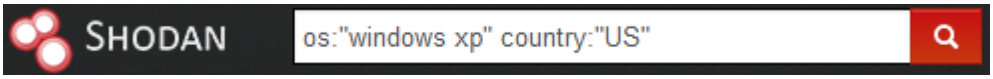


Kuva 42. Shodan haku 1.

2. Netgear kyselyllä voidaan etsiä kaikki Netgear merkkiset reitittimet, jotka ovat yhteydessä verkkoon. Lisättynä vielä port:23 komento saadaan kaikki reitittimet, joissa on portti 23, eli TELNET avoinna.

Kuva 43. Shodan haku 2.

3. Kyselyllä voidaan etsiä USA:sta kaikki tietokoneet, jotka vielä käyttävät Windows XP käyttöjärjestelmää. Käyttöjärjestelmän tuki loppui vuonna 2014, jonka seurauksena käyttöjärjestelmä on todella haavoittuvainen haitalliselle toiminnalle.



Kuva 44. Shodan haku 3.

4. Kyselyllä etsitään Singapore kaupungista kaikki laitteistot, jotka hyödyntävät Apache 2.2 http web-serveriä. Apache 2.2 version tuki on lopetettu 2017 joulukuussa, jonka seurauksena sen käyttämistä ei enää suositella. Vanhentunut versio sisältää haavoittuvuuksia, joita voidaan hyväksikäyttää.

Kuva 45. Shodan haku 4.

Shodan on yksi hyödyllisimmistä hakukone palveluista, jonka seurauksena se on myös saatavilla monissa muissa työkaluissa lisähaku asetuksena, kuten Maltegossa ja Recon-ng:ssä. Shodanin tapa mallintaa verkossa olevien laitteistojen tiedot ja haavoittuvuudet, antaa valtavasti tietoa tietoturva-asiantuntijalle ja valitettavasti myös verkkorikollisille. Saatujen tietojen avulla voidaan turvata yrityksien ja yksilöiden laitteistojen tietoturva, ennen kun niitä pystytään hyödyntämään haitallisiin tarkoituksiin. Ainoana huonona puolena voidaan pitää Shodanin ilmaisversion rajallista hakujen suorittamista, mutta ilmaisella versiolla pystytään suorittamaan pienimuotoisia tiedon etsintöjä, joka riittää useimmissa tapauksissa.

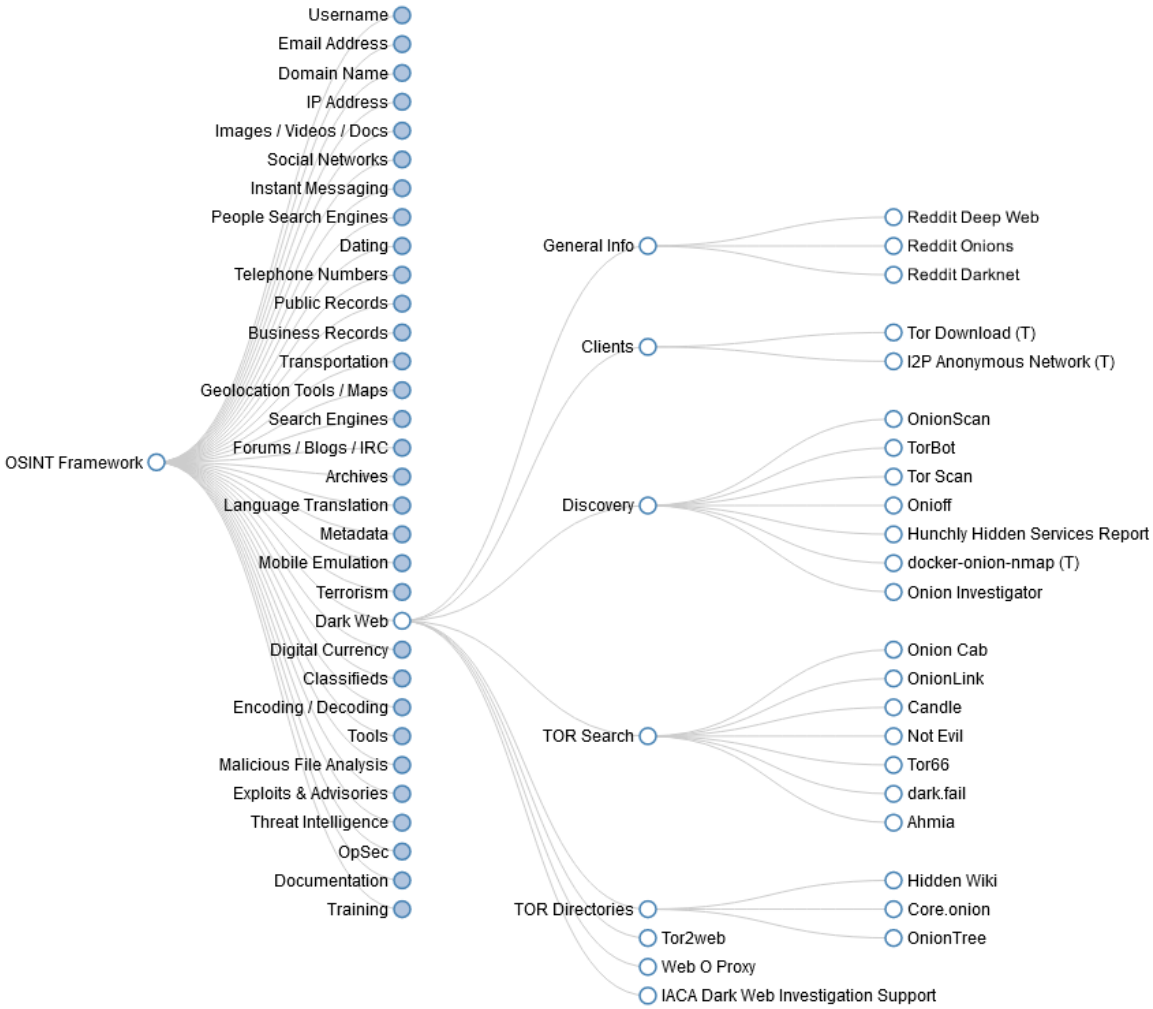
5.2 Avointen lähteiden tiedustelu pimeässä verkossa ja käytetyt työkalut

Tiedonkeräys pimeistä verkoista on oma taiteenlajinsa, jota Justin Nordine on pyrkinyt kuvaamaan avointen lähteiden rakenteessa (Kuva 46). Kuten aikaisemmin mainittiin pimeän verkon toiminta, perustuu anonyymiin ja salattuun toimintaan. Sivustoja ei ole indeksoitu hakukoneisiin, jonka seurauksena tiedon etsijän on tiedettävä juuri oikea URL-osoite tiedon etsintää varten. Pimeää verkkoa ei ole myöskään rajattu pelkästään tietylle kielelle ja tietyt pimeän verkon sivustot vaativat, että käyttäjätili on ollut pitkään aktiivisena. Tämän avulla voidaan suodattaa oikeat käyttäjät lainviranomaisista ja tiedustelijoista. (Ozkaya ja Islam 2019, 223-224.) Monikielisyys ja käyttäjätilien laillisuustarkastukset luovat haasteita tiedustelijoille. On huomattava myös, että pimeän verkon henkilöiden etsintään ei ole saatavilla täsmällisiä työkaluja tiedon löytymisen kannalta. Tämä hankaloittaa tiedon etsintää ja analysointia. Sen seurauksena henkilöiden etsintä pimeän verkon syvyyksistä, on pääsääntöisesti mahdotonta. Mikäli pimeää verkkoa käyttävä henkilö on tietoinen, kuinka pysyä nimettömänä, eikä tee virheitä pimeässä verkossa. Tällöin hänen identifioimisensa on todella vaikeaa. (Bertram 2015, 92.)

Pimeän verkon tietojen etsinnän tuomien haasteiden takia tiedustelijan tulee mukautua pimeän verkon nimettömyyteen. Tiedustelijan on pyrittävä hyödyntämään perinteisiä avointen lähteiden tiedustelun tekniikoita, henkilöiden tunnistamisessa ja löytämisessä. Suurin osa pimeän verkon laittomasta toiminnasta rahoitetaan kryptovaluutan avulla, kuten Bitcoin- ja Litecoin valuuttojen avulla. Kryptovaluutan rahavirtoja seuraamalla voidaan saada tietoa sivuston käyttäjän henkilöllisyydestä ja pimeän verkon toiminnasta. Toinen tekniikka on pohtia, kuinka pimeän verkon käyttäjät löytävät pääsynsä näille salaisille onion-sivustoille. Usein miten salaisten sivustojen URL-osoitteet jaetaan harkitusti pinta verkossa luotettaville pimeän verkon sivustoista kiinnostuneille henkilöille. Tämä on yksi mahdollisuus selvittää pimeän verkon käyttäjän identiteetti. Tätä periaatetta hyödyntäen tiedustelijan täytyy omata vahva epäaito verkkoprofiili, jota hyödyntäen hän voi päästä käsiksi salattuihin pimeän verkon sivustoihin ja sitä käyttäviin henkilöihin. (Bertram 2015, 93-95.)

Pimeän verkon avointen lähteiden tiedon etsintä, pohjautuu pääsääntöisesti manuaaliseen tiedon etsintää salatuilta sivustoilta. Saatavilla olevia salattuja sivustoja voidaan etsiä pintaverkossa olevien foorumeiden ja sivustojen avulla, joita havainnollistetaan kuvasta 46 ”General Info” kohdasta. Pimeän verkon tietojen etsinässä voidaan myös hyödyntää pimeän verkon hakukoneita. Hakukoneet, kuten esimerkiksi DuckDuckGo, Ahmia ja Kilo, ovat indeksoineet onion-sivustoja. Näiden avulla voidaan löytää muun muassa myynnissä olevia tuotteita, foorumeita, blogeja, arvosteluita ja kaikkea mikä vain täsmää hakukriteereihin. Hakukoneet sisältävät vain suosituimmat pimeän verkon sivustot, jonka seurauksena tiedon saatavuus saattaa olla suppea. Pimeän verkon indeksointia hyödyntävät myös lainviranomaiset, jotka ovat kehitelleet pimeän verkon indeksointirobotteja, joiden avulla he pystyvät seuraamaan ja valvomaan sivustoja. Pääasiallisesti indeksointirobotit pyrkivät tutkimaan ja todentamaan sivustoilta kerättyä tietoa ja analysoimaan saadut tiedot. (Akhgar, Bayerl ja Sampson 2016, 123-128.)

Pimeän verkon käyttäjiä ja sivustojen ylläpitäjiä voidaan myös pyrkiä identifioimaan valvonta ja hyökkäys tekniikoiden avulla. Nämä pohjautuvat Tor-selaimen ja pimeiden sivustojen haavoittuvuuksiin ja niiden hyödyntämisiin. Salattujen pimeiden sivustojen toiminnan, sisällön ja haavoittuvuuksien kartoittamiseen on luotu työkaluja, joita on esitelty tulevissa kappaleissa.



Kuva 46. Pimeänverkon OSINT rakenne (Nordine s. a.)

5.2.1 Kuinka liikkua turvallisesti pimeässä verkossa?

Ennen kuin tiedon keräys aloitetaan pimeästä verkosta, on hyvä ymmärtää siinä piilevät riskit ja kuinka tiedon etsintä voidaan suorittaa suojaten oma verkkoliikenne ja identiteetti. Pimeän verkon selaus on pääsääntöisesti turvallista, kun selaus ei johda epäilyttäville sivustoille. Sivustojen luotettavuudesta ei voida olla täysin varmoja, sillä Teribium Labs:in suorittaman tutkimuksen mukaan noin 55 % pimeän verkon sisällöstä on laillisia sivustoja, ja loput sivustot perustuvat jonkinlaisesta rikollisesta toiminnasta tai pyrkimyksenä huijata verkon käyttäjiä (Wilson ja Gollnick 2017, 42). Pimeän verkon ideologia perustuu mahdollisuuteen käyttää Internettiä yksityisesti, turvallisesti ja mahdollistaen täyden sanavapauden (Renewable Freedom Foundation s. a.). Pimeän verkon kehittyessä ja rikollisuuden leviämisen pimeän verkon syövereihin, viranomaiset ja tiedustelujärjestöt ovat liittyneet taisteluun pimeässä verkossa piilevää rikollisuutta vastaan. Taistelussa pyritään paljastamaan henkilöitä, jotka hyödyntävät pimeää verkkoa rikolliseen toimintaan. (Ozkaya ja Islam 2019, 18-23.) Rikollisuuden ja lainvalvojien toiminnan seurauksena pimeän verkon nimettömyyden raja on alkanut hämärtyä ja pimeän verkon käyttäminen ei ole enää niin nimetöntä ja turvallista, kuin se on ennen ollut.

Yksi merkittävimmistä pimeän verkon ajojahdeista tapahtui vuonna 2017. Pimeässä verkossa toimineen pimeän marketin AlphaBay:n ylläpitäjä etsittiin maailmanlaajuisesti Yhdysvaltojen, Thaimaan, Alankomaiden, Liettuan, Kanadan, Ison-Britannian ja Ranskan lainvalvontaviranomaisten sekä Euroopan lainvalvontaviraston Europolin yhteistyöllä. Sivusto haluttiin sulkea, sillä oletettavasti sivustolla oli noin 200 000 käyttäjää ja 40 000 myyjää. Sivustolla myytiin maailmanlaajuisesti tappavia laittomia huumeita, varastettuja ja vilpillisiä henkilötodistuksia ja käyttölaitteita, väärennettyjä tuotteita, haittaohjelmia ja muita tietokoneiden hakkerointityökaluja, tuliaseita ja myrkyllisiä kemikaaleja. (The United States Department of Justice 2017.) Tämän merkittävän tapahtuman seurauksena saatiin kaadettua yksi suurin laittomuuksien välittäjä sivusto. Kaatumisen seurauksena pimeään verkkoon muodostui tyhjiö, joka mahdollisti uusien pimeiden markettien nousun. Viranomaisten ja rikollisten taistelun seurauksena on alettava kyseenalaistamaan pimeän verkon luotettavuutta ja sen tapaa turvata käyttäjien nimettömyys. Kuten huomataan Yhdysvaltain huumepoliisin järjestelmävalvoja Chuck Rosenbergin kommentista alapuolella, heti Alpha Bayn kaatumisen jälkeen.

> *"Pimeän verkon niin sanottu nimettömyys on kuviteltua"*
> (Mimoso 2017).

Yleisin tapa muodostaa yhteys pimeään verkkoon on Tor selaimen kautta. Kuten aikaisemmin mainittiin Tor selain hyödyntää yhteyspisteitä eripuolilla maailmaa suojaten täten käyttäjän identiteetin. Kuinka viranomaiset ja tiedustelujärjestöt pystyvät sitten paikantamaan tällaiset käyttäjät? Tor-selain pystyy hankaloittamaan käyttäjiensä identiteetin selvittämistä melkein täydellisesti, mutta harva järjestelmä on täysin haavoittumaton. Näitä haavoittuvuuksia ja heikkouksia Erdal Ozkaya ja Rafiqul Islam ovat kuvanneet kirjassaan. Kirjassa kuvataan kuinka haavoittuvuuksia ja heikkouksia hyödyntämällä voidaan paljastaa henkilöiden identiteetti Tor-selainta käytettäessä. Useat eri tavat vaativat, kuitenkin suuria resursseja pystyäkseen suorittamaan henkilöiden identiteettien paljastamisen. Näitä erilaisia tapoja on havainnollistettu alla olevassa taulukossa (Taulukko 6).

TAULUKKO 6. Tavat paljastaa Tor-selainta käyttävän identiteetti (Ozkaya ja Islam 2019, 53-59.)

Tapa	Määritelmä
Website fingerprinting	Tämän kaltainen valvonta keskittyy datapaketteihin ennen, kun ne saapuvat Tor-verkon ensimmäiselle solmulle, jolloin niitä ei ole vielä salattu. Kolmannet osapuolet, kuten Internet-palveluntarjoajat pystyvät näkemään tämän toiminnan ja tietävät käyttäjän yhdistäneen Tor-verkkoon, mutta he eivät näe käyttäjän verkkotoimintaa. Saatua tietoa voidaan kuitenkin hyödyntää käyttäjän tunnistamisessa.
Eavesdropping	Tällä toiminnalla salakuunnellaan Tor-verkon solmuja. Solmut salaavat ja siirtävät datapaketteja muille solmuille, kunnes ne saapuvat poistosolmulle. Poistosolmu pystyy purkamaan datapakettien salauksen ja lähettämään sen luettavassa muodossa kohteelle. Solmut koostuvat pääasiallisesti vapaaehtoisten reitittimistä, eli kuka tahansa voi ilmoittautua solmupisteeksi ja alkaa reitittämään Tor verkon toimintaa oman reitittimen kautta. Tämän seurauksena solmu pisteen ylläpitäjä voi seurata sen läpi kulkevaa tietoliikennettä ja paikantaa Tor-verkonkäyttäjiä. Tietoliikennettä voidaan salakuunnella myös ajoitusanalyysin avulla, jolloin tarkkaillaan datapakettien poistumisen kuluvaa aikaa palvelimelta ja kunnes ne saapuvat kohteelle. Tässä tapahtuvan ajan perusteella voidaan muodostaa korrelaatio ja paljastaa verkon käyttäjän tietoliikenne. Tämä vaatii vain, että tarkkailija pystyy valvomaan, kun datapaketit lähtevät käyttäjältä, saapuvat Tor verkkoon ja silloin kun datapaketit poistuvat Tor verkosta kohteelle. Tahot, joilla on mahdollisuus suurien resurssien käyttöön, pystyy hyödyntämään tätä haavoittuvuutta, kuten lainvalvontaviranomaiset.
Traffic Analysis	Liikenneanalyysi perustuu samaan periaatteeseen, kuin Tor verkon salakuuntelu. Ylläpitäessä Tor verkon tulo- ja poistumissolmuja voidaan määrittää henkilöiden tietoliikenne. Hyödyntäen läpi kulkevan tietoliikenteen viestinnän välillä tapahtuvaa aikaeroa voidaan suorittaa korrelaatio. Tämän ansiosta voidaan mahdollisesti erotella tietoliikenne ryppäistä keskenään keskustelevat henkilöt. Tämä kaltainen analyysi on todella haastavaa ja kallista suorittaa, jonka seurauksena tällaista analyysiä pystyy suorittamaan vain valtiolliset toimijat ja varakkaat rikollisjärjestöt.
Exit Node Block	Tämän kaltainen toiminta perustuu Tor verkon poistumissolmusta lähtevän tietoliikenteen torjumiseen tietyillä sivustoilla. Sivustot haluavat olla tietoisia mistä IP-osoitteesta heidän sivulleen yhdistetään. Tämän seurauksena Tor verkon kautta tuleva yhteyspyyntö havaitaan epänormaaliksi tietoliikenteeksi ja pääsy sivustolle evätään.
Browser Vulnerabilities	Tor-selain on muunnettu versio Mozilla Firefox-selaimesta. Muuntamisella on pyritty vaikeuttamaan verkonkäyttäjän jäljittämistä muun muassa JavaScript-koodien ja evästeiden avulla. Tor-selain pitää sisällään samat haavoittuvuudet, kuten Mozilla-selain, joita voidaan hyödyntää Tor-selaimen käyttäjiä vastaan.
The Bad apple attack	Tämän kaltaisessa hyökkäyksessä pyritään paljastamaan Tor-verkon käyttäjien IP-osoitteet epäluotettavalla sovelluksella. Hyökkäys aloitetaan hyödyntämällä epäluotettavaa sovellusta IP-osoitteiden paljastamiseksi. Tämän jälkeen paljastettu IP-osoite yritetään yhdistää suojattuun sovellukseen. Jonka jälkeen tietoliikenne virrat voidaan jäljittää Tor-verkkoa käyttäviin henkilöihin. Tämä on mahdollista, sillä Tor-selain ei suojele käyttäjiään sovellustason hyökkäyksiltä.

Tor-selaimen haavoittuvuuksien ymmärtämisen jälkeen voidaan pyrkiä hankaloittamaan tietoliikenteen kuuntelua ja henkilöiden paljastumisen riskiä entistä enemmän. Haavoittuvuuksien löytämisen riskiä voidaan pienentää esimerkiksi poistamalla käytöstä evästeiden, JavaScript-koodien, Adobe Flash Playerin, Javan ja kaikkien muiden aktiivisten palveluiden hyödyntämisen verkkoselailussa. Näiden poiskäytön seurauksena Tor-verkon käyttäjästä jäävien tietojen määrä pyritään minimoimaan ja estämään verkkoselailun seurannan. Tor-verkkoa selatessa on myös pyrittävä välttämään HTTP sivustojen käyttöä. Tiedot, jotka välittyvät tällaisten sivustojen kautta ovat vailla suojausta. Tämän seurauksena, kun tietoliikenne kulkeutuu poistosolmun läpi tiedot voivat olla kolmansien osapuolien tarkkailussa. (James 2018.)

On olemassa myös kaksi tapaa pyrkiä lisäämään Tor-verkon selailun turvallisuutta. Ensimmäinen tapa on Tor-verkon yli VPN yhteys ja toinen on VPN yhteys Tor-verkon kautta. Näiden tekniikoiden avulla voidaan lisätä yksi solmu lisää Tor-verkko rakenteeseen. VPN yhteyden avulla voidaan pyrkiä suojautumaan kolmannen osapuolen tiedustelulta, mutta se voi myös heikentää Tor-verkon nimettömyyttä. VPN-palveluntarjoajilla on omat sopimusehtonsa muun muassa tietoliikenteen lokitietojen säilyttämisessä. Lainvalvontaviranomaisilla on tietyissä maissa mahdollista saada salatut tietoliikenne tiedot VPN-palveluntarjoajalta oikeuden avulla. Tämän takia näiden tekniikoiden käyttö luotettavan VPN-palveluntarjoajan kautta voi lisätä turvallisuutta, mutta VPN käyttäjien kannalta haitallista on se, etteivät he tiedä mitä VPN-palveluntarjoajien palvelimilla oikeasti tapahtuu. (Hoi 2017.)

Luotettavan VPN-palveluntarjoajan löydyttyä voidaan perehtyä kahteen eri tapaan yhdistyä Tor-verkkoon. Tor-verkon yli VPN yhteydellä, tarkoitetaan, että aluksi tietokone yhdistetään VPN-palveluntarjoajaan. VPN:än salattu tietoliikenne ohjataan tämän jälkeen Tor-verkon solmujen kautta Internettiin (Kuva 47). (Aazean 2017.)

KUVA 47. Tor-verkon yli VPN yhteys

Puolestaan, kun luodaan VPN yhteys Tor-verkon kautta, niin tietokone yhdistetään ensin Tor-selaimeen, jonka jälkeen tietoliikenne ohjataan solmujen läpi VPN-serverille. VPN-serveri salaa poistosolmuista lähtevän tietoliikenteen (Kuva 48). (Aazean 2017.)

KUVA 48. VPN yhteys Tor-verkon kautta.

Näiden tekniikoiden avulla pyritään pääasiallisesti piiloittamaan IP-osoite ja yhteys Tor-verkkoon Internet- tai VPN-palveluntarjoajalta. Näillä molemmilla on hyvät ja huonot puolensa liittyen nimettömänä pysymiseen ja tietoliikenteen reitittymisen kannalta. Näitä on esitelty taulukossa (Taulukko 7).

TAULUKKO 7. Hyvät- ja huonot puolet Tor-verkon yli VPN yhteydessä ja VPN yhteydessä Tor-verkon kautta (Aazean 2017.)

	Vahvuudet	Heikkoudet
Tor verkon yli VPN yhteys	- Internet-palveluntarjoaja näkee vain yhteyden VPN-palveluntarjoajaan, eikä tiedä Tor-yhteydestä.	- Tor viimeisestä solmusta lähtevä liikenne on salaamaton ja sitä voidaan valvoa.

	- VPN-palvelun tarjoaja näkee vain yhteyden Tor-solmuihin, mutta ei pysty valvomaan Tor verkon toimintaa. - Sisääntulo solmu Tor-verkkoon yhdistyessä ei sisällä käyttäjän todellista IP-osoitetta, vaan VPN-palveluntarjoajan IP-osoitteen. - Mahdollistaa pääsyn piilotetuille .onion-pääte sivustoille, sillä Tor on viimeinen yhdyskäytävä, ennen Internet-yhteyden muodostamista. - Nopea ja hyvä suorituskyky.	- VPN-palveluntarjoaja näkee mahdollisesti todellisen IP-osoitteen. - Jos VPN-palveluntarjoaja säilyttää verkkovierailu lokit, niin se vastaa samaa asiaa, kun olisi ilman VPN pimeässä verkossa. Tämän seurauksena tietoliikenne voidaan linkittää takaisin todelliseen IP-osoitteeseen. - VPN yhteyden sammuminen kesken pimeän verkon selailun. VPN-tarjoajalla täytyy olla omat DNS-palvelimet ja kill-switch-järjestelmät, jotta yhteys ei katkea kesken Tor-selaimen käyttämisen.
VPN yhteys Tor-verkon kautta	- VPN-palveluntarjoaja ei näe todellista IP-osoitetta, vaan Tor-verkon poistumissolmun IP-osoitteen. - Internet-palveluntarjoaja näkee vain yhteyden Tor-verkkoon, mutta ei tiedä yhteyttä VPN-palveluntarjoajaan. - VPN-palveluntarjoajan ansiosta voidaan määrittää IP-osoitteen maantieteellinen sijainti. - Mahdollisuus yhdistää sivustoihin, jotka rajoittavat Tor poistumissolmun IP-osoitteita.	- Tor verkon piilotettuja .onion-pääte sivustoja ei voida käyttää. - Ei suojausta kolmannen osapuolen vakoilulta Tor poistumissolmuilla. Eikä salaa verkkovierailua Tor-verkossa Internet palveluntarjoajalta. - Haavoittuvainen kokonaisvaltaiselle ajoitusanalyysi hyökkäykselle. - Tor-verkko ei saata hyväksyä tätä asetusta, sillä VPN-palvelimen uskotaan keräävän tietoa Tor-verkon käyttäjistä ja heidän toiminnastansa. - VPN-palveluntarjoaja voi jäljittää käyttäjän taloudellisten tietueitten läpi. Sillä he pystyvät tunnistamaan käyttäjän IP-osoitteen, Tor-poistumissolmun IP-osoitteen kautta. Tämän kaltainen tunnistautuminen, voidaan estää ostamalla VPN-palvelu nimettömänä kryptovaluutalla. Hyödyntäen esimerkiksi Bitcoinia. - Hidas ja huono suorituskyky.

VPN ja Tor-verkon hyödyntäminen yhdessä hankaloittaa käyttäjän identifiointia pimeässä verkossa. Tor verkon yli VPN yhteys antaa käyttäjälleen enemmän turvallisuutta Tor-verkon hyödyntämisessä ja mahdollisuudessa päästä käsiksi piilotettuihin .onion-pääte sivustoihin. Puolestaan VPN yhteys Tor-verkon kautta suojaa paremmin käyttäjän nimettömänä pysymistä, jos vain VPN-palveluntarjoaja on hankittu nimettömästi. Jos pohditaan, kumpi sopii paremmin OSINT-tiedustelun näkökulmaan niin pohjautuen mahdollisuuteen päästä käsiksi piilotetuille sivustoille ja saaden paremman suorituskyvyn, niin paremmaksi vaihtoehdoksi osoittautuu Tor-verkon yli VPN yhteys.

5.2.2 OnionScan modifioitu versio

OnionScan on pimeän verkon tiedusteluun ja tutkimiseen suunniteltu työkalu, ja sen on luonut Sarah Jamie Lewis (@s-rah). Työkalu on avoimen lähdekoodin työkalu, jonka avulla voidaan tunnistaa pimeän verkon sivustojen konfiguraatio- ja tietoturva ongelmat. Tällä pyritään suojelemaan henkilöiden yksityisyyttä ja nimettömyyttä pimeässä verkossa. Työkalu on myös luotu myös automatisoimaan tutkijoiden ja tiedustelijoiden pimeän verkon valvomista ja seuraamista. (S-rah 2017.)

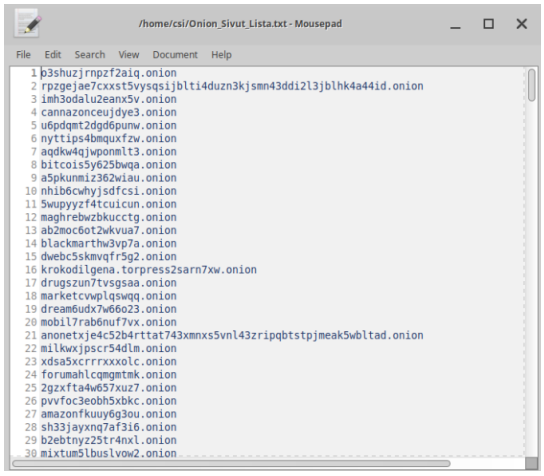
OnionScan työkalun avulla voidaan löytää pimeän verkon sivustoilta haavoittuvuuksia, jotka saattavat paljastaa sivuston ylläpitäjän identiteetin. Tällaisia käyttäjän nimettömyyttä vaarantavia löydöksiä ovat muun muassa paljastuneet verkkopalvelin asetukset, PGP-salausavaimet, SSH-sormenjäljet ja Bitcoin-osoitteet. (TOKYONEON 2017.) PGP-salausavaimilla tarkoitetaan tietojen salausohjelman purkamisavaimia, joiden avulla voidaan purkaa tekstien ja viestien salaus (Brook 2018). Puolestaan SSH-fingerprint, eli Secure Shell:in käyttämä "sormenjälki", on pääasiallisesti salausavain, jonka avulla voidaan tunnistaa palvelin, johon voidaan yhdistää (HEESCHEN 2018). Bitcoin-osoitteilla tarkoitetaan kryptovaluutan tunnistetta, jotka ovat henkilökohtaisia tunnisteita (Bitcoin s. a.).

Tässä työssä modifioidaan alkuperäistä OnionScan työkalua, jotta voidaan hallita sitä järjestelmällisemmin ja käsittelemään sen tuloksia. Modifiointi perustuu Justinin, kirjoittamaan sivustoon, jossa perehdytään aihekohtaisesti, kuinka OnionScan ladataan ja kuinka sitä voidaan muokata haluamalla tavalla. (Justin 2016.) OnionScan latausohje ja kirjoitettu Python sovellus sijaitsevat liitteessä (Liite 1).

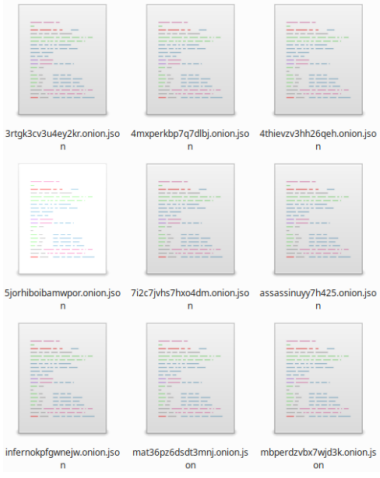
Ohjelmistojen latautumisten ja sovelluksen kirjoituksen jälkeen voidaan aloittaa työkalun käyttö. Sovelluksen suorittaminen tapahtuu "python onionrunner.py" komennolla (Kuva 49). Tämän jälkeen sovellus aloittaa onion-sivustojen läpikäymisen pohjautuen "Onion_Sivut_Lista.txt" tekstitiedostossa oleviin onion-sivusto osoitteisiin (Kuva 50). Tekstitiedoston sisältämät onion-osoitteet on saatu Hunchly Daily Dark Web raportista, joka sisältää tämän päivän aikana syntyneet-, kaatuneet- ja käynnissä olevat pimeän verkon sivustot. Hunchly on suunniteltu keräämään automatisoidusti käytyjen verkkosivustojen tiedot ja dokumentit ja tallentamaan sivustojen tiedot muistiin (Hunchly s. a. a). Tässä työssä hyödynnetään käynnissä olevien pimeän verkon sivustoja, joita on 3531 kappaletta (Kuva 49).

```
csi@csi-analyst:~$ python onionrunner.py
[*] Total onions for scanning: 3531
[*] Running 0 of 3531.
```

Kuva 49. Python-sovelluksen suoritus komento.



Kuva 50. Tekstitiedosto, sisältäen suoritettavat onion-sivustot.



Kuva 51. Onion-sivustojen tulokset.

Tiedonkeräyksen jälkeen tulokset on tallennettu ”OnionScan_Tulokset” nimiseen kansioon. Kaikista läpikäydyistä sivustoista on luotu oma JSON-tiedosto, joka pitää sisällään sivustolta löytyneet tiedot ja mahdolliset haavoittuvuudet (Kuva 51). JSON-tiedoston sisältämät tiedot voivat koostua aikaisemmin mainituista paljastuneista verkkopalvelin asetuksista, PGP-salausavaimista, SSH-sormenjäljistä, Bitcoin-osoitteista sekä uusista onion-sivustojen osoitteista, IP- ja sähköpostiosoitteista.

Modifioidulla OninScan työkalulla voidaan käsitellä suuria määriä onion-sivustoja ja saada merkittäviä tietoa sivustojen haavoittuvuuksista, sekä sivuston sisältöön tallennettuja käyttäjien identifiointi tietoja. Työkalu on ohjelmoitu python-ohjelmointikielellä, joten sen muokkaaminen ja kehittäminen on mahdollista. Tietojen kerääminen on yllättävän hidasta, johtuen tietoliikenteen reitittymisen Tor-verkon kautta, mutta se on automatisoitua. Tietojen analysointi ja prosessointi on haastavaa, sillä työkalun keräämät tiedot ovat JSON-tiedostoissa staattisessa muodossa. Tätä osa-aluetta on mahdollista kehittää ja luoda selkeämpi tietojen analysointityökalu, jolla pystytään käsittelemään näitä tietoja paremmin.

5.2.3 Hunchly

Hunchly on pinta-, syvän- ja pimeän verkon sivustojen dokumentointi työkalu, joka tallentaa automaattisesti selattujen sivustojen sisällön tietokoneelle. Hunchly toimii pelkästään Chrome-selaimella, mutta sen voi konfiguroida reitittymään Tor-verkoston kautta pimeään verkkoon. Hunchly tallentaa sivustolta kaiken sisällön muun muassa sivuston javascript- ja lähdekoodit, sivustolla olevat kuvat, tekstit ja julkaisut. Hunchly työkalun saa ilmaiseksi kuukauden ajaksi, jonka jälkeen lisenssi maksut ovat taulukon mukaiset (Taulukko 8). (Hunchly s. a. a.)

TAULUKKO 8. Hunchly lisenssien hinnat (Hunchly s. a. b.)

Lisenssit	1 - Lisenssi	3 - Lisenssiä	Tiimi lisenssi
Hinta vuodessa	129,46€	324,32€, sisältää 10% alennuksen.	Yli 3 lisenssin ostajat saavat 20% alennuksen hinnasta.

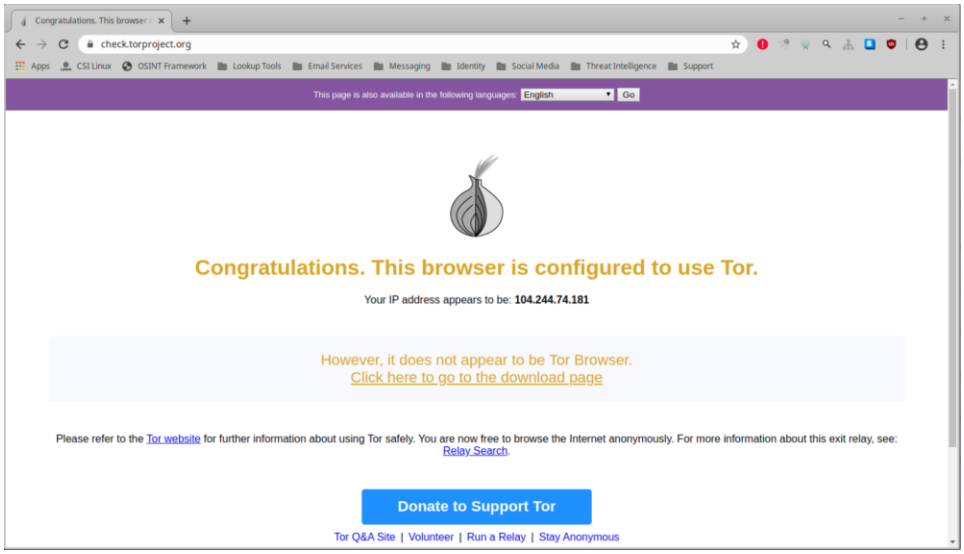
Hunchlyn hyödyntäminen pimeän verkon sivustoilla tapahtuu konfiguroimalla Chrome-selaimen tietoliikenne kulkeutumaan Tor-verkoston kautta. Chrome-selaimen käyttäminen jättää, kuitenkin erilaisen jäljen pimeään verkkoon, kuin Tor-selaimen käyttö. Tor-verkko näkee käyttäjän olevan yhteydessä pimeään verkkoon Chrome-selaimella ja myös selaimen sisältämät laajennukset ja liitännäiset. Tämä voi johtaa siihen, että käyttäjän oikea ulkoinen IP-osoite paljastuu pimeään verkkoon. Tämän seurauksena turvallisempi tapa käyttää Hunchly työkalua pimeässä verkossa on, kun ohjataan tietoliikenne Linux CSI Investigator Gateway virtuaalikoneen kautta Tor verkkoon. (Information Warfare Center s. a. d.) Tämä lisää kerroksellisen suojauksen, verkkoliikenteeseen, joka pohjautuu samaan tekniikkaan, jota Whonix Gateway hyödyntää (Whonix s. a.).

Tämän asetuksen määrittäminen tapahtuu aukaisemalla molemmat Linux CSI Investigator Analyst ja Gateway virtuaalikoneet (Kuva 52). Tämän jälkeen kirjaudutaan sisään Analyst- ja Gateway virtuaalikoneelle. Tämän jälkeen siirrytään Analyst-virtuaalikoneelle ja klikkaamalla vasemmasta alareunasta CSI Gateway kuvaketta, joka ohjaa Analyst virtuaalikoneen tietoliikenteen Gateway virtuaalikoneen kautta Tor-verkkoon. Tämän jälkeen voidaan konfiguroida Chrome-selaimen reitittymään Tor-verkoston kautta Linux terminaaliin syötettävän komennon avulla (Kuva 53). Tämä mahdollistaa pääsyn Chrome-selaimen kautta pimeän verkon sivustoille. Konfiguroimisen toiminnan voi testata yhdistämällä ”check.torproject.org” sivustoon Chrome-selaimen kautta (Kuva 54). Näin luotiin turvallinen ympäristö Hunchly työkalun käyttöön pimeässä verkossa. (Information Warfare Center s. a. d; Seitz s. a.)

Kuva 52. Käynnissä olevat virtuaalikoneet.

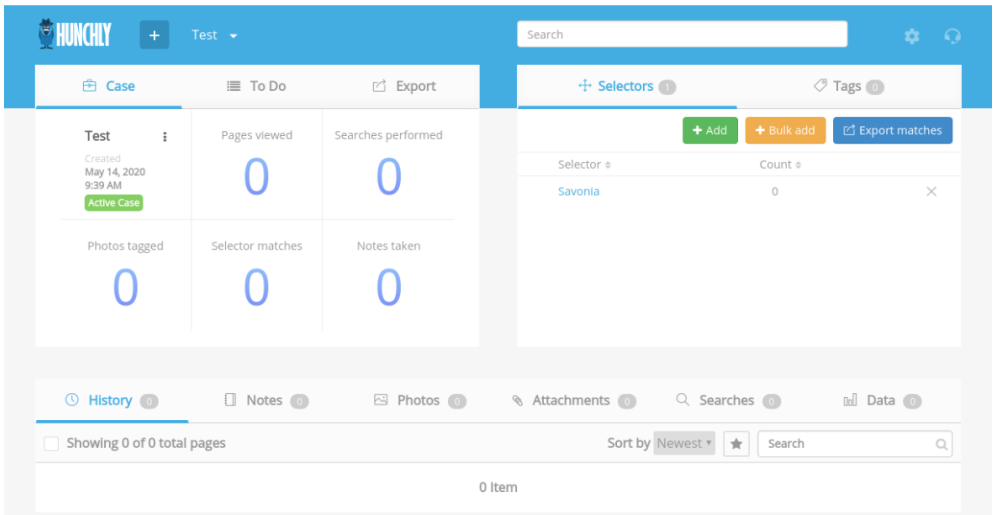
```
csi@csi-analyst:~$ sudo google-chrome --proxy-server="socks5://localhost:9150" -
-host-resolverrules="MAP * ~NOTFOUND , EXCLUDE localhost"/Applications/Google\ C
hrome.app/Contents/MacOS/Google\ Chrome --proxy-server="socks5://localhost:9150"
 --host-resolverrules="MAP * ~NOTFOUND , EXCLUDE localhost"
```

Kuva 53. Reititetään Chrome-selaimen tietoliikenne Tor-verkostoon.



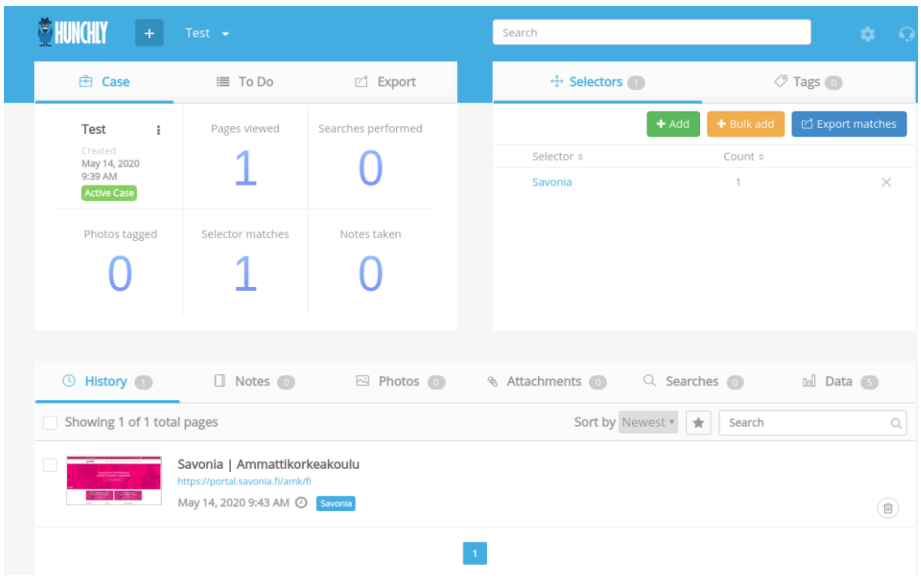
Kuva 54. Onnistunut Chrome-selaimen konfigurointi.

Hunchly työkalun on helppokäyttöinen ja visuaalisesti selkeästi rakennettu (Kuva 55). Vasemmasta yläreunasta plus merkkiä painamalla voidaan lisätä uusi ”tapaus”, jolla tarkoitetaan projektia. Oikeasta reunasta voidaan lisätä valintoja ”Selectors” add nappia painamalla. Näiden avulla Hunchly seuraa sivustoilta löytyviä tietoja ja jos sivustolta löytyy valintaa vastaava tekstikenttä Hunchly korostaa sen näkyviin (Kuva 57). Sivustojen tietojen tallentaminen tapahtuu aukaisemalla Chrome-selain ja klikkaamalla sovelluspalkista Hunchly kuvaketta. Tämän jälkeen aukeaa kuvan mukainen ikkuna (Kuva 56). Tästä ikkunasta täytyy hyväksyä ”Capture” valinta, jotta Hunchly alkaa keräämään sivustojen tietoja. Ikkunasta voidaan myös valita projekti, johon selaustiedot halutaan tallentaa. Hunchly kerää kaiken tiedon sivustoilta, joissa käyttäjä on vieraillut ja esittää ne käyttöliittymän alareunassa (Kuva 57).



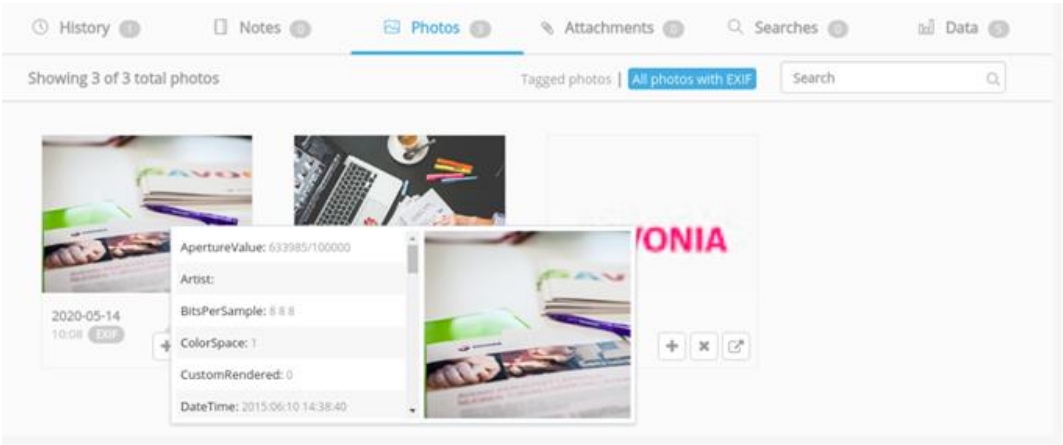
Kuva 55. Hunchly työkalun käyttöliittymä.

Kuva 56. Chrome-selaimen Hunchly asetukset.



Kuva 57. Hunchlyn tiedon etsintä Savonian sivustolle.

Tallennettuja tietoja voidaan tarkastella ”History” välilehdeltä (Kuva 58). Täältä nähdään yleistiedot- ja esikatselu sivustosta. ”Photos” välilehdeltä voidaan tarkastella sivustoilta löytyneitä kuvia ja kuvien EXIF-metadataa (Kuva 58). EXIF-metadatalla tarkoitetaan kuvan sisältäviä tietoa, joita voi olla muun muassa laite, jolla kuva on otettu, aikaleima, sijaintitieto ja henkilö kuka on ottanut kuvan (Ray 2018).

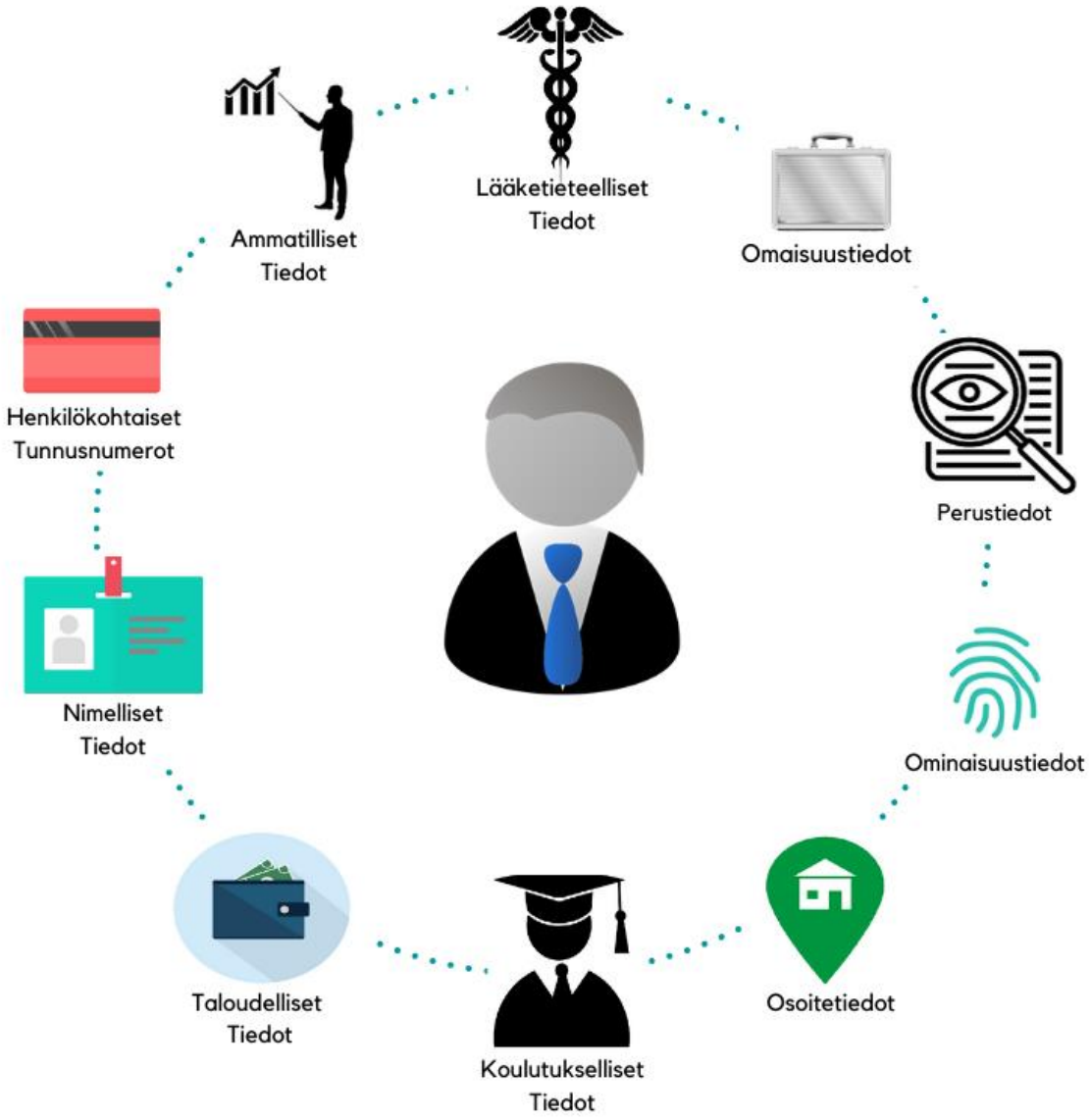


Kuva 58. Hunchly työkalun keräämän kuvan EXIF-metadata.

Hunchly on käytännöllinen verkkoselailun tallentamistyökalu, jota voidaan hyödyntää pinta-, syvässä- ja pimeässä verkossa. Työkalu tallentaa kaiken verkkosivuilta, joissa käyttäjä on vieraillut Hunchly sovellukseen. Tallennettuja tietoja voidaan jälkeenpäin tarkastella ja analysoida. Hunchly on erityisesti käytännöllinen pimeän verkon sivustojen valvomisessa ja analysoimisessa. Näiden tietojen pohjalta voidaan seurata sivuston toimintaa ja mahdollisesti identifioida onion-sivuston käyttäjiä julkaisujen ja kuvien EXIF-metadatan perusteella.

6 AVOITEN LÄHTEIDEN TIEDUSTELU HENKILÖPROFILOINNISSA

Avointen lähteiden tiedustelun avulla on mahdollista kerätä runsaasti henkilökohtaisia tietoja kohdehenkilöstä. Profiloinnin tarkoituksena on luoda kohteena olevasta henkilöstä haavoittuvuusprofiili, jonka avulla pohditaan henkilön alttiutta tulla kohteeksi tietoturvahyökkäykselle. Tietoja etsitään pelkästään julkisesti saatavilla olevista lähteistä hyödyntäen automatisoituja työkaluja. Tietoturvallisuuden näkökulman mukaan henkilöprofiloinnin tarkoituksena on minimoida hyökkääjän mahdollisuus löytää kohteen henkilökohtaisia tunnistetietoja. Tämä toiminta ei kuitenkaan poista henkilön mahdollisuutta joutua henkilötietojen väärinkäytön kohteeksi, vaan sillä pyritään pienentämään väärinkäytön riskiä. Hyökkääjät pyrkivät pääsemään kohteen henkilökohtaisiin tunnistetietoihin (Kuva 59). Näitä tietoja hyväksi käyttäen verkkorikolliset pyrkivät luomaan uusia luottotilejä ja identiteettejä. He pystyvät myös tekemään verkko-ostoja kohteen nimissä. (Akamai 2015. 3-7.)



Kuva 59. Henkilöstä etsittävät henkilökohtaiset tunnistetiedot.

6.1 Henkilöstä kerättävät tiedot

Henkilönprofilointiin kerättävät tiedot koostuvat pääosin henkilökohtaisista tunnistetiedoista (PII; Personally Identifiable Information) kohteena olevasta henkilöstä. Henkilökohtaiset tunnistetiedot määritellään linkitetyiksi- tai linkitettäviksi tiedoiksi. Tietoja hyödyntämällä voidaan jäljittää yksilö ja sen identiteetti. (U.S. General Services Administration 2019.) Linkitetyt tiedot ovat yksilöä koskevia tai siihen liittyviä tietoja, joiden avulla voidaan tunnistaa yksilö. Puolestaan linkitettävät tiedot ovat yksilöä koskevia tai siihen liittyviä tietoja, joiden avulla ei yksinään voida tunnistaa yksilöä. Nämä tiedot ovat epäsuoria, joita yhdistelemällä voidaan luoda kattavampi kuvaus henkilöstä, jonka seurauksena tiedoista tulee identifiointi kelpoisia tietoja. (Infiniwiz s. a.) Henkilökohtaisia tunnistetietoja ei ole mallinnettu mihinkään yksittäiseen informaatio- tai tekniikkaluokkaan. Pikemminkin tietojen käyttö edellyttää tapauskohtaista riskiarviointia, jonka perusteella yksilö voidaan tunnistaa hyödyntämällä linkitettyjä- ja linkitettävissä olevia lähteitä. (U.S. General Services Administration 2019; Erbschloe 2019.)

Yleinen palveluhallinto (GSA; General Services Administration) määrittelee henkilökohtaiset tunnistetiedot yksilöä koskevien tietojen perusteella. Tietoja voivat olla muun muassa henkilökohtaiset tunnusnumerot, nimelliset-, ominaisuus-, osoite-, henkilön perus-, koulutukselliset-, ammatilliset-, lääketieteelliset-, taloudelliset- ja omaisuustiedot. Käytännössä kaikki henkilöön linkitetyt tai linkitettävät tiedot. (Erbschloe 2019, 106). Henkilökohtaiset tunnistetiedot sisältävät valtavasti erityyppisiä tietoja, joita voidaan käyttää erottamaan tai jäljittämään yksilön identiteetti. Tämän takia henkilökohtaisten tunnistetietojen kokonaisuus on todella laaja (U.S. General Services Administration 2019). Ala puolella olevassa taulukossa on pyritty havainnollistamaan tätä kokonaisuutta linkitetyillä ja linkitettävillä tiedoilla (Taulukko 9).

TAULUKKO 9. Linkitetyt- ja linkitettävät tiedot (Erbschloe 2019, 106-107; Infiniwiz s. a.; Mcdonald s. a.; Korolov 2019.)

Henkilökohtaiset tunnistetiedot	
Linkitetyt	Linkitettävät
<u>Henkilökohtaiset tunnusnumerot ja nimelliset tiedot</u> Sosiaaliturvatunnus, passinnumero, matkapuhelimen- ja työpuhelimennumero, ajokortin-, vero-, pankkitilin- ja luottokortin numero. Henkilön kokonimi, puolison-, lapsien- ja vanhempien nimet, sekä läheisenpien ystävien nimet ja mahdolliset peitenimet.	<u>Henkilön perustiedot</u> Syntymäaika, syntymäpaikka, siviilisääty, postinumero, rotu, uskonto, poliittinen suuntautuminen, pituus, paino, säännölliset aktiviteetit arkena ja viikonloppuisin, henkilön elämäntavat, sosiaaliset suhteet, tilien käyttäjätunnukset: (sosiaalinen media, pelit, foorumit ja blogit) ja lemmikkien tiedot.
<u>Osoite- ja Taloudelliset tiedot</u> Katuosoite, maantieteelliset koordinaatit, IP- (Internet Protocol) tai MAC (Media Access Control) osoite ja nimellinen sähköpostiosoite. Omistetut kiinteistöt, verotiedot ja laskut.	<u>Koulutukselliset- ja Ammatilliset tiedot</u> Valmistumistiedot ensimmäisen-, toisen- ja kolmannen asteenkoulutuksesta, koulutapahtumat, mahdolliset luokkakaverit, nykyinen työpaikka, aiemmat työpaikat, työkaverit, sertifikaatit ja lisäkoulutukset.
<u>Henkilökohtaiset ominaisuudet</u> Biometriset tietueet, kuten sormenjäljet, kasvonmallit, muotokuva henkilöstä, verkkokalvo-, röntgenkuvat, käsiala ja ääninäyte.	<u>Lääketieteelliset- ja Omaisuustiedot</u> Mitä tahansa yksilöiviä terveystietoa. Tiedot voivat liittyä yksilön nykyiseen, menneeseen tai tulevaan terveyteen, joko fyysisesti tai henkisesti. Ajoneuvojen- ja elektronisten laitteistojen tiedot (puhelimet, tietokoneet, älykellot ja tabletit).

6.2 Henkilötietojen keräyksen taustat ja motiivit

Henkilöprofiloinnin pääasiallisena tarkoituksena on kartoittaa kohteena olevan henkilön alttius tulla hyödynnetyksi tietoturvahyökkäyksessä. Yleensä yrityksien tietovuodon syynä on inhimillinen virhe, jonka seurauksena tietovuotoja syntyy. Verkkorikolliset usein miten profiloivat korkea-arvoisia henkilöitä, joilla on valtuudet päästä käsiksi luottamuksellisiin tietoihin. (Green ja Ng 2020; Reed 2019.) Verkkorikolliset keräävät samalla tavalla kohteen tunnistetietoja, kuten tietoturva-asiantuntijat. Verkkorikolliset pyrkivät vaikuttamaan kohteeseen luottamuksen, kiristyksen tai uhkailun avulla. Näin he pyrkivät saamaan kohteen luovuttamaan salassa pidettäviä tietoja tai saadakseen taloudellista hyötyä. Tämän kaltaisen manipuloinnin seurauksena uhri saattaa syyttää itseään tapahtuneesta ja tuntea syyllisyyttä ja häpeää vuosi kaudet. (Reed 2019.) Manipuloiduksi joutuminen ei kuitenkaan ole uhrin syy, vaan yritysten, jotka eivät mahdollista henkilökuntansa kehittää heidän tietoturvallisuustietoisuuttansa. (Alashe 2020.) Tämän seurauksena yritykset ovat alkaneet palkkaamaan tietoturva-asiantuntijoita luennoimaan tietoturvahyökkäysten vaaroista ja tekemään henkilöprofilointeja yrityksen korkea-arvoisista henkilöistä.

Kuten aikaisemmin mainittiin henkilökohtaiset tunnistetiedot, nousevat hyvin merkittävään rooliin tämän kaltaisessa tiedustelussa. On kuitenkin huomioitava tiedonkeräyksessä, että osa tiedoista ei ole keräyksen alku hetkellä henkilökohtaisia tunnistetietoja. Kuitenkin näitä tietoja yhdistämällä eri lähteisiin voidaan muodostaa henkilöä yksilöiviä tietoja. Ihmiset paljastavat omin tahoin itsestään henkilökohtaisia tunnistetietoja sosiaalisen median palveluihin kuten Facebook, Instagram ja LinkedIn. Jaetut tiedot sisältävät yleensä nimen, sähköpostiosoitteen, kotiosoitteen, henkilön sijainnin, siviilisäädyn, koulutus- ja ammatilliset tiedot. Tietojen määrä on kuitenkin täysin riippuvainen henkilöiden yksilöllisestä halusta jakaa tietoja sosiaaliseen mediaan. Jaettujen tietojen pohjalta identiteettiin kohdistuvat rikokset on nykyään helppo suorittaa. (SEORG 2019; Erbschloe 2019, 114.)

Avointen lähteiden tiedustelun osalaji SOCMINT (Social Media Intelligence) on sosiaalisen median tiedustelua. Tämän kaltaisessa tiedustelussa kerätään vain tietoja sosiaalisesta mediasta. OSINT:in ja SOCMINT:in erona on vain, että sosiaalisen median tiedustelussa voidaan hyödyntää, joko yksityisiä tai julkisia lähteitä toisin, kuin avointen lähteiden tiedustelu perustuu pelkästään avoimiin lähteisiin. Sosiaalisen median tiedustelun avulla voidaan seurata sosiaalisia kanavia, keskusteluja, julkaistuja viestejä ja kuvia. Näiden tietojen avulla voidaan saada tietoa henkilökohtaisista tunnistetiedoista sekä henkilöiden- ja ryhmien välisistä julkisista ja yksityisistä vuorovaikutuksista. Tiedustelu menetelminä voi olla manuaalinen tiedon seuranta ja etsintä. Tiedustelussa voidaan myös hyödyntää automatisoituja työkaluja tiedon analysoimiseksi ja havainnollistamiseksi. (Privacy International 2017.)

6.3 Henkilöprofiloinnin suorittamisen havainnollistaminen

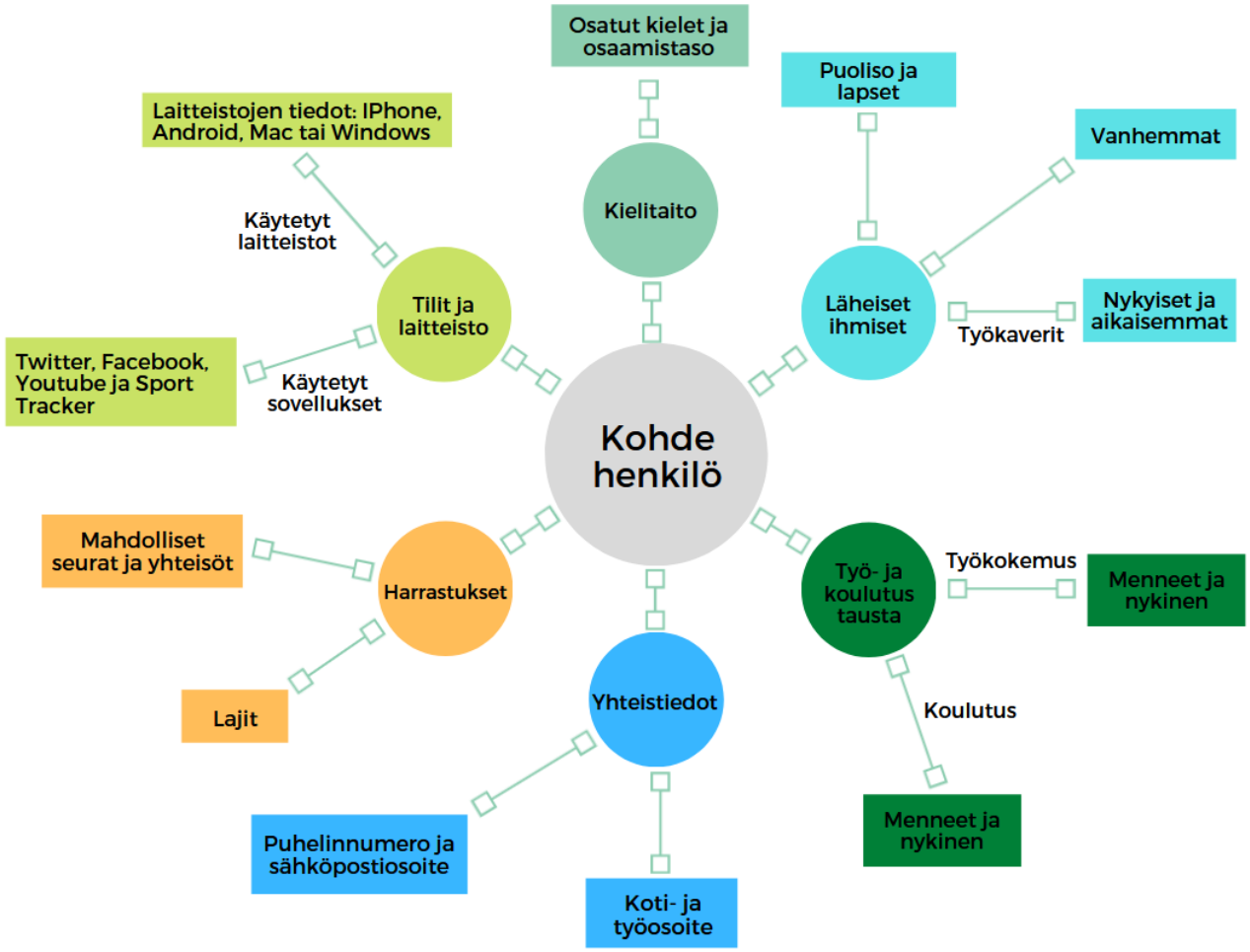
Henkilöprofilointi suoritettiin kohteelle, jonka henkilöllisyyttä ei voi paljastaa salassapitosopimuksen vuoksi. Tässä kappaleessa kuitenkin havainnollistetaan anonyymisti henkilöprofiloimisessa hyödynnettyjä käytäntöjä, tiedon etsintätyökaluja, sekä tiedon etsinnällisesti merkittävimpiä löydöksiä.

Henkilönprofiloiminen aloitettiin perehtymällä valittuun kohteeseen. Ensin pohdittiin mitä merkittäviä tietoja henkilöstä voidaan löytää ja millä tiedoilla ei tule olemaan niin suurta merkitystä tiedustelussa. Määriteltyjen tietojen perusteella, etsittiin lähteet, joista tiedot voitiin löytää. Sosiaalisen median, kuten LinkedIn, Facebook, Instagram ja Twitter alustat toimivat hyvin henkilökohtaisien-, sosiaalisten suhteiden-, ammatillisien- ja koulutuksellisien tietojen lähteinä. Googlen indeksoimat sivustot, pitävät sisällään valtavasti tietoa henkilön ammatillisesta toiminnasta, julkisuus kuvasta ja sivustot toimivat merkittävänä apuvälineenä henkilön linkitettävien tietojen yhdistämisessä. Tietolähteiden määrittelyn jälkeen, voitiin aloittaa kohteen henkilötietojen tiedustelu tiedon keräys- ja analysointityökalujen avulla.

Tiedustelussa hyödynnetyt työkalut pohjautuivat Google Dorks hakuparametreihin ja Hunchly verkkoselailun dokumentointi ja analysointityökaluun. Tiedustelussa hyödynnettiin myös Maltego CE:tä ja Sherlock avoimen lähdekoodin sovellusta, jonka avulla voitiin etsiä kohteen käyttäjätunnuksen perusteella sosiaalisetpalvelut, joissa on käytetty henkilön käyttäjätunnusta (Dushantha 2020). Kohteen tietoja kartoitettiin myös pimeän verkon puolelta hyödyntäen saatavilla olevia hakukoneita ja tiedonetsintä palstoja.

Merkittävimpiä löydöksiä henkilöstä saatiin hyödyntäen Google Dorks:ia, jonka avulla saatiin selville kohteen henkilökohtaisia tunnistetietoja. Näitä tietoja olivat muun muassa lähipiirin henkilötiedot, sekä koulutukselliset- ja ammatilliset tiedot. Google Dorks:in avulla löydettyjen tietojen pohjalta voitiin myös vahvistaa tiettyjen tietojen todenperäisyys. Hunchly nopeutti tietojen analysointia käytettyjen valintojen avulla, sillä valinnat indeksoivat sivustoilta löytyvät tiedot selkeään muotoon. Tiedustelussa kohdattiin ongelma paikantaa kohteen kotiosoite. Sherlock sovelluksen kautta saatiin henkilön Sport Tracker käyttäjätunnus, joka mahdollisti pääsyn henkilökohtaisiin kuntoilusovelluksen tietoihin. Näiden tietojen avulla pystyttiin paikantamaan kohteen kodin oletettu sijainti hyvinkin tarkasti. Tiedustelun aikana havaittiin, että tiettyjä tietoja kohteesta oli vaikea löytää. Tämän seurauksena hyödynnettiin lähipiiristä löytyviä tietoja, joiden avulla pystyttiin linkittämään tiettyjä tietoja kohteeseen. Kohteesta ei löytynyt mitään merkittäviä tietoja Maltegon tai pimeän verkon avulla. Kohteen henkilöprofiloinnin jälkeen muodostettiin löydettyjen tietojen pohjalta raportti ja mindmap-pohjainen solmuanalyysi tilaajayritykselle (Kuva 60).

Henkilöprofloinnin aikana havaittiin, ettei tällä hetkellä ole saatavilla kokonaisvaltaisia automatisoituja työkaluja henkilöidenprofiloimiseen. Tietojen etsintä ja analysointi täytyy suorittaa suurimmaksi osaksi manuaalisesti hyödyntäen hakukoneita ja verkkosivustoja. On kuitenkin olemassa tiettyjä työkaluja, joilla voidaan nopeuttaa ja automatisoida sivustojen sisältöjen läpi käymistä ja analysointia. Kuitenkin lopulta tietojen analysointi ja luotettavuuden todentaminen on tehtävä manuaalisesti. Henkilöprofloiminen toi tilaajayritykselle kehitysideoita heidän omaan henkilöprofilointiin. Erityisesti mielenkiintoa herätti käyttäjätunnus sivustojen kokonaisvaltainen läpikäyminen ja henkilökohtaisten tunnistetietojen havainnollistaminen mindmap-pohjaisessa solmuanalyysissä.



Kuva 60. Kohteen henkilökohtaisten tunnistetietojen mindmap mallipohja.

7 JATKOTUTKIMUS

Opinnäytetyössä käsiteltiin avointen lähteiden tiedustelun teoreettista puolta ja kuinka henkilöprofilointi suoritetaan. Tämän seurauksena jatkotutkimusaiheet pohjautuvat käsitteisiin, joihin ei olla perehdytty opinnäytetyössä. Jatkotutkimuksena voisi pohtia, kuinka avointen lähteiden tiedustelu vaikuttaa kuva-, mittaus-, tunnusmerkki- ja signaalitiedustelussa ja millaisia työkaluja OSINT tarjoaa näillä osa-alueilla. Lisäksi voisi pohtia, kuinka teknologian kehityksen myötä avointen lähteiden tiedustelun tietojen keräys- ja analysointimenetelmät kehittyvät. Kehittyvätkö tekoälylliset sovellukset siihen pisteeseen, että ne voivat automatisoidusti suorittaa avointen lähteiden tiedustelua. Pystyvätkö koneoppimisalgoritmit keräämään ja analysoimaan avointen lähteiden tietoja tehokkaammin ja luotettavammin, mihin ihminen on ikinä pystynyt? Tämän ohella on myös hyvä pohtia, kuinka ihmisten käyttäytyminen sosiaalisessa mediassa muuttuu. Vähentyykö tietojen jakaminen avoimiin lähteisiin? Lisääntyykö ihmisten tietämys tietojen jakamisen riskeistä ja mahdollistaako tämä uusien lakien muodostumisen henkilöiden tietosuojelun ja yksityisyyden parantamiseksi? Kuinka nämä toiminnat vaikuttavat saatavilla oleviin avoimiin lähteisiin ja voiko toimet johtaa avointen lähteiden tiedustelun joutumisen taas muiden tiedusteluluokkien varjoon.

8 YHTEENVETO

Opinnäytetyö onnistui kokonaisuudessaan hyvin ja työssä käsiteltiin työsuunnitelman mukaiset asiat. Teoria osuudessa käsiteltiin, kuinka avointen lähteiden tiedustelu on kasvanut historian saatossa nykyiseksi moderniksi tiedusteluluokaksi. Työssä tarkasteltiin, kuinka laaja avointen lähteiden tiedustelun rakenne on. Työssä kerrotaan mistä, millä ja miten avointen lähteiden tietoja voidaan etsiä ja analysoida nykypäivänä. Työ havainnollisti samalla mitä haasteita, hyötyjä ja eettisyys kysymyksiä avointen lähteiden tiedustelu pitää sisällään. Työssä esiteltiin myös suosituimpia avointen lähteiden tiedustelussa hyödynnettyjä työkaluja eri Internetin tasoilla. Tutkimuskysymyksenä pohdittiin, kuinka henkilöprofilointi suoritetaan ja voidaanko tiedon etsintä- ja analysointimenetelmiä automatisoida ja kehittää jollain tasolla.

Henkilöprofiloinnin aikana havaittiin, ettei henkilöidenprofiloimiseen ole saatavilla kokonaisvaltaisia automatisoituja työkaluja. Tietojen etsintä ja analysointi täytyy tehdä suurimmaksi osaksi manuaalisesti hyödyntäen hakukoneita ja verkkosivustoja. Tiettyjen työkalujen avulla voidaan nopeuttaa ja automatisoida sivustojen sisältöjen läpikäymistä ja analysointia. Kuitenkin lopulta tietojen analysointi ja luotettavuuden todentaminen on tehtävä manuaalisesti. Tilaajayritys sai kehitysideoita suoritetusta henkilöprofiloimisesta. Yksi merkittävistä kehitysideoista oli käyttäjätunnuksen avulla saadut kohteen rekisteröitymiset tiettyihin palveluihin ja sivustoihin. Toinen merkittävä kehitysidea oli kohteesta löydettyjen henkilökohtaisten tunnistetietojen mallintaminen mindmap-pohjaisessa solmuanalyysissä. Tämän avulla pystytään helpottamaan kohteena olevan henkilön välisten suhteiden ymmärtämistä ja havainnollistamista.

Opinnäytetyö eteni aikataulullisesti työsuunnitelman mukaan. Työn edetessä halu oppia ja perehtyä uusiin asioihin laajensi hieman rajattua suunnitelmaa. Työhön tulleet lisäykset, kuten avointen lähteiden tiedustelun eettisyys, Internetin tasot ja tietojen etsintä pimeästä verkosta laajensivat työn kuvausta. Lisäyksistä huolimatta työ suoritettiin määritellyssä aikataulussa ja lisäyksien avulla opinnäytetyöstä tuli laaja ja kattava kokonaisuus.

LÄHTEET JA TUOTETUT AINEISTOT

1. Aazean 2017. TOR OVER VPN & VPN OVER TOR: WHICH IS BETTER? [Blogi]. Bolehvpn [Viitattu 2020-04-23]. Saatavissa: https://blog.bolehvpn.net/tor-over-vpn-vpn-over-tor-which-is-better/
2. Akamai 2015. Weighing Risk Against the Total Cost of a Data Breach: Can You Afford a Web Application Layer Attack? [Verkkojulkaisu]. Akamai [Viitattu 2020-05-21]. Saatavissa: https://www.akamai.com/us/en/multimedia/documents/secure/weighing-risk-against-the-total-cost-of-a-data-breach-white-paper.pdf?campaign_id=F-MC-24907
3. AKHGAR, Babak, BAYERL, P.Saskia, SAMPSON, Fraser 2016. Open Source Intelligence Investigation. [Verkkokirja]. Springer International Publishing. [Viitattu 2020.05.10]. Saatavissa: https://ebookcentral-proquest-com.ezproxy.savonia.fi/lib/savoniafi/reader.action?docID=4774801&query=Open+Source+Intelligence+Investigation%3A+From+Strategy+to+Implementation
4. ALASHE, Oz 2020. Stop saying employees are the weakest link in cybersecurity. [Verkkojulkaisu]. TNW. [Viitattu 2020-05-23]. Saatavissa: https://thenextweb.com/growth-quarters/2020/03/10/stop-saying-employees-are-the-weakest-link-in-cybersecurity/
5. Alexis 2020. Google Hacking for Penetration Testing. [Verkkojulkaisu]. Hackersploit. [Viitattu 2020-04-07]. Saatavissa: https://hsploit.com/google-hacking-for-penetration-testing/
6. APNIC s. a. Autonomous System numbers – FAQs. [Verkkojulkaisu]. APNIC. [Viitattu 2020-05-14]. Saatavissa: https://www.apnic.net/get-ip/faqs/asn/
7. Arin s. a. a. Using Whois. [Verkkojulkaisu]. Arin. [Viitattu 2020-04-20]. Saatavissa: https://www.arin.net/resources/registry/whois/
8. Arin s. a. b. Point of Contact (POC) Records. [Verkkojulkaisu]. Arin. [Viitattu 2020-04-20]. Saatavissa: https://www.arin.net/resources/guide/account/records/poc/
9. BERTRAM, Steward K. 2015. The Tao of Open Source Intelligence. [Verkkokirja]. IT Governance Publishing. [Viitattu 2020-04-29]. Saatavissa: http://web.b.ebscohost.com.ezproxy.savonia.fi/ehost/ebookviewer/ebook/ZTAwMHR3d19fMTAzMDA5NF9fQU41?sid=443c41c8-8a92-4d41-827e-22d34a0deee8@pdc-v-sessmgr06&vid=2&format=EB&rid=1
10. Bitcoin s. a. Protect your privacy. [Verkkojulkaisu]. Bitcoin. [Viitattu 2020-05-13]. Saatavissa: https://bitcoin.org/en/protect-your-privacy
11. BROOK, Chris 2018. What is PGP Encryption? Defining and Outlining the Uses of PGP Encryption. [Verkkojulkaisu]. DIGITALGUARDIAN. [Viitattu 2020-05-13]. Saatavissa: https://digitalguardian.com/blog/what-pgp-encryption-defining-and-outlining-uses-pgp-encryption
12. CLOUDFLARE s. a. a. What Is DNS? | How DNS Works. [Verkkojulkaisu]. CLOUDFLARE. [Viitattu 2020-05-14]. Saatavissa: https://www.cloudflare.com/learning/dns/what-is-dns/
13. CLOUDFLARE s. a. b. What Is A DNS MX Record? [Verkkojulkaisu]. CLOUDFLARE. [Viitattu 2020-05-14]. Saatavissa: https://www.cloudflare.com/learning/dns/dns-records/dns-mx-record/
14. COLQUHOUN, Cameron 2016. A Brief History of Open Source Intelligence. [Verkkojulkaisu]. Bellingcat. [Viitattu 2020-02-18]. Saatavissa: https://www.bellingcat.com/resources/articles/2016/07/14/a-brief-history-of-open-source-intelligence/
15. CONGLETON, Nick 2018. Debian vs Ubuntu. [Verkkojulkaisu]. LINUXCONFIG.org. [Viitattu 2020-03-31]. Saatavissa: https://linuxconfig.org/debian-vs-ubuntu

16. Cyber Army 2020. Exploring Google Hacking Techniques using Dork. [Verkkojulkaisu]. Medium. [Viitattu 2020-04-08]. Saatavissa: https://medium.com/nassec-cybersecurity-writeups/exploring-google-hacking-techniques-using-google-dork-6df5d79796cf
17. Datareportal s. a. GLOBAL SOCIAL MEDIA OVERVIEW. [Verkkojulkaisu]. [Viitattu 2020-03-10]. Saatavissa: https://datareportal.com/social-media-users
18. Dnsimple s. a. What's an NS Record? [Verkkojulkaisu]. dnsimple. [Viitattu 2020-05-14]. Saatavissa: https://support.dnsimple.com/articles/ns-record/
19. DUSHANTHA, Siddharth 2020. Sherlock-project. [Verkkojulkaisu]. Github. [Viitattu 2020-05-24]. Saatavissa: https://github.com/sherlock-project/sherlock
20. EIJKMAN, Quirine, WEGGEMANS, Daan 2013. Open source intelligence and privacy dilemmas: Is it time to reassess state accountability? [PDF] ResearchGate. [Viitattu 2020-05-17]. Saatavissa: https://www.shrmonitor.org/assets/uploads/2017/09/03-Eijkman-Weggemans-v2.pdf
21. ERBSCHLOE, Michael 2019. Social Engineering: Hacking Systems, Nations, and Societies. [Verkkokirja]. Taylor & Francis Group. [Viitattu 2020-05-20]. Saatavissa: https://ebookcentral-proquest-com.ezproxy.savonia.fi/lib/savoniafi/detail.action?docID=5890629&query=Social+Engineering%3A+Hacking+Systems%2C+Nations%2C+and+Societies
22. ExamCollection s. a. Overview of common TCP and UDP default ports. [Verkkojulkaisu]. ExamCollection. [Viitattu 2020-04-09]. Saatavissa: https://www.examcollection.com/certification-training/network-plus-overview-of-common-tcp-and-udp-default-ports.html
23. Expert System Team 2017. Advantages and disadvantages of open source intelligence. [Blogi]. EXPERT SYSTEM. [Viitattu 2020-03-27]. Saatavissa: https://expertsystem.com/advantages-disadvantages-open-source-intelligence/
24. GeekWire 2020. Recon.ng v5 Tutorial. [Verkkojulkaisu]. GeekWire. [Viitattu 2020-04-20]. Saatavissa: https://geekwire.eu/recon-ng-v5-tutorial/
25. GELNAW, Angela 2019. Open Port Vulnerabilities: What's the Big Deal. [Verkkojulkaisu]. BITSIGHT. [Viitattu 2020-05-17]. Saatavissa: https://www.bitsight.com/blog/open-port-vulnerabilities-whats-the-big-deal
26. GREEN, Andy, NG Cindy 2020. CEO Phishing: Hackers Target High-Value Data. [Verkkojulkaisu]. Varonis. [Viitattu 2020-05-23]. Saatavissa: https://www.varonis.com/blog/ceo-phishing-hackers-target-high-value-data/
27. HASSAN, Nihad A., HIJAZI, Rami 2018. Open Source Intelligence Methods and Tools. [Verkkokirja]. New York: Apress L. P. [Viitattu 2020-03-18]. Saatavilla: https://ebookcentral-proquest-com.ezproxy.savonia.fi/lib/savoniafi/detail.action?docID=5446001&query=Open+Source+Intelligence+Methods+and+Tools
28. HEESCHEN, Jered 2018. Rackspace Cloud Essentials - Checking a server's SSH host fingerprint with the web console. [Verkkojulkaisu]. Rackspace. [Viitattu 2020-05-13]. Saatavissa: https://support.rackspace.com/how-to/rackspace-cloud-essentials-checking-a-server-s-ssh-host-fingerprint-with-the-web-console/
29. HOI, Sunny 2017. Which is Better: TOR over VPN or VPN over TOR? [Verkkojulkaisu]. 1337pwn. [Viitattu 2020-05-06]. Saatavissa: https://www.1337pwn.com/which-is-better-tor-over-vpn-or-vpn-over-tor/
30. HU, Evanna 2016. Responsible Data Concerns with Open Source Intelligence. [Verkkojulkaisu]. Responsible Data. [Viitattu 2020-05-15]. Saatavissa: https://responsibledata.io/2016/11/14/responsible-data-open-source-intelligence/

31. HUFF, Josh 2018. OSINT: OPEN Source Intelligence. [PDF]. SANS. [Viitattu 2020-03-05]. Saatavissa: https://www.sans.org/cyber-security-summit/archives/file/summit-archive-1533737204.pdf
32. Hunchly s. a. a. The only Web Capture Tool Designed for Online Investigations. [Verkkojulkaisu]. Hunchly. [Viitattu 2020-05-14]. Saatavissa: https://www.hunch.ly/
33. Hunchly s. a. b. How Many Hunchly Licenses Do You Need? [Verkkojulkaisu]. Hunchly. [Viitattu 2020-05-14]. Saatavissa: https://hunch.ly/pricing
34. Infiniwiz s. a. What Is PII, Non-PII, and Personal Data? [Verkkojulkaisu]. Infiniwiz. [Viitattu 2020-05-15]. Saatavissa: https://www.infiniwiz.com/what-is-pii-non-pii-and-personal-data/
35. Information Warfare Center s. a. a. Cyber Investigations. [Verkkojulkaisu]. Csilinux.com. [Viitattu 2020-03-31]. Saatavissa: https://csilinux.com/
36. Information Warfare Center s. a. b. Download. [Verkkojulkaisu]. Csilinux.com. [Viitattu 2020-03-31]. Saatavissa: https://csilinux.com/download.html
37. Information Warfare Center s. a. c. Features. [Verkkojulkaisu]. Csilinux.com. [Viitattu 2020-03-31]. Saatavissa: https://csilinux.com/features.html
38. Information Warfare Center s. a. d. Tutorials. [PDF]. Csilinux.com. [Viitattu 2020-05-14]. Saatavissa: https://csilinux.com/Documents/A_beginners_guide_to_downloading_and_getting_started_with_CSI_Linux.pdf
39. JAMES, Luke 2018. 7 Tips for Using the Tor Browser Safely. [Verkkojulkaisu]. MakeUseOf. [Viitattu 2020-05-06]. Saatavissa: https://www.makeuseof.com/tag/tor-browser-safety-tips/
40. JavierOlmedo 2018. shodan-filters. [Verkkojulkaisu]. GitHub. [Viitattu 2020-04-13]. Saatavissa: https://github.com/JavierOlmedo/shodan-filters
41. Justin 2016. Dark Web OSINT With Python and OnionScan: Part One. [Verkkojulkaisu]. Automating OSINT. [Viitattu 2020-05-13]. Saatavissa: http://www.automatingosint.com/blog/2016/07/dark-web-osint-with-python-and-onionscan-part-one/
42. KINZIE, Kody 2019. Find Passwords in Exposed Log Files with Google Dorks. [Verkkojulkaisu]. WONDER HOW TO. [Viitattu 2020-03-11]. Saatavissa: https://null-byte.wonderhowto.com/how-to/find-passwords-exposed-log-files-with-google-dorks-0198557/
43. KINZIE, Kody 2020. How Hackers Use OSINT to Find Business Data. [Blogi]. Varonis. [Viitattu 2020-05-12]. Saatavissa: https://www.varonis.com/blog/best-sources-of-business-data-in-2019/
44. KOROLOV, Maria 2019. What is biometrics? 10 physical and behavioral identifiers that can be used for authentication [Verkkojulkaisu]. CSO. [Viitattu 2020-05-20]. Saatavissa: https://www.csoonline.com/article/3339565/what-is-biometrics-and-why-collecting-biometric-data-is-risky.html
45. Maltego s. a. a. What is Maltego? [Verkkojulkaisu]. Maltego. [Viitattu 2020-03-09]. Saatavissa: https://buy.maltego.com/shop/page/adaf0f4e-d531-45a1-8645-d63e9ea60fcc
46. Maltego s. a. b. BUY MALTEGO [Verkkojulkaisu]. Maltego. [Viitattu 2020-03-09]. Saatavissa: https://docs.maltego.com/support/solutions/articles/15000019166-what-is-maltego-
47. Maltego s. a. c. DOWNLOADS. [Verkkojulkaisu]. Maltego. [Viitattu 2020-05-18]. Saatavissa: https://www.maltego.com/downloads/
48. MCDONALD, Rob s. a. Personal Identifiable Information: HIPAA Best Practices. [Verkkojulkaisu]. Virtru. [Viitattu 2020-05-20]. Saatavissa: https://www.virtru.com/blog/personally-identifiable-information-hipaa/
49. Media Sonar 2020. OSINT Success: Efficacy While Remaining Legal and Ethical. [Verkkojulkaisu]. MEDIA SONAR. [Viitattu 2020-05-15]. Saatavissa: https://mediasonar.com/2020/04/30/legal-ethical-osint/

50. MIMOSO, Michael 2017. US, European Law Enforcement Shutter Massive AlphaBay Market. Threathpost. [Verkkojulkaisu]. [Viitattu 2020-04-28]. Saatavissa: https://threatpost.com/us-european-law-enforcement-shutter-massive-alphabay-market/126947/
51. MORGAN, Connor 2019. Humans On The Internet Will Triple From 2015 To 2022 And Hit 6 Billon. [Verkkojulkaisu]. CYBERCRIME MAGAZINE. [Viitattu 2020-03-10]. Saatavissa: https://cybersecurityventures.com/how-many-internet-users-will-the-world-have-in-2022-and-in-2030/
52. mySafety s. a. IDENTITEETTIVARKAUS. [Verkkojulkaisu]. mySafety. [Viitattu 2020-03-18]. Saatavissa: https://www.mysafety.fi/identiteettivarkaus
53. NAINI, Anjaneyulu 2019. 7 Popular Open Source Intelligence Tools for Penetration Testing. [Verkkojulkaisu]. GEEKFLARE. [Viitattu 2020-04-15]. Saatavissa: https://geekflare.com/osint-tools/
54. NATO Open Source Intelligence Handbook 2001. [Verkkokirja]. NATO. [Viitattu 2020-02-18]. Saatavissa: https://archive.org/details/NATOOSINTHandbookV1.2/mode/2up
55. NORDINE, Justin s. a. OSINT Framework. [Verkkojulkaisu]. [Viitattu 2020-04-10]. Saatavissa: https://osint-framework.com/
56. ODNI s. a. WHAT IS INTELLIGENCE? [Verkkojulkaisu]. Office of the Director of National Intelligence. [Viitattu 2020-03-18]. Saatavissa: https://www.dni.gov/index.php/what-we-do/what-is-intelligence
57. Oracle s. a. Welcome to VirtualBox.org! [Verkkojulkaisu]. VirtualBox. [Viitattu 2020-03-31]. Saatavissa: https://www.virtualbox.org/
58. OSINT solutions, Inc 2016. Advantages & Limitations of Open-Source Intelligence. [Verkkojulkaisu]. OSINTsolution. [Viitattu 2020-03-27]. Saatavissa: https://www.edocr.com/v/d4yjxo0a/osintsolutionseo/Advantages-and-Limitations-of-Open-Source-Intellig
59. OZKAYA, Erdal, ISLAM, Rafiqul 2019. Inside the Dark Web. [Verkkokirja]. CRC PRESS. [Viitattu 2020-04-29]. Saatavissa: https://ebookcentral-proquest-com.ezproxy.savonia.fi/lib/savoniafi/reader.action?docID=5793703&query=Inside+the+Dark+Web+
60. PASSI, Harpreet 2018. Top 10 Popular Open Source Intelligence (OSINT) Tools. [Blogi]. GreyCampus. [Viitattu 2020-05-22]. Saatavissa: https://www.greycampus.com/blog/information-security/top-open-source-intelligence-tools
61. PentestIT 2020. List of Operating Systems for OSINT (Open-Source Intelligence). [Verkkojulkaisu]. PentestIT. [Viitattu 2020-03-31]. Saatavissa: https://pentestit.com/operating-systems-open-source-intelligence-osint-list/
62. PETTERS, Jeff 2020. IDS vs. IPS: What is the Difference? [Blogi]. Varonis. [Viitattu 2020-05-12]. Saatavissa: https://www.varonis.com/blog/ids-vs-ips/
63. PÖHLMANN, Markus 2017. Abteilung III b. [Verkkojulkaisu]. International Encyclopedia of the First World War. [Viitattu 2020-03-16]. Saatavissa: https://encyclopedia.1914-1918-online.net/article/abteilung_iii_b
64. Privacy International 2017. Social Media Intelligence. [Verkkojulkaisu]. Privacy International. [Viitattu 2020-05-23]. Saatavissa: https://privacyinternational.org/explainer/55/social-media-intelligence
65. RAY, Goddy 2019. What Is Metadata and Why You Should Care. [Verkkojulkaisu]. SurfShark. [Viitattu 2020-5-14]. Saatavissa: https://surfshark.com/blog/what-is-metadata-and-why-you-should-start-caring-about-it
66. REED, Eric 2019. Psychology of a hack. [Verkkojulkaisu]. The Boston Globe. [Viitattu 2020-05-23]. Saatavissa: http://sponsored.bostonglobe.com/kaspersky/psychology-of-a-hack/

67. Renewable Freedom Foundation s. a. Tor: Enabling Anonymity, Fighting Censorship. [Verkkojulkaisu]. Renewable Freedom Foundation [Viitattu 2020-05-19]. Saatavissa: https://renewablefreedom.org/projects/tor/
68. ROSE, Margaret 2020. TCP (Transmission Control Protocol). [Verkkojulkaisu]. TechTarget. [Viitattu 2020-04-13]. Saatavissa: https://searchnetworking.techtarget.com/definition/TCP
69. RYTE WIKI s. a. Tracking Code. [Verkkojulkaisu]. RYTE WIKI. [Viitattu 2020-05-14]. Saatavissa: https://en.ryte.com/wiki/Tracking_Code
70. SCHULTZ, Jeff 2019. How Much Data is Created on the Internet Each Day? [Verkkojulkaisu]. MicroFocus. [Viitattu 2020-02-21]. Saatavissa: https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/#
71. SecurityTrails Team 2018 a. What is OSINT? How can I make use of it? [Blogi]. SecurityTrails. [Viitattu 2020-05-15]. Saatavissa: https://securitytrails.com/blog/what-is-osint-how-can-i-make-use-of-it
72. SecurityTrails Team 2018 b. Top 20 OSINT Tools. [Blogi]. SecurityTrails. [Viitattu 2020-05-17]. Saatavissa: https://securitytrails.com/blog/top-20-intel-tools
73. SecurityTrails Team 2019. Exploring Google Hacking Techniques. [Blogi]. SecurityTrails. [Viitattu 2020-04-07]. Saatavissa: https://securitytrails.com/blog/google-hacking-techniques
74. SEITZ, Justin s. a. Hunchly DARK WEB INVESTIGATION GUIDE. [PDF]. Huncly. [Viitattu 2020-05-14]. Saatavissa: https://www.hunch.ly/resources/Hunchly-Dark-Web-Setup.pdf
75. SentinelOne 2019. What is OSINT? (And How Is It Used?). [Blogi]. SentinelOne. [Viitattu 2020-02-22]. Saatavissa: https://www.sentinelone.com/blog/what-is-osint-how-is-it-used/
76. SEORG 2019. Protect Yourself from Social Media Scams. [Verkkojulkaisu]. SECURITY THROUGH EDUCATION. [Viitattu 2020-05-22]. Saatavissa: https://www.social-engineer.org/general-blog/protect-yourself-from-social-media%e2%80%afscams/
77. SHEILS, Conor 2020. THE DEEP WEB AND THE DARK WEB. [Verkkojulkaisu]. Digital.com. [Viitattu 2020-04-15]. Saatavissa: https://digital.com/blog/deep-dark-web/
78. Shodan 2020a. What is Shodan? [Verkkojulkaisu]. Shodan. [Viitattu 2020-04-08]. Saatavissa: https://help.shodan.io/the-basics/what-is-shodan
79. Shodan 2020b. Explore. [Verkkojulkaisu]. Shodan. [Viitattu 2020-04-13]. Saatavissa: https://www.shodan.io/explore
80. Shodan 2020c. Choose Your Plan. [Verkkojulkaisu]. Shodan. [Viitattu 2020-05-17]. Saatavissa: https://developer.shodan.io/billing/signup
81. SLUIJTER, Niklas 2020. How OSINT is Used Against Your Employees. [Verkkojulkaisu]. HOXHUNT. [Viitattu 2020-04-15]. Saatavissa: https://www.hoxhunt.com/blog/how-osint-is-used-against-your-employees/
82. SOBERS, Rob 2020. The World in Data Breaches. [Blogi]. Varonis. [Viitattu 2020-05-10]. Saatavissa: https://www.varonis.com/blog/the-world-in-data-breaches/
83. S-rah 2017. onionscan. [Verkkojulkaisu]. GitHub. [Viitattu 2020-05-13]. Saatavissa: https://github.com/s-rah/onionscan
84. Statista 2020. Global digital population as of January 2020. [Verkkojulkaisu]. Statista. [Viitattu 2020-03-10]. Saatavissa: https://www.statista.com/statistics/617136/digital-population-worldwide/
85. SWINHOE, Dan 2020. The 15 biggest data breaches of the 21st century. [Verkkojulkaisu]. CSO. [Viitattu 2020-05-18]. Saatavissa: https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html

86. The PTES Team s. a. Intelligence Gathering. [Verkkojulkaisu]. Pentest-standard. [Viitattu 2020-03-10]. Saatavissa: https://pentest-standard.readthedocs.io/en/latest/intelligence_gathering.html
87. THE RECORDER FUTURE TEAM 2019. What Is Open Source Intelligence and How Is it Used. [Verkkojulkaisu]. Recorded Future. [Viitattu 2020-03-05]. Saatavissa: https://www.recordedfuture.com/open-source-intelligence-definition/
88. The United States Department of Justice 2017. AlphaBay, the Largest Online 'Dark Market,' Shut Down. [Verkkojulkaisu]. The United States Department of Justice. [Viitattu 2020-04-29]. Saatavissa: https://www.justice.gov/opa/pr/alphabay-largest-online-dark-market-shut-down
89. TOKYONEON 2017. Detect Misconfigurations in 'Anonymous' Dark Web Sites with OnionScan. [Verkkojulkaisu]. WONDER HOW TO. [Viitattu 2020-05-13]. Saatavissa: https://null-byte.wonderhowto.com/how-to/detect-misconfigurations-anonymous-dark-web-sites-with-onionscan-0181366/
90. TOMES, Tim 2020. recon-ng. [Verkkojulkaisu]. GitHub. [Viitattu 2020-04-20]. Saatavissa: https://github.com/lanmaster53/recon-ng
91. TOTOUNJI, Ayman 2017. VIRTUAL MACHINES: A CLOSER LOOK. [Verkkojulkaisu] CYNEXLINK. [Viitattu 2020-05-15]. Saatavissa: https://www.cynexlink.com/2017/08/18/virtual-machines-pros-cons/
92. TYSON, Matthew, JavaWorld 2019. What is Tomcat? The original Java servlet container. [Verkkojulkaisu]. JAVAWORLD. [Viitattu 2020-04-13]. Saatavissa: https://www.javaworld.com/article/3510460/what-is-apache-tomcat-the-original-java-servlet-container.html
93. U.S. General Services Administration 2019. 2180.2 CIO GSA Rules of Behavior for Handling Personally Identifiable Information (PII). [Verkkojulkaisu]. GSA. [Viitattu 2020-05-19]. Saatavissa: https://www.gsa.gov/directive/gsa-rules-of-behavior-for-handling-personally-identifiable-information-(pii)-
94. U.S. Government Printing Office 2006. NATIONAL DEFENCE AUTHORIZATION ACT FOR FISCAL YEAR 2006. [Verkkojulkaisu]. Congress Public Law. [Viitattu 2020-03-24]. Saatavissa. https://www.congress.gov/109/plaws/publ163/PLAW-109publ163.htm
95. VACCA, John 2013. Computer and Information Security Handbook 2nd Edition. [Verkkokirja]. Morgan Kaufmann. [Viitattu 2020-04-15]. Saatavissa: https://www.sciencedirect.com/topics/computer-science/onion-router
96. Webroot s. a. What is Social Engineering? [Verkkojulkaisu]. WEBROOT. [Viitattu 2020-05-10]. Saatavissa: https://www.webroot.com/us/en/resources/tips-articles/what-is-social-engineering
97. Whonix s. a. Software That Can Anonymize Everything You Do Online. [Verkkojulkaisu]. Whonix. [Viitattu 2020-05-14]. Saatavissa: https://www.whonix.org/
98. WILLIAMS, Heather J., BLUM, Ilana 2018. Defining Second Generation Open Source Intelligence (OSINT) for the Defence Enterprise. [PDF]. RAND CORPORATION. [Viitattu 2020-03-24]. Saatavissa: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=14&ved=2ahUKEwiDxqqh3bLoAhXE-yoKHXRPAMUQFjANegQIBBAB&url=https%3A%2F%2Fwww.rand.org%2Fcontent%2Fdam%2Frand%2Fpubs%2Fresearch_reports%2FRR1900%2FRR1964%2FRAND_RR1964.pdf&usg=AOvVaw2FBYLg197InRwO7z4NIPuf
99. WILSON, Emily, GOLLNICK, Clare 2017. SEPARATING FACT FROM FICTION THE TRUTH ABOUT THE DARK WEB. [Verkkolehti]. Cyber Defense Magazine. [Viitattu 2020-04-21] Saatavilla: https://www.cyberdefense-magazine.com/annual-editions/RSA-2017/

100. WILSON, Jim 2019. What Shodan Is and How to Use It Most Effectively. [Blogi]. SafetyDetectives. [Viitattu 2020-04-08]. Saatavissa: https://www.safetydetectives.com/blog/what-is-shodan-and-how-to-use-it-most-effectively/

101. Wondersmith_rae 2019. A Beginner's Guide to OSINT Investigator with Maltego. [Verkkojulkaisu]. Medium. [Viitattu 2020-03-09]. Saatavissa: https://medium.com/@raebaker/a-beginners-guide-to-osint-investigation-with-maltego-6b195f7245cc

102. z3roTrust 2018. Open-Source Intelligence (OSINT) Reconnaissance. [Verkkojulkaisu]. Medium. [Viitattu 2020-04-15]. Saatavissa: https://medium.com/@z3roTrust/open-source-intelligence-osint-reconnaissance-75edd7f7dada

103. z3roTrust 2019. The OSINT-ification of ISIS on the Dark Web. [Verkkojulkaisu]. Medium. [Viitattu 2020-04-15]. Saatavissa: https://medium.com/@z3roTrust/the-osint-ification-of-isis-on-the-dark-web-19644ec90253

104. ZENG, Meela 2018. Difference between HTTPS Port 443 and Port 8443. [Verkkojulkaisu]. Router-Switch.com. [Viitattu 2020-04-13]. Saatavissa: https://www.router-switch.com/faq/difference-between-https-port-443-and-8443.html

LIITE 1: ONIONSCAN ASENNUSOHJE JA MODIFIOITU KOODI

OnionScan asennetaan CSI Linux Investigator käyttöjärjestelmälle, hyödyntäen VirtualBox-käynnistysalustaa. Ennen OnionScanin asennusta asennetaan tarvittavat sovelluspäivitykset ja ladataan tarvittavat sovelluspaketit, joista OnionScan on riippuvainen. Asennukset suoritetaan Linux terminaalia käyttäen, alla olevien komentojen avulla.

Ensiksi tarkistetaan sovelluspäivitykset.

- sudo apt-get update

```
csi@csi-analyst:~$ sudo apt-get update
```

Tämän jälkeen asennetaan OnionScanin vaatimat sovellus paketit.

- apt-get install tor git bison libexif-dev

```
csi@csi-analyst:~$ sudo apt-get install tor git bison libexif-dev
```

- apt-get install python-pip

```
csi@csi-analyst:~$ sudo apt-get install python-pip
```

- pip install stem

```
csi@csi-analyst:~$ pip install stem
```

Seuraavaksi asennetaan Go-ohjelmointikieli versio 1.9, jonka avulla pystytään suorittamaan OnionScan ohjelmisto.

- bash < <(curl -s -S -L https://raw.githubusercontent.com/moovweb/gvm/master/binscripts/gvm-installer)

```
csi@csi-analyst:~$ bash < <(curl -s -S -L https://raw.githubusercontent.com/moov
web/gvm/master/binscripts/gvm-installer)
```

- [[-s "$HOME/.gvm/scripts/gvm"]] && source "$HOME/.gvm/scripts/gvm"

```
csi@csi-analyst:~$ [[ -s "$HOME/.gvm/scripts/gvm" ]] && source "$HOME/.gvm/scrip
ts/gvm"
```

- gvm install go1.9 --binary

```
csi@csi-analyst:~$ gvm install go1.9 --binary
```

- gvm use go1.9

```
csi@csi-analyst:~$ gvm use go1.9
```

Tämän jälkeen voidaan asentaa OnionScan.

- go get github.com/s-rah/onionscan

```
csi@csi-analyst:~$ go get github.com/s-rah/onionscan
```

- go install github.com/s-rah/onionscan

```
csi@csi-analyst:~$ go install github.com/s-rah/onionscan
```

OnionScan on nyt asennettu ja kaikki tarvittavat sovellukset, joista OnionScan on riippuvainen. OnionScan toimivuus voidaan testata onionscan komennolla.

- onionscan

```
csi@csi-analyst:~$ onionscan
Usage of onionscan:
    onionscan [flags] hiddenservice | onionscan [flags] --list list | onionscan --mode analysis
  -batch int
        number of onions to scan concurrently (default 10)
  -cookie string
        if provided, onionscan will use this cookie
  -crawlconfigdir string
        A directory where crawl configurations are stored
  -dbdir string
        The directory where the crawl database will be stored (default "./onionscandb")
  -depth int
        depth of directory scan recursion (default: 100) (default 100)
  -fingerprint
        true disables some deeper scans e.g. directory probing with the aim of just getting a fingerprint of the service. (default true)
  -jsonReport
        print out a json report providing a detailed report of the scan.
  -jsonSimpleReport
        print out a simple report as json, false by default
  -list string
        If provided OnionScan will attempt to read from the given list, rather than the provided hidden service
  -mode string
        one of scan or analysis. In analysis mode, webport must be set. (default "scan")
  -reportFile string
        the file destination path for report file - if given, the prefix of the file will be the scanned onion service. If not given, the report will be written to stdout
  -scans string
        a comma-separated list of scans to run e.g. web,tls,... (default: run all)
  -simpleReport
        print out a simple report detailing what is wrong and how to fix it, true by default (default true)
  -timeout int
        read timeout for connecting to onion services (default 120)
  -torProxyAddress string
        the address of the tor proxy to use (default "127.0.0.1:9050")
  -verbose
        print out a verbose log output of the scan
  -webport int
        if given, onionscan will expose a webserver on localhost:[port] to enabled searching of the database (default 8080)
```

Näin ollen, kun OnionScan on saatu toimimaan, voidaan alkaa modifioimaan OnionScania. Tämä mahdollistaa OnionScan:in järjestelmällisen hallinnan ja tuloksien käsittelyn. Tätä varten täytyy konfiguroida Tor-verkkoa, jotta voidaan kutsua uusia Tor-verkon IP-osoitteita.

Aluksi luodaan salattu todennusavain.

- tor --hash-password Savonia

```
csi@csi-analyst:~$ tor --hash-password Savonia
16:F3DBD4F9A75E07FF60BF6AB8FE94A4DE1DDCEA19A21460052847B046B8
```

Seuraavaksi luodaan Tor-verkon ohjaus

- nano -w /etc/tor/torcc

```
csi@csi-analyst:~$ nano -w /etc/tor/torrc
```

- ControlPort 9051
- ControlListenAddress 127.0.0.1
- HashedControlPassword 16:F3DBD4F9A75E07FF60BF6AB8FE94A4DE1DDCEA19A21460052847B046B8

```
ControlPort 9051
ControlListenAddress 127.0.0.1
HashedControlPassword 16:F3DBD4F9A75E07FF60BF6AB8FE94A4DE1DDCEA19A21460052847B046B8
```

Tämän jälkeen käynnistetään Tor-palvelin uudelleen.

- service tor restart

```
csi@csi-analyst:~$ service tor restart
```

Konfigurointi on suoritettu, jonka jälkeen voidaan aloittaa kirjoittamaan automatisoidumman version OnionScan työkalusta python-ohjelmointi kielellä. Ohjelmiston nimi on onionrunner.py.

```
#  Rivit 3-14:  Tuodaan sciptissä käytettävät moduulit.

#-*- coding: utf-8 -*-
from stem.control import Controller
from stem import Signal
from threading import Timer
from threading import Event
```

```
import codecs
import json
import os
import random
import subprocess
import sys
import time

#  Rivit 17-18: Luodaan kaksi tyhjää listää, johon tuodaan käsiteltävät .onion-sivus-
tot ja toiseen haun aikana käsiteltävät .onion-sivustot.
onions = []
session_onions = []

#  Rivit 21-22: Luodaan Event objekti, joka mahdollistaa kahden suoritettavan kejun käy-
tön. Set Event asetetaan ensin, jotta pääketju toteutetaan myöhemmin.
identity_lock = Event()
identity_lock.set()

# Rivit 26-40: Valitaan käytettävä .onion-sivusto lista. Avataan käytet-
tävä lista, jos listaa ei löydy lopetetaan sovelluksen toiminta.
# Luetaan listan sisältö läpi ja lisätään ne stored_onions listaan. Esitetään komentoke-
hotteessa skannauksen kulku kappaleina.
def get_onion_list():

   if os.path.exists("Onion_Sivut_Lista.txt"):

        with open("Onion_Sivut_Lista.txt", "rb") as fd:

            stored_onions = fd.read().splitlines()
    else:

        print "[!] Can't find onion list!"
        sys.exit(0)

    print "[*] Total onions for scanning: %d" % len(stored_onions)

    return stored_onions

# Rivit 43-50: Kirjoitetaan löydetyt .onion-sivustot käytettyyn onion listaan
def stored_onion(onion):

    print "[++] Storing %s in master list." % onion

    with codecs.open("Onion_Sivut_Lista.txt","ab",encoding="utf8") as fd:
        fd.write("%s\n" % onion)

    return

# Rivit: 54-75: Suoritetaan sivuston tiedonetsintä. Asetetaan ajastin 300 sekunt-
tiin. Aliprosessin Popenin avulla saadaan sivuston tiedot JSON-muodossa ja kommuni-
koida stdout:in, eli standard output:in ja stderr:in Standard streams kanssa.
```

```
# Nämä mahdollistavat tiedon lukemisen ja kirjoittamisen. Jos sivustolta löydetään tar-
vittavat tiedot, nollataan ajastin ja palautetaan JSON-muotoinen stdout. Jos vii-
dessä minuutissa ei saada tietoja palautetaan vain None.
def run_onionscan(onion):

    print "[*] Onionscanning %s" % onion

    # Käynnistetään onionscan
    process = subprocess.Popen(["onionscan","--webport=0","--jsonReport","--simpleRe-
port=false",onion],stdout=subprocess.PIPE,stderr=subprocess.PIPE)

    # Ajastin
    process_timer = Timer(300,handle_timeout,args=[process,onion])
    process_timer.start()

    # Odotetaan onionscanin tuloksia
    stdout = process.communicate()[0]

    # Saatiin tulokset, lopetetaan ajastin.
    if process_timer.is_alive():
        process_timer.cancel()
        return stdout

    print "[!!!] Process timed out!"

    return None

# Rivit: 78-114: handle_timeout funktiolla, pyritään lopettamaan vanha suori-
tus ja sen jälkeen alustamaan uusi yhteys Tor-verkkoon.
def handle_timeout(process,onion):

    global session_onions
    global identity_lock

    # Tyhjennetään pääketju, ennen kun aloitamme uuden suorituksen.
    identity_lock.clear()

    # Lopetetaan prosessiobjekti, jonka toteuttamisessa kesti liian kauan.
    try:
        process.kill()
        print "[!!!] Killed the onionscan process."
    except:
        pass

    # Vaihdetaan Tor-identiteettiä varmistaaksemme hyvän yhteyden.
    with Controller.from_port(port=9051) as torcontrol:

        # Todentaminen paikalliselle TOR-ohjaimelle.
        torcontrol.authenticate("Savonia")

        # Lähetetään uusi signaali uudeksi identiteetiksi.
        torcontrol.signal(Signal.NEWNYM)
```

```
        # Odotetaan uuden identiteetin alustamista.
        time.sleep(torcontrol.get_newnym_wait())

        print "[!!!] Switched TOR identities."

    # Lisätään sivuto uudellee listaan ja sekoitetaan session_onions lista.
    session_onions.append(onion)
    random.shuffle(session_onions)

    # Mahdollistetaan pääketjun palata suoritukseen.
    identity_lock.set()

    return

# Rivit 117-142: process_results funktiolla määritellään saadut tulokset.
def process_results(onion,json_response):
    global onions
    global session_onions

    # Luodaan tuloksille oma kansio.
    if not os.path.exists("OnionScan_Tulokset"):
        os.mkdir("OnionScan_Tulokset")

    # Kirjoitetaan JSON-muotoiset tulokset kansioon.
    with open("%s/%s.json" % ("OnionScan_Tulokset",onion), "wb") as fd:
        fd.write(json_response)

    # Etsitään uusia .onion-sivustoja skannaus listaan.
    scan_result = ur"%s" % json_response.decode("utf8")
    scan_result = json.loads(scan_result)

    if scan_result ['identifierReport']['linkedOnions'] is not None:
        add_new_onions(scan_result['identifierReport']['linkedOnions'])

    if scan_result['identifierReport']['relatedOnionDomains'] is not None:
        add_new_onions(scan_result['identifierReport']['relatedOnionDomains'])

    if scan_result['identifierReport']['relatedOnionServices'] is not None:
        add_new_onions(scan_result['identifierReport']['relatedOnionServices'])

    return

# Rivit 145-161: add_new_onions funktio käsitteleen uudet löydetyt .onion-sivustot
def add_new_onions(new_onions_list):

    global onions
    global session_onions

    # Käsitellään new_onions_list lista läpi. Tarkistetaan, ettei kyseistä sivus-
toa ole jo listassa ja että sivusto on .onion päätteinen. Tämän jälkeen lisätään si-
vusto listaan ja sekoitamme listan.
    for linked_onion in new_onions_list:
```

```
        if linked_onion not in onions and linked_onion.endswith(".onion"):

            print "[++] Discovered new .onion => %s" % linked_onion

            onions.append(linked_onion)
            session_onions.append(linked_onion)
            random.shuffle(session_onions)
            stored_onion(linked_onion)

    return

# Kutsutaan get_onion_list funktiota joka lataa kaikki .onion-sivustot.
onions = get_onion_list()

# Sekoitetaan sivustot ja luodaan kopio listasta ja tallennetaan se session_onions lis-
taan. Tämän jälkeen alustetaan laskurimuuttuja, joka mahdollistaa käsiteltävien .onion-
sivustojen loppumisajankohdan.
random.shuffle(onions)
session_onions = list(onions)

count = 0
 # Rivit: 173-199: Luodaan silmukka, joka lopettaa suorittamisen, kun kaikki .onion-si-
vustot on käyty läpi.
while count < len(onions):

    # Odotetaan Event-objektion asetusta, ennen suorittamisen jatkamista. Muis-
tat, että tämä pysähtyy täällä vain, jos handle_timeout-funktiomme käsittelee uuden Tor-
identiteetin tarttumista.
    # Kun tunnuslukko on tyhjennetty, siirrymme tämän rivin ohi
    identity_lock.wait()

    # Otetaan uusi .onion-sovusto käisttelyyn.
    print "[*] Running %d of %d." % (count,len(onions))
    onion = session_onions.pop()

    # Tarkistetaan ollaanko saatu jo tulokset kyseiseltä .onion-sivustolta. Jos on ski-
pataan ja siirrytään seuraavaan muuten jatketaan.
    if os.path.exists("OnionScan_Tulokset/%s.json" % onion):

        print "[!] Already retrieved %s. Skipping." % onion
        count += 1

        continue

    # Suoritetaan .onion-sivuston skannaus.
    result = run_onionscan(onion)

    # Prosessoidaan tulokset.
    if result is not None:

        if len(result):
            process_results(onion,result)
            count += 1
```