Viljami Töhönen

Sosiaalisen median käyttäjän digitaalinen jalanjälki ja sen hallinta

Tietotekniikan Kandidaatintutkielma

5. toukokuuta 2024

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Viljami Töhönen

Yhteystiedot: vijumato@jyu.fi

Ohjaaja: Timo Tiihonen

Työn nimi: Sosiaalisen median käyttäjän digitaalinen jalanjälki ja sen hallinta

Title in English: Social media user's digital footprint and its management

Työ: Kandidaatintutkielma

Sivumäärä: 24+0

Tiivistelmä: Avointen lähteiden tiedustelu on noussut keskeiseksi keskustelunaiheeksi etenkin internetin ja Ukrainan sodan takia. Yksi avointen lähteiden tiedustelun sovelluskohde on sosiaalisen median käyttäjän digitaalisen jalanjäljen tarkastelu. Pahimmassa tapauksessa digitaalinen jalanjälki voi kääntyä käyttäjää vastaan. Sosiaalisen median käyttäjien onkin hyvä olla tietoinen omasta digitaalisesta jalanjäljestään ja mahdollisista hallintatavoista.

Avainsanat: Avointen lähteiden tiedustelu, Digitaalinen jalanjälki, Digitaalisen jalanjäljen hallinta,

Abstract: Open source intelligence has risen to central point of discussion after emerging of the internet and war in Ukraine. One of the application of open source intelligence is the inspection of social media user's digital footprint. In worst case scenario digital footprint can turn against the user. Social media user's therefore should know about their digital footprint and possible ways to control it.

Keywords: Open source intelligence, Digital footprint, Digital footprint management

Termiluettelo

Avointen lähteiden tiedustelu	(engl. Open Source Intelligence, OSINT) tiedustelulaji, jossa dataa kerätään avoimista lähteistä.
Bellingcat	Bellingcat on riippumaton verkkojulkaisualusta, joka koostuu tutkijoista ja journalisteista. Bellingcat on erikoistunut erityisesti avointen lähteiden tutkimiseen.
IP	Lyhenne sanoista Internet Protocol. Verkkokerroksella yleisesti käytetty protokolla.
NATO	Sanoista North Atlactic Treaty Organization suomeksi Pohjois-Atlantin liitto. NATO on poliittinen ja sotilaallinen liitto, jonka tarkoitus on turvata jäsenmaidensa vapaus ja turvallisuus.
Ohjelmointirajapinta	(engl. application programming interface, API) mahdollistaa kommunikoinnin sovellusten välillä.
Sosiaalinen manipulointi	(engl. Social Engineering) on erityisesti rikollisten käyttämä tekniikka, jolla huijataan uhri paljastamaan arkaluonteista sisältöä tai huijataan uhri tekemään muuta rikollisen toiveesta.
Sosiaalinen media	Internetin palvelu, joissa käyttäjillä on mahdollista kommunikoida toistensa kanssa ja tuottaa omaa sisältöä.
Sosiaalisen median käyttäjä	Henkilö, joka käyttää sosiaalisen median palveluita.
TCP	Lyhenne sanoista Transmission Control Protocol. Kuljetuskerroksella yleisesti käytetty prokolla.

Kuviot

Kuvio 1. TCP/IP-protokollapino 9

Sisällys

1 JOHDANTO 1

2 AVOINTEN LÄHTEIDEN TIEDUSTELU 3
2.1 Avoimet lähteet 3
2.2 Avointen lähteiden tiedustelu 4
2.3 Avointen lähteiden tiedustelija 5
2.3.1 Julkiset toimijat 5
2.3.2 Yksityiset toimijat 6
2.4 Tiedonkeräys sosiaalisesta mediasta 7

3 DIGITAALINEN JALANJÄLKI 9
3.1 Passiivinen digitaalinen jalanjälki 9
3.1.1 TCP/IP-protokollapino 9
3.1.2 Evästeet ja muut tunnistetiedot 10
3.2 Aktiivinen digitaalinen jalanjälki 11

4 DIGITAALISEN JALANJÄLJEN HALLINTA 12
4.1 Miksi omaa digitaalista jalanjälkeä olisi hyvä hallita? 12
4.2 Hallinta 13

5 YHTEENVETO 15

LÄHTEET 16

1 Johdanto

Internetin aikakausi on tuonut ihmisille mahdollisuuden kommunikoida ja verkostoitua uudella tavalla. Suuri avoimen informaation määrä voi olla suo, mutta oikealle etsijälle varsinainen kultakaivos. Pew Research Centerin tekemässä kyselytutkimuksessa vuosina 2005-2015 käy ilmi, että lähes 65% yhdysvaltalaisista aikuisista käyttää ainakin yhtä sosiaalisen median alustaa. Nuorilla (18v - 29v) luku on noin 90% (Perrin 2015). On syytä olettaa, että sosiaalisen median käyttö Suomessa ja Yhdysvalloissa eivät merkittävästi poikkea toisistaan.

Digitaalinen jalanjälki koostuu siitä kaikesta tiedosta, jota jätetään jälkeen käytettäessä internetiä. Oma ja muiden digitaalinen jalanjälki kiinnostaa muita internetin käyttäjiä. Madden ym. (2007) kertovat, että vuonna 2007 tietoa itsestään internetistä oli hakenut 47%. Madden (2013) kertoo, että vuoteen 2013 mennessä luku oli noussut 56%.

Avointen lähteiden tiedustelu (engl. OSINT) on tiedustelulaji, jossa informaatiota kerätään julkisista lähteistä vastaamaan tiedustelijan asettamaan tiedustelukysymykseen (Steele 2007). Sosiaalisen median kontekstiin sovellettua avointen lähteiden tiedustelua kutsutaan joskus nimellä SOCMINT (sanoista Social Media Intelligence). Sosiaalisen median alustoina voidaan pitää esimerkiksi Facebookkia, Instagramia ja LinkedIniä (Şuşnea ym. 2018). Kerätty tieto voi mahdollisesti vaikuttaa käyttäjän elämään, hyvinvointiin ja turvallisuuteen. Taitava tiedustelija osaa tehdä datan perusteella johtopäätöksiä kohteestaan (Pastor-Galindo ym. 2020).

Samalla kun sosiaalisen median käyttö on suosittua ja dataa jaetaan paljon sosiaalisen median alustoilla, siitä koituvat lieveilmiöt, kuten kyberrikollisuus, identiteettivarkaudet ja häirintä kulkevat sen rinnalla (Saridakis ym. 2016).

Noin 86% prosenttia internetin käyttäjistä on yrittänyt piilottaa digitaalista jalanjälkeään. (Rainie ym. 2013). Internetin käyttäjille siis löytyy kiinnostusta hallita netistä löytyvää digitaalista jalanjälkeä. Tämän tutkielman on tarkoitus kartoittaa yksityishenkilön digitaalisen jalanjäljen merkitystä ja sen hallintaa. Tutkielmassa lähestytään näkökulmaa avointen lähteiden tiedustelun kautta, sillä se on tehokas tapa kerätä tietoa yksityishenkilön digitaalisesta jalanjäljestä. Yksityishenkilön on hyvä tiedostaa avointen lähteiden tiedustelun merkitys

tehdessään päätöksiä oman digitaalisen jalanjäljen suhteen. Tutkielman tutkimuskysymykset ovat: Mitä tietoja voidaan internetistä kerätä henkilöstä ja miten niiden pääsy ulkopuolisten käsiin voidaan estää?

Tutkielma toteutetaan kirjallisuuskatsauksena ja se sisältää kolme päälukua. Toisessa luvussa perehdytään avointen lähteiden tiedusteluun. Kolmannessa luvussa tarkastellaan käyttäjän synnyttämään digitaalista jalanjälkeä. Neljännessä luvussa pohditaan käyttäjän digitaalisen jalanjäljen merkitystä ja hallintaa.

2 Avointen lähteiden tiedustelu

Tässä luvussa perehdytään avointen lähteiden tiedusteluun. Ensimmäisessä alaluvussa käsitellään avoimia lähteitä. Toisessa alaluvussa avataan avointen lähteiden tiedustelua yleisesti. Kolmannessa alaluvussa käsitellään avointen lähteiden tiedustelijaa, alaluvussa vastataan kysymykseen; kuka käyttää avointen lähteiden tiedustelua hyväkseen. Viimeisessä alaluvussa käsitellään minkälaista tietoa voidaan hakea avointen lähteiden tiedustelun avulla sosiaalisesta mediasta.

2.1 Avoimet lähteet

Hribar ym. (2014) määrittelevät neljä eri luokkaa informaatiolle perustuen informaation saatavuudelle; "valkoinen informaatio", "harmaa informaatio", "musta informaatio"ja "olematon informaatio"(suomennokset minun).

Hribar ym. (2014) mainitsevat, että valkoinen informaatio muodostaa informaatiokokonaisuudesta 90%. Valkoinen informaatio on informaatiota, jonka saatavuus on täysin julkista ja laillista. Se on informaatiota, jonka levikki on tyypillisesti laajaa. Esimerkkejä tällaisesta informaatiosta ovat sanomalehdet, vertaisarvioidut artikkelit ja kirjat (Hribar ym. 2014). Williams ym. (2018) määrittelevät sosiaalisen median konteksissa avoimet lähteet kahteen eri luokkaan; hidas sosiaalisen median sisältö (engl. *Long-Form Social Media Content*) ja nopea sosiaalisen median sisältö (engl. *Short-Form Social Media Content*) (suomennokset minun).

Hidas sosiaalisen median sisältö on usein tekstipainotteista sisältöä. Se voi olla esimerkiksi verkossa julkaistuja blogitekstejä tai Reddit-postauksia. (Williams ym. 2018). Sanamäärältään hidas sosiaalisen median sisältö on tyypillesti pidempää, kuin nopea sosiaalisen median sisältö. Edellisen artikkelin kirjoittajat mainitsevat, että sosiaalisen median analyysi painottuu nopeaan sosiaalisen median sisältöön. Jolloin hidas sisältö jää usein alihyödynnetyksi.

Nopea sosiaalisen median sisältö on tarkoitettu nopeasti kulutettavaksi. Esimerkiksi Twitterissä julkaistu twiitti tai Facebookissa julkaistu päivitys ovat nopeaa sosiaalisen median sisäl-

töä (Williams ym. 2018). Twitterissä twiittien maksimipituus on 280 merkkiä (Twitter 2024). Williams ym. (2018) mainitsevat, että yksittäisellä nopean sosiaalisen median sisällön viestillä on harvoin tiedusteluarvoa. Tiedusteluarvoa saadaan usein tällaisten viestien koosteesta. Poikkeuksena artikkelin kirjoittajat mainitsevat viestit käyttäjiltä, jotka ovat yhteiskunnallisessa erityisasemassa.

Harmaa informaatio muodostaa koko informaatiokokonaisuudesta 9%. Harmaaksi informaatioksi sanotaan sellaista informaatiota, jonka saatavuus on kyllä laillista ja julkista, mutta informaatiota ei ole levitetty tai julkaistu laajasti. Siltä puuttuu viralliset julkaisukanavat. Harmaan kirjallisuuden voidaan katsoa olevan harmaata informaatiota (Hribar ym. 2014). Harmaa kirjallisuus on kirjallisuutta, jota on vaikea löytää harmaan informaation tapaan, koska sillä ei ole laajaa julkaisualustaa. Harmaalta kirjallisuudelta puuttuu yleensä vertaisarvionti, joten tietoon tulee suhtautua kriittisesti. Tieto voi olla kuitenkin ajankohtaista, koska julkaisemista ei hidasta vertaisarviontiprosessi (LibGuides, n.d.).

Musta informaatio on salaiseksi luokiteltua informaatiota ja se muodostaa 0,9% informaatiokokonaisuudesta(Hribar ym. 2014). Steele (2007) mainitsee, että avointen lähteiden tiedustelussa tietoa kerätään avoimista lähteistä, joten salaiseksi luokiteltu informaatio ei sovellu käytettäväksi avointen lähteiden tiedusteluun.

Williams ym. (2018) mainitsevat, että internetin aikakausi on tuonut uusia haasteita valkoisen ja harmaan informaation määrittelyyn. Enää ei ole yhtä helppoa erottaa valkoista ja harmaata informaatiota toisistaan. Edellä mainitussa artikkelissa esitetäänkin, että informaatio voitaisiin luokitella yksilöiden tuottamaan informaation ja instituutioiden tuottamaan informaation.

2.2 Avointen lähteiden tiedustelu

Avointen lähteiden tiedustelu (engl. OSINT) on julkisista lähteistä kerättyä informaatiota, jonka avulla voidaan käsittelyn ja tulkinnan jälkeen voidaan vastata asetettuun tiedustelukysymykseen (Steele 2007). Avointen lähteiden tiedustelun vahvuudeksi voidaan mainita avointen lähteiden määrän, työkalujen ja tiedon helppo saatavuus. Avoimista lähteistä informaation kerääminen on yleensä halvempaa kuin suljetuista. Tehokas avointen lähteiden tie-

dustelu kuitenkin vaatii, että informaation purkaminen, kokoaminen ja oikeaan kysymykseen vastaaminen suorittaa taho, jolla on alan erityistaitoja (Hribar ym. 2014). Edellisen artikkelin kirjoittajat mainitsevat, että ammattimainen tiedustelija osaa etsiä oikeaa informaatiota oikeaan kysymykseen ja käyttää oikeita työkaluja. Ammattimainen tiedustelija osaa hakea myös relevanttia tietoa oikeasta paikasta. (NATO 2001) mainitsee myös operaatioturvallisuuden (engl. Operational Security) ja operaation resurssit kulmakiviksi tehokkaan avointen lähteiden tiedustelun toimintaan

Avointen lähteiden tiedustelu on ollut käytössä jo pitkään. Journalistit, yliopistot ja muut toimijat ovat ajan saatossa tutkineet avoimia lähteitä (Hwang ym. 2022). Avoimia lähteitä ennen internetiä olivat esimerkiksi sanomalehdet, radiolähetykset ja kirjat. Internetiä voidaan pitää suurena avointen lähteiden varastona. Internetin kasvun myötä myös lähteiden saatavuus on parantunut ja alalle on tullut uusia toimijoita (Mercado 2009).

Avointen lähteiden tiedustelu ei voi aina korvata muiden tiedustelulajien tuottamaa informaatiota. Sillä voidaan kuitenkin tuottaa vankka pohja tiedusteluoperaatioille ja täydentää jo olemassa olevia tietoja (Steele 2007). On hyvä muistaa, että avointen lähteiden tiedustelulla on paljon päällekäisyyksiä muiden tiedustelulajien kanssa. Esimerkiksi yleisesti avointen lähteiden tiedustelussa käytettävä kuvien paikannustekniikka voidaan tulkita geotiedusteluksi (engl. GEOINT) (Williams ym. 2018; Bellingcat 2024).

2.3 Avointen lähteiden tiedustelija

Avointen lähteiden tiedustelijat voidaan karkeasti jakaa kahteen luokkaan; julkisiin sekä yksityisiin toimijoihin. Julkisella puolella avointen lähteiden tiedustelua tekevät esimerkiksi valtiolliset toimijat, kuten tiedustelupalvelut, puolustusorganisaatiot, tutkimuslaitokset ja poliisivoimat. Yksityiset toimijoista esimerkkinä mainittakoon journalistit, rikolliset ja yksityishenkilöt (Krilov ym. 2022).

2.3.1 Julkiset toimijat

Julkiset toimijat käyttävät avointen lähteiden tiedustelua esimerkiksi rikollisuuden torjumiseen, ulkoisten ja kotimaisten mielipiteiden tarkkailuun, vastaterrorismiin ja muihin operaa-

tioihin, jotka vaativat tiedustelutietoa. Avointen lähteiden ja suljettujen lähteiden käyttäminen mahdollistaa yksityiskohtaisen tiedonkeräämisen julkisille toimijoille (Krilov ym. 2022).

Avointen lähteiden tiedustelua pidetään tärkenä osana NATOn (North Atlantic Treaty Organization) tiedustelua (Steele 2007). Suurten organisaatioiden laajat resurssit mahdollistavat asiantuntevan henkilöstön palkkaamisen ja sitä kautta tehokkaamman tiedustelun. Toisaalta NATOn asettamat tiedustelutarpeet poikkeavat mahdollisesti yksityishenkilön tekemän tiedustelun tiedustelutarpeista. Esimerkiksi yksityishenkilön ei mahdollisesti tarvitse miettiä yhtä monimutkaisia valitiollisia suhteita ja näkökulmia.

2.3.2 Yksityiset toimijat

Sanomalehdet, verkkojulkaiset ja tutkivaa journalismia tekevät toimittajat ovat jo pitkään käyttäneet avointen lähteiden tiedustelua etsiäkseen tietoa. Eräs tällainen verkkojulkaisualusta on Bellingcat, jonka suosio on kasvanut erityisesti paljastettuaan venäläisten separatistien osallisuuden lennon MH-17 pudotukseen (Higgins 2021). Venäläisten separatistien osallisuuden todistamisessa käytettiin erityisesti kuvan paikannusmenetelmää, jossa vertaillaan kuvasta erottuvia maaston muotoja, rakennuksia ja muita erityispiirteitä julkisesti saatavilla olevaan karttamateriaaliin. Julkisia karttatietoja internetissä tarjoaa esimerkiksi Google earth ja Google maps (Bellingcat 2024). Kotimainen esimerkki on YLEn tuottama MOT-ohjelma, joka keskittyy tutkivaan juornalismiin. Ohjelmien tutkimuksien tukena käytetään ainakin osittain avointen lähteiden tiedustelua (“YLE - MOT”, n.d.).

Yritykset käyttävät avointen lähteiden tiedustelua rekrytointiin, kilpailijoiden tarkkailuun, markkinointiin ja uusien markkinoiden tutkimiseen. (Garrido-Pintado ym. 2023; Hwang ym. 2022). Yritykset voivat tehdä avoimista lähteistä kerätylle datalle analyysia, jonka avulla ne voivat saada parempaa tietoa markkinoista (Chen ym. 2012). Internetin ansiosta pienemmätkin yritykset voivat käyttää avointen lähteiden tiedustelua apunaan (Hwang ym. 2022).

Yksi huolestuttavista ilmiöstä on rikollisten tekemä avointen lähteiden tiedustelu. Rikolliset voivat kerätä pohjatietoja avointen lähteiden tiedustelun avulla ja voivat käyttää saatua dataa rikosten tekemiseen. Mahdollisia rikoksia, joissa avointen lähteiden tiedustelusta on hyötyä ovat esimerkiksi palvelunestohyökkäykset, virusten levittäminen, identiteettivarkaudet ja

valheellisen informaation levittäminen (Hwang ym. 2022).

Yksityishenkilötkin tekevät avointen lähteiden tiedustelua internetissä. Madden ym. (2007) esittävät, että yksityishenkilöt etsivät enimmäkseen internetistä toisten henkilöiden yhteystietoja. Toisekseen he halusivat selvittää tietoja henkilöiden saavutuksista ja kiinnostuksenkohteista. Luvussa 1 mainittiin, että 56% oli hakenut itsestään tietoa internetistä vuonna 2013. Yksityishenkilöiden tiedonkeruu voi siis kohdistua myös heihin itseensä.

2.4 Tiedonkeräys sosiaalisesta mediasta

Sosiaalinen media tarjoaa paljon tietoa ja tietoon on suhtellisen helppoa päästä käsiksi (Akhgar ym. 2017). Sosiaalisesta mediasta voidaan saada arvokasta tietoa. Esimerkiksi Bellingcat-verkkosivusto on tullut tunnetuksi avointen lähteiden tiedustelusta, joissa käytetään sosiaalisesta mediasta saatavilla olevaa dataa. Bellingcat on julkaissut blogeja, joissa tutkitaan venäläisten separatistien yhteyttä Venäjän hallintoon. Bellingcat on myös ollut tutkimassa kaasuaseiden käyttöä Syyrian sisällissodassa (Higgins 2021). Sosiaalisessa mediassa avointen lähteiden tiedustelulla voidaan tarkkailla käyttäjien välistä toimintaa, kuten käyttäjien välistä viestittelyä kommenttikentissä. Kaikki julkisesti henkilöstä saatava tieto kumuloituu ja sitä voidaan pahimmassa tapauksessa käyttää henkilöään itseään vastaan (Pastor-Galindo ym. 2020).

Tiedustelijan tulee keskittyä informaatiolähteiden keräämiseen, eikä niinkään informaatioon itsessään, sillä informaation tarve muuttuu aina kun siirrytään vastaamaan toiseen tiedustelukysymykseen (Steele 2007). Tiedonkeräyksen avuksi on kehitetty monenlaisia työkaluja ja informaatiolähteitä on listattu erilaisille verkkosivuille kaikkien saataville (Bellingcat 2024).

Tiedonkeräys voi olla hidas prosessi ihmisille suoritettavaksi, siksi tiedonkeräykseen on kehitty monenlaisia työkaluja, jotka helpottavat hakuprosessia. Yksi tällainen työkalu on Web crawler -botti, joka aloittaa tietystä nettisivusta ja tallentaa nettisivun, jos tietyt ehdot täyttyvät. Web crawler etenee seuraavaan nettisivuun, kunnes sen toiminta pysähtyy. Tällaiset työkalut ovat hyödyllisiä ja yleensä nopeuttavat tiedonhakuprosessia, mutta eivät korvaa ihmistä prosessissa (Akhgar ym. 2017).

Ohjelmointirajapinnat (engl. API) ovat yleisesti käytettyjä hakumenetelmiä. Twitter, Facebook ja monet muut sosiaalisen median alustat tarjoavat käyttäjilleen tällaisia ohjelmointirajapintoja. Ohjelmointirajapinnat mahdollistavat tarkennettujen hakujen tekemisen ja suuren datamäärän. Sosiaalisen median alustat yleensä kuitenkin asettavat rajoitteita datan keräämiselle, kuten enimmäismäärän hauille (Akhgar ym. 2017). Google tarjoaa käyttäjilleen myös *Advanced Search*-palvelua, jonka avulla käyttäjä voi hakea tarkennetuilla hakuehdoilla tietoa (Google, n.d.).

Valheellisen tiedon erottaminen oikeellisesta tiedosta on sosiaalisessa mediassa erittäin vaikeaa ja hankaloittaa tiedonkeräysprosessia. Ihmiset voivat jakaa sosiaalisessa mediassa valheellisia tietoja itsestään, joka hankaloittaa prosessia entisestään. Tiedustelijan tehtävä on tulkita onko lähde luotettava vai ei (Williams ym. 2018).

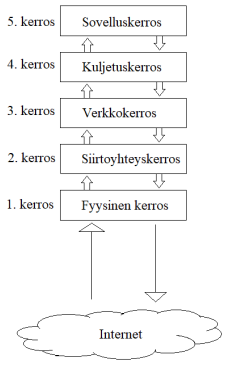
3 Digitaalinen jalanjälki

Internetin käytöstä jää aina jälki internetiin. Tätä jälkeä kutsutaan digitaaliseksi jalanjäljeksi. Digitaalinen jalanjälki voidaan jakaa kahteen kategoriaan sen muodostumisen perusteella; aktiiviseen sekä passiiviseen digitaalisen jalanjälkeen. Aktiivinen digitaalinen jalanjälki tarkoittaa tietoisesti internetiin jätettyä dataa. Passiivinen digitaalinen jalanjälki on digitaalisen jalanjäljen osa, jota ei tietoisesti jaeta internetiin (Madden ym. 2007). Yksi esimerkki tällaisesta on IP-osoite.

3.1 Passiivinen digitaalinen jalanjälki

Tässä alaluvussa käsitellään passiivista digitaalista jalanjälkeä. Ensimmäisessä alaluvun alaluvussa käsitellään tietoliikenteen osalta syntyvää passiivista digitaalista jalanjälkeä. Toisessa alaluvun alaluvussa käsitellään selaimen osalta syntyvää passivista jalanjälkeä. Passiivisen digitaalisen jalanjäljen määritelmä lyhyesti on: Se osa digitaalisesta jalanjäljestä, jota käyttäjä ei jaa tietoisesti. Siksi passiivinen digitaalinen jalanjälki saattaa jäädä käyttäjältä vähemmälle tarkastelulle, kuin aktiivinen digitaalinen jalanjälki.

3.1.1 TCP/IP-protokollapino



Kuvio 1. TCP/IP-protokollapino

TCP/IP-protokollapino koostuu viidestä "kerroksesta"; sovelluskerroksesta, kuljetuskerroksesta, verkkokerroksesta, siirtoyhteyskerroksesta ja fyysisestä kerroksesta. TCP/IP-protokollapinon

nimi tulee siitä, että TCP- ja IP-protokolla ovat yleisimmät kuljetus- ja verkkokerroksen protokollat. Sitä kutsutaan pinoksi, koska se koostuu useasta eri protokollasta, joista tieto valuu pinon päältä alas tai nousee pinon alta ylös ja jokainen kerros toimii erillään muista kerroksista (Goralski 2017).

Bernardos ym. (2015) mainitsevat, että kaikkia kerroksia voidaan käyttää käyttäjän yksilöimiseen, mutta käyttäjän sijainti voidaan saada selville etenkin kerroksen 2 ja kerroksen 3 osoitteista. Kerroksella 2 tarkoitetaan siirtoyhteyskerrosta ja kerroksella 3 tarkoitetaan verkkokerrosta. Siirtoyhteyskerroksella on MAC-osoite ja verkkokerroksella IP-osoite.

IP-osoitetta, MAC-osoitetta ja sovelluskerroksen tunnisteita, kuten evästeitä ja mediasoittiminen sarjanumeroa voidaan käyttää käytetyn laitteen tunnistamiseen. Myös käyttäjän tunnistaminen on mahdollista (Aura ym. 2006). Verkkosivulla vierailu jättää aina jäljen palveluntarjoajan palvelimelle. Palvelin voi tallentaa esimerkiksi käyttäjän IP-osoitteen, käyttäjätunnuksen, sähköpostiosoitteen ja muita tunnistavia tietoja (Akhgar ym. 2017). Suomessa pidättämiseen oikeutetun virkamiehen on oikeus saada tällaisia tunnistetietoja palvelimen ylläpitäjältä, jos viestin sisällön toimittaminen yleisölle on määrätty rangaistavaksi (*Laki sananvapauden käyttämisestä joukkoviestinnässä,* 2003/460 § 17). Verkkotunnisteita voidaan käyttää myös todisteena oikeudenkäynneissä. Esimerkiksi Suomessa IP-osoitteet ovat olleet todisteina Vastaamon tietomurron yhteydessä käydyssä oikeudenkäynnissä (Salumäki ym. 2023).

3.1.2 Evästeet ja muut tunnistetiedot

Evästeet mahdollistavat sovelluskerroksen HTTP-protokollassa tilan ylläpidon. HTTP-protokolla ei säilytä itsessään käyttäjän tilaa. Esimerkiksi evästeet helpottavat nettisivujen tekemistä, jotka vaativat kirjautumista, sillä käyttäjän ei tarvitse kirjautua jatkuvasti sisään mennessään nettisivulla uuteen resurssiin (Barth 2011).

Evästeitä on kritisoitu, koska niillä voidaan seurata milloin käyttäjä palaa nettisivulle tai käyttää toista nettisivua. Evästeet helpottavat käyttäjien seurantaa, koska niitä voidaan jakaa palvelimien välillä ja ne ovat pysyviä koko istunnon ajan (Barth 2011). Barth (2011) mainitsee, että erityisen huolestuttavia ovat kolmannen osapuolen evästeet. Erityisesti sen takia,

että käyttäjän ei tarvitse vieraille nettisivuilla, jotka keräävät kolmannen osapuolen evästeitä. Kolmannen osapuolen evästeitä käyttävät esimerkiksi markkinointiverkostot.

Nettisivut keräävät myös dataa käyttäjien käytöksestä. Nettikaupat ja markkinoijat ovat hyödyntäneet jo pitkään tällaista dataa. Nettisivujen käytöksen avulla saadaan tietoa esimerkiksi käyttäjän oletusta iästä, sukupuolesta ja kiinnostuksenkohteista. Yleisimmin tällaista dataa käytetään markkinoinnin ja tarjonnan kohdentamiseen (Abramson ym. 2013). (Abramson ym. 2013) osoittavat, että tällaiset tiedot eivät ole yksinään tarpeeksi vahvoja erottamaan käyttäjiä toisistaan. Kuitenkin tutkimuksen mukaan jonkinlaista osviitta nettisivun käyttäjästä saadaan.

3.2 Aktiivinen digitaalinen jalanjälki

Aktiivinen digitaalinen jalanjälki on tietoisesti internetiin jaettua tietoa. Yleensä aktiivinen digitaalinen jalanjälki jätetään tietyssä kontekstissa, kuten keskustelufoorumilla käyty väittely nimimerkin takaa tai Instagramissa jaettu kuva seuraajille. Mitä enemmän julkiseen internetiin käyttäjä antaa tietojaan itsestään tietoisesti, sitä enemmän käyttäjän aktiivinen digitaalinen jalanjälki kasvaa. Internetistä on vaikeampi hakea lisää tietoa henkilöstä, jos haettavalla henkilöllä on yleinen etu- tai sukunimi tai muita yhteisiä piirteitä monen muun henkilön kanssa (Madden ym. 2007).

Sosiaalisen median alustalla julkaistu valokuva on aktiivista digitaalista jalanjälkeä. Pelkästä valokuvasta voidaan saada paljon tietoa aktiivisen sekä passivisen digitaalisen jalanjäljen muodossa. Bellingcat (2024) esittelee työkaluja, joilla kuvista voidaan löytää sijainti käyttäen geopaikannusta. Kuvista voidaan myös erottaa sormenjälkiä. Vastaamo tietomurron yhteydessä Ylilaudalla julkaistusta valokuvasta pystyttiin Poliisin toimesta löytämään epäillyn sormenjälki (Salumäki ym. 2023). Kuva sisältää myös metadataa ellei sitä ole erikseen poistettu. Kuvan metadataan voidaan tallentaa sijainti, kameran malli, merkki ja muita tunnistavia tietoja (JEITA ym. 2023). Netissä on nettisivuja ja työkaluja, joilla voidaan kuvasta hakea tietoja kuvan metadatasta (Bellingcat 2024).

4 Digitaalisen jalanjäljen hallinta

Tässä osassa perehdytään miksi digitaalista jalanjälkeä olisi hyvä hallita ja esitetään tapoja hallita digitaalista jalanjälkeä.

4.1 Miksi omaa digitaalista jalanjälkeä olisi hyvä hallita?

Digitaalisen jalanjäljen hallintaan on monenlaisia syitä. Useimmiten syyt liittyvät kuitenkin oman maineen hallintaan. Jotkin työpaikat myös vaativat, että oma digitaalinen jalanjälki on pidettävä työpaikan ohjeiden mukaisena. Toisaalta joissain työnkuvissa henkilö joutuu markkinoimaan itseään sosiaalisessa mediassa. Tällaiset henkilöt kuitenkin seuraavat omaa digitaalista jalanjälkeä aktiivisemmin (Madden ym. 2007). Sosiaalisessa mediassa erityisesti nuoret ovat kiinnostuneet oman verkkomaineensa hallinasta (Madden ym. 2010).

Oman digitaalisen jalanjäljen hallinnalla voidaan lievittää yksilöön kohdistuvia uhkia. Mahdollsia henkilöön kohdistuvia uhkia ovat identiteettivarkaudet, häirintä ja yksityisyyden loukkaaminen (Marwick ym. 2010).

Sosiaalisessa mediassa oleva digitaalinen jalanjälki voi olla kynnysehto palkkaamiselle. Jotkin työnantajat seuraavat myös työntekijöidensä sosiaalisen median digitaalista jalanjälkeä luokkaavan sisällön varalta (Hidy ym. 2013). Edellä mainitussa tutkimuksessa esitetään, että työnantajista oli jättänyt palkkaamatta 8% työntekijöistä johtuen huonosta käytöksestä sosiaalisen median alustoilla.

Nettirikollisille on hyödyllistä, että uhrista löytyy mahdollisimman paljon tietoa. Esimerkiksi sosiaaliseen manipulaation (engl. Social Engineering) perustuvat hyökkäykset hyötyvät uhrin digitaalisesta jalanjäljestä. Rikollisille on paljon helpompaa kehittää tarina tai keksiä valheita, jos he tietävät minkälainen tarina tulee kertoa uhrille (Heikkinen 2006).

4.2 Hallinta

Tietoisilla valinnoilla ja työkaluilla voidaan digitaalista jalanjäljen muodostumista rajoittaa. Aina digitaalisen jalanjäljen rajoittaminen ei ole käytännöllistä, sillä työkalujen käyttö voi hidastaa internetin käyttämistä suhteettomasti. Käyttäjän tulee harkita tapauskohtaisesti omien intressien ja päämäärien perusteella sopivat työkalut ja valinnat oman digitaalisen jalanjäljen suhteen.

Madden ym. (2007) mainitsevat, että yleinen käytäntö sosiaalisessa mediassa on rajoittaa oman sosiaalisen median sisältö vain kavereiden saataville. Ongelma tästä muodostuu, jos sosiaalisen median kaveripiiriin pääsee ulkopuolinen taho. Sisällön rajoittaminen voi antaa myös käyttäjälle vääränlaisen turvallisuuden tunteen, jolloin sosiaalisessa mediassa sisältöä jaetaan vapaammin. Hengstler (2011) esittää aktiivisen digitaalisen jalanjäljen osalta nyrkkisäännön: sosiaalisessa mediassa sisältöä jakaessa tulee olettaa, että kaikki näkevät sen jonain päivänä.

Sosiaalisen median käyttäjän on oltava erityisen tarkkana julkaistaessaan sosiaaliseen mediaan valokuvia, jos käyttäjä haluaa piilottaa oman sijaintinsa. Alaluvussa 3.1 esiteltiin, että valokuvista on mahdollista erottaa sijainti ympäristön perusteella. Käyttäjä voi yrittää tehdä sijainnin tunnistamisesta mahdollisimman hankalaa poistamalla ympäristöstä tunnistettavat erityispiirteet.

Joskus jo jaettua tietoa voidaan poistaa tai ainakin rajoittaa. Google (n.d.) tarjoaa palvelua, jossa hakuehdotuksista voidaan poistaa käyttäjän haluamaa tietoa. Aina tämä ei kuitenkaan toimi, sillä Google ei poista tietoja hakukoneen tuloksista, jos tiedot ovat arvokkaita julkiselle yleisölle. Toinen asia joka tulee ottaa huomioon on, että tieto on vielä olemassa internetissä ja mahdollisesti nähtävillä muilla hakukoneilla.

Kuvissa on metadataa, joista voidaan mahdollisesti erottaa sijainti (JEITA ym. 2023). On siis hyvä käytäntö poistaa metadata valokuvista ennen niiden julkaisemista. Bellingcat (2024) esittelee mahdollisia työkaluja metadatan poistamiseen.

Kappaleessa 3.1 kerrotaan, että käyttäjän sijainti voidaan saada selville etenkin TCP/IP-protokollapinon kerroksista 2 ja 3. Näiden osoitteiden piilottamiseen on kehitetty työkaluja.

Microsoft (2024) esittelee kolme mahdollista tapaa piilottaa kerroksen 3 osoite eli IP-osoite; VPN (Virtual Private Network), TOR-selain (lyhenne The Onion Router) ja välityspalvelimet (engl. Proxy). VPN-palvelut ovat yleensä maksullisia ja niiden turvallisuus on riippuvainen myös VPN-tarjoajien luotettavuudesta. TOR-selainta käytettäessä siirtoyhteysnopeus on yleensä heikko ja siten epäkäytännöllinen päivittäisessä käytössä. MAC-osoite eli kerroksen 2 osoitteen piilottamiseksi on kehitetty myös työkaluja. Bernardos ym. (2015) mainitsevat, että MAC-osoite voidaan generoida ilman globaalia mekanisimia. Käyttäjä voi siis luoda itselleen satunnaisen MAC-osoitteen omalta koneeltaan.

Alaluvussa 3.1 esitettiin, että evästeet ovat osa passiivista digitaalista jalanjälkeä. Samassa kappaleessa mainittiin, että erityisen huolestuttavia ovat kolmannen osapuolen evästeet. Traficom (n.d.) esittelee tapoja, joilla evästeiden käyttöä voidaan rajoittaa. Edellä mainitulla verkkosivulla mainitaan, että kolmannannen osapuolen evästeet voidaan poistaa oletuksena selaimen asetuksista. Hyvä käytäntö on myös poistaa väliajoin selaushistoria ja evästeet selaimen asetuksista. Suosituimmat selaimet, kuten Chrome ja Firefox mahdollistavat incognito-tilan. Incognito-tilassa selain ei tallenna käyttäjän evästeitä eikä selaushistoriaa.

5 Yhteenveto

Tutkielmassa perehdyttiin digitaaliseen jalanjälkeen ja sen hallintaan. Digitaalisen jalanjäljen hallinta on mahdollista tiettyyn pisteeseen saakka. Etenkin nykyinternetissä, jossa tehokkaat työkalut avointen lähteiden tutkimiseen ovat kaikkien saatavilla. Alaluvussa 3.1 mainittiin, että verkkokäyttymisen perusteella on mahdollista saada jonkinlaista osviittaa nettisivun käyttäjästä. Tällaiseen tunnistamiseen ei ole oikeastaan muita suojatumiskeinoja, kuin internetin ja sosiaalisten median alustojen käytön lopettaminen.

Aktiivisen digitaalisen jalanjäljen hallintaan voidaan sosiaalisessa mediassa pääasiassa vaikuttaa tietoisilla valinnoilla. Käyttäjän tulee olla tietoinen passiivisen digitaalisen jalanjäljen muodoustumisesta, jotta siihen voidaan erikseen puuttua. Passiivisen digitaalisen muodostumista voidaan hallita erilaisilla työkaluilla, mutta työkalujen käyttö vaatii käyttäjältä jonkinlaista tietoteknistä osaamista. Jolloin passiivisen digitaalisen jalanjäljen hallinta saattaa jäädä pienemmälle huomiolle.

Avointen lähteiden tiedustelusta on tehty viimeaikoina paljon tutkimusta. Digitaaliseen jalanjälkeen liittyvää tutkimusta on myös tehty paljon, etenkin suurten sosiaalisten verkostojen synnyttyä. Vaikka tutkimusta digitaalisesta jalanjäljestä oli tehty, ei näistä löytynyt kattavia ohjeita tai parhaita käytänteitä oman digitaalisen jalanjäljen hallintaan.

Lähteet

Abramson, Myriam ja David Aha. 2013. “User authentication from web browsing behavior”. *FLAIRS 2013 - Proceedings of the 26th International Florida Artificial Intelligence Research Society Conference,* 268–273.

Akhgar, Babak, P Saskia Bayerl ja Fraser Sampson. 2017. *Open source intelligence investigation: from strategy to implementation.* Springer.

Aura, Tuomas ja Alf Zugenmaier. 2006. “Privacy, control and internet mobility”. Teoksessa *Security Protocols: 12th International Workshop, Cambridge, UK, April 26-28, 2004. Revised Selected Papers 12,* 133–145. Springer.

Barth, Adam. 2011. *HTTP State Management Mechanism.* Tekninen raportti, Request for Comments 6265. https://doi.org/10.17487/RFC6265.

Bellingcat. 2024. “Bellingcat toolkit”. Viitattu 29. huhtikuuta 2024. https://docs.google.com/spreadsheets/d/18rtqh8EG2q1xBo2cLNyhIDuK9jrPGwYr9DI2UncoqJQ/edit#gid=930747607.

Bernardos, Carlos J., Juan Carlos Zúñiga ja Piers O’Hanlon. 2015. “Wi-Fi internet connectivity and privacy: Hiding your tracks on the wireless Internet”. Teoksessa *2015 IEEE Conference on Standards for Communications and Networking (CSCN),* 193–198. https://doi.org/10.1109/CSCN.2015.7390443.

Chen, Hsinchun, Roger HL Chiang ja Veda C Storey. 2012. “Business intelligence and analytics: From big data to big impact”. *MIS quarterly,* 1165–1188.

Garrido-Pintado, Pablo, Juan Gabriel García Huertas ja Diego Botas Leal. 2023. “Identity and virtuality: The influence of personal profiles on social media on job search”. *Business Information Review.*

Google. n.d. “Find and remove personal contact info in Google Search results”. Viitattu 29. huhtikuuta 2024. https://support.google.com/websearch/answer/12719076?hl=en.

Google. n.d. “Google Advanced Search”. Viitattu 29. huhtikuuta 2024. https://www.google.com/advanced_search.

Goralski, Walter. 2017. *The illustrated network: how TCP/IP works in a modern network.* Morgan Kaufmann.

Heikkinen, Seppo. 2006. “Social engineering in the world of emerging communication technologies”. Teoksessa *Proceedings of Wireless World Research Forum,* 10. Citeseer. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=be5a68ba31989b6d224dd5666a6b2392b067b886.

Hengstler, Julia. 2011. “Managing your digital footprint: Ostriches v. Eagles”. *Education for a digital world* 2:89–139.

Hidy, Kathleen McGarvey ja Mary McDonald. 2013. “Risky Business: The Legal Implications of Social Media’s Increasing Role in Employment Decisons”. *Available at SSRN 3188753,* https://doi.org/10.2139/ssrn.3880912.

Higgins, Eliot. 2021. *We are Bellingcat: An intelligence agency for the people.* Bloomsbury Publishing.

Hribar, Gašper, Iztok Podbregar ja Teodora Ivanuša. 2014. “OSINT: A “Grey Zone”?” *International Journal of Intelligence and CounterIntelligence* 27 (3): 529–549. https://doi.org/10.1080/08850607.2014.900295.

Hwang, Yong-Woon, Im-Yeong Lee, Hwankuk Kim, Hyejung Lee, Donghyun Kim ym. 2022. “Current status and security trend of osint”. *Wireless Communications and Mobile Computing,* https://doi.org/10.1155/2022/1290129.

Exchangeable image file format for digital still cameras: Exif Version 3.0. 2023. Camera Imaging Products Association. Viitattu 29. huhtikuuta 2024. https://www.cipa.jp/std/documents/download_e.html?DC-008-Translation-2023-E.

Krilov, Brittany, Ryan Gough, Patricia Kickland, Katherine Wolfe, Danielle Waters, William Nasuti, Chianna Mirtaheri, Rosangie Paiz ja Shani Spivak. 2022. Viitattu 4. toukokuuta 2024. https://www.dhs.gov/sites/default/files/2022-09/Ethical%20Frameworks%20in%20OSINT%20Final.pdf.

Laki sananvapauden käyttämisestä joukkoviestinnässä. 2003/460 § 17. https://www.finlex.fi/fi/laki/ajantasa/2003/20030460.

LibGuides. n.d. “Harmaa kirjallisuus”. Viitattu 29. huhtikuuta 2024. https://libguides.oulu.fi/c.php?g=689390&p=4934741.

Madden, Mary. 2013. “Majority of online Americans ‘Google themselves’”. *Pew research center,* viitattu 29. huhtikuuta 2024. https://www.pewresearch.org/short-reads/2013/09/27/majority-of-online-americans-google-themselves/.

Madden, Mary, Susannah Fox, Aaron Smith ja Jessica Vitak. 2007. “Digital Footprints”. *Pew research center,* https://www.pewresearch.org/internet/2007/12/16/digital-footprints/.

Madden, Mary ja Aaron Smith. 2010. “Reputation management and social media”. *Pew Internet & American Life Project* 26.

Marwick, Alice E, Diego Murgia-Diaz ja John G Palfrey. 2010. “Youth, privacy and reputation”.

Mercado, Stephen C. 2009. *Sailing the Sea of OSINT in the Information Age.* Nide 78. Routledge Abingdon, UK. https://doi.org/10.1037/e741272011-005.

Microsoft. 2024. “How to hide your IP address”. Viitattu 30. huhtikuuta 2024. https://www.microsoft.com/en-us/microsoft-365-life-hacks/privacy-and-safety/why-and-how-to-hide-ip-address.

NATO. 2001. “NATO Open Source Intelligence Handbook”. Viitattu 29.4.2024, https://archive.org/details/NATOOSINTHandbookV1.2/page/n21/mode/2up.

Pastor-Galindo, Javier, Pantaleone Nespoli, Félix Gómez Mármol ja Gregorio Martínez Pérez. 2020. “The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends”. *IEEE Access* 8:10282–10304. https://doi.org/10.1109/ACCESS.2020.2965257.

Perrin, Andrew. 2015. “Social media usage”. *Pew research center* 125:52–68.

Rainie, Lee, Sara Kiesler, Ruogu Kang, Mary Madden, Maeve Duggan, Stephanie Brown ja Laura Dabbish. 2013. “Anonymity, privacy, and security online”. *Pew research center* 5. Viitattu 5. toukokuuta 2024. https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_AnonymityOnline_090513.pdf.

Salumäki, Tiina ja Teemu Hallamaa. 2023. “Vastaamo-kuulustelut paljastavat: Epäillyn jäljille päästiin Onlyfans-maksujen kautta”. Viitattu 29. huhtikuuta 2024. https://yle.fi/a/74-20058473.

Saridakis, George, Vladlena Benson, Jean-Noel Ezingeard ja Hemamali Tennakoon. 2016. “Individual information security, user behaviour and cyber victimisation: An empirical study of social networking users”. *Technological Forecasting and Social Change* 102:320–330. ISSN: 0040-1625. https://doi.org/https://doi.org/10.1016/j.techfore.2015.08.012.

Steele, Robert David. 2007. “Open source intelligence”. *Handbook of intelligence studies* 42 (5): 129–147.

Şuşnea, Elena ja Adrian Iftene. 2018. “The significance of online monitoring activities for the social media intelligence (SOCMINT)”. Teoksessa *Conference on mathematical foundations of informatics,* 230–240.

Traficom. n.d. Viitattu 3. toukokuuta 2024. https://www.kyberturvallisuuskeskus.fi/fi/toimintamme/saantely-ja-valvonta/evasteet?toggle=Ev%C3%A4steiden%20k%C3%A4ytt%C3%B6%C3%B6n%20vaikuttavat%20muutokset%20selaimen%20asetuksilla.

Twitter. 2024. “Counting characters”. Viitattu 29. huhtikuuta 2024. https://developer.twitter.com/en/docs/counting-characters.

Williams, Heather J. ja Ilana Blum. 2018. *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise.* Santa Monica, CA: RAND Corporation. https://doi.org/10.7249/RR1964.

“YLE - MOT”. n.d. Viitattu 29. huhtikuuta 2024. https://yle.fi/aihe/mot.