Toni Pietiläinen

Käyttäjän digitaalisen jalanjäljen hallinta

Tietotekniikan Kandidaatintutkielma

15. joulukuuta 2022

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Toni Pietiläinen

Yhteystiedot: pieton@student.jyu.fi

Ohjaaja: Timo Tiihonen

Työn nimi: Käyttäjän digitaalisen jalanjäljen hallinta

Title in English: Controlling the digital footprint of a user

Työ: Kandidaatintutkielma

Opintosuunta: Tietotekniikka

Sivumäärä: 20+0

Tiivistelmä: Tässä tutkielmassa selvitetään, miten käyttäjien digitaalinen jalanjälki koostuu esimerkiksi internetin selaamisesta, profiilien luomisesta, julkisista lausunnoista ja laitetiedoista. Myös erilaiset tietovuodot ja rajoitetun tiedon leviäminen voivat vaikuttaa jälkeen. Tarkastelussa ovat myös erilaiset keinot vaikuttaa jalanjäljen muodostumiseen, kuten evästeja yksityisyysasetukset, selaamisen suojaamiseen tarkoitetut sovellukset ja julkisten julkaisujen rajoittaminen mediassa. Tutkimus toteutetaan kirjallisuuskatsauksena, jossa käydään läpi käyttäjän näkökulmaa tiedonkeruussa sekä sitä, mitä käyttäjän tietämättä taustalla tapahtuu, ja mitä käyttäjä voi tehdä tiedonkeruuasetusten muuttamiseksi.

Avainsanat: digitaalinen jalanjälki, tiedonkeruu, tiedon käsittely

Abstract: The purpose of this Bachelor's Thesis is to clarify how the digital footprint of users compound for example by browsing the internet, creating profiles, publishing posts or from device diagnostics or by data breaches and the spreading of restricted data. Various ways of influencing the formation of footprints, such as cookie and privacy settings, applications intended to protect browsing and limiting public posts in the media, are also being considered and clarified. This thesis is conducted as a literature review, in which the focus is in the standpoint of the user in the data collection process and also the events that occur without the users awareness and how the user can have an affect on the data collection process.

Keywords: data collection process, data handling process, digital footprint

Termiluettelo

GDPR	Euroopan Unionin tietosuoja-asetus. Kaikissa EU-maissa sovellettuna oleva laki, joka määrittelee ja sääntelee henkilötietojen käsittelyä. Sen tarkoitus on suojata käyttäjien henkilötietoja sekä taata käyttäjille enemmän keinoja hallita heitä koskevien tietojen käsittelyä.
HTTP-Pyyntö	Asiakasohjelman kommunikoi palvelimen kanssa lähettäen sille pyyntömuotoisesti erilaisia metodikutsuja, joilla halutaan suorittaa jokin tietty toiminto resurssille.
IoT	Esineiden Internet. Yhä useampi laite ja esine on liitetty internettiin, jolloin laitteet pystyvät verkon avulla kommunikoimaan keskenään, ja kerättyjen sekä käsiteltyjen tietojen avulla laitteet voivat toimia itsenäisesti tai osana erilaisia järjestelmiä.
IP-osoite	IP-protokolla määrittelee yksilöivän osoitteen laitteille (tai verkoille) internetissä, jonka avulla voidaan tunnistaa dataa lähettävät ja vastaanottavat laitteet.
OSI-Malli	Tiedonsiirtoprotokollien yhdistelmää kuvaava seitsemänkerroksinen mallinnus. Lähimpänä käyttäjää on sovelluskerros. Sen alapuolella on esitystapakerros, istuntokerros, kuljetuskerros, verkkokerros, siirtokerros ja alimpana fyysinen kerros.
VPN	Virtuaalinen erillisverkko. Tekniikka, jolla käyttäjäasiakkaan verkkoliikenne ohjataan erillisen palvelimen kautta, jolloin asiakkaan kohdepalvelin näkee käyttäjän oikean IP-osoitteen sijasta välityspalvelimen osoitteen.

Kuviot

Sisällys

1 Johdanto

Ihmisistä kerättävän informaation määrä kasvaa kokoajan. Jokainen pieni vahingossa tai tarkoituksellisesti jaettu tietolohko on osa isompaa koontia, jolla jokaiselle käyttäjälle luodaan heitä profiloiva digitaalinen jalanjälki. Jos käyttäjä haluaa varmistaa itselleen turvallisemman ympäristön, on syytä kiinnittää jatkuvaa huomiota kaikkeen toimintaan verkossa sekä pysyä ajan tasalla ja kiinnostuneena siitä, millaiseksi oma henkilökohtainen jalanjälki muodostuu.

Tiedonkeräys asettaa ihmiset haavoittuvaan asemaan, sillä kerättävän tiedon joukossa saattaa olla myös jotain, jota yksityinen käyttäjä ei välttämättä haluaisi julkisesti jakaa. Tämä data on monella tavalla arvokasta ja tästä syystä datan suojaamiseksi, käytön rajoittamiseksi tai salassapitämiseksi on olemassa erilaisia ratkaisuja. Informaatiota käytetään yhä useammin käyttäjien intressien vastaisesti. Valtaosa käyttäjistä ilmaisee heille olevan tärkeää, että he pystyvät itse vaikuttamaan siihen, kuka tai mikä heitä koskevia tietoja saa käsitellä ja käyttää. Mutta vain pieni osa käyttäjistä on oikeasti kartalla siitä, miten heidän tietojaan tosiasiassa käytetään ja jaetaan eteenpäin.

Tämän tutkielman tavoitteena on löytää ja selvittää erilaisia hallintatyökaluja ja käytänteitä, joilla yksityinen käyttäjä pystyy rajoittamaan, hallitsemaan tai tutkimaan hänen omaa digitaalista jalanjälkeänsä. Usein nämä tiedot ovat käyttäjän itsensä verkkoon jakamia joko tietoisesti tai vahingossa, joten on syytä tarkastella myös eri käyttäjien toimintaperiaatteita ja käytänteitä silloin, kun he jotain tietoa itsestään tietoisesti jakavat.

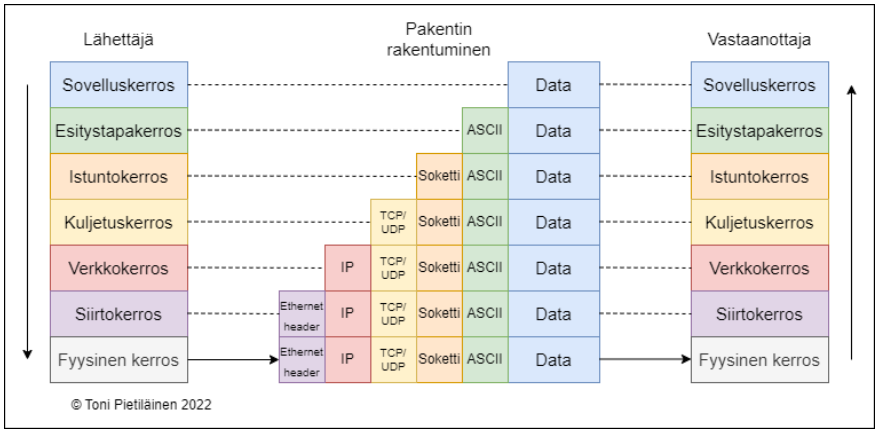
Tutkielman alussa käsitellään sitä, miten käyttäjästä kerätty tieto kasautuu käyttäjän digitaaliseksi jalanjäljeksi kerroksittain ja miten tätä jalanjälkeä käsitellään erilaisten tietoa keräävien tai muokkaavien tahojen toimesta. Tämän jälkeen selvitetään, miten käyttäjän näkökulmat yleisesti eroavat todellisuudesta, millaisia tapoja ja käytänteitä ihmisillä on tiedonkeruuseen liittyen sekä pohditaan mahdollisia syitä siihen, miksi osa käyttäjistä on epätietoisia tietojensa todellisesta keräämisestä ja käsittelystä. Samassa luvussa käsitellään lyhyesti myös tietovuotoja ja tietomurtoja, sekä niihin liittyviä riskejä käyttäjän näkökulmasta. Lopuksi on tarkoitus muodostaa käyttötottumusaineistoa hyödyntäen mahdollisia ratkaisukokonaisuuksia, joilla käyttäjät voivat hallita jalanjälkeään.

2 Käyttäjän digitaalinen jalanjälki

Nykyisin arkeemme kuuluu paljon kanssakäymistä ja vuorovaikutusta erilaisten verkkoihin kytkettyjen laitteiden kanssa, joskus jopa ilman että tiedostamme asiaa. Interaktioista suuri osa tallennetaan pitkäksi aikaa usein hajautettuihin sijainteihin, joista erilaisia algoritmejä ja kaavoja hyödyntäen voidaan muodostaa käyttäjästä profiili, digitaalinen jalanjälki, ja tämän tallentamisen välttäminen on usein hankalaa tai mahdotonta. (Haimson ym. 2016). Tässä luvussa havainnollistetaan aluksi jalanjäljen syntymistä viestinvälityksen eri tasoilla tasokohtaisten esimerkkien avulla. Tämän jälkeen syvennytään jalanjäljen tarkastelumahdollisuuksiin ja muokkaukseen käyttäjän näkökulmasta. Selvitetään esimerkiksi, kuinka usein tiedonkeruussa kysytään käyttäjän omaa suostumusta tietojenkeruuseen tai käyttäjää koskevien tietojen käsittelyyn sekä miten käyttäjä voisi teknisesti mahdollisesti tarkastella oman jalanjälkensä osia.

2.1 Digitaalisen jalanjäljen muodostuminen

Jalanjälkiä ja niiden muodostumista voidaan tarkastella usealta eri kerrokselta esimerkiksi OSI-mallia mukailevan viestipaketin muodostumista esittävän kuvion mukaisesti (kuvio 1). Jalanjälkeen kertyy tietoa käytännössä kaikilta mallin tasoilta kuten viestinvälityksessä yhteen pakettiin, ja tässä alaluvussa tarkastellaan erilaisia jalanjälkeen kontribuoivia osia eri kerroksilta käyttäen apuna muutamaa erilaista tietosisältöesimerkkiä.



Kuvio 1. Viestinvälitysprosessin paketin muodostuminen (mukaillen Networkel.com (2016))

Sovelluskerroksella eli käyttäjän päädyssä arkkitehtuuria sosiaalisen median palvelut ovat yksi esimerkki palvelutyypistä, jotka kykenevät luomaan hyvinkin yksityiskohtaisen, yksilöitävän ja monipuolisen jalanjäljen yksittäisestä käyttäjästä (Grover ja Mark 2017). Palveluiden käyttö perustuu erilaisten profiilien luontiin, joiden on tarkoitus tehdä jokaisesta käyttäjästä uniikki ja tunnistettava, ja profiilien perusteella käyttäjät löytävät toisensa palvelun käyttäjäkunnasta. Käyttäjäprofiilin luomiseen usein kuuluu esimerkiksi omien henkilötietojen lisääminen profiiliin, tai muiden yksilöivien tietojen lisäys, mutta myös esimerkiksi tilannepäivitykset, mielipiteiden ilmaisut ja julkiset mieltymystottumukset kertyvät usein samaisen profiilin jalanjälkeen (Deeva 2019).

Käyttäjäprofiileja luodaan myös esimerkiksi erilaisiin verkkokauppoihin tai keskustelupalstoille, ja jälkimmäisissä usein osa käyttäjän tiedoista on suoraan toisten käyttäjien nähtävissä ilman tietojen näkyvyysasetusten muuttamista. Henkilökohtaisiakin tietoja tuodaan julki siis tarkoituksellisesti tai vahingossa palveluiden tehokkaamman toiminnan takaamiseksi, mutta niiden mahdollinen pitkäaikainen olemassaolo ja merkittävyys muihin tietoihin yhdistettäessä jää usein pohtimatta silloin, kun ollaan täyttämässä tietoja omaan profiiliin tai luomassa julkaisua.

Tutkimuksessa (Azucar, Marengo ja Settanni 2018) havaittiin, että sosiaalisen median profiilien ominaisuuksista pystyttiin merkittävällä tarkkuudella päätellä esimerkiksi käyttäjien viiden suuren persoonallisuuspiirteen ominaisuuksia. Piirteitä analysoimalla voidaan ennakoida entistä tarkemmin esimerkiksi käyttäjän yleistä käyttäytymistä ja toimintaa verkossa (Wang 2013). Jo yksittäisen profiilin luominen saattaa aiheuttaa prosessin, jonka lopputuloksena käyttäjälle saattaa välittyä jo hyvinkin spesifisti kohdennettua toimintaa.

Toinen tunnettu nykypäivän esimerkki tietojen keräyksestä ovat evästeet, jotka sijoittuisivat mallin mukaisesti sovelluskerroksen ja istuntokerroksen väliin. Evästeet ovat käyttäjän laitteelle tallennettuja tietoja, jotka palvelin lähettää käyttäjän päätelaitteen selaimelle. Palvelin voi myöhemmässä vaiheessa pyytää tätä tietoa takaisin, ja selaimet yleisesti lähettävät tietoja takaisin vain samalle palvelimelle, jolta selain on ne alunperin tallentanut (Kristol 2001). Yksittäisen palvelun ja käyttäjän välillä liikkuvia evästeitä kutsutaan ensimmäisen osapuolen evästeiksi. Verkkosivusto saattaa käyttää myös ulkopuolisia palveluita, jotka saattavat lähettää omia evästeitään ja joita kutsutaan kolmannen osapuolen evästeiksi.

Evästeitä on olemassa kahta eri tyyppiä. Istuntokohtaiset evästeet poistuvat aina, kun istunto web-palvelussa lopetetaan. Toinen tyyppi ovat pysyvät evästeet, jotka ovat tallessa johonkin aikarajaan asti, tai kunnes käyttäjä ne itse poistaa. Evästeet kulkevat osana HTTP-pyyntöjä selaimen ja palvelimen välillä yhteyden muodostuksen jälkeen, jolloin palvelimelle päätyy tarkkaa dataa mm. käyttäjän selaushistoriasta tai personointivalinnoista (Kristol 2001). Myös esimerkiksi erilaiset vastauslomakkeet ja muut tietokentät saattavat tallentua selaimen evästeisiin, ja ne voivat sisältää myös esimerkiksi salasanoja tai muita julkisuudeltaan rajattuja tai suojausta vaativia tietueita.

Aktiivisen jäljen ohella palvelimille voi tallentua erilaisia kuljetus- ja verkkokerrokselle ominaisia ja kuuluvia teknisiä tietoja selauksesta itsestään, joita käyttäjä ei usein täysin edes ymmärrä tai tiedosta tuovansa julki. Esimerkiksi palvelimelle tallentuvalla käyttäjän IP-osoitteella voidaan saada selville käyttäjän sijaintitietoja. Myös jo tiedon liikkumisesta tiettyjen laitteiden kautta saattaa jäädä talteen siirtokerrokselle ominaisia laitetietoja, kuten verkkosovitinlaitteiden fyysisiä tunnisteita, joiden perusteella voidaan kyetään verkosta riippumatta tunnistamaan laite, jolla käyttäjä on selausta suorittanut. Jalanjälkeen kertyy osia siis käytännössä datapakettien kaikista osista.

Käyttäjällä ei kuitenkaan yleensä ole mahdollisuuksia päästä käsiksi heitä koskeviin jo kerättyihin tietoihin, sillä tiedot ovat yleensä jonkun ison teknologiayrityksen hallussa, esimerkiksi Apple, Google ja Facebook (nyk. Meta) (Sjöberg ym. 2016). Kerättyä ja käsiteltyä tietoa voidaan siis käyttää käyttäjän kustannuksella yrityksissä esimerkiksi voiton tuottamiseen, josta voi aiheutua turvallisuusriskejä tai ongelmia käyttäjälle. Tietoja voidaan käsitellä ja levittää ilman käyttäjän suostumusta tai tietoisuutta, vaikka esimerkiksi GDPR sääntelee yksityishenkilöitä koskevien tietojen keräystä ja käsittelyä. Käytännössä ainoa tapa vaikuttaa kerättyyn tietoon ja sitä kautta jalanjälkeen on siis ennaltaehkäistä ja rajoittaa sitä omilla toimilla. Yksikin klikkaus ja tietopaketin lähetys voi olla ratkaisevassa asemassa käyttäjän tietosuojatasoa ja turvallisuutta arvioidessa.

Digitaalista jalanjälkeä voidaan käyttää siis esimerkiksi henkilön verkkotoiminnan seurantaan ja laitteiden seuraamiseen ja paikantamiseen, tai vaikkapa käyttäjän toiminnan ennakointiin esimerkiksi kohdennetun markkinoinnin ja mainonnan tavoin (Arya, Sethi ja Paul 2019).

2.2 Jalanjäljen tarkastelu ja käsittely

Huomattava osa käyttäjistä vähintään aliarvioi digijalanjälkensä merkityksellisyyden sekä kyvyn kontrolloida sitä. Koska käyttäjällä harvoin on mahdollista vaikuttaa jo kerättyyn tietoon, on vastuu tiedon käsittelystä ja jakamisesta usein siis isommilla tekijöillä ja organisaatioilla. Näille toimijoille on tyypillistä se, että yksittäisen käyttäjän preferenssit tietojen levitykselle ja käsittelylle ovat usein toissijaisia, ja käyttäjien suostumusta toimenpiteisiin ei aina erikseen hankita (Sjöberg ym. 2016).

Sovelluskerroksen tasolla käyttäjien digitaalisia jalanjälkiä käytetään esimerkiksi palveluiden ja sovellusten monitorointiin kehittämistarkoituksessa. Kun käyttäjä kirjautuu palveluun, voidaan syntyneistä loki- ja tunnistetiedoista tallentaa käyttäjän laitteiston tietoja palvelun omille palvelimille. Myös esimerkiksi kirjautumistiedot voidaan usein tallentaa selaimeen, ja käyttäjä pystyy selaimen asetuksista esimerkiksi tarkastamaan eri sivustoille tallennettuja kirjautumistietoja, mukaanlukien salasanoja dekryptatussa, selkokielisessä muodossa. Käyttäjän ja palveluntarjoajan välille on mahdollista määritellä palvelutasosopimus, jolla määritellään palvelulle tietyt vaatimustasot esimerkiksi suojauksille niin, että se täyttää asiakkaan vaatimukset (Karadsheh 2012). Sopimukset voivat olla hankalia ymmärtää, mutta tarjoavat paljon tietoa palvelun toiminnasta ja tuesta, myös ongelmatilanteissa (Grinavich 2016).

Käyttäjien tietojen käsittelyä säännellään Euroopan alueella GDPR:n avulla, jonka periaatteina ovat tietojen käsittelyn tarpeellisuuden arviointi sekä tiedon tyypin, käyttötarkoituksen, käsittelyn keston ja oikeudellisten perusteiden määrittely (Guinchard 2021). Asetus mahdollistaa käyttäjälle oikeuden tarkistaa häntä koskevia tietoja sekä saada tieto siitä, miten tiedot ovat kerätty, miten niitä käsitellään ja miten niitä jaetaan. 2020-luvun pandemian aikana asetus nousi keskustelun aiheeksi pohdittaessa ihmisten liikkumisesta ja terveydestä kerättäviä tietoja ja niiden suojausta, ja yksittäisen käyttäjän todellinen kyky seurata tiedonkeruuta hänestä itsestään kyseenalaistettiin (Guinchard 2021). Sovelluksien, kuten sosiaalisten medioiden tai suoratoistopalveluiden liikenne saattaa kulkeutua palvelimelta asiakkaalle käyden Euroopan unionin alueen ulkopuolella, jolloin tiedonsiirto ei ole enää GDPR:n alaista.

3 Käyttäjä tiedonkeruun mahdollistajana

Tässä luvussa perehdytään kuluttajakäyttäytyjän toimintaan tiedonkeruuseen liittyen. Blank, Dutton ja Lefkowitz (2019) totesivat kyselynsä tulosten perusteella, että noin 70 prosenttia vastaajista tuntee olonsa epämukavaksi joutuessaan kohdennetun mainonnan kohteeksi, mutta vain noin neljännes vastaajista ylläpitää toimia tämän toiminnan estämiseksi. Kohdennetun mainonnan edellytyksenä on esimerkiksi pitkäkestoinen käyttäjän selaushistorian tallennus-, analysointi- ja käsittelyprosessi. Prosessiin saattaa kuulua mukaan myös tietoja, joita käyttäjä ei ole erikseen sallinut käytettävän, ja yleinen ongelma onkin se, että käyttäjät sallivat tiedoilleen suurempia ja monipuolisempia toimia, kuin mitä oikeasti ajattelevat tai haluavat. Käyttäjän yksityisyyden suojaamisen suurimpia haasteita onkin suojauskäytäntöjen ja toiminnallisuuksien haastavuus ja monimutkaisuus (Shillair ym. 2015).

3.1 Käyttäjän aktiivisesti jakama tieto

Käyttäjällä on loppupelissä hyvin vähän mahdollisuuksia vaikuttaa itse prosesseihin, mutta käyttäjällä on käsissään koko yhteisprosessin alkuaineisto. Jos käyttäjä ei rajoita tätä alkuaineiston käyttöä alusta saakka, hän hyvin todennäköisesti joutuu tilanteeseen, jossa aineistosta on luotu hänelle henkilökohtaistettuja mainoksia tai ehdotuksia. On siis tärkeää, että käyttäjä on jatkuvasti ajan tasalla siitä, mitä tietoja hän itse tietoisesti jakaa, ja mikä palvelin tiedot saa haltuunsa (Micheli, Lutz ja Büchi 2018). Yhtä tärkeää on myös tiedostaa tekniikan käytöstä johtuvien tietojen keräys ja niiden luonne. Tekniikan ominaisuuksien ja luonteen vuoksi on lähes mahdotonta saavuttaa tilannetta, jossa yksilö voisi käyttää palveluita toimivasti, mutta samalla estää esimerkiksi yksilöiväksi tiedoksi leimattua tietoa kulkemasta palvelimelle (Haimson ym. 2016).

Jokaisella käyttäjällä on omat mielipiteensä ja ajatuksensa riittävästä turvallisuustasosta, mutta jokaisen teknologiaa hyödyntävän käyttäjän tulisi kiinnittää erityistä huomiota siihen, että tietoja jaettaessa on aina olemassa riski joutua esimerkiksi verkkorikoksen uhriksi. Yeh ym. (2018) mukaan käyttäjien halukkuus jakaa omia tietojaan oli suoraan riippuvainen siihen liitetystä saavutuksesta tai palkinnosta. Esimerkiksi verkkopalvelun profiilia luodes-

sa ne käyttäjät, joiden tietoisuus tietosuojariskeistä ja yksityisyydestä oli suurempaa, yleensä arvioivat saatavia ja kustannuksia tarkemmin, kun taas vähemmän ymmärtävät monesti käyttivät esimerkiksi oletusasetuksia tai noudattivat suoraan yleiseiä ohjeita (Yeh ym. 2018).

Goad, Collins ja Gal (2021) totesivat artikkelissaan, että ihmisten antama arvo omalle yksityisyydensuojalleen on usein riippuvainen käyttäjän saamasta hyödystä, jos hän päättää pitää tietojaan yksityisenä. Kun tähän yhdistetään Yeh ym. (2018) esiintuomat havainnot tiedon jakamiseen liittyen, syntyy käsitys, jonka mukaan valtaosa käyttäjien toiminnasta on jollain tavalla aina riippuvainen käyttäjän saamasta palkinnosta, toimi hän tiedon jakamisen suhteen miten tahansa.

3.2 Tiedostamaton tietojen jakaminen

Kuten artikkelissa (Goad, Collins ja Gal 2021) mainitaan, että käyttäjien käytössä olevat tietosuojapreferenssit ovat usein alhaisemmat kuin mitä käyttäjät itse väittävät. Tämä juontaa mahdollisesti juurensa juuri siitä, että käyttäjät eivät todellisuudessa tiedä täysin, miten järjestelmät tietoa keräävät, eivätkä sitä myötä osaa suojautua tiedonkeruulta oikein. Yksi selitys tälle on, että käyttäjät väittävät tietoisesti omaavansa korkeamman tason säännöstelymetodit tiedonjakamiselle kuin mitä he oikeasti omaavat, jotta he voisivat kuvitteellisesti paikata tilannetta (Goad, Collins ja Gal 2021). Evästeasetukset ovat nousseet esille 2020-luvulla monissa tiedonkeruuseen liitetyssä artikkelissa, ja niiden ymmärtämisessä on ihmisillä vaikeuksia edelleen. (Kristol 2001) mukaan asetusten kanssa kannattaa kuitenkin olla erittäin tarkka alusta saakka, jos pyritään rajoittaan niiden osuutta jalanjälkeen.

Micheli, Lutz ja Büchi (2018) jakaa passiivisen tiedonkeräämisen kahteen erilaiseen osaan. Ensimmäinen osa koskee alustojen ja palveluiden käyttäjistä keräämää tietoa, ja toinen osa koskee käyttäjien toisista henkilöistä julkaisemaa tietoa. Shillair ym. (2015) toteavat artikkelissaan, että käyttäjien esimerkiksi sosiaalisissa medioissa yleisesti jakamat henkilötiedot, kuvat ja nimet kasvattavat merkittävästi yksilön lisäksi myös koko verkoston riskiä joutua hyökkäyksen kohteeksi. Verkosto voi tässä vaiheessa kattaa yksittäisen laiteverkoston, jos kyseessä on suora hyökkäys käyttäjää itseensä vastaan, tai vaihtoehtoisesti käyttäjän sosiaalisen verkoston. Jaettujen kuvien ja tietojen joukossa on monesti myös tietoja esimerkiksi

julkaisuun merkatusta henkilöstä tai julkaisut voivat koskea nimenomaan kohdekäyttäjää itse. Micheli, Lutz ja Büchi (2018) tuovat artikkelissaan esiin, että kohteeksi joutuneiden käyttäjien jalanjälki kasvaa julkaisujen myötä, ja niiden sisältämä mahdollinen epätoivottu tieto tai materiaali voi tuottaa merkittäviä haittavaikutuksia käyttäjälle.

3.3 Riskit liittyen tietovuotoihin ja murtoihin

Yksittäisen käyttäjän tiedot voivat esimerkiksi huolimattoman toiminnan (Sen ja Borle 2015) takia päätyä osaksi laajempaa tietomurtoa, jotka yleensä kohdistuvat johonkin tiettyyn tahoon. Tällöin usean käyttäjän tiedot ovat teknisesti kenen tahansa nähtävillä ja käsiteltävissä, mikä luo merkittävän riskin yksittäiselle käyttäjälle, jos mukana on esimerkiksi terveys- tai henkilötietoja. Käyttäjät eivät yleensä kykene tehdä asialle tässä tapauksessa mitään, jonka vuoksi vuotojen estämiseksi määritellään jatkuvasti uusia asetuksia ja lakeja (Sen ja Borle 2015) sekä kohteiksi joutuneita tahoja painostetaan tuomaan asia julki, jotta selvittelytoimet ja korjaukset pystytään suorittamaan tehokkaammin.

Riskejä yksittäiselle käyttäjälle muodostuu myös tilanteissa, joissa esimerkiksi yksi hänen heikoista salasanoista murretaan. Murron suorittanut taho voi omata samasta tunnistettavasta käyttäjästä tietovuodon seurauksena esimerkiksi muiden palveluiden käyttäjätilien tietoja, joihin salasanaa tai sen variaatioita voidaan kokeilla (Poornachandran ym. 2016). Tilien suojaamiseen käytettävien salasanojen tulisi aina olla tyypiltään vahvoja, ja erilaisia kaikissa eri palveluissa. Poornachandran ym. (2016) mukaan noin kolmannes tutkituista verkkosivujen käyttäjätileistä olivat vaarassa joutua salasanan uudelleenkäytön vuoksi käyttäjätilin varastuksen kohteeksi. Myös reilusti yli puolet käyttäjistä ilmaisivat käyttävänsä samaa salasanaa useamman eri palvelun käyttäjätileissä (Poornachandran ym. 2016).

4 Työkaluja tiedonhallintaan

Perera ym. (2020) mainitsevat artikkelissaan, että sovelluskehittäjien resurssien suurempi tarve muilla osa-alueilla kuin tietosuojan selkeyden ja ymmärrettävyyden parissa aiheuttaa käyttäjälle hankaluuksia ymmärtää tietosuojan toimintaa. Käyttäjien tulisi siis käyttää enemmän omia resurssejaan oman tietosuojan ylläpitämiseen, ja merkittävä osa käyttäjistä jättää siksi tämän tekemättä (Shillair ym. 2015). Huomattava osa Internetin käyttäjistä on myös ilmaissut, että heidän turvallisuudestaan verkosta eivät vastaa käyttäjät itse, tai vaihtoehtoisesti havaitsevat olevansa itse kyvyttömiä huolehtimaan siitä (LaRose ja Rifon 2007). Palveluiden suojauksessa resursseja allokoidaan useammin järjestelmiin itsessään, koska niiden sisältämän tiedon arvokkuus ja suojaus on usein huomattavasti suurempaa (Sen ja Borle 2015), joten kaikkia käyttäjien tarpeita ei yksinkertaisesti pysty asettamaan kehittäjien harteille. Tässä kappaleessa pohditaan vaihtoehtoja, joita yksittäisellä käyttäjällä on olemassa koskien hänen aktiivista tiedon jakamista eri palveluille, sekä tutkitaan mahdollisuuksia, joilla tiedostamatonta tai passiivista tiedonkeruuta voidaan rajoittaa.

4.1 Vaikuttaminen tietoisesti jaettavaan tietoon

Symanovich (2018) mukaan hyvä alkuaskel digitaalisen jalanjäljen hallintaan on selvittää, mitä hakutuloksia ja tietoja omalla nimellä internetistä haettaessa löytyy. Jos haku tuottaa käyttäjää itseään koskevia tuloksia, joiden julkisuus ei ole hyvä asia, käyttäjä voi pyytää palvelun ylläpitäjältä tietoten poistamista. Sosiaalisen median puolella Symanovich (2018) mainitsee palveluiden profiilien yksityisyysasetusten tarkastamisen sekä huolellisuuden kaikissa julkaisuissa. Ellison ym. (2011) tuovat esiin myös yksityisyysasetusten tarkastamisen ja huolellisuuden merkityksellisyyden, koska tosiassa käyttäjän on hankala tarkasti määritellä, että kenen on mahdollista päästä julkaisuihin tai profiilien tietoihin käsiksi.

Shillair ym. (2015) ja Micheli, Lutz ja Büchi (2018) toivat esiin sosiaalisen verkostojen mahdollisuuden toimia tietojen levityksen välineenä. Käyttäjän on siis pysyttävä ajan tasalla omasta verkostostaan, sekä usein järkevää on hyväksyä osaksi omaa sosiaalista verkostoa vain välttämättömät tai tunnetut henkilöt. Vaihtoehtoisesti mukaan voi hyväksyä kaikki pyr-

kivät, jolloin kyseisessä verkostossa ei pidä jakaa mitään, mitä ei julkisesti voisi jakaa (Symanovich 2018). Jos verkosto koostuu vain omasta lähipiiristä tai tunnetuista henkilöistä, on myös asiallista kysyä lupa muilta käyttäjiltä, jos aikomuksena on julkaista heitä koskevia julkaisuja tai tietoja.

Symanovich (2018) korostaa myös, että käyttäjien on tärkeää käyttää uniikkeja ja vahvoja salasanoja palveluissa. Vaikka käyttäjätili itsessään ei olisi erityisen tärkeä, voi profiilin tiedoista löytyä jotain, jota käyttäjän muihun tietoihin yhdistämällä voidaan hyödyntää (Poornachandran ym. 2016). Salasanojen hallintaan on olemassa sovelluksia, jolla voi luoda turvallisesti säilytettäviä vahvoja salasanoja, jotka säilytetään suojatussa tietokannan tyyppisessä tiedostossa. Myös turhat tai vanhentuneet profiilit ja käyttäjätunnukset on hyvä poistaa palveluista.

4.2 Tiedostamattoman tiedonkeräyksen rajoitus

Symanovich (2018) mukaan käyttäjien tulisi tarkistaa kaikkien käyttämiensä sovellusten käyttöoikeudet koskien henkilökohtaisia tietoja. Oikeuksia rajoittamalla voidaan estää esimerkiksi sähköpostien tai yhteystietojen päätyminen sovelluksen tai palveluntarjoajan haltuun, tai sijaintitietojen julkisuus esimerkiksi puhelimissa, kelloissa tai muissa IoT-laitteissa. Toisena välineenä Symanovich (2018) mainitsee VPN-yhteyksien käyttämisen, joilla voidaan estää esimerkiksi palveluita asettamasta selaushistoriaa keräämiä evästeitä. Myös käyttäjän alkuperäinen IP-osoite voidaan peitää VPN:n avulla, jolloin käyttäjän paikantaminen on myös haastavampaa.

Suomessa digi- ja väestötietovirasto ylläpitää väestötietojärjestelmää, jossa henkilöt voivat tietoja suojatakseen asettaa rajoitteita tietojensa luovutukseen esimerkiksi suoramarkkinointikielto, turvakielto ja yhteystietojen luovutuskielto. Teleoperaattorit mahdollistavat oman puhelinnumeron salaamisen, jolloin numeron pitäisi kadota numero- ja hakupalveluista. Näin käyttäjä kykenee soittamaan muihin kuin viranomaisnumeroihin ilman, että käyttäjän puhelinnumero näkyy vastaajalle (Digi- ja väestötietovirasto 2022). Myös esimerkiksi terveydenhuollon verkkopalveluissa tai muissa rekisteröintijärjestelmissä on usein mahdollista tarkastella ja rajoittaa omien tietojen keräystä ja julkisuutta, ja GDPR:n noudattamiseksi järjestel-

mät ovat nykyään vielä osittain hajautettu, sillä kaikki palvelut eivät todellisuudessa säännöstä täysin noudata.

Shillair ym. (2015) ja Micheli, Lutz ja Büchi (2018) mukaisesti esimerkiksi läheiset tai perheenjäsenet voivat huomaamattaan vaikuttaa käyttäjän oman jalanjäljen muodostumiseen julkaisemalla tätä koskevia julkaisuja. Tämän rajoittamiseksi on tärkeää sopia sosiaalisessa verkostossa yhteiset pelisäännöt sille, mitä kenestäkin saa jakaa julkisesti ja milloin on syytä kysyä lupa. Kyberturvallisuuskeskus kiteyttää tämän kehoitukseen olla julkaisematta netissä materiaalia, joka koskee toisia henkilöitä, ja jota ei itse haluaisi itsestä julkaistavan (Kyberturvallisuuskeskus 2022).

5 Yhteenveto

Verkostoituneet laitteet keräävät meistä nykypäivänä tietoa yhä kasvavissa määrin, ja tätä tietoa pyritään hyödyntämään elämän helpottamiseksi luomalla käyttäjille digitaalisia jalanjälkiä, mutta tietoja käytetään myös ikäviin tarkoituksiin. Goad, Collins ja Gal (2021) vahvistivat artikkelissaan havaintoa siitä, että suurempi panostus tiedon suojaamiseen ei aina ole paras ratkaisu silloin, kun tiedon suojaamiseksi käytettävät resurssit voittavat arvoltaan suojaamisesta saadut hyödyt. Käyttäjien tulisikin aina luoda arvio siitä, mitä tietojen julkaisu käytännössä heille maksaa ja millaista hyötyä hän siitä saa (Micheli, Lutz ja Büchi 2018), koska tiedonkeruuta on hankala estää tai rajoittaa merkittävästi. Jalanjäljen tarkastelusta ja käsittelystä pyritään luomaan erilaisten säännösten avulla käyttäjälle helpompaa samalla parantaen myös osittain olemassaolevan tiedon poistomahdollisuuksia (Guinchard 2021).

Tärkeimpiä järjestelmällisiä suojaustoimia käyttäjille itselleen tietoja jakaessa ovat vahvat ja uniikit salasanat (Poornachandran ym. 2016), jotka pienentävät riskiä joutua tietovarkauden uhriksi sekä selaustietoja peittävien sovelluksien käyttö, kuten välityspalvelimen kautta verkkoliikenteen ohjaava VPN (Symanovich 2018). Tämän ohella käyttäjien kannattaa seurata omaa julkista jalanjälkeä hakukoneilla (Symanovich 2018) ja rajoittaa sosiaalista verkostoa omien preferenssien mukaan (Shillair ym. 2015), (Symanovich 2018). Selainten evästetietojen ja laitteiden diagnostiikkatiedojen kanssa tulee myös olla tarkkana, jotta esimerkiksi oman sijainnin voi suojata.

Tutkielman aihealueesta on olemassa jonkin verran erilaisia tutkimuksia ja artikkeleita, joissa on yleensä tarkastelunäkökulmaksi valittu esimerkiksi vaikutukset liiketoimintaan tai lailliset näkökulmat. Koska ihmisistä kerääntyvän datan määrä kasvaa suurella nopeudella, ja yhä useampi arkinen palvelu ottaa teknologian osaksi toimintaansa, on myös tutkimuksia kerättävään tietoon liittyen suoritettava lisää käyttäjän näkökulmasta, esimerkiksi vertailemalla tietojen keräämistä aiheena eri ikäryhmien tai yhteiskuntaluokkien sisällä. Yleisesti tutkimustuloksien luotettavuus on suhteellisen hyvä, vaikka tutkimustuloksien tarkastelunäkökulmien erot luovat lieviä ristiriitoja tulosten välille. Täydellistä ratkaisua ongelmaan ei ole olemassa, mutta hyvä ratkaisu muodostuu suhteellisen rajallisesta määrästä erilaisia ohjeita ja toimintaperiaatteita.

Lähteet

Arya, Vikas, Deepa Sethi ja Justin Paul. 2019. “Does digital footprint act as a digital asset?–Enhancing brand experience through remarketing”. *International Journal of Information Management* 49:142–156. https://doi.org/10.1016/j.ijinfomgt.2019.03.013.

Azucar, Danny, Davide Marengo ja Michele Settanni. 2018. “Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis”. *Personality and individual differences* 124:150–159. https://doi.org/10.1016/j.paid.2017.12.018.

Blank, Grant, William H Dutton ja Julia Lefkowitz. 2019. “Perceived Threats to Privacy Online: The Internet in Britain, the Oxford Internet Survey, 2019”, https://oxis.oii.ox.ac.uk/896-2/.

Deeva, Irina. 2019. “Computational personality prediction based on digital footprint of a social media user”. *Procedia computer science* 156:185–193. https://doi.org/10.1016/j.procs.2019.08.194.

Digi- ja väestötietovirasto. 2022. “Tietojen luovuttamisen kieltäminen”, https://dvv.fi/tietojen-luovuttamisen-kieltaminen.

Ellison, Nicole B, Jessica Vitak, Charles Steinfield, Rebecca Gray ja Cliff Lampe. 2011. “Negotiating privacy concerns and social capital needs in a social media environment”. Teoksessa *Privacy online,* 19–32. Springer. https://doi.org/10.1007/978-3-642-21521-6_3.

Goad, David, Andrew T Collins ja Uri Gal. 2021. “Privacy and the Internet of Things- An experiment in discrete choice”. *Information & Management* 58 (2): 103–292. https://doi.org/10.1016/j.im.2020.103292.

Grinavich, Jim. 2016. “How to Evaluate Security Company Service Level Agreements”, https://www.vectorsecurity.com/blog/how-to-evaluate-security-company-service-level-agreements.

Grover, Ted, ja Gloria Mark. 2017. “Digital footprints: Predicting personality from temporal patterns of technology use”. Teoksessa *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers,* 41–44. https://doi.org/10.1145/3123024.3123139.

Guinchard, Audrey. 2021. “Our digital footprint under Covid-19: should we fear the UK digital contact tracing app?” *International Review of Law, Computers & Technology* 35 (1): 84–97. https://doi.org/10.1080/13600869.2020.1794569.

Haimson, Oliver L, Jed R Brubaker, Lynn Dombrowski ja Gillian R Hayes. 2016. “Digital footprints and changing networks during online identity transitions”. Teoksessa *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* 2895–2907. https://doi.org/10.1145/2858036.2858136.

Karadsheh, Louay. 2012. “Applying security policies and service level agreement to IaaS service model to enhance security and transition”. *computers & security* 31 (3): 315–326. https://doi.org/10.1016/j.cose.2012.01.003.

Kristol, David M. 2001. “HTTP Cookies: Standards, privacy, and politics”. *ACM Transactions on Internet Technology (TOIT)* 1 (2): 151–198. https://doi.org/10.1145/502152.502153.

Kyberturvallisuuskeskus. 2022. “Netiketti - Verkossa liikkujan työkalupakki”, https://www.kyberturvallisuuskeskus.fi/fi/ajankohtaista/ohjeet-ja-oppaat/netiketti-verkossa-liikkujan-tyokalupakki.

LaRose, Robert, ja Nora J Rifon. 2007. “Promoting i-safety: effects of privacy warnings and privacy seals on risk assessment and online privacy behavior”. *Journal of Consumer Affairs* 41 (1): 127–149. https://doi.org/10.1111/j.1745-6606.2006.00071.x.

Micheli, Marina, Christoph Lutz ja Moritz Büchi. 2018. “Digital footprints: an emerging dimension of digital inequality”. *Journal of Information, Communication and Ethics in Society,* https://doi.org/10.1108/JICES-02-2018-0014.

Networkel.com. 2016. “Osi Model : 7 Layer Of The Network Communication”, https://networkel.com/osi-model-7-layer-network-communication/.

Perera, Charith, Mahmoud Barhamgi, Arosha K Bandara, Muhammad Ajmal, Blaine Price ja Bashar Nuseibeh. 2020. “Designing privacy-aware internet of things applications”. *Information Sciences* 512:238–257. https://doi.org/10.1016/j.ins.2019.09.061.

Poornachandran, Prabaharan, M Nithun, Soumajit Pal, Aravind Ashok ja Aravind Ajayan. 2016. “Password reuse behavior: How massive online data breaches impacts personal data in web”. Teoksessa *Innovations in Computer Science and Engineering,* 199–210. Springer. https://doi.org/10.1007/978-981-10-0419-3_24.

Sen, Ravi, ja Sharad Borle. 2015. “Estimating the contextual risk of data breach: An empirical approach”. *Journal of Management Information Systems* 32 (2): 314–341. https://doi.org/10.1080/07421222.2015.1063315.

Shillair, Ruth, Shelia R Cotten, Hsin-Yi Sandy Tsai, Saleem Alhabash, Robert LaRose ja Nora J Rifon. 2015. “Online safety begins with you and me: Convincing Internet users to protect themselves”. *Computers in Human Behavior* 48:199–207. https://doi.org/10.1016/j.chb.2015.01.046.

Sjöberg, Mats, Hung-Han Chen, Patrik Floréen, Markus Koskela, Kai Kuikkaniemi, Tuukka Lehtiniemi ja Jaakko Peltonen. 2016. “Digital me: Controlling and making sense of my digital footprint”. Teoksessa *International Workshop on Symbiotic Interaction,* 155–167. Springer, Cham. https://doi.org/10.1007/978-3-319-57753-1.

Symanovich, Steve. 2018. “What is a digital footprint? And how to help protect it from prying eyes”, https://us.norton.com/internetsecurity-privacy-clean-up-online-digital-footprint.html.

Wang, Shaojung Sharon. 2013. ““I share, therefore I am”: Personality traits, life satisfaction, and Facebook check-ins”. *Cyberpsychology, Behavior, and Social Networking* 16 (12): 870–877. https://doi.org/10.1089/cyber.2012.0395.

Yeh, Ching-Hsuan, Yi-Shun Wang, Shin-Jeng Lin, Timmy H. Tseng, Hsin-Hui Lin, Ying-Wei Shih ja Yi-Hsuan Lai. 2018. “What drives internet users’ willingness to provide personal information?” *Online Information Review* 42 (6): 923–939. https://doi.org/10.1108/OIR-09-2016-0264.