VAASAN YLIOPISTO
TEKNILLINEN TIEDEKUNTA
TIETOTEKNIIKKA

Jarkko Antintupa

KÄYTTÄJIEN SEURANTA INTERNETISSÄ

Tietotekniikan
pro gradu –tutkielma

Teknisen viestinnän koulutusohjelma

VAASA 2015

SISÄLLYSLUETTELO

KUVIOLUETTELO **sivu**

VAASAN YLIOPISTO
Teknillinen tiedekunta
Tekijä: Jarkko Antintupa
Tutkielman nimi: Käyttäjien seuranta Internetissä
Ohjaajan nimi: Johanna Aalto
Tutkinto: Kauppatieteiden maisteri
Oppiaine: Tietotekniikka
Koulutusohjelma: Teknisen viestinnän koulutusohjelma
Opintojen aloitusvuosi: 2011
Tutkielman valmistumisvuosi: 2015 **Sivumäärä:** 71

TIIVISTELMÄ

Tietopaljastukset ja huoli yksityisyydestä Internetissä on viime vuosina kasvanut. Teknologian nopean kehityksen myötä myös Internet-käyttäjän on hankalampi pysyä mukana kehityksessä ja olla perillä Internet-selailuun liittyvistä vaaroista. Tämän työn tarkoituksena on tutkia, kuinka käyttäjiä seurataan Internetissä ja mitä uhkia käyttäjän seuranta tuottaa. Työssä esiintyy erilaisia tekniikoita, joita käytetään Internet-käyttäjän seurantaan sekä tutkitaan, ketkä seuraavat ja mitä tarkoitusta varten. Kaikkia teknologioita ei kuvata työssä yksityiskohtaisesti, vaan työssä tarkastellaan yleisimpiä ja työn kirjoitushetkellä vallitsevia tekniikoita. Viruksien tai haittaohjelmien avulla tehdyt tiedonkalastelut jäävät työn ulkopuolelle.

Ensimmäisenä työ johdattaa lukijan markkinointiin, jonka tarkoituksena on selvittää tavalliset syyt käyttäjien seurantaan. Seuraavaksi esitellään seurantaan käytetyt tekniikat ja niiden toimintaperiaatteet. Tämän jälkeen työ keskittyy itse seurantaan ja sen uhkiin. Viimeisenä havainnollistetaan empiirisellä tutkimuksella, kuinka usealle sivustolle verkkoselailu voi lähettää tietoa selailijasta. Tätä tarkastellaan Firefox -selaimen Lightbeam nimisen lisäosan avulla ja tarkastelun kohteena on 20 suomalaisten suosimaa sivustoa.

Työn tuloksena paljastui, että seurantateknologioita on lukuisia ja omien selailutietojen paljastuminen on todellinen uhka, johon käyttäjän tulisi kiinnittää huomiota. Käyttäjän vieraillessa verkkosivustoilla, voi tietoa lähteä useammalle vierailemattomalle sivustolle mainosten ja erilaisten verkkoteknologioiden kautta. Seurantateknologiat ovat tärkeitä varsinkin markkinointiyrityksille. Uudet Internet-teknologiat tuovat uudenlaisia uhkia käyttäjien seurantaan ja kaikkien seurantateknologioiden estäminen tai niiden tunnistaminen saattaa olla käyttäjälle hankalaa.

AVAINSANAT: Internet-seuranta, yksityisyys, evästeet

UNIVERSITY OF VAASA
Faculty of technology
Author: Jarkko Antintupa
Topic of the Master's Thesis: Käyttäjien seuranta Internetissä
Instructor: Johanna Aalto
Degree: Master of Science in Economics and Business Administration
Major Subject: Computer Science
Degree Programme: Degree Programme in Technical Communication
Year of Entering the University: 2011
Year of Completing the Master's Thesis: 2015 **Pages:** 71

ABSTRACT

Concern over data seepage and privacy in the Internet has risen recently through different media. Because of fast advancement of the Internet technology, it is harder than ever to stay informed in every aspect that concerns Internet browsing and dangers what a user may face. The purpose of this thesis is to study how Internet users are being tracked and what are the associated threats. Different techniques or technology related to tracking Internet users are also presented in this thesis. This thesis also examines who uses tracking technology and for what reason. Not every technology is presented in detail but the main focus is around the most commonly used and the most prevalent technology. Computer viruses and malicious software are not part of this thesis.

At the beginning this thesis examines what marketing means and how it relates to tracking. Next part explains the tracking technologies and how they work. After that this paper tries to explain what Internet-tracking means and what kind of threats an Internet user may face. The last part illustrates how web browsing can send information about the Internet user to multiple other websites or trackers with empirical study made with Firefox add-on called Lightbeam.

This research found that there are several different tracking technologies and revealing one's own browsing data is a real threat where everyone should pay attention. When a user visits different websites, information may be leaked to several other unvisited websites from advertisements or through different web technologies. Tracking technologies are important tools at least for marketing companies. New web technologies draw more concerns over privacy risks and it might be hard for a user to block or recognize all tracking technologies.

KEYWORDS: Internet tracking, privacy, cookies

1 JOHDANTO

1.1 Työn taustaa

Aiemmin Internet-sivustojen sisältö toteutettiin ja ylläpidettiin yhden henkilön tai toimijan avulla. Nykyään sisältöä tulee myös kolmansilta osapuolilta, kuten mainoksista tai markkinointiyrityksiltä, analyyttisistä työkaluista, sosiaalisista verkoista jne. Kolmannen osapuolen palveluissa on kuitenkin riskinsä, mikä on kiinnittänyt tutkijoiden, yhteiskuntaorganisaatioiden ja yhteiskuntapäättäjien huomion. Huomio kohdistuu Internet-selailijan selailutietojen seurantaan. (Mayer ja Mitchell 2012.)

Lukuiset yritykset hyödyntävät erilaisia tapoja käyttäjien seurantaan ja tunnistamiseen Internetissä (Privacy Rights Clearinghouse 2014). Uusi Web kieli tai hypertekstin merkintäkieli (Hypertext Markup Language) ja sen tuomat toiminnot tarjoaa vielä enemmän seurantamahdollisuuksia (Tanzina 2010).

Tiedon valtava määrä on luonut uuden käsitteen nimeltä Big data. Tämän hallitseminen hyödyntävästi ja turvallisesti on vaikeaa johtuen tiedon kasvuvauhdista, koosta ja monimutkaisuudesta (Navint 2012). Asiantuntijat arvioivat, että vuonna 2015 tulee olemaan 25 miljardia yhteyksiin pystyvää laitetta ja määrän odotetaan kasvavan 50 miljardiin vuoteen 2020 mennessä. Tästä on syntynyt käsite *esineiden Internet* (FTC 2015). Näiden asioiden seurauksena syvällisempi tutkimus Internet-selailijan yksityisyyteen ja heidän seurantaan on ajankohtaista.

1.2 Työn tarkoitus ja toteutus

Tämän työn tarkoituksena on tutkia kuinka käyttäjiä seurataan Internetissä ja myös pohtia, mitä uhkia tähän aiheeseen liittyy. Työssä esiintyy erilaisia tekniikoita, joita käytetään Internet-käyttäjän seurantaan, sekä ketkä seuraavat ja mitä tarkoitusta varten. Kaikkia teknologioita ei kuvata työssä yksityiskohtaisesti, vaan työssä tarkastellaan yleisimpiä ja työn kirjoitushetkellä vallitsevia tekniikoita.

Tietovuotouutisia ja muita tämän kaltaisia tiedotteita on tullut jo kauemman aikaa. Moni asia hoidetaan Internetissä ja sieltä voi lukea uutisia, hoitaa pankkiasioita, etsiä erilaisia ohjeita tai vaikka opiskella. Työpaikoillakin käytetään monesti Internetiä. Nykyään Internet on monella myös ns. liikenteessä käytössä, sillä monet matkapuhelimet kykenevät muodostamaan yhteyden Internetiin. Jos seuranta on mukana jokaisessa näissä toimenpiteissä, on yksityisyyden menettämisen uhka jo suuri. Seurantatietoja voidaan käyttää hyvää tarkoitusta varten, kuten tarjoamalla käyttäjille verkkosivustoa heidän omalla kielellä perustuen heidän maantieteelliseen sijaintiin tai tarjoamalla tuotteita, joista käyttäjä on kerättyjen tietojen mukaan kiinnostunut (Cranor, Sleeper & Ur 2013). Työ käsittelee asiaa etsien vastausta onko tietojen kerääminen yleistä ja mihin tarkoitukseen tietoa käytetään. Työssä pyritään hyödyntämään uusimpia uutisia, artikkeleita ja tutkielmia löytääkseen vastauksia seurantaan liittyen.

Luvussa kaksi tarkastellaan seurannan syntyä. Tässä osiossa kerrotaan markkinoinnin historiasta, mitkä asiat tai tapahtumat ovat ajaneet seurantateknologioiden syntymistä sekä mitä varten niitä on kehitetty. Luvussa kolme perehdytään tarkemmin seurantaan liittyviin tekniikoihin ja seurannan osapuoliin. Seuraavassa osiossa käsitellään itse seurantaa. Tässä kohtaa esiintyy usein erilaisia seurantaan liittyviä työkaluja ja teknologiaa, minkä takia ne on selostettu aiemmassa luvussa kattavammin.

Lähteistä saatuja tietoja verrataan työn empiiriseen tutkimusosioon. Tutkimuksessa havainnollistetaan Firefox -selaimeen asennettavan Lightbeam -lisäosan avulla sitä, kuinka monelle eri sivustolle tai toimijalle tietoa kulkeutuu suosituimmissa suomalaisten

käyttämistä sivustoista. Tämä empiirinen osio antaa konkretiaa aiemmin työssä käsitel-tyihin asioihin. Työn lopussa on yhteenveto käsitellyistä asioista.

2 INTERNET-MARKKINOINTI SEURANNAN EDISTÄJÄNÄ

Vasta Internet-selaimet siirsivät Internetin graafisten käyttöliittymän hallitsevaan "web-biaikaan". Tekstipohjainen Internet edelsi vuosien 1993 - 1994 selaininnovaatioita. Internet-markkinointi yleistyi Internet-selainten johdosta, koska selaimet kiihdyttivät Internetin kehittymistä ja liiketaloudellisten mallien kasvua. Ensimmäinen graafinen verkkoliittymä, Mosaic, tuli ilmaiseksi jakeluun Internetissä vuonna 1993. Internetin tuomaa muutosta nopeutti huomattavasti lokakuussa 1994 Netscape Navigator-selainohjelman ilmestyminen. (Steinbock 1998: 53-54.)

2.1 Internet-mainonnan ja seurannan synty

Markkinoinnin kannalta suuri kehitys oli Wired Venturesin oman verkkosivuston avaaminen. Erilaisia kotisivuja ja verkkomainontaa oli kehitetty jo ennen tätä, mutta Wired Venturesin verkkosivuston vahvuus oli sen liiketaloudellinen malli. Malli sai paljon huomiota lehtien otsikoissa ja moni alkoi jäljitellä tätä mallia. Internetin kasvaessa, sen kaupallistuminen lähti vahvimmin liikkeelle kuluttajamarkkinoiden sijasta yritysmarkkinoilla. Tämä mahdollisti yritysten nopeamman verkostoitumisen. (Steinbock 1998: 54 - 57.)

Mediayhtiöt eivät voineet pysytellä Internetin ulkopuolella, koska johtavien ja globaalien markkinoijien verkostoituminen olisi voinut tarkoittaa mediayhtiöille merkittäviä menetyksiä mainostuloissa. Medioiden siirtyminen Internetiin kiihdytti verkkopalveluiden rahoittamista mainostuloin. Luotettavia tutkimuslaitoksia tarvittiin tuottamaan pätevää tutkimustietoa mainosyleisöstä ja jo huhtikuussa 1994 markkinoille tuli kaksi seurantaan keskittynyttä yritystä. Vaatimukset tunnuslukujen tarkkuudesta ja tehokkuudesta kasvoivat. Yritykset alkoivat kehittää erilaisia liiketoimintamalleja ja hyödyntää seu-

rantaa avustavia tekniikoita. Eväste -tekniikkaa alettiin hyödyntää käyttäjien seurannassa ja tekniikalle tuli useita kokeilijoita. Tämän seurauksena alkoi esiintyä kiistoja yksityisyyden suojasta. (Steinbock 1998: 72-80.)

Markkinointiyritykset yrittävät jatkuvasti kehittää parempia seurantatekniikoita, koska käyttäjän on mahdollista poistaa esimerkiksi edellä mainitut HTTP -evästeet, joita yritys voi käyttää seurantaan. Evästeiden poistaminen vaikuttaa markkinointiyrityksiin siten, että he eivät pysty tarkasti laskemaan uniikkien kävijöiden lukumäärää ja näin ollen kävijämäärä voi vaikuttaa todellista suuremmalta. (Soltani, Canty, Mayo, Thomas ja Hoofnagle 2009.)

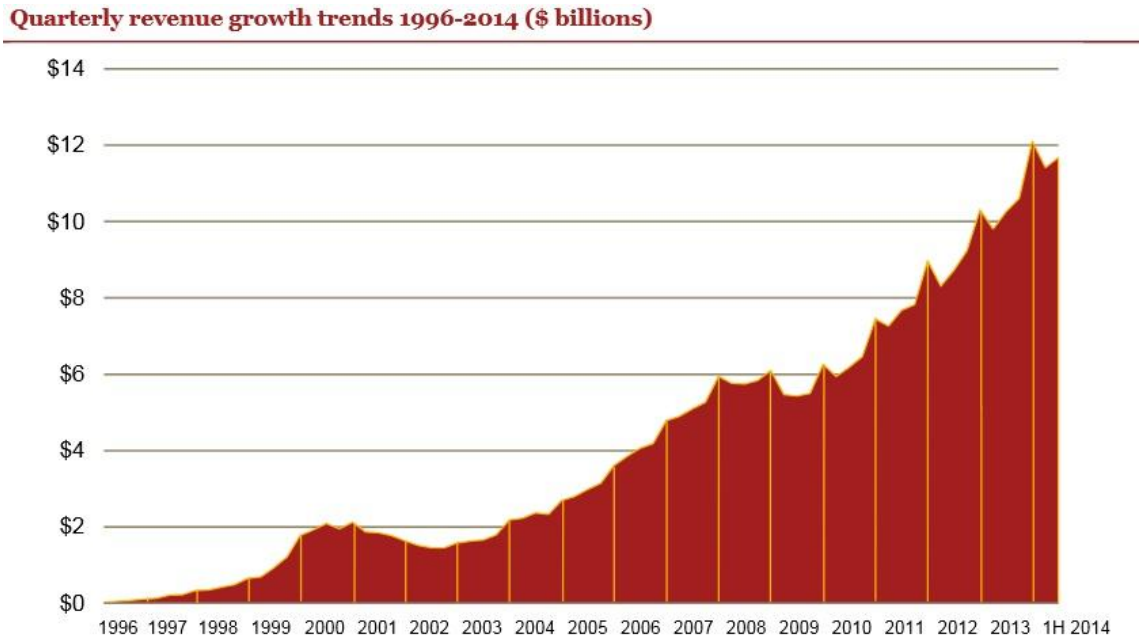
Markkinointitapoja on monenlaisia, kuten esimerkiksi suoramarkkinointi, sisältöön perustuva markkinointi ja käyttäytymiseen perustuva markkinointi. Suorassa markkinoinnissa jotain tuotetta markkinoidaan henkilölle hänen aiempien valintojen perusteella. Verkkokauppa voi esimerkiksi tarjota samankaltaisia kirjoja, mitä käyttäjä on aiemmin katsellut. Tämä tietenkin tarkoittaa sitä, että tietoa on kerätty aiemmista tapahtumista. Sisältöön perustuvassa markkinoinnissa käyttäjälle mainostetaan tuotteita sivuston sisällön tai aiheen perusteella. Käyttäytymiseen perustuva markkinointi taas viittaa ajan kuluessa kerättyyn ja kasattuun kokonaisuuteen. Tiedon keruu ei rajoitu vain Internetiin, vaan myös Internetin ulkopuolelta kerätty tieto on voitu yhdistää selailun, sosiaalisten median, verkko-ostosten tai vaikkapa Internet-hakujen tietoihin. Internetin ulkopuolisia tietoja voidaan kerätä esimerkiksi kuluttajien käyttämien matkapuhelimien kautta, jonka perusteella on mahdollista saada käyttäjän fyysinen sijainti selville. Tämän avulla voi siis tehdä käyttäjästä yksityiskohtaisemman profiilin markkinointia varten. (Privacy Rights Clearinghouse 2014.)

2.2 Internetin taloudellinen puoli

Parempien tai osuvampien mainosten esille tuonti on tärkeää monien toiminnan rahoittamiseen, joka palvelee myös niiden palvelujen käyttäjiä. Suosittuja Internet-sivustoja ja palveluita, kuten uutis-sivustoja, blogeja, sähköposteja ja sosiaalisten verkostojen sivustoja on lähivuosina rahoitettu kasvavassa määrin mainostuloin, eikä suoraan käyttäjiltä. Monien sivustojen toimijat ovat riippuvaisia näistä mainostuloista voidakseen kehittää ja tarjota sisältöä sivuston käyttäjille. (Beales 2010.)

Mainostulot ovat tärkeä määrittävä tekijä sivustojen sisällön laatuun ja määrään niin kauan, kuin mainostulot pysyvät tuottajien ensisijaisena rahanlähteenä. Ilman näitä tuloja, käyttäjän saama hyöty palveluista pienenee. Vuonna 2009 käyttäytymiseen perustuva kohdennettu mainonta tuotti 2.68 kertaa enemmän tuloa, kuin ei kohdennettu mainonta. Kohdennettu mainonta oli myös kaksi kertaa tehokkaampi saamaan mainosten katsojasta asiakkaita. (Beales 2010.)

Moni verkkosivusto ei olisi pystyssä ilman Internet-markkinointia. Esimerkiksi viisi suurinta Yhdysvaltalaista sivustoa, kuten Google, Facebook, Yahoo, Youtube ja Amazon.com käyttää Internet-markkinoinnin tuloja rahoittaakseen toimintaansa. (Dagar, Endo, Gupta, Li, Pabla, Ramaswamy, Sidhu 2013). Nämä sivustot ovat luultavasti myös monelle suomalaiselle tuttuja. Interactive Advertising Bureau (2014) mukaan pelkästään Yhdysvaltojen Internet-markkinoinnin liikevaihto oli 23.1 miljardia Yhdysvaltain dollaria vuoden 2014 ensimmäisiltä kuudelta kuukaudelta. Tämä on 15.1% enemmän kuin vuonna 2013 samalta ajanjaksolta tarkastettuna. Kasvua ei ole tapahtunut pelkästään vuoden tai parin aikana, vaan kasvua on tapahtunut jo pidemmän aikaa. Kuvassa 1 on esitetty Internet-markkinoinnin kasvu 29 vuoden ajalta.



Kuva 1. Internet-markkinoinoinnin kasvu (Lähde: Interactive Advertising Bureau, IAB)

Kuviossa on nähtävissä, että kasvun trendi on kovassa kasvussa. Mainostulojen trendin onkin hyvä olla kasvavaa. Beales (2010) mainitsee mainostulojen olevan tärkeä määrittävä tekijä sivustojen sisällön laatuun ja määrään niin kauan, kuin mainostulot pysyvät tuottajien ensisijaisena rahanlähteenä. Dagar ym. (2013) toteavat, että yksi tapa lisätä Internet-markkinoinnin arvoa, on tarjota osuvampia mainoksia.

3 SEURANTATEKNIIKAT JA NIIDEN OSAPUOLET

Tässä luvussa selvennetään työn myöhemmässä vaiheessa esiintyviä tekniikoita ja seurannan osapuolia hieman tarkemmin. Luvun tarkoitus on antaa lukijalle perustietoa seurantaan liittyvien tekniikoiden toimintaan, joka helpottaa havainnollistamaan tietojen tallennusta ja liikkumista Internetissä.

3.1 Evästeet

Käyttäjän selatessa Internetissä, moni verkkosivu tallentaa vierailuista erilaisia tietoja. Tietoja tallennetaan mm. evästeiden avulla. Tässä verkkopalvelu lähettää tietoa käyttäjän selaimelle ja tieto tallentuu käyttäjän koneelle evästeeksi. Verkkopalvelu voi myöhemmin pyytää tätä tietoa takaisin käyttäjältä. Evästeet voivat sisältää erilaista tietoa, kuten verkkosivustojen kirjautumistietoa, sivustoasetuksia tai tietoa ostoskorien sisällöstä. (Privacy Rights Clearinghouse 2014.)

Tavallisia evästeitä on kahdenlaisia. Ensimmäinen on istuntokohtainen eväste (session cookie), joka tallentuu käyttäjän keskusmuistiin. Toinen on pysyvä eväste (persistent cookie), joka tallentuu käyttäjän kiintolevylle. Evästeelle voidaan asettaa voimassaoloaika. Jos voimassaoloaika on pidempi, kuin käyttäjän oleskelu sivustolla, eväste tallennetaan käyttäjän kiintolevylle. (Helopuro, Perttula ja Ristola 2009: 231-232.)

Helopuro ym. (2009: 232-233) mainitsevat, että evästeen voi lukea vain sen lähettänyt verkkopalvelu tai palvelin. Tämä ei kuitenkaan takaa, että evästeen asettaja olisi aina sivusto, jolla vieraillaan. Heidän mukaan vierailtavan sivuston mainokset voivat olla peräisin toiselta palvelimelta. Kun mainos latautuu toiselta palvelimelta, voi evästeen asettajana toimia tämä toinen palvelin. Evästeiden päätarkoituksena on palvelujen käytön helpottaminen tai mainonnan tehostaminen.

Käyttäjän on kuitenkin mahdollista muuttaa verkkoselaimensa asetuksia ja estää evästeiden käytön. Tästä huolimatta monet verkkosivustot vaativat toimiakseen, että evästeet ovat käytössä. Tallennetut evästeet on mahdollista poistaa selaimen omalla poistotoiminnolla tai erillisellä ohjelmalla. Evästeitä käyttävät sivustot pyrkivät asettamaan uuden evästeen, jos aiempaa evästettä ei löydy. (Helopuro ym. 2009: 235.)

3.2 Flash -eväste

Local shared objects eli Flash -eväste tuli Adobe Flash Player 6 -version mukana. Flash -eväste toimii samalla tavalla kuin tavallinen eväste, mutta pystyy sisältämään tavallista evästettä monimutkaisempaa tietoa, eikä vain pelkkää tekstiä. (Adobe Systems Incorporated 2014.) Flash -evästeet pystyvät sisältämään 100 kilobittiä tietoa, kun tavallinen eväste pystyy sisältämään vain 4 kilobittiä. Tavallinen eväste myös vanhenee ilman erillistä määrittelyä, kun käyttäjä sulkee verkkoselaimen, mutta Flash -eväste ei vakioasetuksilla vanhene ollenkaan. (Soltani ym. 2009).

Flash -eväste on myös tavallista evästettä uudempi ja sitä käyttävät monet verkkosivut. Sitä on vaikeampi poistaa ja tavalliset poistotoiminnot, kuten selaushistorian poistaminen, välimuistin tyhjentäminen tai selaimen henkilökohtaisten tietojen poistotoiminto ei poista Flash -evästettä. Flash -evästeiden poistaminen vaatii erillisen sovelluksen ja tällaisia sovelluksia on saatavilla esimerkiksi Mozilla Firefox selaimeen lisäosana. (Privacy Rights Clearinghouse 2014.) Flash -evästeessä olevia tietoja pystyy hyödyntämään myös laitteen jokaisessa selaimessa, joka ei onnistu tavallisen HTTP -evästeen kanssa. Muitakin Flash -evästeen kaltaisia tekniikoita on olemassa, kuten Microsoftin Silverlight tekniikka, joka kilpailee Adobe Flashin kanssa. (Cranor ym. 2013).

3.3 Supercookie ja Evercookie

Evästeitä käytetään kaikkialla, mutta ne on mahdollista poistaa. Seuranta teknologioiden kehitys on tuonut evästeet, joita sanotaan superevästeiksi (supercookie). Yksi tällainen eväste on esimerkiksi aiemmin esitelty Flash -eväste. Selaimet eivät tavallisesti tarjoa käyttöliittymää näiden evästeiden tarkasteluun tai poistamiseen, sillä selaimet eivät hallitse näitä evästeitä. (Eurpean Network and Information Security Agency, ENISA 2012.)

HTML5-merkintäkieli ja sen tuomat ominaisuudet tarjoaa kattavammat seurantamahdollisuudet, kuin aiemmat versiot. Merkintäkielen avulla on mahdollista kerätä ja tallentaa erittäin suuria määriä tietoa käyttäjistä. Kalifornialainen ohjelmoija Samy Kamkar on tehnyt uudenlaisen evästeen, jota hän kutsuu nimellä evercookie. Jopa alan ammattilaisten on hankala poistaa tätä evästettä ja eväste on saanut eri blogeissa kommentteja kuten "erittäin vaikea poistaa" ja "kauhistuttava". Tämä eväste pystyy hyödyntämään perinteisiä seurantateknologioita uuden merkintäkielen rinnalla ja tallentaa tietoa ainakin 10 eri paikkaan käyttäjän koneella. (Tanzina 2010.)

Ennen HTML5 merkintäkieltä tuli verkkosivustojen tiedot tallentaa evästeisiin. HTML5 version tuoman paikallisen tallennuksen avulla tietoa pystyy tallentamaan käyttäjän selaimeen ainakin viiden megatavun verran. Tämä on huomattavasti enemmän, mitä tavallinen eväste pystyy sisältämään. (W3schools.com 2015.)

3.4 Laitteen sormenjälki

Laitteen sormenjälki (device fingerprint) on käyttäjän asetusten ja ohjelmien kokonaisuudesta tehty yhteenveto. Eri laitteilla on tunnusomaisia piirteitä tai asetuksia. Erilaiset asetukset kellolle, eri fontit ja ohjelmat tekevät laitteesta ainutlaatuisen. Verkossa on mahdollista tunnistaa näitä ainutlaatuisia piirteitä ja muodostaa niistä yhteinen kokonai-

suus. Tälle yhteenvedolle voidaan asettaa tunnistenumero, jonka avulla kyseisen laitteen voi erottaa muista laitteista verkossa. (Privacy Rights Clearinghouse 2014.)

3.5 Kolmannen osapuolen palvelut

Mayer ym. (2012) mainitsevat, että kolmannen osapuolen verkkosivustoilla on karkeasti kuusi erilaista liiketoimintamallia. Niitä ovat markkinointiyritykset, analyyttiset palvelut, sosiaaliset verkostot, sisällöntuottajat, käyttöliittymäpalvelut ja verkkopalvelun tarjoajat. Alempana on esitelty Mayer ym. (2012) esittämät liiketoimintamallit hieman tarkemmin.

3.5.1 Markkinointiyritykset

Markkinointiyrityksillä tai tarkemmin sanottuna Internet-markkinoinnissa on kolme johtavaa toimintamallia: suoraosto (direct buy), mainosverkosto (advertising network) ja mainosvaihtokauppa (advertising exchanges).

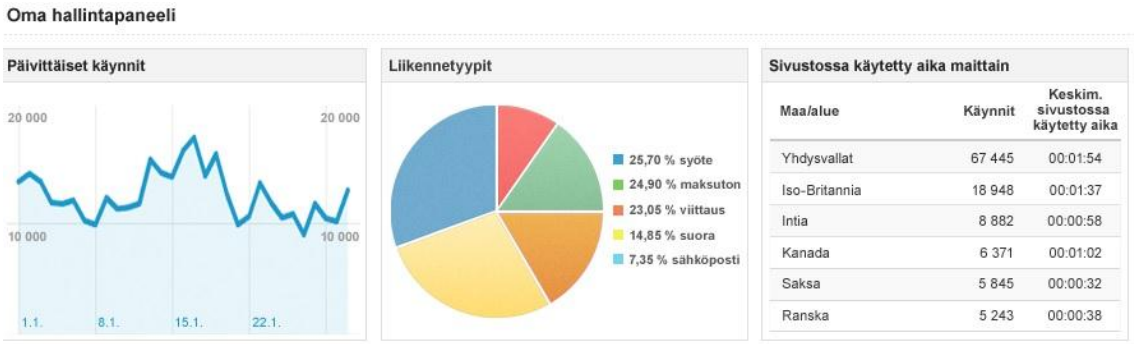
Suoraosto on näistä kolmesta vanhin toimintamalli. Suoraostossa markkinoijat tekevät mainossopimukset suoraan verkkosivuston ylläpitäjien tai sisällönjulkaisijoiden kanssa. Kasvavan mainonnan takia mainosten sopimisesta tuli epäkäytännöllistä. Tätä asiaa helpotti 90 -luvun lopussa tullut **mainosverkosto**, jonka avulla markkinoijien oli helpompaa asettaa mainoksia julkaisijoiden saataville ja julkaisijoiden oli taas helpompaa hyödyntää useampien markkinoijien mainostarjontaa. Mainosverkostossa näiden kahden osapuolen ei tarvinnut olla suoraan yhteydessä toisiinsa.

Verkosto mahdollisti mainoksien kohdentamisen. Mainoksien kohdentaminen tietyille käyttäjille onnistui esimerkiksi sijainnin, sivun sisällön tai selailijan selaushistorian perusteella. Mainokset eivät kuitenkaan aina menneet kaupaksi ja 2000 -luvun puolella esiintyi **mainosvaihtokauppa**, jossa jäljelle jääneet mainokset myytiin tarjousten perusteella. Sen avulla saatiin tehtyä rahaa mainoksilla, jotka eivät aiemmin menneet kaupak-

si. Etuna tässä oli reaaliaikainen kaupankäynti useammassa mainosverkostossa. Tämä menetelmä synnytti myös erilaisia toimijoita kuten tiedon tarjoajia. Tiedon tarjoajat tarjosivat markkinoijille käyttäjätietoja mainoksien kohdentamista varten. (Mayer ym. 2012.)

3.5.2 Analyyttiset palvelut

Analyyttiset palvelut tarjoavat työkaluja verkkosivustojen ylläpitäjille, joiden avulla on mahdollista tietää enemmän sivuston vierailijoista. Näiden avulla on mahdollista tietää esimerkiksi käyttäjien maantieteellinen sijainti, käyttöliittymätietoja, mitä sisältöjä käyttäjä on katsonut ja kuinka hän on sivustoa käyttänyt. Kuvassa 2 on nähtävissä esimerkki analyyttisen työkalun tiedoista.



Kuva 2. Esimerkki analyyttisen työkalun tiedoista (Lähde: Google.com 2015b)

Yllä esitetty kuva havainnollistaa minkälaista tietoa on mahdollista saada sivuston käytöstä analyyttisten työkalujen avulla. Analyyttisia palveluita on saatavilla maksullisina että maksuttomana. Maksuttomat palvelut saavat kuitenkin rahallista korvausta keräämistään tiedoista mm. käyttämällä tietoa kohdemarkkinointiin tai avustamaan ymmärtämään kohdemarkkinoita paremmin. (Mayer ym. 2012.)

3.5.3 Sosiaalisten palveluiden integrointi

Sosiaalisten palveluiden tarjoamien piensovellusten hyödyntäminen verkkosivustolla mahdollistaa sisällön personoinnin ja yhtenäisen kirjautumistavan sosiaalisten medioiden käyttäjille. Tunnetuimpia sosiaalisten medioiden pienois-sovelluksia ovat Facebookin tykkää -nappi ja kommentointi lisäosa, Googlen +1 -nappi ja Twitterin oma piensovellus. Kuvassa 3 on esitetty esimerkkejä sosiaalisten medioiden piensovelluksista.

Kuva 3. Sosiaalisten medioiden piensovellukset (Lähde: Facebook.com 2015; Google.com 2015a; Twitter.com 2015.)

Sosiaaliset mediat ovat myös synnyttäneet useita erilaisia toimijoita, jotka avustavat verkkosivuston ylläpitäjiä ottamaan sosiaalisten medioiden sovelluksia käyttöön. Nämä palvelut rahoittavat toimintansa keräämällä käyttäjistä tietoa ja myyvät nämä tiedot taas eteenpäin. (Mayer ym. 2012.)

3.5.4 Käyttöliittymäpalvelut

JavaScript koodauskielellä tehtyjä sovelluksia käytetään verkkopalvelun toimivuuden parantamiseen tai lisäämään verkkosivuston toiminnollisuuksia. Näitä JavaScript sovelluksia on saatavilla esimerkiksi Googlen sovellus kirjastosta. (Mayer ym. 2012.)

JavaScript on ohjelmointikieli, jota yleensä käytetään selaimissa tai sen sisällössä. Sen avulla voidaan lisätä vuorovaikutteisuutta, hallita verkkoselaimen käyttäytymistä tai esimerkiksi muuttaa sisältöä, joka näkyy verkkoselaimessa. (Flanagan 2006: 1.)

3.5.5 Julkaisualustat

Jotkin ulkopuoliset palvelut tarjoavat julkaisualustoja, jotka helpottavat sisällöntuottamista tai sisällön jakamista. Esimerkkinä tällaisesta on blogin hallintaan tarkoitettu Wordpress järjestelmä ja sisällön jakeluverkosto (content distribution network) Akamai. (Mayer ym. 2012.)

Wordpress on tällä hetkellä suosituin julkaisualusta Internetissä. Yli 20% verkkosivuista käyttää Wordpress alustaa. Aluksi Wordpress kehitetiin blogien hallintaan, mutta nykyään siitä on tullut laaja sisällönhallintajärjestelmä, johon on saatavilla tuhansia erilaisia lisäosia. (Wordpress 2015a; Wordpress 2015b.)

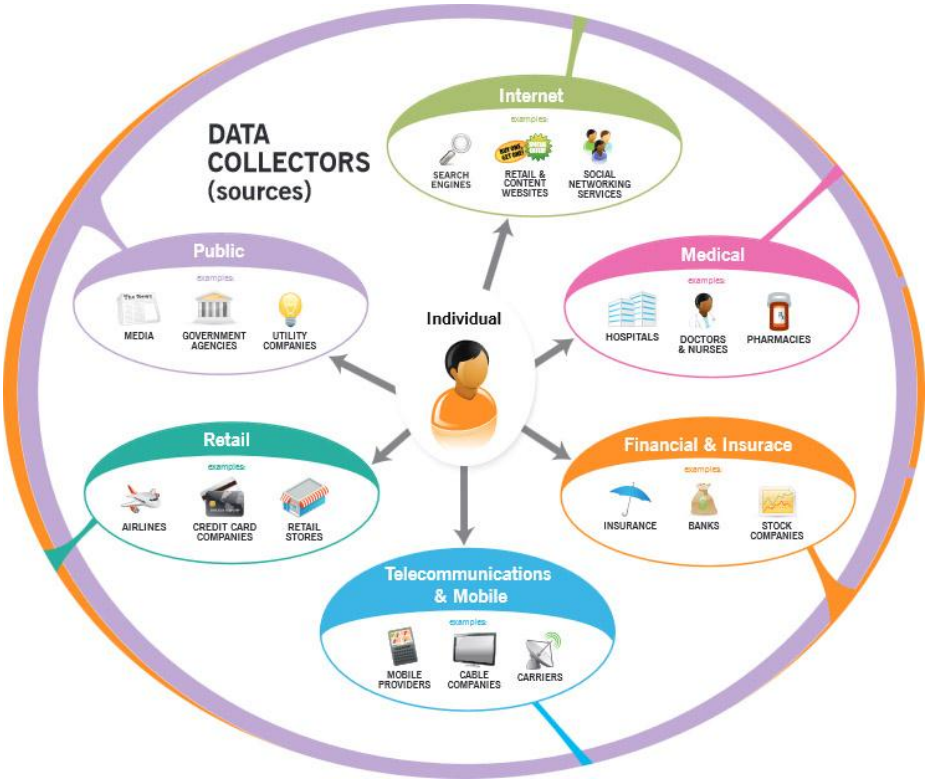
4 KÄYTTÄJIEN SEURANTA

Aiemmin Internet-sivustojen sisältö toteutettiin ja ylläpidettiin yhden henkilön tai toimijan avulla. Nykyään sisältöä tulee myös kolmansilta osapuolilta, kuten mainoksista tai markkinointiyrityksiltä, analyyttisistä työkaluista, sosiaalista verkoista jne. Ulkopuolisten osapuolien palveluissa on riskinsä, mikä on kiinnittänyt tutkijoiden, yhteiskuntaorganisaatioiden ja yhteiskuntapäättäjien huomion. Huomio kohdistuu Internet-selailijan selailutietojen seurantaan. (Mayer ym. 2012.)

4.1 Ketkä keräävät tietoa ja miksi

Tiedonkerääjät ovat tavallisesti niitä toimijoita, jotka keräävät tietoa jonkin palvelun käyttäjästä. Tiedon välittäjät taas saavat tietonsa tiedon kerääjiltä. Tiedon välittäjät analysoivat tietoa ja tekevät siitä hyödynnettävän kokonaisuuden, jonka jälkeen tarjoavat sen tiedon käyttäjille. Katso Kuva 4. Tiedonkerääjät.

Tiedon välittäjiä ovat esimerkiksi verkkosivustot, mainosverkostot, analytiikkapalvelut ja luottotietopalvelut. Viimeisenä tulevat tiedon käyttäjät, joita ovat markkinoijat, mediat, pankit, työnantajat, yksityiset, lainvalvojat, valtiot, juristit ja postituspalvelut. He ostavat tiedonvälittäjiltä kasatut tiedot ja käyttävät niitä omassa liiketoiminnassaan tai palvelussaan hyödyksi. Kuvassa 4 on nähtävissä, että tiedonkeräys ei rajoitu pelkästään Internetiin. (Dagar ym. 2013.)



Kuva 4. Tiedonkerääjät. (Lähde: Federal Trade Comission, FTC 2009)

Taulukossa 1 on listattuna kuluttajatiedon käyttökohteita, joissa tieto on henkilökohtaista tai kootussa muodossa. Lueteltuna ovat esimerkiksi rahoituspalvelut, luoton myöntäminen, vakuutuksen myöntäminen, tarjouskupongit jne. Tiedolle on paljon käyttötarkoitusta ja taulukkoa katsomalla voi todeta tiedolla olevan tärkeä rooli yhteiskunnassa. Ongelmana on kuitenkin, jos tieto päätyy vääriin käsiin tai kun tieto ymmärretään väärin.

Esimerkkejä tunnistettujen tai yhdistettyjen henkilötietojen käytöstä	
Rahoituspalvelut kuten pankit tai sijoitustilit.	Luottopalvelut kuten luotto- tai pankkikortit, asuntolainat, auto- tai käyttöluotot, autovuokraukset tai puhelinpalvelut.
Vakuutuksen myöntäminen kuten terveyteen, autoihin ja elämään liittyvät vakuutukset.	Vähittäismyyntikupongit ja erikoistarjoukset.
Kuvastojen ja lehtien yhteydenotot.	Verkko- ja mobiilipalvelut kuten sisältö-, e-mail- ja hakupalvelut sekä sosiaaliset verkostot.
Tuotteiden ja palveluiden toimitukset kuten video-, kaapeli-, ja pakkauslähetykset.	Asianajotoimistot esimerkiksi tutkimuksissa.
Journalismi esimerkiksi totuuden tarkastuksissa.	Sähköisessä-, suorassa-, ja puhelinmarkkinoinnissa.
Työpaikan tai isännöitsijän taustatutkimukset.	Tietojen välittäjät voivat käyttää tietoa niiden yhdistämiseen tai myyntiin.
Henkilöiden etsintään esimerkiksi katoamistapauksissa.	Lainvalvonnassa.
Tutkimuksissa kuten terveys- ja rahoitusalalla.	Petosten etsinnöissä tai estämisessä.
Valtioiden palveluissa.	

Taulukko 1. Tiedon käytön esimerkit. (Lähde: FTC 2009)

Canadian Internet Policy and Public Intrest Clinic, Cippic (2008) jakaa syyt miksi käyttäjien henkilötietoja yritetään kerätä. He jakavat motivaatiotekijät kolmeen laajaan kategoriaan, joita ovat petos, henkilökohtaiset hyökkäykset ja kaupallisuus. Petos on suurin uhka Internet-käyttäjän yksityisyyteen. Henkilökohtaisilla tiedoilla, kuten taloudellisilla tiedoilla, henkilöllisyystodistuksilla, salasanoilla ja tilaustiedoilla on kaikilla arvoa. Huijauksen ollessa hyökkääjän motivaation tekijänä, tekijä yleensä tietää tekohetkellä tekevänsä eettisesti väärin ja tietää luultavasti rikkovansa lakia.

Petoksen tekijät ovat kiinnostuneet henkilökohtaisista tiedoista juuri niiden arvon takia pimeillä markkinoilla. Verkkohuijaajien pimeästä kaupasta on tullut hienostunut ja tuotteistettu. Laiton kaupankäynti on helpompaa kuin koskaan. Ei tarvitse olla enää erikoinen tietokoneenkäyttäjä löytääkseen kaupankäyntipaikkoja tai hankkia teknologiaa aloittaakseen oman identiteettihuijausoperaation. Varastettuja luottokortteja voi ostaa suoraan netistä, kunhan tietää minne sivustolle mennä. Näiden sivustojen löytäminen on nykyään myös helppoa.

Henkilökohtaiset syyt kuten pariskuntien eroaminen tai työtilanne voi olla henkilökohtaisten tietoon kohdistuvan hyökkäyksen syynä. Nämä hyökkäykset harvoin kohdistuu henkilön taloudellisiin varoihin kuten petoksissa. Henkilökohtaisissa hyökkäyksenä tekijällä saattaa olla tavallisemmat syyt kuten kosto, kateus tai viha. Kolmantena syynä henkilötietojen keräämiseen on kaupallisuus. Internet-markkinointi on suuri bisnes. Asiakkaiden tiedot ovat arvokkaita yrityksille.

Uudet teknologiat sallivat henkilötietojen keräämisen odottamattomilla tavoilla ja saattavat välillä liikkua lain harmaalla alueella. Kaupallisessa tarkoituksessa kerätty tieto on houkutteleva kohde verkkohuijaajalle, joka tuo tietovuodot uhkaksi henkilötiedoille. Verkkosivuston keräämät tiedot onkin yhtä turvattu, kuin palveluntarjoajan verkkopalvelu. Verkkosivustot voivat olla haavoittuvia tietoturvaloukkauksille ja ulkopuolisten palveluiden hyväksikäyttämiseen. (CIPPIC 2008.)

Eurpean Network and Information Security Agency, ENISA (2012) listaa motivaatiotekijät taas seuraavanlaisesti. He sanovat verkkosivustojen ylläpitäjien seuraavan käyttäjiä tarjotakseen personoituja palveluita käyttäjälle, kuten säilyttävän käyttäjien verkkokaupan ostoskoria tai suosikkeja. Ensimmäisen osapuolen seurantaa käytetään myös petosten havaitsemiseen ja lain valvontaan. Useat säädökset vaativat verkkosivustoja tallentamaan käyttäjätietoa petosten estämiseen, rahanpesua vastaan, kansallisen turvallisuuden ja lainvalvonnan vuoksi. Kolmannen osapuolen syyksi ENISA (2012) mainitsee kaksi pääsyytä, joita ovat profilointi sekä analytiikka ja mittaukset.

Käyttäytymiseen perustuvan tiedonkeruun tarkoituksena on seurata käyttäjän tekemisiä pidemmällä aikavälillä ja luoda profiili käyttäjän kiinnostusten kohteista, ominaisuuksista kuten iästä ja sukupuolesta sekä ostostottumuksista. Verkostomarkkinoijat käyttävät käyttäytymiseen perustuvaa kohdennettua markkinointia näyttääkseen mainoksia, jotka heijastaa käyttäjän mielenkiinnon kohteita. Verkkokaupat käyttävät myös tätä seurantaa ja profilointia suosittelemaan käyttäjälle kiinnostavia tuotteita esimerkiksi perustuen käyttäjän aiemmin ostamien tuotteiden, samankaltaisten käyttäjien ostamien tuotteiden ja suosikkituotteiden perusteella. Verkkoanalytiikkaa ja mittaamista taas käytetään saadakseen tietoa verkkosivuston liikehdinnästä ja kuinka tehokasta mainonta on ollut. Verkkosivustojen ylläpitäjien on mahdollista itse kerätä ja mitata näitä tietoja, mutta useat käyttävät kolmannen osapuolen palveluita analyyttisena työkaluna. (ENISA 2012.)

Tiedonkerääjä ei aina ole jokin yritys. Järvinen (2014: 11) mainitsee, että uutena tekijänä digitaalisen elämän seurantaan ja tallentamiseen on tullut eri maiden tiedustelupalvelut ja viranomaiset. Digitalisoituminen on heille vain vanhan toiminnan jatkamista uusin keinoin. Yhdysvaltojen urkinta on saanut eniten julkisuutta, mutta urkintaa tekevät muutkin organisaatiot. Julkisuutta on tullut mm. Edvard Joseph Snowdenin tekemistä julkaisuista medioille. Järvinen (2014: 21-22) antaa Snowdenin julkaisuista yhtenä esimerkkinä uutisen, jossa kerrottiin, että kansalaisten puhelutietojen oli luovutettu Yhdysvaltain tiedusteluorganisaatiolle NSA:lle (National Security Agency). Hän myös mainitsee, että Senaatin tiedustelukomitean demokraattipuheenjohtaja Dianne Feinstein sanoi, ettei puheluita kuunneltu, vaan kyse oli teletunnistetiedoista. Julkaisu nostatti keskusteluja Yhdysvalloissa. Yhdysvalloissa teletunnistetietoja ei kuitenkaan lasketa viestintäsalaisuuden piiriin kuten Suomessa.

Yhdysvalloilla on Suomeen nähden etu kansainväliselle seuraamiselle. Yhdysvaltojen lävitse kulkevaa viestintää voi seurata ja sitä on mahdollista poimia maan omista runkoverkoista. Yhdysvalloissa sijaitsee lähes kaikki suosituimmat nettipalvelut, kuten Facebook, Twitter ja Google, joka on etu heidän tekemälle seurannalle. Mm. evästeet toimivat sen lähettäneen palvelun sisällä, jonka takia yksin evästeen avulla ei voi seurata käyttäjien liikkumista palvelusta toiseen. Runkoverkkojen kautta kaikkien evästeet on

mahdollista nähdä ja tämä mahdollistaa käyttäjien liikkeiden seuraamista palvelusta toiseen. (Järvinen 2014: 66, 68-69.)

Järvinen (2014: 174-175) mainitsee, että kansainvälisen urkinnan kiinnostavuuteen vaikuttaa suuresti yksittäisen henkilön ammatti. Merkittävien yritysten johtajat, poliitikot ja valtion virkamiehet ovat kiinnostavia kohteita. Muita maita kiinnostaa mm. miten eri päätökset syntyvät ja miten niihin voi vaikuttaa. Järvinen toteaa, että viestinnässä on aina kaksi osapuolta. Henkilö saattaa viestiä kiinnostavan henkilön kanssa, jolloin molempien viesteistä tulee kiinnostavia. Hän mainitsee kolmen loikan periaatteen, joka voi johtaa kenet vain urkinnan kohteeksi. Ystävillä voi olla salattu elämä, mistä ei aina tiedä mitään. Toiseksi ystävien ystävät muodostavat jo niin laajan joukon, että mukaan mahtuu kaikenlaista. Jos huomioidaan vielä näiden ystävät, niin voi olla takuuvarma jonkun olevan tiedustelun urkinnassa.

4.2 Miten käyttäjää seurataan ja mitä ongelmia tähän liittyy

Petteri Järvinen kirjoitti jo vuonna 2002 evästeistä julkaisemassaan kirjassa. Hän kertoo evästeiden olevan surffauksen paljastajia ja pystyvän välittämään kaikenlaista tietoa Internet-selailijasta mainostajille. Hän myös mainitsee kolmannen osapuolen palvelut ja niiden mahdollisuudet seurata käyttäjien vierailuja eri sivustoilla, jos kaikilla vierailtavilla sivustoilla on sama mainoksien tarjoaja ja näin ollen myös evästeiden asettaja. Silloin mainostaja pystyy yhdistämään sivustot ja evästeet yhteen käyttäjään. Myös käyttäjän henkilöllisyys paljastuu, jos sivusto vaatii rekisteröintiä ja luovuttaa nämä tiedot mainostajalle. (Järvinen 2002: 152,156.) Evästeiden käyttö ei ole loppunut vuosien kuluessa ja muitakin tekniikoita on kehitetty, kuten seuraavaksi tullaan näkemään.

4.2.1 Seurantateknologioiden käyttö

Tavalliset verkkosivustot käyttävät evästeitä apuna tarjoamaan mainoksia käyttäjälle, joka vierailee uudelleen heidän sivustolla. He käyttävät evästeitä myös seuraamaan mainonnan tehokkuutta. Näitä evästeitä kutsutaan ensimmäisen osapuolen evästeiksi. Kolmannen osapuolen evästeistä puhutaan, kun evästeen asettajana toimii ulkopuolinen sivusto tai palvelu. Privacy Rights Clearinghousen (2014.)

Digitaaliset palvelut ja laitteet ovat kasvavassa määrin mukana ihmisten sosiaalisissa kanssakäymisissä, viihteessä, ostoksissa ja tiedonkeruussa. Nämä tiedot ovat helposti tallennettavissa ja niitä pystyy sitten helposti analysoimaan. Tämä on auttanut laskennallisten yhteiskuntatieteiden, uusien palveluiden kuten personoitujen hakukoneiden, suositus palveluiden sekä kohdistetun markkinoinnin kehitystä. Tämä laaja tiedon määrä ja halu tietää käyttäjistä enemmän tuottaa vakavia haasteita käyttäjien yksityisyyteen ja siihen, kuka tiedon omistaa. Saatavilla olevien tietojen avulla voidaan tehdä tilastollisia havaintoja ja ennustaa käyttäjistä tiettyjä asioita (Kosinski, Stillwell & Graepel 2013.)

Yhteisten säädöksien määrittämiseen on olemassa sääntelijöitä. Yhteensä 8 eurooppalaista sääntelijää, joiden vastuulla on evästeiden sääntelyjen määrittäminen, tarkasteli evästeiden käyttöä eri verkkokaupoissa, mediassa ja julkisella sektorilla. Euroopan sääntelijöiden yhteistyössä tekemä tutkimus 478 Euroopan kansalaisten käyttämistä sivustoista osoittaa, että verkkosivuston ylläpitäjät tiedottavat käyttäjiä evästeiden käytöstä, mutta niitä käytetään silti ilman käyttäjien suostumusta. Evästeitä asetetaan suuria määriä, niiden voimassaolo on turhan pitkä ja löytyy parantamisen varaa niiden käytön tiedottamisessa ja suostumusten pyynnössä. Tutkimus paljasti, että yhteensä asetettiin yli 16000 evästettä. 70 % evästeistä oli kolmansien osapuolien asettamia ja yli puolet evästeistä tuli vain 25:ltä verkkotunnukselta. Keskiarvoisesti eväste oli voimassa 1-2 vuotta. Kuitenkin 20 %:lla evästeistä oli voimassaoloaika kahden ja viiden vuoden välillä ja 374 evästeellä oli voimassaoloaika yli 10 vuotta. Kolmella evästeellä oli voimassaoloaika melkein 8000 vuotta. Näistä sivuista 26 % ei antanut mitään tiedotetta evästeiden käytöstä. (European Comission 2015.)

Selaushistoria voi paljastaa selailijasta monenlaista tietoa. Se voi paljastaa esimerkiksi mistä käyttäjä on kiinnostunut, tietoa ostoksista, terveydentilasta, taloudellisesta ja työllisyystilanteesta. (Mayer ym. 2012.) Krishnamurthy, Naryshkin ja Wills (2011) tekemän tutkimuksessa suurin osa suosituimmista sivustoista suoraan vuoti arkaluonteista ja tunnistettavissa olevia käyttäjätietoja kolmansille osapuolille. Mukana oli mm. kymmenen suosituinta terveydenhuolto ja lennonvaraus -sivustoa, joista yhdeksän vuoti tietoa käyttäjän tekemistä hakusanoista. He valitsivat sivustot tutkimukseensa alexa.com sivuston tarjoamilta listoilta. Kriteereinä heillä oli, että sivustolla on vähintään 100 000 rekisteröityä käyttäjää. He ottivat kuitenkin mukaan sivustoja joiden rekisteröityjen käyttäjien määrä ei ollut tiedossa, jos sivusto oli suosittu ja muilla saman kategorian sivustoilla oli suuria käyttäjämääriä. Mukana oli sivustoja seuraavista kategorioista: taide, työllistyminen, videopeliuutiset, valokuvien jakaminen, uutiset, matkailu, shoppailu, suhteet, sukupolvet ja ikäryhmät, urheilu, sosiaaliset verkostot ja terveys. Mukaan tuli 120 sivustoa. Näistä sivustoista 67 eli 56 % vuoti yksityistä tietoa kolmansille osapuolille.

Verkkosivustolta voi löytyä tietosuojalausunto, jossa sivuston ylläpitäjät kertovat kuinka sivusto kerää tai käyttää vierailijoista kerättyä tietoa. Tämä ei kuitenkaan kata aivan kaikkea tietoa, jota sivustolla mahdollisesti kerätään. Privacy Rights Clearinghouse (2014) mainitsee web bugit tai "webbiötökät", joita käytetään seuraamaan sivustolla vierailijaa. Web bugi voi olla jokin graafinen elementti tai kuva sivustolla ja sen avulla voi saada tietoon esimerkiksi mistä IP -osoitteesta on tietty sivu luettu. Sen asettajana voi olla ulkopuolinen palvelu, jonka takia verkkosivuston tietoturvalausunto ei koske tämän tekniikan keräämää tietoa.

Myös Dagar ym. (2013) mainitsevat, että moni markkinointiyritys sisällyttää evästeen mainoksen yhteyteen, kuten edellä mainitun kuvan yhteyteen. Tämä mahdollistaa mainosyrityksen seurata käyttäjää jokaisella sivustolla, missä heidän mainos on esillä. Sivuston ylläpitäjä ei myöskään välttämättä tällaisessa tilanteessa tiedä, että heidän sivuston käyttäjiä seurataan.

Välityspalvelimia on ollut tarjolla jo 90 –luvun alkupuoliskolta. Tämä mahdollistaa IP-osoitteen piiloutumisen. Vierailu esimerkiksi verkkosivustolla näyttäisi siltä, kuin vie-

railija olisi välityspalvelimen sijainnissa. Välityspalvelimet siis kierrättävät yhteyden eri palvelimen kautta, joka saa yhteyden näyttämään tulevan toisesta paikkaa, kuin se todellisuudessa tulee. Valitettavasti tämä keino ei kokonaan piilota käyttäjää seurannalta tai estä seurantateknologioita kuten evästeitä tai palveluita, johon kirjaudutaan erikseen. (Dagar ym. 2013.)

Tor on Internetin sisällä toimiva verkko. Tor -verkossa liikkuva tieto on salattua tiedon siirron aikana. Käyttäjän kirjoittaessa selaimeen verkkosivuston osoitteen, Tor ohjelma arpoo kolme verkossa olevaa konetta, jonka kautta lähetettävä tieto kiertää. Näitä välipisteitä kutsutaan solmuiksi. Solmukohtia on mahdollista lisätä, mutta tämä hidastaisi liikennettä. Solmut eivät näe paketin sisällä olevaa tietoa, eivätkä sen lähtö- tai kohdeosoitetta. Jokaisessa solmukohdassa lähetettävästä tiedosta puretaan yksi salauskerros, joita on kolme. Yhden salauskierron purkaminen paljastaa aina seuraavan määränpään osoitteen. Viimeinen kone purkaa viimeisen salauskerroksen ja näkee, mihin osoitteeseen sivupyyntö on osoitettu. Sivuston palauttamat tiedot välittyvät taas täysin käänteisesti. Vasta solmujen viimeinen kone näkee varsinaisen liikenteen, mutta ei pysty selvittämään käyttäjän tai ensimmäisen koneen IP -osoitetta. Näin Tor piilottaa viestinnän osapuolet, joka tekee yhteyden seuraamisesta ja kontaktien analysoinnista mahdotonta. Tor-käyttäjän henkilöllisyys voi kuitenkin paljastua selaimissa tai verkkosivustoissa olevien toimintojen kautta. Esimerkiksi evästeet, javascript-ohjelmat tai Googlen mainokset voivat olla vaarallisia, jos samoja palveluita käytetään myös ilman Tor-verkkoa. (Järvinen 2014.)

Cybersecurity & Privacy foundation (2013) mainitsee, että Tor palvelun tarjoamaan suojaan vaikuttaa käyttäjän omat tekemiset. Suojattu yhteys ei auta, jos käyttäjä luovuttaa vapaaehtoisesti tietoa omasta identiteetistään. Esimerkiksi kuvallisen henkilöllisyystodistuksen lataaminen Internetiin paljastaa henkilöllisyyden suojatusta selauksesta huolimatta.

Internet-selailun lisäksi tiedonkeruuta voi tapahtua mobiililaitteissa. Mobiililaitteille on erilaisia sovelluksia, jotka voivat kerätä erilaisia tietoja. Sovellusten käyttäminen tai asentaminen yleensä vaatii toimiakseen, että käyttäjän on hyväksynyt eri tietojen käyt-

tämisen. Näitä tietoja voi olla esimerkiksi puhelimeen tallennetut kontaktit, puhelinlokit, kalenteritiedot, Internetin käytön tiedot, laitteen tunnistenumero, sijaintitiedot tai vaikka kuinka sovellusta on käytetty. Jotkin sovellukset tarvitsevat toimiakseen näitä tietoja. On kuitenkin sovelluksia, jotka hyödyntävät näitä tietoja, vaikka niiden toiminta ei tätä tarvitse. (Privacy Rights Clearinghousen 2014.)

Lainsäätäjien ja virkamiesten huolia on nostanut käyttäytymiseen perustuva mainonta (behavioral advertising). Uudenlaisia säädöksiä kuluttajien yksityisyyden turvaamiseen on aloitettu miettimään. Pääaiheena ovat käyttäjien mahdollisuudet seurannan välttämiseen, mutta esimerkiksi Flash -evästeiden vaikutuksista yksityisyyteen on sivuutettu. (Soltani ym. 2009.) Flash -eväste toimii samalla tavalla kuin tavallinen HTTP -eväste, mutta pystyy sisältämään tavallista evästettä monimutkaisempaa tietoa, eikä vain pelkkää tekstiä, mistä on tässä työssä aiemmin mainittu. (Adobe Systems Incorporated 2014.)

Soltani ym. (2009) tekivät tutkimuksen Flash -evästeistä sadalla suositulla verkkosivustolla. Suosituimmat sivustot he ottivat QuantCastin tarjoamalta listalta. Heidän tutkimuksen tuloksena esiintyi, että Flash -eväste oli yleinen tapa erilaisten sivustoasetusten ja tietojen tallennukseen. He pitivät tätä yksityisyyden kannalta ongelmallisena, sillä Flash -evästeitä käytettiin yhdessä tavallisten HTTP -evästeiden kanssa tallentamaan tunnistetietoa käyttäjästä. Tämä mahdollistaa HTTP -evästeen varmuuskopioinnin. Näin sivusto voi luoda Flash -evästeen avulla tavallisen evästeen uudelleen, jos käyttäjä oli tämän poistanut. Yhteensä sadasta sivustosta 54 sivustolla asetettiin Flash -evästeitä, joita löytyi kaikkiaan 281 kappaletta. Kaikkien evästeiden yhteenlaskettu määrä oli 3602 kappaletta 98 sivustolla sadasta.

Soltani ym. (2009) pyrkivät tutkimuksessaan tarkastelemaan kuinka Flash -evästeitä käytettiin katsomalla evästeiden tallentamien muuttujien nimiä, jotka sisältävät evästeen keräämiä tietoja. Heidän tutkimuksessaan useimmiten esiintyvien muuttujien nimet olivat: volume, userid, user, id, lts, _tpf, _fpf, uid, perf ja computerguid. He sanovat, että äänen voimakkuuden (volume) asetuksia tallennetaan evästeihin, koska Flash -videoita on Internetissä paljon, joten tämä muuttuja ei ole yllätys. Yllättävänä he pitävät "userid-,

user- ja id-" muuttujien käyttöä, sillä niihin tallennetaan yksilön tunnisteita, joita voidaan käyttää yksilön seurantaan. Myös "_tpf- ja fpf-" muuttujiin oli tallennettu yksilöllinen tunniste, joka oli samanlainen kuin tavallisissa evästeissä käytetty muuttujan arvo "uid" tai "userid". Samojen arvojen käyttö näissä kahdessa evästeessä nostaa huomion evästeiden uudelleen luontiin, sillä Flash -evästeen avulla olisi mahdollista luoda poistettu HTTP -eväste uudelleen.

Soltani ym. (2009) myös mainitsivat, että Flash -tekniikka tallensi asetustiedostoja domainin eli verkko-osoitteen mukaan käyttäjän koneelle. Tässä tallentamisessa on samankaltaisuuksia selaimen selaushistorian kanssa, mutta nämä asetustiedostot eivät poistu selaimen selaushistorian poistotoiminnolla. Näin ollen käyttäjä voi virheellisesti kuvitella poistaneensa selainhistoriansa, vaikka asia onkin aivan toisin. Flash -eväste voi tallentua käyttäjän koneelle jo pelkästään verkkosivustolla vierailtaessa ja vaikkei käyttäjä paina esimerkiksi Flash -tekniikalla tehtyä mainosta. Sivustolle upotettu mainos voi olla ulkopuolisen palvelun omistama, jolloin evästeen asettajakin on tämä ulkopuolinen palvelu.

Cheng (2010) mainitsee Evercookien tallentavan kahdeksaan eri paikkaan tietoa käyttäjän koneelle käyttäjätunnisteesta ja evästiedoista. Hän kertoo lukumäärän olevan vielä kasvussa. Evercookiessa tallennuksessa on mukana mm. tavallisia HTTP -evästeitä, Flash -evästeitä, sivuhistoriaa ja HTML5 -tallennus. Ongelma on, jos näistä tallennuspaikoista tuhoaa yhden tai useamman, niin jäljelle jääneen avulla pystytään luomaan uudet poistettujen tilalle. Jos Flash -evästettä ei ole poistettu, niin Evercookie voi toimia myös muissakin koneen selaimissa.

Seuranta voi perustua myös eri laitteiden tunnusomaisiin piirteisiin tai asetuksiin. Erilaiset asetukset kellolle, eri fontit ja ohjelmat tekevät laitteesta ainutlaatuisen. Verkossa on mahdollista tunnistaa näitä ainutlaatuisia piirteitä ja muodostaa niistä yhteinen kokonaisuus. Tälle yhteenvedolle voidaan asettaa tunnistenumero, jonka avulla kyseisen laitteen voi erottaa muista laitteista verkossa. Tätä kutsutaan laitteiston sormenjäljeksi (device fingerprint). (Privacy Rights Clearinghouse 2014.)

Nikiforakis, Kapravelos, Joosen, Kruegel, Piessens ja Vigna (2013) sanoo yhden kolmasosan käyttäjistä poistavan laitteen evästeet kuukauden sisällä selailusta. Verkkoselaimiin on tarjolla lisäosia, jotka paljastavat kolmannen osapuolen seurannan ja osa nykyaikaista selaimista kykenee kolmannen osapuolen evästeiden estämiseen vakiona. Hän mainitsee myös verkkoselaimen yksityisyystilan estävän selailun jälkien jäämisen käyttäjän koneelle. Tämä on motivoinut markkinoijia ja seuraajia etsimään uudenlaisia seurauskeinoja.

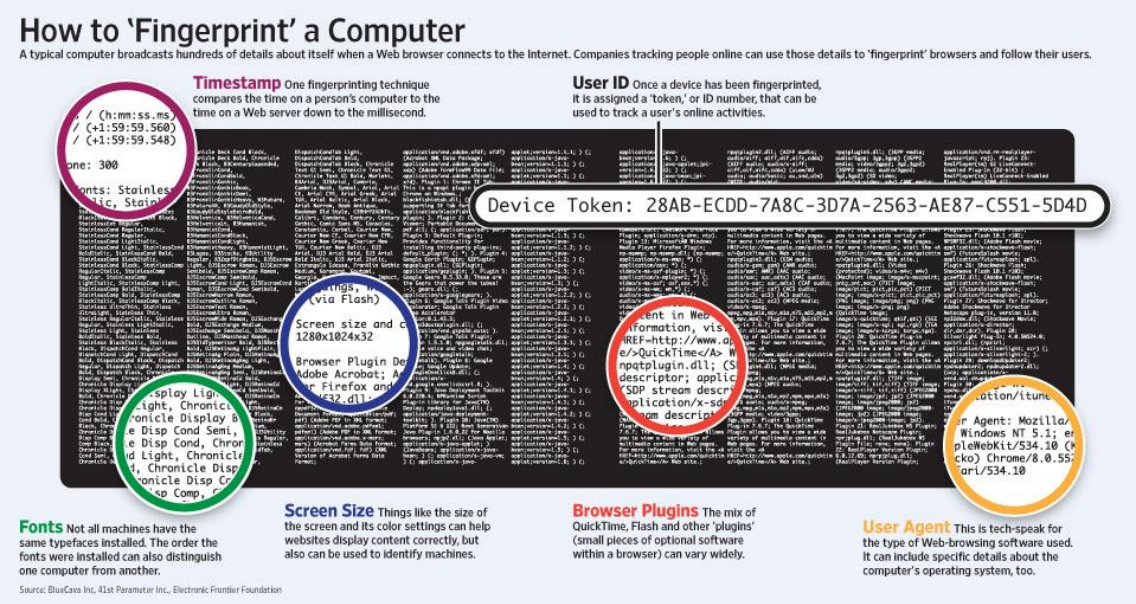
Laitteiston sormenjälkiä voidaan käyttää rakentavasti ja tuhoavasti. Oikein tunnistetun laitteen avulla on mahdollista tunnistaa petosyritys esimerkiksi tilanteissa, jossa käyttäjältä on varastettu kirjautumistunnukset tai evästetiedot. Väärästä laitteesta kirjautuvan käyttäjän on mahdollista tunnistaa hyökkääjäksi. Tuhoavasti laitteen sormenjälkeä voidaan käyttää käyttäjien seurantaan ilman heidän tietämystä tai tapaa jättäytyä seurannasta pois. Haittaa haluavien toimijoiden on lisäksi mahdollista hyödyntää laitteen tunnistusta räätälöimään tietomurtotoimet toimimaan tietyissä sovelluskombinaatioissa. (Nikiforakis ym. 2013.)

Privacy Rights Clearinghousen (2014) mukaan on mahdollista, että tämä menetelmä syrjäyttää ajan kuluessa evästeet käyttäjien seurannassa. He sanovat menetelmän olevan evästeitä vaikeampi estää. Se ei jätä koneelle mitään tietoa, eikä käyttäjä voi näin ollen tietää, että häntä seurataan. Käyttäjän on mahdollista torjua tämän menetelmän käytön poistamalla JavaScript käytöstä, mutta tämä taas aiheuttaa joidenkin kotisivujen toiminnollisuuksien menetyksen. (Flanagan 2006: 1). Javascriptiä voidaan käyttää myös tunnistamaan millä sivustoilla käyttäjä on vierailtu sivustolinkkien värityksien perusteella. Verkkoselaimiin on kehitetty ominaisuuksia parantaakseen käyttäjäkokemusta mm. värjäämällä linkit eri värillä sen mukaan, onko käyttäjä vieraillut linkin takana olevalla verkkosivustolla vai ei. Tätä voidaan käyttää hyväksi asettamalla sivustolle tuhansia käyttäjälle näkymättömiä linkkejä ja tarkastelemalla Javascriptin avulla minkä värinen linkki esiintyy. (Cranor ym. 2013).

Laitteiston sormenjälkeen perustuvaa tekniikkaa on myös kehitetty astetta pidemmälle. BlueCava nimisen yrityksen käyttämä tekniikka pystyy yhdistämään käyttäjän käyttä-

mät laitteet toisiinsa. He tallentavat käyttäjän laitteista erilaisia tietoja, kuten mikä laite on kysymyksessä, IP -osoitteen, asennetut fontit, selainversio ja sen lisäosat sekä lukuisia muita ominaisuuksia. Eri laitteille asetetaan tunnistenumero ja myöhemmin kerättyjä tietoja yhdistetään heidän markkinointikumppaneilta saatuihin tietoihin. Näin samat laitteet voidaan yhdistää samaan talouteen kuuluvaksi. BlueCava mahdollistaan käyttäjän ilmoittamaan heidän kotisivuillaan, jos hän haluaa jättäytyä pois tästä seurannasta tai tunnistamisesta. Tämä kuitenkin tulee tehdä jokaiselle laitteelle erikseen. (Privacy Rights Clearinghouse 2014.)

Angwin ja Valentino-DeVries (2010) mainitsee, että BlueaCava on tunnistanut 200 miljoonaa eri laitetta ja odotettavissa on, että tunnistettavien laitteiden määrä kasvaa huomattavasti vuodesta 2010 eteenpäin. Mainostajat eivät tyydy enää pelkkiin mainoksiin, vaan haluavat tavoittaa tiettyjä ihmisiä. BlueCavan tarkoituksena on luoda jokaisen laitteen käyttäjästä käyttäytymiseen, ostostottumuksiin ja maantieteelliseen sijaintiin perustuva profiili, jonka mainostajat voivat ostaa. BlueCava kuitenkin sanoo, ettei heidän keräämistään tiedoista pysty jäljittämään yksittäistä henkilöä. Kuvassa 5 on esimerkkejä laitteen tunnistamisesta.



Kuva 5. Laitteen tunnistamisen esimerkkejä. (Lähde: Angwin ym. 2010)

Kuten kuvasta on nähtävissä, tavallinen tietokone lähettää erilaisia tietoja itsestään Internet-selailun yhteydessä. Kuvassa on poimittu tietoja selailijan laitteen kellonajasta, jota vertaillaan verkkopalvelimen kellonaikaan, asennettuja fontteja, näytön resoluutio, selaimen lisäosat ja mitä selainta vierailija käyttää.

Yksi ongelmallinen tekniikka on verkkoselaimen otsikkokentän tieto, jota kutsutaan ETagiksi (ETag). ETagit ovat HTTP otsikkotietoa, joka mahdollistaa verkkosivun liittää käyttäjän vierailemalle sivustolle pysyvän tunnistenumeron. Verkkosivustojen sisällöstä on mahdollista tallentaa ns. ”kopioita” paikalliselle koneelle. Tällöin sivustovierailun yhteydessä ei tarvitse tehdä uutta sivustolatausta verkon välityksellä, vaan sivuston sisältö voidaan ladata käyttäjän koneelta. Sivu voi yhdistää sivuston sisällön tähän ETagi -tunnisteeseen. Jos sivuston sisältö on muuttunut ja tunniste ei täsmää, niin sisältö haetaan uudelleen. (Cranor ym. 2013.)

ETagin avulla tapahtuva seuranta on mahdollista, vaikka käyttäjä estäisi HTTP -evästeiden, Flash -evästeiden tai HTML5 -tallennuksen. Sen avulla on myös mahdollista luoda myös HTTP -evästeitä uudelleen. Ainoa keino poistaa ETagi on poistaa selaimen välimuisti jokaisen sivustovierailun jälkeen. (Ayenson, Wambachm Soltani, Good & Hoofnagle 2011.)

Ayenson ym. (2011) tekivät tutkimuksen 100 suosituimmalla sivustolla perustuen QuantCast.com sivuston listaan. He löysivät HTTP -evästeitä jokaisella sivustolla, joita löytyi yhteensä 5675 kappaletta. Vuonna 2009 he löysivät niitä vain 3602 kappaletta, joten evästeiden käyttö on lisääntynyt. Vuoden 2011 tutkimuksessa 20 sivustoa asetti evästeitä 100 tai enemmän. Kolmansia osapuolia oli yli 600 ja heiltä tuli yhteensä 4915 evästettä. Google oli mukana 97 sivustossa sadasta. Flash -evästeitä löytyi 100 kappaletta yhteensä 281 sivustolla. HTML5 -tallennusta käytti 17 sivustoa ja HTML5 -tallenteissa käytettiin samoja arvoja, kuin tavallisissa evästeissä. Tutkimus paljasti myös kolmen sivun synnyttävän uudelleen poistetun HTTP -evästeen toisen tallennustekniikan avulla.

Youyou, Kosinki ja Stillwell (2014) tekivät persoonallisuuskyselyn 86 220 henkilölle, johon kuului 100 kohtaa. Kyselyyn osallistuneiden Facebook kavereilta ja perheenjäseniltä kyseltiin osallistujien persoonallisuustietoja 10 kohdan kyselyillä. Osallistujista tehtiin myös tietokonepohjainen persoonallisuus arvio perustuen Facebook tykkäyksiin. Tutkimukseen saatiin tuloksia 17 622 osallistujasta, joita arvioi yksi kaveri tai perheenjäsen ja 14 410 joita arvioi kaksi kaveria tai perheenjäsentä. Tulokset näyttivät, että tietokonepohjainen arviointi oli tarkempi 10 tykkäyksen jälkeen kuin työkaveri, 70 tykkäyksen jälkeen tarkempi kuin kaveri tai kämppäkaveri, 150 jälkeen tarkempi kuin perheenjäsen ja 300 tykkäyksen jälkeen tarkempi kuin puoliso. Persoonallisuuskysely keskittyi henkilön avoimuuteen, tunnollisuuteen, ulospäin suuntautuneisuuteen, ystävällisyyteen ja neuroottisuuteen.

Tiedon valtava määrä on luonut uuden käsitteen nimeltä Big data. Big datan laajuuden ja monimutkaisuuden ansiosta monilla yrityksillä on ongelmia hallita Big dataa hyödyntävästi ja turvallisesti. Big data on termi, joka siis viittaa tietokokonaisuuksiin tai eri aineistojen yhdistelmiin joiden koko, monimutkaisuus ja kasvuvauhti tekevät tiedon hallinasta vaikeaa tavanomaisilla työkaluilla. Monimutkaisen tiedosta tekee se, että tietoa syntyy useiden modernien teknologioiden avulla kuten weblogeista, radiotaajuus tunnisteista, laitteistojen antureista, eri koneista, ajoneuvoista, Internet etsinnöistä, sosiaalisista verkostoista, puhelimista jne. Tietoa kerääntyy myös reaaliaikaisesti eri laitteista, järjestelmistä ja verkoista. Tietovuodot voivat tulla yritykselle erittäin kalliiksi. Esimerkiksi yhdessä tapauksessa terveysalan kanteessa kanteen alullepanija vaatii 1000 dollaria jokaista potilastietoa kohden, johon on päästy käsiksi väärin perustein tai se on hukattu. (Navint 2012.)

Polonetsky ja Tene (2013) sanoo Big datan luovan suurta arvoa maailmantaloudelle kansallisen turvallisuuden, markkinoinnin, luottoriskin analysoinnin, lääketieteellisen tutkimuksen ja kaupunkisuunnittelun muodossa. He kuitenkin toteavat big datan ympärillä pyörivän huolia yksityisyydestä ja tietoturvasta. He sanovat yksityisyyden kannattajien olevan huolissaan muuttuvista valtasuhteista hallitusten, yritysten ja yksityisten välillä. He lisäävät tämän voivan johtaa erilaisiin profilointeihin, syrjintään, kriminalisointiin tai muuhun vapauksien rajoitteisiin.

Big datan hyödyn ja sen tuoman yksityisyyden riskien tasapainon löytäminen voikin olla yksi suurimmista poliittisista haasteista tällä hetkellä. Se vaatii päätöstä, että onko arvokkaampaa löytää parannus kohtalokkaita tauteja vastaan ja estää terrorismia, mitä altistaa ihmiset kaikkitietävälle valvonnalle ja algoritmiselle päätöksenteolle. (Polonetsky ym. 2013.)

Yllä mainitut radiotaajuustunnisteet, koneet ja laitteet, ajoneuvot ovat kaikki myös esimerkkejä esineiden Internetistä (Internet of things). Esineiden Internetiin periaatteessa lasketaan mukaan kaikki laitteet, jotka ovat Internetiin yhteydessä lukuun ottamatta matkapuhelimia, tabletteja tai tavallisia tietokoneita. Asiantuntijat arvioivat, että vuonna 2015 tulee olemaan 25 miljardia yhteyksiin pystyvää laitetta ja määrän odotetaan kasvavan 50 miljardiin vuoteen 2020 mennessä. Uusien kehitysten odotetaan tuovan kuluttajille suurta hyöytä. Hyötyä yhteyksissä olevista laitteista on esimerkiksi terveysalalla, jossa lääkäri voi olla yhteydessä paremmin potilaaseen tai auto-onnettomuuksissa, jossa auto voi ilmoittaa ensiapuun heti onnettomuuden sattuessa. Nämä yhteydessä olevat laitteet kuitenkin kerää, lähettää, tallentaa ja mahdollisesti jakaa kuluttajien tietoja, jotka ovat yksityisiä ja näin ollen tuovat huolia yksityisyydestä ja tietoturvasta. (FTC 2015.)

Samsung -yhtiö on varoittanut, että Samsungin televisiot saattavat osana puheentunnistusta salakuunnella television lähellä käytyjä keskusteluja. Kun television puheentunnistus on päällä, käyttäjä voi äänikomennoin ohjata television toimintoja. Puheentunnistus lähettää kuulemansa äänet toiseen palveluun, joka etsii lähetetyistä tiedoista oikeat käskyt ja palauttavat ne televisioon. Samsung sanoi, ettei toimintaan kuulumattomia ja lähetettyjä ääniä tallenneta tai myydä eteenpäin. (Samsung 2015, BBC 2015.)

Peppet (2014) mainitsee tuoreiden tutkimuksien näyttäneen, että nykyisillä älypuhelimien sensorien avulla on saatu tuloksia käyttäjän mielentilasta, stressitasosta, persoonallisuus tyypistä, kaksisuuntaisesta mielenhäiriöstä, tupakoinnista, kuinka käyttäjä nukkuu, onnellisuudesta, harrastustottumuksesta, Parkinsonin taudin etenemisestä, yleisestä hyvinvoinnista ja väestötiedoista kuten sukupuolesta, siviilisäädystä, työtilanteesta ja henkilön iästä. Hän myös mainitsee tällä olevan väärinkäytön ongelmia ja mahdollisia vaikeuksia saattaa kerääntyä Big datan analytiikan kautta.

Monien kuluttajille tarkoitettujen anturilaitteiden valmistajat pyrkivät tekemään kerätyistä tiedoista tunnistamattomia eli tekevät tiedosta anonyymia. Tällaisia laitteita ovat esimerkiksi aktiivisuusrannekkeet, autot ja älypuhelimet. Valmistajat myös lupaavat, että tietoa jaetaan tavalla, ettei siitä pysty tunnistamaan yksittäistä henkilöä. Tässä on kuitenkin muutama tekninen ja laillinen ongelma yksityisyyden kannalta. Esimerkiksi tekniseltä kannalta katsottuna aktiivisuusrannekkeen käyttäjä on kohtalaisen helposti tunnistettavissa. Vaikka käyttäjän nimi, osoite ja muu selvästi käyttäjän tunnistamiseen helpottava tieto saataisiin piilotettua, niin käyttäjän pystyy tunnistamaan sen ainutlaatuisten ominaisuuksien kautta. Jos tietää vaikka aktiivisuusrannekkeen käyttäjän ainutlaatuisen kävelytavan, niin loput tiedoista pystyy yhdistämään tähän käyttäjään. (Peppet 2014.)

Hunt (2013) sanoo, että vain katsomalla aktiivisuusrannekkeen keräämiä tietoja, on mahdollista tunnistaa kuinka pitkä henkilö on ja onko henkilö ylipainoinen vai ei. Hän myös toteaa, että pelkkä askeltyyli riittää käyttäjän tunnistamiseen. Peppet (2014) mainitsee lailliseksi ongelmaksi sen, että yksityisyyden laki ei ole vielä valmis suojaamaan yksityisyyttä esineiden Internetissä, sillä moni laki katsoo henkilötiedoiksi nimen, osoitteen, sosiaaliturvatunnuksen ja puhelinnumeron. Monen muun tiedon ei oleteta paljastavan henkilön identiteettiä.

Peppet (2014) toteaa, että lähiaikana oikeustieteilijät ovat tulleet enemmän huolestuneemmaksi siitä, että voiko suuria tietomassoja koskaan tehdä täysin tunnistamattomaksi. Tiedon tekeminen täysin tunnistamattomaksi tulee entistä vaikeammaksi, kun tietoa tulee eri toiminnoista pitkin päivää. Myös toisen käyttäjän lähettämiä tietoja pystytään vertaamaan toisen käyttäjän tietoihin, milloin voidaan saada selville toisesta käyttäjästä puutuvia tietoja. Jos sensorit välittävät tietoa liikkeestä, mutta ei tarkasta sijainnista, niin on mahdollista katsoa toisen käyttäjän sijaintitietoja, jonka sensorit välittävät samanlaisia liiketietoja.

Tietoa siis kerääntyy paljon, mutta seurannasta voi tulla myös häiritsevä tekijä. Tivi (2014b) uutisoi mahdollisesta tarjousten ilmoittamisesta mobiililaitteisiin. Esimerkkinä uutisessa on, että kuluttaja saisi viestin mobiililaitteeseensa häntä kiinnostavien tuottei-

den tarjouksista. Kiinnostuksen kohteet olisi kerätty käyttäjän selaustietojen perusteella. Joillekin tämä voi olla kannattavaa, mutta kaikki eivät välttämättä olisi samaa mieltä. Jos asiakas tunnistetaan heti liikkuessa kivijalkakaupan ohitse, niin ensinnäkin tämä tarkoittaisi, että asiakkaan liikkeitä seurataan eli tiedetään, missä henkilö on liikkunut. Tässä työssä on jo esiintynyt usein tiedon jakamisesta ja jos nämäkin tiedot jaetaan muille, on ihmisten jokaisen liikkeen seuranta jo lähellä.

Toiseksi matkapuhelimesta saattaisi tulla jatkuva "roskapostin" kerääjä. Miten käy asiakkaan liikkuessa ostoskeskuksissa. Tuleeko tulevaisuudessa viestejä jatkuvalla syötöllä jokaisesta liikkeestä, jonka ohitse kävelee. Toiminta saattaisi hankaloittaa muiden viestien lukemista, sillä mobiililaite täyttyisi kaikenlaisilla mainos- ja markkinointiviesteillä. Erilaisten havaintojen tulokset ja ennustukset voivat tuoda kaikenlaisia ongelmia. Yhteiskunnallinen asema, elämäntilanteet tai kulttuuri voi vaikuttaa millaiseen asemaan ihminen joutuu, jos henkilöstä paljastuu tietoa, mitä hän itse ei olisi halunnut paljastuvan.

Kosinski ym. (2013) mainitsevat amerikkalaisen vähittäismyyntiketjun käyttäneen asiakkaiden ostostietoja ennustaakseen naispuolisten asiakkaiden raskautta. Tiedon avulla he pystyivät lähettämään oikeanlaisia tuotetarjouksia. He sanovat, että tällä voi olla huonoja seuraamuksia, sillä raskaus voi paljastua esimerkiksi naisen läheisille ja kaikissa kulttuureissa naimattoman naisen raskautta ei katsota hyvällä. Ennustettu raskaus ostosten perusteella ei myöskään ole varmaa. He mainitsevat, että yksilöiden piirteiden ja ominaisuuksien ennustamisella on pitkä historia. Ihmisten siirtyminen digitaalisen aikaan on mahdollistanut ennustamisen digitaalisista tallenteista.

Yksittäisen henkilön piirteiden ennustamisella digitaalisista tallenteista voi olla siis erittäin huonoja seuraamuksia, koska niitä voi tehdä lukuisille henkilölle ja ilman heidän suostumustaan tai, että ne edes huomaisivat tämän toiminnan. Eri toimijat pystyisivät käyttämään ohjelmia, joiden avulla voisi päätellä toisen henkilön piirteitä, kuten älykkyyttä, seksuaalista suuntautumista tai poliittisia näkemyksiä. Tämänlaisilla ennusteilla voi olla vakavia seurauksia kohteena olevalle henkilölle. (Kosinski ym. 2013.)

4.2.2 Tietojen yhdistäminen käyttäjän identiteettiin

Narayanan Arvind (2011) mainitsee viisi tapaa kuinka käyttäjän identiteetin on mahdollista yhdistää aiemmin kerättyihin tietoihin.

1. Kolmas osapuoli on joskus ensimmäinen osapuoli

   Käyttäjä tunnistetaan, kun hän kirjautuu esimerkiksi Facebookin tai Googlen sivustolle. Joillain sivustoilla on Facebookin like -painike tai Googlen +1 -painike. Näiden avulla he voivat yhdistää tiedot käyttäjän vierailusta ulkopuolisilla sivustoilla sitten käyttäjän käyttäjätiliin.

2. Tunnistetietojen siirtyminen ulkopuoliselle sivustolle

   Ainakin neljä tapaa on tunnistettavissa, kuinka käyttäjätietoa voi siirtyä vierailtavalta sivustolta ulkopuoliselle sivustolle. Tietoa voi liikkua hypertekstin siirtoprotokollan (HTTP) käyttämän referer -kentän, verkkosivuosoitteen, ulkopuolisten palveluiden asettaman evästeen ja sivuston otsikkokentän mukana. Järvinen (2002: 147) mainitsee referer -kentän sisältävän tiedon sivustosta, jossa aiemmin vierailtiin. Sivuston ylläpitäjä voi referer -kentän avulla saada selville, minkä sivuston linkin takaa käyttäjän on sivustolle saapunut. Mayer (2011) antaa esimerkin verkkosivuosoitteen ja otsikkokentän mukana liikkuvasta tiedosta. Hän sanoo ulkopuolisten palveluiden saavan tietoa käyttäjistä, jos sivustolla on heidän tarjoamaa sisältöä. Esimerkkinä hän käyttää seuraavaa verkkosivuosoitetta, jonka tyylinen osoite voi muodostua käyttäjän rekisteröityessä Internet -palveluun.

   ```
   "http://example.com/register?
   username=GoCardinal
   &name=Leland%20Stanford
   &email=leland%40stanford.edu
   &..."
   ```

   Ulkopuolinen palvelu voi saada tästä poimittua käyttäjänimen, oikean nimen ja sähköpostiosoitteen referer -kentän tai vastaavan avulla. Ulkopuolisten palvelui-

den ohjelmistokoodi taas voi palauttaa sivuston otsikkokentän tiedot, jossa voi lukea esimerkiksi verkkosivustolla tunnistetun nimi. (Mayer 2011.)

3. Ulkopuolinen palvelu voi ostaa käyttäjän tunnistetietoja

   Eri sivustot voivat tehdä yhteistyötä ja yhdistää kerätyt tiedot toistensa kanssa. Esimerkiksi kyselysivustot voivat kerätä käyttäjistä tietoa ulkopuolisilla sivustoilla erilaisten tekniikoiden avulla. Käyttäjä paljastaa itsestään tietoja, kun hän suorittaa kyselysivustojen kyselyitä tai rekisteröityy sivustolle. Tämän jälkeen kyselysivusto voi yhdistää aiemmin ja muilta sivustoilta kerätyt tiedot käyttäjään. Sivusto voi tarjota keräämiään tietoja myös muille sivustoille.

   On myös mahdollista, että kyselysivustolla on seurantatekniikkaa käytössä ulkopuolisilta palveluilta. Kyselysivusto voi vapaaehtoisesti sallia erilaisten tunnistetietojen paljastamisen näille ulkopuolisille seurantasivustoille. Tällä seurantasivustolla saattaa olla kerättynä kattava määrä selailijan selaushistoriaa, mutta käyttäjän tunnistamiseen vaativat tiedot puuttuvat. Tämä kyselysivusto voi luovuttaa nämä käyttäjän tunnistetiedot seurantasivustolle, jos käyttäjä on nämä kyselyssä paljastanut.

4. Tietojen murtaminen

   Erilaisia ohjelmointivirheitä on mahdollista hyödyntää ja saada näiden avulla selville käyttäjän tunnistetietoja tai selaushistorian. Tunnetut ohjelmointivirheet yleensä korjataan, mutta uusia esiintyy jatkuvasti lisää.

5. Tunnistaminen

   Algoritmien avulla on mahdollista selvittää tietoja Internet-selailijasta. Algoritmi voi katsoa ajankohtaa milloin käyttäjä on vieraillut tietyillä sivustoilla. Kun käyttäjä on kommentoinut esimerkiksi omalla nimellään kaverin blogissa ja selannut saman päivän aikana kymmenellä eri sivustolla. Algoritmi voi tunnistaa käyttäjän helposti, sillä on pieni todennäköisyys, että jokin muu on vieraillut samoilla si-

vuilla juuri samoihin aikoihin. Mitä suurempaa selaushistoriaa algoritmi tarkastelee, sitä enemmän selainkuvio on tunnistettavissa.

4.2.3 Sosiaalinen media ja kasvojen tunnistus

Sosiaalisissa medioissa jaettu kuva tai kommentti voi lähteä jakoon muiden toimesta, vaikka se olisi tarkoitettu vain kavereiden nähtäväksi. Monesti onkin tärkeää tutustua palvelun yksityisyys-selosteeseen. Identiteettivarkaat, huijarit, vaanijat, velanperijät ja yritykset ovat kiinnostuneita näistä jaetuista tiedoista. Yritykset pyrkivät saamaan kuluttajatiedoilla etua markkinoilla. He voivat myös myydä kerättyjä tietoja muille markkinoijille. (Privacy Rights Clearinghousen 2014.)

Acquisti, Gross & Stutzman (2011) tekemä tutkimus osoittaa, että kasvon kuvia pystytään vertaamaan verkossa olevaan tietoon kasvojentunnistuksella. He yhdistivät julkisesti saatavilla olevia tietoja kasvojentunnistusohjelmistoon. Tutkimuksessa otettiin valenimen eli nimimerkin takana olevien henkilöiden profiilikuvia nettideittailu -sivustoilta ja verrattiin Facebookin julkisen profiilin kuvaan kasvojentunnistuksella. Acquisti ym. (2011) sanovat pystyneen tutkimuksessaan tunnistamaan tilastollisesti merkittävän osan jäsenistä. Toisena tutkimuksena he kuvasivat henkilöitä nettikameran avulla ja vertasi näitä kuvia taas Facebook profiilikuvaan. Tämän avulla he pystyivät tunnistaman noin joka kolmannen henkilön.

Erilaiset yksityisyysaiheiset tapaukset ovat päässeet uutispalstoille. Iltasanomat (2011) uutisoi Itävaltalaisen Max Schremsin pyytäneen Facebookilta kaikkia hänestä kerättyjä tietoja. Hän oli ollut Facebookissa vuoden ajan ja hänellä oli ollut reilut 200 kaveria. Max sai Facebookilta CD -levyn, jossa oli 1222 PDF -sivullista tietoa. CD -levyltä löytyi hänen lähettämänsä ja vastaanottamansa viestit, tykkäämiset, kaverihistoria ja puhelimella otettujen kuvien GPS -tiedot. Tiedoissa oli mukana myös Maxin poistamat tiedot.

Max Schrems on myös nostanut Facebookia vastaan joukkokanteen 1.8.2014, johon on osallistunut jo yli 20 000 henkilöä. Osallistujia on useista eri Euroopan maasta, mukaan lukien Suomesta. Amerikkalaiset ja Kanadalaiset eivät voi osallistua kanteeseen johtuen maiden lainsäädäntöjen eroavaisuuksista. Kanteessa keskeisenä asiana on mm. tietojen luvaton luovutus ulkopuolisille sovelluksille, käyttäjien seuranta ulkopuolisilla sivustoilla tykkää -nappien avulla ja tiedon käyttöehtojen virheellisyys. Tuomioistuin kokoontuu kuulemaan kanteesta 9.4.2015. (Theguardian 2014; Tivi 2014a; Techcrunch 2014; Facebook class action 2015.)

Muitakin tiedon keräämiseen liittyviä uutisia on julkaistu Facebookkiin liittyen. Mashable (2014) uutisoi Facebookin Atlas mainosverkostosta. Se on uusi mainosverkosto, jonka avulla voidaan tarjota verkkoselailijalle mainoksia Facebookin ulkopuolisilla sivustoilla sen perusteella, mitä Facebook käyttäjästä tietää. Sen avulla markkinoijat pystyvät seuraamaan käyttäjiä verkossa ja mobiililaitteissa, koska Atlasta käyttävä markkinoija voi tarjota mainoksia niin Internet-sivuilla, kuin mobiilisovelluksissa. Uutisessa mainitaan myös, että Twitterin mukaan lähteminen tähän mainosverkostoon on mahdollista.

Image.fi (2015) testasi kuunteleeko Facebookin älypuhelinapplikaatio sen lähettyvillä käytyjä keskusteluja. He sanovat Facebookin sivuilla olevan tiedote, että jos käyttäjä antaa luvan tunnistaa käyttäjän kuuntelemia ja katsomia asioita, niin Facebook käyttää mikrofonia käyttäjän kirjoittaessa tilapäivitystä. Toiminto pitäisi myös olla käytettävissä vain Yhdysvalloissa. Image.fi:n usean testin perusteella kuuntelua kuitenkin tapahtuu. He kertovat kokeilleensa useiden eri sanojen lausumista puhelimen lähettyvillä. Lausunnan jälkeen Facebook tarjosi sanaan liittyviä mainoksia. Yhtenä esimerkkinä he kertovat sanan sosiaalidemokraatit, milloin Facebook tarjosi Helsingin demareiden Facebook ryhmää. Toisen esimerkkinä sanan "lehti" jälkeen mainoksia tuli Seiska lehden tilauksesta. He myös huomasivat puheenaiheiden ilmestyneen Facebookin uutisvirtaan, vaikka älypuhelimen sovellus oli kiinni.

Kokeilussa käytettiin kahta eri puhelinta ja puhelinta pidettiin alle metrin päässä ja myöhemmin useiden metrien päässä puhujista. Molemmissa tapauksissa puhelimen ap-

plikaatio poimi sanoja keskustelusta. Facebook on ilmoittanut jo vuonna 2014 lisäävänsä käyttäjän ympäristöä kuuntelevan ominaisuuden, joka kuuntelee ainoastaan taustamelua, eikä ihmisten keskusteluja. Testi kuitenkinkin osoittaa nykyisen ominaisuuden käyttävän muutakin kuin taustamelua. (Image.fi 2015.)

Monilla sivustoilla näkyy Facebookin tykkäys- tai suosituspainikkeet. Tämän avulla Facebookilla on mahdollisuus seurata käyttäjien selailua Facebookin ulkopuolisilla sivustoilla. Näitä painikkeita ei tarvitse erikseen painaa, että tieto käyttäjän vierailusta ulkopuolisella sivustolla siirtyy. Käyttäjän tarvitsee vain olla kirjautuneena Facebookiin. Selainikkunan ei tarvitse olla auki, vaan riittää ettei käyttäjä ole kirjautunut ulos Facebookista. (Järvinen 2012: 298-299.)

Myös Krishnamurthy ja Wills (2009) mainitsevat, että sosiaalisen median käyttäjän tunnistetietoja voi päätyä ulkopuolisille mainostajille tai tiedonkerääjille. He näkevät tässä ongelmia, sillä käyttäjästä on voitu kerätä tietoa useita vuosia ja nyt yhdistää sosiaalisesta mediasta saatuun tietoon. Tämä tarkoittaa, että myös tulevaisuudessa kerättyjä tietoja voidaan yhdistää tähän identiteettiin. He sanovat, että suurin osa tiedonkerääjistä itse sanoo keräävänsä tietoa käyttäjän käyttäytymisestä, eikä tallenna käyttäjien henkilökohtaisia tunnistetietoja. He kuitenkin toteavat, että henkilökohtaisia tunnistetietoja on kerääjien saatavilla, vaikka varmoja niiden tallentamisesta tai yhdistämisestä muuhun kerättyyn tietoon ei olekaan.

4.2.4 Käyttäjätilien ongelmat

Yksi mainittava ongelma on toisen nimellä esiintymisen helppous. Toisen nimellä esiintyminen Internetissä on helppoa, eikä laki vielä tällä hetkellä estä sitä. Nettipalveluiden on hankala tarkistaa käyttäjän henkilöllisyyttä, joten blogin, Facebook-profiilin tai Twitter-tilin perustaminen toisen tiedoilla on mahdollista. (Järvinen 2012: 256-258.)

Oikeusministeriön (2014) mukaan Suomen lainsäädännössä ei ole tällä hetkellä itsenäistä rangaistussäännöstä identiteettivarkautta kohtaan. Tekijä kuitenkin voidaan tuomita esimerkiksi väärän henkilötiedon antamisesta (Rikoslaki luku 16 pykälä 6 §), rekisteri-

merkintärikoksesta (luku 16 pykälä 7 §), väärennyksestä (luku 33 pykälä 1§), yksityiselämää loukkaavan tiedon levittämisestä (luku 24 pykälä 8§) ja kunnianloukkauksesta (luku 24 pykälä 9§). Eduskunnalle on 13.11.2014 annettu esitys rikoslakimuutoksesta, jonka mukaan identiteettivarkaus olisi rangaistava itsenäisenä rikoksena. Muutokset koskisivat erityisesti vahingontekoa, vaaran aiheuttamista tietojenkäsittelylle, viestintäsalaisuuden loukkausta, tietojärjestelmän häirintää ja tietomurtoa. (Oikeusministeriö 2014; Finlex 2015)

Vaikka rikoksen kriteerit eivät toteutuisikaan, harhaan johdetulle uhrille voi aiheutua suurta mielipahaa. Laillisin keinoin tämän toiminnan ennaltaehkäisy on vaikeaa, juurikin sen vuoksi, kun palveluntarjoajat eivät pysty tarkistamaan kaikkien henkilöllisyyttä. Nettikeskusteluissa toisella henkilöllä voi myös oikeasti olla sama nimi. Suurimmaksi ongelmaksi asia koituu silloin, kun tekijä ylittää lailliset rajat ja alkaa käyttää toisen tietoja esimerkiksi petoksien tekemiseen. (Järvinen 2012: 256-258).

Security Intelligence (2014) kertoo yhden kirjautumismenetelmän aukosta, jonka avulla käyttäjä pystyy kirjautumaan eri sivustoille toisen henkilön nimellä käyttäen sosiaalisen median tunnistusta. He sanovat, että osa sivuista tarjoaa tällaista mahdollisuutta kirjautumiseen ja käyttäjän tunnistamiseen. Tietoturva-aukon on havaittu vain tietyssä ja tietyllä lailla integroiduissa kirjautumisjärjestelmissä ja vaatii sosiaalisen median sähköpostiosoitteen tietämistä. Tämä kuvastaa kuitenkin hyvin, millaisia seurauksia esimerkiksi tämän kaltaisella tiedon jakamisella voi olla. Monesti tarkoitukset ovat hyvät, mutta ongelmien sattuessa, haitoista kärsii monesti loppukäyttäjä. Oletusarvona voisi olla, ettei mikään sosiaalinen verkosto tai vastaava antaisi ulkopuolisen sivuston kommunikoida käyttäjän käyttäjätilin kanssa. Käyttäjälle voitaisiin antaa mahdollisuus valita tilinsä asetuksista tällainen mahdollisuus erikseen, mutta ei oletuksena.

> "EU:n tiukat tietosuojalait ovat vaatimattomia niin kauan kuin eurooppalaiset antavat suostumuksen tietojensa käytölle ja jopa auttavat niiden luovuttamisessa. Tämä muuttuu vasta, kun Euroopassa syntyy kilpailijoita Facebookille ja Googlelle. Eli käytännössä ei koskaan." (Järvinen 2012: 311).

Tietojen siirtyminen Internetiin on myös ongelma siinä, ettei ikinä voi olla varma, saako siirtyneitä tietoja poistettua. Erilaiset tiedot voivat tulla esiin myöhemmin ihmisten elämässä ja aiheuttaa henkilölle ongelmia esimerkiksi työhaastattelussa. (Järvinen 2012: 312.)

5 LIGHTBEAM TUTKIMUS ULKOPUOLISISTA YHTEYKSISTÄ

Tämä osio koostuu kolmesta erillisestä tutkimuksesta. Tutkimuksilla havainnollistetaan tiedon siirtyminen ulkopuolisille sivustoille. Tutkimuksissa käytettiin Firefox -selaimen Lightbeam -lisäosaa. Mozilla:n (2014) Internet-sivujen mukaan lisäosa näyttää käyttäjälle, kuinka moneen eri sivustoon kohteena oleva tai vierailtava sivusto on yhteydessä. Lightbeam tuottaa graafisen kuvion yhteyksistä, joita muodostuu sivustovierailun yhteydessä. Nähtävillä on myös tarkempi kuvaus vierailluista sivustoista ja siitä muodostuneista yhteyksistä.

Lightbeam julkaistiin syksyllä 2013. Sen kehitys aloitettiin heinäkuussa 2011 yhden Mozilla:n ohjelmistokehittäjän toimesta. Vuonna 2012 Mozilla aloitti yhteistyön Emily Carr yliopiston kanssa lisätäkseen sovellukseen visuaalisia ominaisuuksia. Mozillan mukaan kaikkien tulisi saada tehdä omat johtopäätöksensä omasta yksityisyydestään ja siitä, kuka näitä tietoja kerää. Avoimen tutkimusalustan avulla Mozillan tarkoitus on lisätä käyttäjien tietoisuutta, edistää tutkimusta ja viime kädessä aiheuttaa muutoksia seurannan ja yksityisyyden sääntelyyn. (Mozilla 2014.)

5.1 Toteutus

Tämä tutkimus ei paljasta, mitä tietoa eri sivustoille konkreettisesti liikkuu. Se kuitenkin osoittaa, että tiedon siirtyminen on mahdollista ja näyttää kuinka yleisesti se on käytössä suosituimmissa sivustoissa. Mozilla:n (2014) Internet-sivujen mukaan kaikki seuranta ei ole kuitenkaan pahasta, sillä moni palveluntarjoaja käyttää seurantatietoja avuksi tarjoamaan parempaa käyttökokemusta sivustolla sekä näyttämään aiheeseen sopivaa sisältöä.

Tutkimuksissa käytetty tietokone alustettiin tehdasasetuksille, ettei aiempien käyttäjien selailu vääristä tuloksia. Ennen selailun aloittamista tietokone tyhjennettiin ja käyttöjärjestelmä asennettiin uudelleen. Tämän jälkeen asennettiin tarvittavat sovellukset tutkimuksen toteuttamista varten. Tutkimuksissa käytettiin selaimena Mozilla Firefox 36.0 versiota, johon oli asennettu Lightbeam 1.2.1.

Ensimmäiseen tutkimukseen otettiin mukaan 20 suosituinta sivustoa Suomessa Alexa.com sivuston tarjoamalta listalta. Jokaisella sivustolla vierailtiin erikseen ja tarkasteltiin tuloksia Lightbeam -lisäosan avulla. Tarkastelussa on vain 20 sivustoa, koska yhteyksien hahmottaminen kuvasta ja niiden listaaminen on helpommin hahmotettavissa pienemmällä määrällä.

Alexa:n (2014) Internet-sivustolla kerrotaan heidän saavan tiedot käyttäjistä useiden selainten lisäosien avulla sekä oman sivustokoodin avulla, jonka verkkosivuston omistaja on vapaaehtoisesti asentanut omalle sivustolle. Näistä tiedoista he muodostavat suosituimpien sivustojen listan.

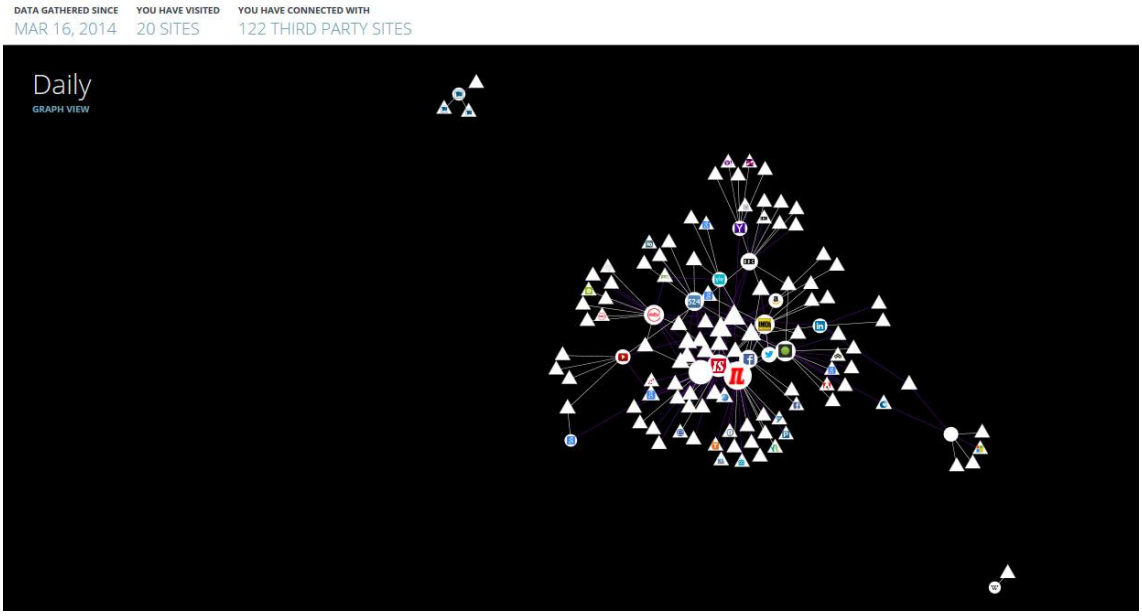
Toisena tutkimuksena Alexa:n tarjoamalta listalta otettiin mukaan viisi sivustoa, joilla selailtiin noin kahden minuutin ajan. Sivustot olivat iltalehti.fi, iltasanoma.fi, yle.fi, hs.fi ja yahoo.com, koska näillä sivustoilla on paljon sisältöä, joka mahdollistaa helposti kahden minuutin selailun eri aihepiireissä. Iltalehti.fi, iltasanomat.fi, yle.fi ja hs.fi ovat myös neljä suosituinta sivustoa TNS Metrix vuoden 2015 viikon 10 listalla TNS Gallup (2015). Ainoa sivusto, jota ei löydy TNS Metrix -listalta on yahoo.com sivusto. Myös tässä toisessa tutkimuksessa käytettiin tehdasasetuksille palautettua tietokonetta. Ensimmäisessä tutkimuksessa tehtiin ainoastaan vierailu sivustolla eli odotettiin, että sivusto latautui, jonka jälkeen siirryttiin seuraavalle sivustolle. Tämän toisen tutkimuksen tarkoituksena on näyttää, että syvempi selailu sivustolla tuottaa lisää yhteyksiä ulkoisille sivustoille. Toiseen tutkimukseen otettiin mukaan 20 sivun listalta sivustoja, joilla on useampia linkkejä alasivuille. Sivustoista tehtiin alustava tutkimus, jonka avulla selvitettiin, että sivustoilla pystyy selailemaan kahden minuutin ajan ilman liikkumista samoilla alasivuilla. Näin sivuston sisältö vaihtuu ja sivustolla mahdollisesti esiintyy uusia mainoksia tai muuta sisältöä, jotka luovat ulkopuolisen yhteyden.

Kolmannessa tutkimuksessa testattiin eri selainasetuksien ja lisäosien vaikutuksia ulkopuolisten sivustojen määrään erikseen ja yhdessä. Sivustovierailu oli ensimmäisen tutkimuksen mukainen eli jokaisella 20 sivustolla vierailtiin lyhyesti. Tutkimuksessa mukana olevat lisäosat olivat AdBlock Plus 2.6.8 ja HTTPS-Everywhere 4.0.3. Selaimen asetuksista testattiin evästeiden käytön pois ottamista ja JavaScript tuen poistamista.

HTTPS Everywhere salaa käyttäjän liikenteen. Monet sivustot tarjoavat yhteyttä HTTPS (Hypertext Transfer Protocol Secure) ylitse vain rajallisesti. Esimerkiksi sivustolla saattaa olla linkkejä sivuston osiin, jotka eivät ole HTTPS salattuja. HTTPS Everywhere muuttaa tällaisissa tapauksissa käyttäjän sivustolatauspyynnön menemään HTTPS:n ylitse. HTTPS Everywhere pystyy muuttamaan yhteyden menemään HTTPS:n ylitse vain silloin, kun sivusto tarjoaa HTTPS yhteyttä. (Electronic Frontier Foundation 2015a; Electronic Frontier Foundation 2015b.) HTTPS on HTTP:n turvattu versio, jota monesti käytetään verkkopankeissa ja verkkokaupoissa. HTTP -protokollaa käytetään lähettämään verkossa liikkuva tieto selaimen ja käytetyn palvelun välillä. HTTPS salaa tämän liikenteen (Instant SSL 2015.) Adblock Plus taas on selaimeen asennettava lisäosa, joka torjuu verkossa näkyviä mainoksia. Se pystyy myös torjumaan seurantaa, haitallisia verkkosivustoja, bannereita, ponnahdusikkunoita ja videomainoksia. (Adblock Plus 2015.)

5.2 Tulokset

20 sivustovierailun jälkeen ensimmäisessä tutkimuksessa yhteyksien määrä oli 142. Tämä tarkoittaa, että yhteyksiä lähti 122 kappaletta ulkopuoliselle sivustolle eli keskiarvolta hieman yli 6 yhteyttä ulkopuolisiin sivustoihin per sivusto. Liiteessä 1 on listattuna tässä tutkimuksessa esiintyneet sivustot. Lukema on kohtalaisen suuri huomioiden, että vierailtuja sivustoja oli vain 20 kappaletta. Yhteydet ja niistä Lightbeam -lisäosan tekemä graafinen kuvio on nähtävissä alempaa (kuva 6).



Kuva 6. Yhteydet ulkopuolisille sivustoille

Yläpuolella esitetty kuva osoittaa laajan verkoston yhteyksistä, joita sivustojen välillä vallitsee. Kuvassa ympyrä-kuviot esittävät niitä sivustoja, joilla suoranaisesti vierailtiin. Kolmiot ovat ulkopuolisia sivustoja, johon luotiin yhteys vierailtaessa jollain muulla sivustolla. Viivat taas osoittavat sivustojen väliset yhteydet. Yhteydet ovat suurimmaksi osaksi menneet yhteen kasaan. Tämä näyttää kuinka laajasti eri sivustot ovat toisiinsa yhteydessä, vaikka suoria yhteyksiä ei olisikaan. Vain muutamat sivustot ovat jääneet ulkopuolelle, eivätkä ole yhteydessä kovinkaan useaan sivustoon. Liitteissä 1 on nähtävissä jokaisen sivuston osoite, kuinka moneen sivustoon kyseinen sivusto oli yhteydessä ja oliko kyseessä ulkopuolinen vai vierailtu sivusto.

Ayenson ym. (2011) tekemä 100 sivuston tutkimus evästeiden käytöstä osoitti kolmansia osapuolia olleen yli 600, joilta tuli yhteensä 4915 evästettä. Tämä tutkimus ei paljasta evästeiden määrää, mutta kolmansien osapuolien määrä on suhteessa melko sama. 20 sivuston tutkimuksessa paljastui 122 yhteyttä ulkopuoliseen sivustoon, joka antaa keskiarvoksi hieman yli 6. Ayenson ym. (2011) tekemä tutkimus antaa kolmansien osapuo-

lien määrän keskiarvoksi 100 sivustolla noin. 6 kappaletta, koska kolmansia osapuolia oli yli 600.

Tämän työn toisen tutkimuksen tulokset ovat myös mielenkiintoisia. Tässä vierailtiin vain viidellä sivustolla. Jokaisella sivustolla selailtiin tasan kaksi minuuttia. Selailun aikana siirryttiin jatkuvasti uuteen uutiseen tai katsomaan seuraavaa sisältöä. Aluksi kirjoitettiin jokaisen sivuston osoitteen selaimen osoiteriville ja ladattiin sivusto valmiiksi. Näin saadaan viitearvo yhteyksien määrästä, kun verkkosivustolla ei oltu vielä ollut varsinaisesti liikkunut. Yhteyksiä oli ulkopuolisiin sivustoihin 55 kappaletta, kun kaikki viisi sivustoa oli latautunut. Kahden minuutin selailun jälkeen, yhteyksien määrä ulkopuolisiin sivustoihin oli kasvanut 172 kappaleeseen.

Tässä tutkimuksessa yhteyksien määrä kasvoi suuremmaksi, mitä ensimmäisessä tutkimuksessa, jossa avattiin 20 sivustoa. Yhteensä aikaa kului 10 minuuttia. Kymmenen minuutin aikana selailu viestitti siis 172 eri paikkaan tietoa. Huomion arvoista on, että selailun aikana mainokset vaihtuivat sivustoilla sisällön perusteella. Jos siirtyi esimerkiksi asumiseen liittyvään aihepiiriin, niin mainoksia tuli vuokra-asunnoista ja taas sääst ä liittyvässä sisällössä oli mainoksia talvipipoista. Työssä on aiemmin esiintynyt, että mainoksia kohdennetaan mm. aihepiirin mukaan, ja että mainoksen asettaja voi olla ulkopuolinen sivusto. Tämä voi ainakin osaltaan, jos ei jopa kokonaan selittää yhteyksien kasvavan määrän selatessa sivustoja kauemman aikaa ja eri aihepiireissä.

Liitteissä 1 on nähtävillä kaikki ensimmäisessä tutkimuksessa vieraillut ja ulkopuoliset sivustot. Kaikkien osoitteiden takana ei ole varsinaista sivustoa, jolla voi vierailla. Vierailla kuitenkin voi esimerkiksi sitetat.com, suomi24-static.fi, effectvemeasure.net, adtech.de ja adnxs.com sivustoilla. Sitestat.com sivuston etusivulla mainitaan sivuston olevan osa johtavaa markkinatutkimusyhtiötä, joka tutkii ja raportoi Internet -trendejä ja käyttäytymistä. Effectivemeasure.com sivustolla kerrotaan heidän tarjoavan tietoa mm. markkinointikampanjoita varten kohderyhmistä, verkostokäyttäytymisestä, väestörakenteesta ja markkinoinnin tehokkuudesta. Suomi24-static.fi ohjautuu taas suoraan suomi24.fi sivuston etusivulla. Adnxs kertoo sivustollaan olevansa paikka mainoksien huu-

tokaupalle ja Adtech.de kertoo tarjoavansa eri toimijoille mahdollisuuden hallita, tarjota ja tehdä raportteja markkinointikampanjoista.

Monien sivustojen osoitteet eivät anna Internet -selailijalle vihjettä niiden tarkoituksesta. Sivustot kuten d2o307dm5mqftz.cloudfront.net, 2o7.net, revsci.net, rfhub.com, liijt.com, bkrtx.com, burstnet.net jne. eivät kerro vihjettä sivuston käyttötarkoituksesta sivustonimen perusteella. Sivusto 2o7.net ohjautuu kuitenkin adobe.com sivustolle, jossa kerrotaan Adoben sivuston verkostomarkkinointi ja analytiikka palveluista. Burstnet.net sivusto ohjautuu hostwinds.com sivustolle, joka kertoo tarjoavansa mm. verkkopalvelimia yrityksille. Monille sivustoille ei päässyt, joten käyttäjällä ei ole minkäänlaista tietoa, mitä nämä sivustot voivat olla. Tarkempia tietoja verkkopalveluiden rekisteröinnistä on kuitenkin saatavilla ilmaiseksi esimerkiksi http://www.register.com/whois.rcmx sivustolta. Tämä palvelu kertoo esimerkiksi aiemmin mainitun revsci.net sivuston organisaation olevan AudienceScience Inc ja maan olevan Yhdysvallat. Tämän jälkeen ainakin omistajayhtiön tiedot ovat kohtalaisen helposti saatavilla Internet-etsinnän avulla, mutta vieläkään ei pysty tarkalleen tietämään, mitä esimerkin mukaisella sivustolla tehdään.

Sivustot kuten mtv.fi, mtv3.fi, kauppalehti.fi, telkku.com, facebook.net, katsomo.fi ja makuja.fi ovat varmasti monille tuttuja. Selailun aikana sivustolla näkyi erilaisia mainoksia tai julkaisuja toisista medioista. Näihin ulkopuolisiin sivustoihin luultavimmin muodostui yhteys juuri sen takia, koska näistä palveluista oli sivustolla mainos tai muu media. Taulukossa 2 on esitetty useamman sivuston toimintatarkoitus taulukkomuodossa.

Sivusto	Mitä sivusto tarjoaa
Doubleclick.net	Markkinointipalveluja.
Macromedia.com	Ohjautuu adoben.com sivustolle. Tarjoaa erilaisia tietokonesovelluksia mm. Adobe Flash.
Scorecardresearch.com	Tutkii kuinka ihmiset käyttävät Internetiä, mitä ihmiset siellä tekee ja mistä ihmiset eivät Internetissä pidä.
Casalemedia.com	Tarjoaa työkaluja kuluttajien käyttäytymisen arviointiin.
Betradar.com	Tarjoaa tietoa urheilu ja veikkaamise.en liittyvistä asioista
Qservz.com	Sivustolla tulee vain teksti "It Works!".
Almamedia.fi	Digitaalisiin palveluihin ja julkaisutoimintoihin keskittyvä mediakonserni.
Telkku.com	Näyttää televisiosta tulevat ohjelmat.
Kauppalehti.fi	Kauppalehden verkkopalvelu.
Atemda.com	Mainosvaihtokaupan tarjoaja.
Criteo.com	Tarjoaa markkinointitietoa ja sovelluksia.
Pubmatic.com	Tarjoaa markkinointisovelluksia.
Ibillboard.com	Tarjoaa markkinointisovelluksia
Turn.com	Tarjoaa markkinoinnin työkaluja.
Weatherproof.fi	Ohjautuu ilmatieteenlaitos.fi sivustolle, joka tarjoaa tietoa ilmastosta.
Spring-tns.net, dnn506yrbagrg.cloudfront.net, yieldlab.net, vidicosuite.com, semasio.net, adformdsp.net ja gght.com	Näille sivustoille ei päässyt.

Taulukko 2. Ulkopuolisten sivustojen tarkoitus.

Kolmannessa tutkimuksessa yhteyksien määrän saatiin pienennettyä huomattavasti. Yhteyksien määrä ulkopuolisiin sivustoihin oli vielä 125 kappaletta, kun käytettiin HTTPS-Everywhere lisäosaa. Evästeiden poistaminen käytöstä pudotti yhteydet jo 77 ja AdBlock 67 kappaleeseen. JavaScriptin poistaminen käytöstä vaikutti yksittäisesti eniten. Ulkopuolisten yhteyksien määrä oli enää 33, kun JavaScript oli poistettuna, joka on huomattava ero muihin tutkimuksen keinoihin verrattuna. Kaikkien edellä mainittujen lisäosien ja asetusten yhtäaikainen käyttäminen pudotti yhteyksien määrän lopulta 28 kappaleeseen.

Evästeiden ja JavaScriptin poistaminen käytöstä aiheutti sivustojen toimimattomuutta. Live.com sivusto ei esimerkiksi toimi ollenkaan, jos JavaScript on poistettu käytöstä. Sivusto kuitenkin antaa viestin vierailijalle JavaScript -tuen puuttumisesta. Twitter.com antaa taas huomion, että tämä voi vaikuttaa sivuston käyttökokemukseen. Op.fi ilmoittaa, että verkkopalvelu ei välttämättä toimi oikein, jos JavaScript on poistettu käytöstä. Kuvassa 7 on esimerkki JavaScript huomautuksesta live.com sivustolla. Adblock Plus taas vähensi selvästi mainoksien määrään sivustoilla. Tämän käyttäminen ei tuonut esiin huomautuksia eikä sivustojen toimimattomuutta tässä tutkimuksessa.

Kuva 7. JavaScript huomautus live.com sivustolla.

Myös evästeiden käytöstä huomautettiin viidellä sivustolla. Evästeistä huomautettiin google.fi, youtube.com, twitter.com, nordea.fi ja bbc.co.uk sivustoilla. Taulukossa 3 on esitetty kootusti tässä työssä tehdyt tutkimukset ja niiden tulokset.

Sivustovierailun määrä	Käytetty menetelmä	Ulkopuoliset yhteydet	Ulkopuolisten yhteyksien keskiarvo
20	Vierailu sivustolla.	122	6,1
5	Kahden minuutin selailu jokaisella sivustolla.	172	34,4
20	HTTPS-Everywhere.	125	6,25
20	Evästeet poistettu käytöstä.	77	3,85
20	Adblock Plus.	67	3,35
20	Javascript poistettu käytöstä.	33	1,65
20	HTTPS-Everywhere, adblock plus ja evästeet sekä javascript poistettu käytöstä.	28	1,4

Taulukko 3. Tutkimuksen tulokset.

HTTPS -Everywhere lisäosan käyttö ei näyttänyt vaikuttavan ulkopuolisten yhteyksien määrään. Siinä yhteyksien määrä oli hieman jopa suurempi, mitä ensimmäisessä tutkimuksessa, jossa ei käytetty Lightbeam -lisäosan ohella muita lisäosia. Yhteyksien määrä ei luultavammin lisääntynyt kuitenkaan lisäosan takia, vaan yhteyksien määrät saatavat vaihdella. Sivustojen mainokset vaihtelevat ja niiden ylläpitäjät voivat hyödyntää eri teknologiaa eri aikoina, joka on voinut vaikuttaa yhteyksien määrään.

6 JOHTOPÄÄTÖKSET

Työn tarkoituksena oli tutkia kuinka käyttäjiä seurataan Internetissä ja etsiä vastauksia siitä, millaisia uhkia tähän liittyy. Työn tuloksista huomaa, että seuranta on monessa asiassa mukana. Esineiden Internet jo kuvastaa, että verkosto on todella laaja ja monesti arjessa mukana. Kehitys on kasvavaa ja tulevaisuudessa verkosto on entistä laajempi.

Seurannan alkutekijät juurtuvat markkinointiin. Medioiden siirtyminen Internetiin kiihdytti verkkopalveluiden rahoittamista mainostuloin. Tarvittiin kuitenkin luotettavia tutkimustuloksia mainosyleisöstä, joka synnytti seurantaan keskittyneitä yrityksiä. Vaatimukset tunnuslukujen tarkkuudesta ja tehokkuudesta kasvoivat. Yritykset alkoivat kehittää erilaisia liiketoimintamalleja ja alkoivat hyödyntää seurantaa avustavia tekniikoita. Potentiaalisista asiakkaista halutaan tietää aina vain enemmän, joka kasvattaa kerättyjen tietojen tarkkuusvaatimuksia. Teknologisella kehityksellä on myös suurta potentiaalia avustaa ja turvata eri asioita, joka tarkoittaa älykkäämpiä teknologioita ja tietoa ympäristöstä.

Työstä paljastui, että käyttäjästä kerätään monenlaista tietoa useilla erilaisilla tekniikoilla Internet-selailun yhteydessä. Seuranta on mukana ihmisten arjen toimessa erilaisien uusien teknologioiden kuten antureiden avulla enemmän kuin ennen. Uhkia löytyy paljon. Tietoa voidaan jakaa, myydä, varastaa, kerätä tai vuotaa. Tiedon hallinnasta on myös tullut vaikeaa, johtuen sen suuresta määrästä. Tämän lisäksi ongelmallista on selaustietojen uudelleenluonti. Tekniikoita kuten Flash -evästeitä tai ETag otsikkoja käytetään poistettujen historiatietojen uudelleen synnyttämiseen, joka on ongelma. Tällainen käytös on vastakkainen käyttäjän toivomukseen poistaa selauksen historiatietoja. Kaikkien seuranta teknologioiden estäminen saattaa olla hankalaa, johtuen niiden monimutkaisuudesta ja paljoudesta. Tulevaisuuden kehitykseen vaikuttaisi heikentävästi seurannan tehokas estäminen, jonka takia oikean ratkaisun löytäminen yksityisyysaiheisten säännösten asetteluun on vaikeaa.

Empiirisessä tutkimuksessa esiintyneiden yhteyksien määrä kertoo jo hieman siitä, että tiedon siirtyminen voi tapahtua erittäin nopeasti. Aiemmin tässä työssä on kerrottu tiedon jakamisesta tai sen myymisestä eteenpäin. Tässä havainnollistavassa tutkimuksessa siirtyi tietoa mahdollisesti 142 eri sivustoon. Jos jokainen näistä jakaisi toisilleen kaikki tiedot, mitä ovat selailijasta keränneet, olisi tiedon määrä jo aikamoinen. Yhdeltä sivustolta olisi voitu esimerkiksi saada kirjautumisen yhteydessä joitain henkilötietoja, toiselta lääke- tai terveystuotteiden selaustietoja, kolmannelta tietoja huonekaluostoksista jne. Erilaisten profiilien muodostuminen olisi hyvin jo käynnissä. Näiden sivustojen olisi vielä teoriassa mahdollista myydä vielä kerättyä tietoa eteenpäin muille siitä kiinnostuneille, joten verkosto voi olla vielä huomattavasti laajempi.

Toisessa tutkimuksessa kymmenen minuutin vierailu 20 sivustolla muodosti jo 172 ulkopuolista yhteyttä. Kaikki tieto ei aina vaikuta kovin vakavalta, mutta aina ei voi tietää tiedonkerääjän aikeita ja kuinka paljon todellisuudessa näitä tietoja tallennetaan ja hyödynnetään pidemmällä aikavälillä. Juuri pidemmällä aikavälillä progressiivisesti kerätty tieto voi alkaa huolestuttamaan, sillä vähäistäkin tietomäärää pystytään nykyään vertaamaan julkisesti jaettuun tai aiemmin kerättyyn tietoon. Tässä työssä esiintyvät tutkimukset osoittavat, että jo pienestä tietomäärästä on mahdollista tunnistaa yksittäinen käyttäjä sen ainutlaatuisten piirteiden avulla. Kolmannessa tutkimuksessa huomattiin, että lisäosilla ja selainasetuksilla on mahdollista pienentää ulkopuolisten yhteyksien määrää huomattavasti. Tietyt asetukset voivat kuitenkin aiheuttaa palvelujen toimimattomuutta, jonka takia seurannan laaja estäminen huonontaa käyttäjän saamaa kokemusta palveluista.

Tässä työssä on aiemmin esiintynyt, että sivustoilla on parantamisen varaa evästeiden käyttämisen tiedottamisessa. Tämä saattaa olla syynä, miksi kolmannessa tutkimuksessa vain viidellä sivustolla tuli huomautus evästeistä, kun ne olivat poistettu käytöstä. Toinen mahdollinen syy on, että muut sivustot eivät aseta tai käytä evästeitä. Myös evästeiden käyttöä on käsitelty tässä työssä ja tulokset osoittavat niiden olevan laajassa käytössä. Tästä syystä on todennäköisempää, että syynä ei ole evästeiden käytön vähäisyys vaan tiedottamisen puute.

Monet sivustot saavat tuloja Internet -mainoksista. Markkinointi tai sen avulla saavutettu rahavirta sallii käyttäjille ilmaisia palveluita. Teknologinen kehitys parantaa tiettyjen toimintojen turvallisuutta ja mahdollisesti helpottaa kasvavassa määrin arjen asioita. On vaikeaa arvioida, milloin yksityisyys on tärkeämpää kuin mahdollinen kehitys. Markkinointiyritykset, tietoturvan ammattilaiset ja säädösten asettelijat pyrkivät saamaan kiinni Internet-teknologian räjähtävän kasvun, jolloin tietovuodotkin ja muut uhkatekijät mahdollisesti vähenevät. Seuranta tai tiedonkeruu markkinointitarkoituksessa ja kehityksen edistäjänä voidaan pitää positiivisena asiana, mutta ongelmaksi asia muodostuu, kun väärä toimija pääsee käsiksi näihin tietoihin tai toimii itse tiedon kerääjänä. Sivustolla voi olla mainoksia tai liitännäisiä ulkopuolisilta palveluilta, jonka takia sivuston ylläpitäjät eivät välttämättä tiedä käyttäjän seurannasta.

Työn valmistuttua pystyy helpommin ymmärtämään suuntaa, mihin kehitys on menossa. Internet -yhteyteen pystyvät matkapuhelimet ovat olleet jo jonkin aikaa ihmisten mukana. Kasvavassa määrin ihmisten arkeen tulee kuitenkin antureilla varustettuja laitteita. Nämä laitteet kykenevät erilaisiin yhteyksiin ja täten myös keräämään sekä lähettämään tietoa. Ihmisen jokainen liike saatetaan pian tietää ja tästä herääkin kysymys, että kuinka tämän tiedon väärinkäyttö kyetään tehokkaasti estämään, sillä jopa keskustelu television tai puhelimen lähettyvillä saattaa liikkua ulkopuolelle.

Käyttäjän tulee miettiä, mitä tietoja itsestään paljastaa pitäen mielessä, että vapaaehtoisesti luovutetut tiedot voivat siirtyä myös ulkopuolisille tahoille. Tulee myös muistaa, että Internetissä periaatteessa jokainen hiiren painallus on mahdollisesti tallennettavissa. Arkaluonteisen tai henkilötietojen täyttämistä eri sivustoilla tulee aina miettiä. Verkkosivuston keräämät tiedot ovat yhtä turvattuja kuin palveluntarjoajan verkkopalvelu. Tietoa liikkuu niin moneen paikkaan ja niin nopeasti, ettei Internetissä levinnyttä tietoa pysty välttämättä koskaan poistamaan. On hyvä muistaa, että rikollisuutta tapahtuu niin Internetissä, kuin sen ulkopuolella.

Perehtyminen sivuston yksityisyys-selosteeseen on suotavaa, vaikkei se ulkopuolisten asettamien mainoksien, tietovuotojen tai kaikkien seurantateknologioiden avulla tehtyä seurantaa estä. Halutessa voi perehtyä eri seurantateknologioiden estämiseen ja poista-

miseen tarkoitettuihin sovelluksiin, sillä tallennettujen asetusten poistaminen ei välttämättä onnistu helpolla.

Empiirisessä tutkimuksessa käytettiin 20 suomalaisten eniten suosimaa Internet-sivustoa. Tämä on voinut vaikuttaa tuloksien määrään suurentavasti verrattuna siihen, jos sivustot olisi valittu sattumalta. Nämä sivustot ovat olennaisia niiden suosion takia juuri tämän tutkimuksen kannalta. Tulokset saattavat helposti vaihtua, jos tutkimus tehtäisiin uudelleen, sillä mm. sivustojen ylläpitäjät ja ulkoiset mainostajat voivat lisätä tai poistaa omia asetuksia tulevaisuudessa. Tämä vaikuttaisi yhteyksien määrään. Tutkimus osoitti, että yhteyksiä ulkopuolisille sivustoille lähtee useita kappaleita. Yhteyksien määrä on riippuvainen selailun pituudesta. Pelkkä käynti sivustolla luo selvästi vähemmän yhteyksiä, mitä pidempi vierailu.

Jatkotutkimuksena voisi tehdä tässä työssä tehdyt tutkimukset useampaan kertaan eri ajankohtana. Tällä saadaan tuloksia yhteyksien vaihtelevaisuudesta. Jatkossa voisi myös tarkastella keinoja, millä selvittää kerätyt tai siirtyvät selaustiedot. Esimerkiksi tarkastella eri palveluiden tallentamia evästeitä ja mitä tietoa liikkuu selaimen otsikkokentän mukana. Tämän kaltaisella tutkimuksella on mahdollista saada tarkempia tuloksia tallennetuista tai siirtyvistä tiedoista. Tutkimukset yksityisyyden ja tietoturvan sääntelystäkin olisi paikallaan. Eroavaisuuksia löytyy EU:n ja sen ulkopuolen välillä. Internetissä ei välttämättä aina tiedä kenen sääntelyn piiriin sivusto kuuluu.

LÄHTEET

Adblock Plus (2015). [online]. [Lainattu 22.3.2015]. Saatavilla <https://adblockplus.org/en/about>

Acquisti, Alessandro., Ralph Gross & Fred Stutzman (2011). *Face Recognition Study - FAQ*. [online]. [Lainattu 21.02.2015]. Saatavilla: <www.heinz.cmu.edu/~acquisti/face-recognition-study-FAQ/#Q1.2>

Adobe Systems Incorporated (2014). *What are local shared objects?* [online]. [Lainattu 20.2.2014]. Saatavilla: <http://www.adobe.com/security/ flashplayer/articles/lso/>

Alexa (2014). About Us [online]. [Lainattu 9.9.2014] Saatavilla: <http://www.alexa.com/about>

Angwin, Julia ja Jennifer Valentino-DeVries (2010). *Race Is On to 'Fingerprint' Phones, PCs*. The Wall Street Journal. [online]. [Lainattu 27.12.2014]. Saatavilla: <http://www.wsj.com/news/articles/SB10001424052748704679204575646704100959546>

Ayenson, Mika D., Dietrich J, Wambach., Ashkan, Soltani., Nathaniel, Good & Chris Jay Hoofnagle (2011). *Flash Cookies and Privacy II: Now with HMTL5 and ETag Respawning*. [online]. [Lainattu 14.02.2015]. Saatavilla: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1898390

BBC (2015). *Not in front of the telly: Warning over 'listening' TV*. [online]. [Lainattu 21.02.2015]. Saatavilla: <www.bbc.com/news/technology-31296188>

Beales, Howard (2010). *The Value of Behavioral Targeting*. [online]. [Lainattu 15.11.2014]. Saatavilla: <http://www.networkadvertising.org/pdfs/ Beales_NAI_Study.pdf>

Canadian Internet Policy and Public Interest Clinic, CIPPIC (2008). *Online Privacy Threats: A Review And Analysis Of Current Threats*. [online]. [Lainattu 18.01.2015]. Saatavilla: <https://cippic.ca/sites/default/files/ publications/CIPPIC-Online_Privacy_Threats-Final.pdf>

Cheng, Jacqui (2010). Zombie cookie wars: evil tracking API meant to "raise awareness". [online]. [Lainattu 4.3.2015]. Saatavilla: <http://arstechnica.com/business/2010/09/evercookie-escalates-the-zombie-cookie-war-by-raising-awareness/>

Cranor, Lorrie., Manya Sleeper & Blase Ur (2013). *Tracking and Surveillance.* Privacy and Information Technology, IAPP. [online]. [Lainattu 15.02.2015]. Saatavilla: <http://lorrie.cranor.org/pubs/tracking-and-surveillance-chapter-draft.pdf>

Cybersecurity & Privacy foundation (2013). Anonymous Browsing. [online]. [Lainattu 12.3.2015]. Saatavilla: < http://cybersecurityprivacyfoundation.org/Anonymous%20Browsing.pdf>

Dagar, Anil., Yasuhiro, Endo., Abhay, Gupta., Yan, Li., Kuldip, Pabla., Sridhar, Ramaswamy & Ikhlaq Sidhu (2013). *Internet, Economy and Privacy.* [online]. [Lainattu 15.11.2014]. Saatavilla: < http://www.funginstitute.berkeley.edu/sites/default/files/Internet-Economy-and-Privacy.pdf>

Electronic Frontier Foundation (2015a). *HTTPS Everywhere.* [online]. [Lainattu 21.3.2015]. Saatavilla <https://www.eff.org/https-everywhere>

Electronic Frontier Foundation (2015b). *HTTPS Everywhere FAQ*. [online]. [Lainattu 21.3.2015]. Saatavilla < https://www.eff.org/https-everywhere /faq>

European Comission (2015). *Article 29 Data Protection Working Party - PRESS RELEASE - Joint survey by European regulators on website cookie usage finds improvement in information but cookies still being set without consent.* [online]. [Lainattu 4.3.2015]. Saatavilla: <http://ec.europa.eu/ justice/data-protection/article-29/press-material/press-release/art29_press_ material/20150217__wp29_press_release_on_cookie_sweep_.pdf>

European Network and Information Security Agency, ENISA (2012). *Privacy considerations of online behavioural tracking.* [online]. [Lainattu 14.02.2015]. Saatavilla: <http://www.amedeomaturo.com/wp-content/uploads/2012/11/Privacy-considerations-of-online-behavioural-tracking.pdf>

Facebook.com (2015). [online]. [Lainattu 6.3.2015]. Saatavilla: <https://developers.facebook.com/blog/post/2013/04/03/new-apis-for-comment-replies/>

Facebook class action (2015). [online]. Lainattu 6.4.2015]. Saatavilla: <https://www.fbclaim.com/ui/page/updates>

Federal Trade Comission, FTC (2015). *Internet of things. Privacy & Security in a Connected World*. [online]. [Lainattu 31.01.2015]. Saatavilla: < https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127 iotrpt.pdf>

Federal Trade Comission, FTC (2009). *Personal data ecosystem*. [online]. [Lainattu 13.12.2014]. Saatavilla: < https://www.ftc.gov/sites/default/

files/documents/public_events/exploring-privacy-roundtable-series/personaldataecosystem.pdf>

Finlex (2015). Rikoslaki. [online]. [Lainattu 15.02.2015]. Saatavilla: < https://www.finlex.fi/fi/laki/ajantasa/1889/18890039001>

Flanagan, David (2006). JavaScript: *The Definitive Guide.* 5. painos. Sebastopol, California: O'Reilly Media.

Google.com (2015a). Yritystason verkkoanalyysi. [online]. [Lainattu 6.3.2015]. Saatavilla: <http://www.google.com/analytics/>

Google.com (2015b). Google+ Badge. [online]. [Lainattu 7.3.2015]. Saatavilla: < https://developers.google.com/+/web/badge/>

Helopuro, Sanna., Juha, Perttula & Jukkapekka, Ristola (2009). *Sähköisen viestinnän tietosuoja*. 2. painos. Helsinki: Talentum Media Oy.

Hunt, Ira (2013) *Even the CIA is struggling to deal with the volume of real-time social data. The CIA's Grand Challenges With Big Data*. [online]. [Lainattu 04.02.2015]. Saatavilla <https://gigaom.com/2013/03/20/even-the-cia-is-struggling-to-deal-with-the-volume-of-real-time-social-data/2/>

Image.fi (2015). Facebook kuuntelee keskustelusi ja tarjoaa mainoksia sen mukaan. [online]. [Lainattu 10.3.2015]. [Saatavilla: <http://www.image.fi/image-lehti/image-facebook-kuuntelee-keskustelusi-ja-tarjoaa-mainoksia-sen-mukaan>

Instant SSL (2015). *What is HTTPS?*. [online]. [Lainattu 21.3.2015]. Saatavilla: < https://www.instantssl.com/ssl-certificate-products/https.html>

Interactive Advertising Bureau, IAB (2014). *IAB Internet Advertising revenue report*. [online]. [Lainattu 15.11.2014]. Saatavilla: <

http://www.iab.net/media/file/ IAB_Internet_Advertising_Revenue_Report_HY_2014_PDF.pdf>

Järvinen, Petteri (2014). *NSA Näin meitä seurataan.* Jyväskylä: Docendo Oy.

Järvinen, Petteri (2012). *Arjen tietoturva.* Jyväskylä: Docendo Oy.

Järvinen, Petteri (2002). *Tietoturva & yksityisyys*. 2. painos. Jyväskylä: Docendo Finland Oy, Sanoma WSOY-konserni.

Kosinski, Michal., David, Stillwell & Thore Graepel (2013). *Private traits and attributes are predictable from digital records of human behavior* [online]. [Lainattu 13.09.2014]. Saatavilla: <http://www.pnas.org/content/110/15/5802.full.pdf+html>

Krishnamurthy, Balachander., Konstantin, Naryshkin & Craig E. Wills (2011). *Privacy leakage vs. Protection measures*: the growing disconnect [online]. [Lainattu 01.03.2014]. Saatavilla: <http://www.goodtimesweb.org/documentation/2012/w2sp11.pdf>

Krishnamurthy, Balachander & Craig E. Wills (2009). *On the Leakage of Personally Identifiable Information Via Online Social Network.* [online]. [Lainattu 01.10.2014]. Saatavilla: < http://www2.research.att.com/~bala/papers/wosn09.pdf>

Mashable (2014). *Facebook Ads Will Follow You Around. Starting Now.* [online]. [Lainattu 23.11.2014]. Saatavilla: < http://mashable.com/2014/09/29/facebook-ads-atlas/>

Mayer, Jonathan (2011). *Tracking the Trackers: Where Everybody Knows Your Username.* [online]. [lainattu 18.01.2015]. Saatavilla: < cyberlaw.stanford.edu/blog/2011/10/tracking-trackers-where-everybody-knows-your-username>

Mayer, Jonathan R & John C. Mitchell (2012). *Third-Party Web Tracking: Policy and Technology* [online]. [Lainattu 13.05.2014]. Saatavilla: <http://ieeexplore.ieee.org/ielx5/6233637/6234400/06234427.pdf>

Mozilla (2014). About Lightbeam [online]. [Lainattu 9.9.2014] Saatavilla: <https://www.mozilla.org/en-US/lightbeam/about/>

Narayanan Arvind (2011). *There is no such thing as anonymous online tracking*. [online] [Lainattu 01.03.2014]. Saatavilla: < http://cyberlaw.stanford.edu/node/6701>

Navint (2012). *Why is BIG Data Important?.* [online]. [Lainattu 24.01.2015]. Saatavilla: <http://www.navint.com/images/Big.data.pdf>

Nikiforakis, Nick., Alexandors, Kapravelos., Wouter, Joosen., Christopher, Kruegel., Frank, Piessens & Giovanni Vigna (2013). Cookieless Monster: Exploring the Ecosystem of Web-based Device Fingerprinting. [online]. [Lainattu 4.3.2015]. Saatavilla: <https://lirias.kuleuven.be/bitstream/123456789/393661/1/>

Oikeusministeriö (2014). *Tietoverkkorikoksia koskeviin säännöksiin muutoksia – Identiteettivarkaus rangaistavaksi itsenäisenä rikoksena.* [online]. [Lainattu 15.02.2015]. Saatavilla: < http://oikeusministerio.fi/fi/index/ajankohtaista/tiedotteet/2014/11/tietoverkkorikoksiakoskeviinsaannoksiinmuutoksia-identiteettivarkausrangaistavaksiitsenaisenarikoksena.html>

Peppet Scott R. (2014). *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent.* [online]. [Lainattu 31.01.2015]. Saatavilla: < http://www.texaslrev.com/wp-content/uploads/Peppet-93-1.pdf>

Polonetsky Jules & Omer Tene (2013). *Privacy and Big Data:Making Ends Meet*. [online]. [Lainattu 24.01.2015]. Saatavilla: < http://cyberlaw.stanford.edu/files/publication/files/PolonetskyTene.pdf>

Privacy Rights Clearinghouse (2014). *Online Privacy: Using the Internet safely*. [online]. [Lainattu 06.09.2014]. Saatavilla: <https://www.privacyrights.org/online-privacy-using-internet-safely>

Samsung (2015). *Samsung Global Privacy Policy – SmartTV Supplement.* [online]. [Lainattu 21.02.2015]. Saatavilla: <https://www.samsung.com/uk/info/privacy-SmartTV.html?CID=AFL-hq-mul-0813-11000170>

Security Intelligence (2014). SpoofedMe Social Login Attack Discovered by IBM X-Force Researchers. *IBM X-Force Finds Social Login Attack That Allows Intrusion to Many Websites Local Accounts*. [online]. [Lainattu 04.01.2015]. Saatavilla: <http://securityintelligence.com/spoofedme-social-login-attack-discovered-by-ibm-x-force-researchers/#.VKmJOnuzkoF>

Soltani, Ashkan., Shannon Canty., Quentin Mayo,. Lauren Thomas & Chris Jay Hoofnagle (2009). *Flash Cookies and Privacy* [online]. [Lainattu 21.09.2014]. Saatavilla: <www.aaai.org/ocs/index.php/SSS/SSS10/paper/download/1070/1505>

Steinbock, Dan (1998). *Internet ja markkinointiviestinnän muodonmuutos*. Helsinki: Oy Edita Ab.

Tanzina, Vega (2010). *New Web Code Draws Concern Over Privacy Risks.* The New York Times. [online]. [Lainattu 11.10.2014]. Saatavilla: <http://www.nytimes.com/2010/10/11/business/media/11privacy.html?adxnnl=1&pagewanted=all>

Techcrunch (2014). *European Facebook Class Action Suit Attracts 60k Users As It Passes First Court Hurdle.* [online]. [Lainattu 1.10.2014]. Saatavilla: < http://techcrunch.com/2014/08/21/european-facebook-class-action-suit-attracts-60k-users-as-it-passes-first-court-hurdle/>

TheGuardian (2014). *Lawyer suing Facebook overhelmed with support*. [online]. [Lainattu 1.10.2014]. Saatavilla: <http://www.theguardian.com/technology/2014/aug/06/facebook-privacy-action-austria-max-schrems>

Tivi (2014a*). Jo yli 11 000 käyttäjää vaatii Facebookilta rahaa - "Suosio yllätti".* [online]. [Lainattu 1.10.2014]. Saatavilla < http://www.tivi.fi/kaikki_uutiset/jo+yli+11+000+kayttajaa+vaatii+facebookilta+rahaa++quotsuosio+yllattiquot/a1001292>

Tivi (2014b). *Väite: Suomalaiset ovat valmiimpia suostuttelevaan markkinointiin kuin yritykset uskovat.* [online]. [Lainattu 6.1.2015]. Saatavilla < http://www.tivi.fi/kaikki_uutiset/vaite+suomalaiset+ovat+valmiimpia+suostuttelevaan+markkinointiin+kuin+yritykset+uskovat/a1039529>

Twitter.com (2015). Follow Button. [online]. [Lainattu 6.3.2015]. Saatavilla < https://dev.twitter.com/web/follow-button>

Wordpress (2015a). About Us. [online]. [Lainattu 15.02.2015]. Saatavilla: < https://wordpress.com/about/>

Wordpress (2015b). About Wordpress. [online]. [Lainattu 15.02.2015]. Saatavilla < https://wordpress.org/about/>

W3schools.com (2015). *HTML5 Local Storage*. [online]. [Lainattu 15.02.2015]. Saatavilla < http://www.w3schools.com/html/html5_webstorage.asp>

Youyou, Wu., Michal, Kosinski., & David Stillwell (2014). Computer-based personality judgments are more accurate than those made by humans. [online].

[Lainattu 21.2.2015]. Saatavilla: <www.pnas.org/content/112/4/1036.full.pdf+html>

LIITTEET

Liite 1 tutkimuksen sivustot

	Luokka	Sivusto	Sivuja yhteydessä
1	Vierailtu	google.fi	2
2	Ulkopuolinen	gstatic.com	3
3	Ulkopuolinen	google.com	7
4	Vierailtu	facebook.com	8
5	Ulkopuolinen	akamaihd.net	1
6	Vierailtu	youtube.com	7
7	Ulkopuolinen	ytimg.com	2
8	Ulkopuolinen	doubleclick.net	12
9	Ulkopuolinen	ggpht.com	1
10	Ulkopuolinen	googlesyndication.com	4
11	Ulkopuolinen	macromedia.com	7
12	Ulkopuolinen	googleusercontent.com	1
13	Ulkopuolinen	content.googleapis.com	2
14	Vierailtu	wikipedia.org	1
15	Ulkopuolinen	wikimedia.org	1
16	Vierailtu	iltasanomat.fi	20
17	Ulkopuolinen	snstatic.fi	6
18	Ulkopuolinen	sanoma.fi	3
19	Ulkopuolinen	scorecardresearch.com	12
20	Ulkopuolinen	adtech.de	6
21	Ulkopuolinen	facebook.net	3
22	Ulkopuolinen	adform.net	8
23	Ulkopuolinen	adformdsp.net	1
24	Ulkopuolinen	google.analytics.com	9
25	Ulkopuolinen	360yield.com	4
26	Ulkopuolinen	rvty.net	2
27	Ulkopuolinen	casalemedia.com	7

28	Ulkopuolinen	adnxs.com	5
29	Ulkopuolinen	creative-serving.com	2
30	Ulkopuolinen	adsrcr.org	2
31	Ulkopuolinen	betradar.com	2
32	Ulkopuolinen	wtp101.com	2
33	Ulkopuolinen	spring-tns.net	5
34	Ulkopuolinen	dnn506yrbagrg.cloudfront.net	4
35	Ulkopuolinen	adtlgc.com	4
36	Ulkopuolinen	qservz.com	2
37	Vierailtu	iltalehti.fi	29
38	Ulkopuolinen	almamedia.fi	1
39	Ulkopuolinen	telkku.com	1
40	Ulkopuolinen	emediate.se	3
41	Ulkopuolinen	kauppalehti.fi	1
42	Ulkopuolinen	emediate.eu	4
43	Ulkopuolinen	atemda.com	6
44	Vierailtu	twitter.com	5
45	Ulkopuolinen	criteo.com	1
46	Ulkopuolinen	adscale.de	1
47	Ulkopuolinen	yieldlab.net	1
48	Ulkopuolinen	pubmatic.com	2
49	Ulkopuolinen	vidicosuite.com	1
50	Ulkopuolinen	ibillboard.com	1
51	Ulkopuolinen	smartadserver.com	1
52	Ulkopuolinen	semasio.net	2
53	Ulkopuolinen	turn.com	1
54	Ulkopuolinen	chartbeat.com	1
55	Ulkopuolinen	chartbeat.net	1
56	Vierailtu	yle.fi	7
57	Ulkopuolinen	visualrevenue.com	1
58	Ulkopuolinen	leiki.com	2
59	Ulkopuolinen	weatherproof.fi	3
60	Ulkopuolinen	googletagmanager.com	1
61	Vierailtu	hs.fi	23

62	Ulkopuolinen	scribblelive.com	1
63	Ulkopuolinen	2mdn.net	3
64	Ulkopuolinen	serving-sys.com	3
65	Ulkopuolinen	into-digital.fi	1
66	Ulkopuolinen	interquest.fi	1
67	Ulkopuolinen	fbcdn.net	1
68	Ulkopuolinen	bsuiteads.com	1
69	Vierailtu	nordea.fi	3
70	Ulkopuolinen	cdnnordea.com	1
71	Ulkopuolinen	webtrendslive.com	1
72	Ulkopuolinen	nordea.com	1
73	Vierailtu	yahoo.com	7
74	Ulkopuolinen	yimg.com	1
75	Ulkopuolinen	interclick.com	1
76	Ulkopuolinen	footprint.net	1
77	Ulkopuolinen	yieldmanager.com	1
78	Ulkopuolinen	staticflickr.com	1
79	Vierailtu	linkedin.com	6
80	Ulkopuolinen	licdn.com	1
81	Ulkopuolinen	quantserve.com	2
82	Ulkopuolinen	imrworldwide.com	1
83	Ulkopuolinen	demdex.net	2
84	Vierailtu	suomi24.fi	14
85	Ulkopuolinen	suomi24-static.fi	2
86	Ulkopuolinen	rampanel.com	1
87	Ulkopuolinen	de17a.com	1
88	Ulkopuolinen	bidtheatre.com	2
89	Ulkopuolinen	mtv3.fi	1
90	Vierailtu	mtv.fi	16
91	Ulkopuolinen	katsomo.fi	1
92	Ulkopuolinen	lp4.io	1
93	Ulkopuolinen	mtvmedia.fi	1
94	Ulkopuolinen	makuja.fi	1
95	Ulkopuolinen	planet49.com	1

96	Ulkopuolinen	sitestat.com	1
97	Ulkopuolinen	twing.com	2
98	Vierailtu	imdb.com	12
99	Ulkopuolinen	media-imdb.com	1
100	Ulkopuolinen	amazon-adsystem.com	12
101	Vierailtu	amazon.com	6
102	Ulkopuolinen	openx.net	1
103	Ulkopuolinen	burstnet.com	1
104	Ulkopuolinen	contextweb.com	1
105	Ulkopuolinen	rubiconproject.com	2
106	Ulkopuolinen	amazonaws.com	1
107	Ulkopuolinen	sfo9.cloudfront.net	1
108	Vierailtu	live.com	6
109	Ulkopuolinen	gfx.ms	1
110	Ulkopuolinen	bkrtx.com	1
111	Ulkopuolinen	omtrdc.net	1
112	Ulkopuolinen	bluekai.com	2
113	Ulkopuolinen	microsoft.com	1
114	Vierailtu	op.fi	0
115	Vierailtu	imgur.com	15
116	Ulkopuolinen	fmpub.net	1
117	Ulkopuolinen	googletagservices.com	2
118	Ulkopuolinen	thoughtleadr.com	1
119	Ulkopuolinen	liijt.com	14
120	Ulkopuolinen	crowdscience.com	1
121	Ulkopuolinen	googleadservices.com	1
122	Ulkopuolinen	mathtag.com	1
123	Ulkopuolinen	simpli.fi	1
124	Ulkopuolinen	rfhub.com	1
125	Ulkopuolinen	abmr.net	1
126	Ulkopuolinen	chango.com	2
127	Ulkopuolinen	bidswitch.net	1
128	Ulkopuolinen	media6degrees.com	1
129	Ulkopuolinen	sitescout.com	1

130	Ulkopuolinen	company-target.com	1
131	Ulkopuolinen	rtbidder.net	1
132	Vierailtu	bbc.co.uk	11
133	Ulkopuolinen	bbci.co.uk	1
134	Ulkopuolinen	bbcimg.co.uk	1
135	Ulkopuolinen	revsci.net	1
136	Ulkopuolinen	2o7.net	1
137	Ulkopuolinen	effectivemeasure.net	1
138	Ulkopuolinen	gemius.pl	1
139	Ulkopuolinen	bb.com	1
140	Ulkopuolinen	images-amazon.com	1
141	Ulkopuolinen	d2o307dm5mqftz.cloudfront.net	1
142	Ulkopuolinen	ssl-images-amazon.com	1