jamk.fi

Behavioral Tracking – Käyttäjätietojen kerääminen ja käyttäjän jälki verkossa

Katri Ahlgrén

Opinnäytetyö
Joulukuu 2019
Tekniikan ala
Insinööri (AMK), Tieto- ja viestintätekniikan tutkinto-ohjelma

Jyväskylän ammattikorkeakoulu
JAMK University of Applied Sciences

jamk.fi

Kuvailulehti

Tekijä(t)	Julkaisun laji	Päivämäärä
Ahlgrén, Katri	Opinnäytetyö, AMK	Joulukuu 2019
	Sivumäärä 50	Julkaisun kieli Suomi
		Verkkojulkaisulupa myönnetty: Kyllä

Työn nimi
Behavioral Tracking – Käyttäjätietojen kerääminen ja käyttäjän jälki verkossa

Tutkinto-ohjelma
Tieto- ja viestintätekniikka

Työn ohjaaja(t)
Tero Kokkonen
Pasi Hakkarainen

Toimeksiantaja(t)
Jyväskylän ammattikorkeakoulu JAMK / JYVSECTEC

Tiivistelmä

Tavoitteena oli kartoittaa verkossa tapahtuvaa käyttäjädatan keräämistä, mitä uhkia tästä on käyttäjälle ja kuinka näitä riskejä on mahdollista minimoida. Osana Jyväskylän ammattikorkeakoulun IT-instituutin JYVSECTEC (Jyväskylä Security Technology) kyberturvallisuuden tutkimus-, kehitys- ja koulutuskeskuksen ja Poliisiammattikorkeakoulun yhteishanketta nimeltä CYBERDI, ylemmän tason tavoitteena oli löytää käytäntöjä kyberrikoksien estämiseen, tutkimiseen ja selvittämiseen sekä kasvattaa tietoisuutta kyberrikoksista ja -uhkista.

Tutkimuskysymyksenä oli, kuinka käyttäjädataa seurataan verkossa. Tähän kysymykseen haettiin vastauksia, mutta ratkaisuja ei lähdetty toteuttamaan käytännössä, joten tutkimusmenetelmänä käytettiin laadullista eli kvalitatiivista tutkimusmenetelmää. Aineistona toimivat useammat verkkojulkaisut, aiheeseen liittyvät aiemmat tutkimukset ja konferenssijulkaisut sekä alan muu kirjallisuus.

Käyttäjädataa kerätään mainosten kohdentamiseksi, verkkosivujen käyttökokemusten analysointiin ja sen parantamiseksi, tekoälyn kehitykseen sekä hyödynnetään kaupunkisuunnittelussa, rikosten ehkäisyssä ja terroriuhkien havaitsemisessa.

Käyttäjädatan keräys tuo mukanaan riskejä käyttäjälle, mikäli palvelin, jolla tiedot ovat, hakkeroidaan, mikä saattaa johtaa identiteettivarkauteen tai kiristyksen kohteeksi joutumiseen. Tietoja saatetaan myös myydä eteenpäin ja käyttää profilointiin.

Käyttäjädatan keräämistä vastaan on kehitetty erilaisia selaimen lisäosia, ja esimerkiksi VPN on tehokas keino suojata identiteettiä verkossa. Riskejä käyttäjälle datan keräys kuitenkin aiheuttaa niin kauan, kun se on laillista.

Avainsanat (asiasanat)
yksityisyys, kyberturvallisuus, digitaalinen jalanjälki, kohdennettu mainonta, evästeet

Muut tiedot (Salassa pidettävät liitteet)

jamk.fi

Description

Author(s)	Type of publication	Date
Ahlgrén, Katri	Bachelor's thesis	December 2019
		Language of publication: Finnish
	Number of pages: 50	Permission for web publication: Yes

Title of publication

Behavioral Tracking – Collecting user data and user's track online

Degree programme

Information and Communication Technology

Supervisor(s)

Kokkonen Tero

Hakkarainen Pasi

Assigned by

JAMK University of Applied Sciences / JYVSECTEC

Abstract

The objective was to determine the collection of user data online, what risks this causes to the users and how they might be minimized. As a part of JAMK University of Applied Sciences' Institute of Information Technology and JYVSECTEC (Jyväskylä Security Technology) Cybersecurity Research, Advancement and Training Centre and Police University College's joint venture CYBERDI, the goal was also to find practices to prevent cybercrimes, how to research them and also to raise awareness about cybercrimes and threats.

The research question was, how user data is being tracked online. The answer was studied with qualitative research method; however, no actions to remove the problem were implemented. The material consisted of numerous online articles, previous research studies, conferences and other related literature.

User data is constantly collected for targeted advertising to analyze and improve user experience on websites and to develop Artificial Intelligence. Data is also beneficial to urban planning and preventing crimes and terrorism.

If a server where user data is stored is hacked, this might lead to identity theft or users might be blackmailed. Data might also be sold on and/or used for profiling.

To prevent user tracking there are various kinds of browser extensions and VPN is an effective way to protect identity online; however, the risks that user tracking brings cannot be totally minimized as long as data collection is legal.

Keywords/tags (subjects)

privacy, cybersecurity, digital footprint, behavioral tracking, behavioral targeting, big data, cookies

Miscellaneous (Confidential information)

Sisältö

Kuviot

Lyhenteet

AI	Artificial Intelligence, tekoäly
DNT	Do not track, älä seuraa-pyyntö palvelimelle
EFF	Electronic Frontier Foundation
ETag	Entity tag
GDPR	General Data Protection Regulation, Euroopan Unionin tietosuojakäytäntö
GIF	Graphics Interchange Format
HTML5 API	Application Programming Interface, ohjelmointirajapinta
HTTP	Hypertext Transport Protocol
HTTPS	Hypertext Transport Protocol Secure
IKEv2	Internet Key Exchange
IP	Internet Protocol
IPsec	Internet Protocol Security
JYVSECTEC	Jyväskylä Security Technology
L2TP	Layer Two Tunneling Protocol
MITM	Man-in-the-middle-hyökkäys
PII	Personally Identifiable Information, henkilötiedot
PNG	Portable Network Graphics
POLAMK	Poliisiammattikorkeakoulu
PPTP	Point-to-Point Tunneling Protocol
SSTP	Secure Socket Tunneling Protocol
TOR	The Onion Router
ToS;DR	Terms of Service; Didn’t Read-verkkoprojekti tietosuojalausuntojen selventämiseksi
VPN	Virtual Private Network
XSS	Cross-site Scripting-hyökkäys

1 Johdanto

Behavioral Tracking tarkoittaa internetkäyttäjien toimintojen keräämistä ja tallentamista erilaisten evästeiden ja muiden seurantamenetelmien kautta. Tänä päivänä käyttäjädata on valuuttaa, jota myydään yrityksille muun muassa mainonnan kohdentamiseksi. Tämä mahdollistaa sen, että verkkopalvelut ovat käyttäjille usein ilmaisia, mutta samalla kasvaa riski siitä, että dataa käytetään väärin esimerkiksi profiloitiin. Tässä opinnäytetyössä kerrotaan, mitä käyttäjädataa kerätään, miksi ja miten ja kuinka käyttäjän on mahdollista pienentää omaa digitaalista jalanjälkeään.

Aihe käsittelee kyberturvallisuusuhkaa, sillä käyttäjädatan leviäminen saattaa johtaa muun muassa identiteettivarkauteen, ja oikeus omaan dataan heikentää ihmisen päätäntävaltaa ja oikeuksia. Työ tehtiin Jyväskylän ammattikorkeakoulun IT-instituutin JYVSECTEC (Jyväskylä Security Technology) kyberturvallisuuden tutkimus-, kehitys- ja koulutuskeskuksen ja Poliisiammattikorkeakoulun (POLAMK) yhteishankkeelle nimeltä CYBERDI. Projektin tavoitteena on kehittää käytäntöjä kyberrikoksien estämiseen, tutkimiseen ja selvittämiseen sekä kasvattaa tietoisuutta kyberrikoksista ja -uhkista (CYBERDI 2019). Jokaisella ihmisellä on oikeus tietää, mitä tietoja heistä kerätään ja mihin näitä tietoja käytetään.

Käyttäjätietoja kerätään muun muassa sosiaalisesta mediasta käyttäjien tilien kautta, mikä on aiheuttanut huolta viime vuosina. Suomen kyberturvallisuusstrategiassa (2019) on yhtenä linjauksena kyberturvallisuuden osaamisen kehittäminen kaikkien yhteiskunnan toimijoiden keskuudessa, ja strategiassa on maininta, että ”Kansallisesti on varmistettava, että jokaisella on riittävät valmiudet toimia turvallisesti digitaalisessa toimintaympäristössä.” Koska teknologia kehittyy niin nopeasti, ei voi luottaa siihen, että viranomaiset ja laki pitävät huolta käyttäjän yksityisyydestä, vaan jokaisen on otettava itse vastuuta siitä, mihin antaa luvan, kun puhutaan esimerkiksi kolmannen osapuolen evästehyväksynnästä. Aihe on mielenkiintoinen, koska käyttäjätietoja keräämällä ei pelkästään kohdenneta mainontaa, vaan tiedoilla on jo vaikutettu muun muassa Yhdysvaltojen presidentin vaaleissa vuonna 2016 (Cadwallard & Graham-Harrison 2018). Mitä enemmän toisesta tietää, sitä helpompi on vaikuttaa hänen päätöksiinsä.

Aiheesta on kirjoitettu joitain opinnäytetöitä jo aiemmin, ja artikkeleita ja videoita aiheesta löytyy paljon. Ajankohtaisuuden puolesta aiheesta kuitenkin riittää vielä kirjoitettavaa, sillä teknologiaa, niin käyttäjätietojen etsimiseen kuin piilottamiseenkin, kehitetään jatkuvasti.

Työssä kartoitettiin käyttäjätietojen keruun tuomia uhkia ja miksi ihmisten pitäisi välittää omasta yksityisyydestään. Tarkoituksena ei ole ajaa ihmisiä pois sosiaalisesta mediasta tai kehottaa palaamaan takaisin aikaan ennen, nykyään päivittäin käytössä olevia palveluita, vaan kertoa, mitä tietoja nämä palvelut keräävät käyttäjästä ja mihin niitä käytetään. Kun ihmisillä on tieto siitä, että tällaista tapahtuu, on mahdollisuus vaikuttaa siihen, että asiat muuttuisivat. Jo tieto siitä, miten ja mitä dataa kerätään, riittää siihen, että ihmisistä tulee vastuuntuntoisempia sen jakamisessa ja useammat alkavat vaatia oikeuksia omiin tietoihinsa ja estää niiden jakamista kolmansille osapuolille.

2 Tutkimusasetelma

2.1 Tutkimuskysymys

Tämän tutkimuksen tutkimusongelmana oli käyttäjädatan kerääminen internetissä: Kuinka käyttäjädataa seurataan verkossa? Tutkimuskysymys jakautui seuraaviin alikysymyksiin:

- Kuinka ja miksi käyttäjätietoja kerätään verkossa?
- Mitä riskejä käyttäjätietojen kerääminen tuo käyttäjille?
- Kuinka riskejä on mahdollista minimoida?

2.2 Tutkimusmenetelmä

Tutkimuksen tarkoituksena oli löytää edellisessä kappaleessa mainittuun tutkimuskysymykseen vastaus kuvailemalla käyttäjädatan seurantaan liittyviä ilmiöitä. Tutkimusmenetelmänä käytettiin kvalitatiivista eli laadullista tutkimusmenetelmää.

Kvalitatiivisessa tutkimuksessa harjoitetaan kokonaisvaltaista tiedon hankintaa ja aineisto kootaan luonnollisissa, todellisissa tilanteissa ja toisin kuin kvantitatiivisessa eli määrällisessä tutkimuksessa työn ilmiöitä ei lähdetty kuvaamaan ja tulkitsemaan tilastojen ja numeroiden avulla. Kvalitatiivisessa tutkimuksessa aineistoa tarkastellaan monelta taholta, yksityiskohtaisesti ja pyritään paljastamaan odottamattomia seikkoja ilmiöstä (Hirsjärvi, Remes & Sajavaara 2009, 164). Näiden lisäksi kvalitatiivinen tutkimus on valittu tämän työn tutkimusmenetelmäksi, koska siinä tutkimussuunnitelmaa muutetaan olosuhteiden mukaisesti ja aineisto määrää prosessin pituuden. Kvalitatiivisen tutkimuksen tuloksista halutaan saada syvällinen näkemys ja ilmiöstä hyvä kuvaus (Kananen 2015, 69-71). Kvalitatiivisessa tutkimuksessa tutkija ei itse määrää sitä, mikä on tärkeää (Hirsjärvi, Remes & Sajavaara 2009, 164).

Kvalitatiivisessa tutkimuksessa aineistona toimivat havainnointi eri muodoissa, dokumentit ja erilaiset haastattelut, jotka liittyvät tutkittavaa ilmiöön (Kananen 2015, 81). Tässä opinnäytetyössä aineistoa analysoitiin siinä muodossa kuin ne on löydetty, eli kirjallisena tai suullisena, ja eri lähteistä saatu aineisto on yhteismitallistettu eli muutettu tekstimuotoon, jota on vuorostaan analysoitu työssä.

3 Käyttäjän jälki verkossa

3.1 Yksityisyys verkossa

Kun puhutaan yksityisen datan keräämisestä, on syytä määritellä, mikä on yksityistä dataa. PII (Personally Identifiable Information) eli henkilötiedot ovat yksityistä dataa. Euroopan Unionin tietosuoja-asetuksen mukaan henkilötieto on sellaista tietoa, josta henkilön voi tunnistaa suoraan tai välillisesti kuten nimi, kotiosoite, sähköposti, sijaintitiedot tai IP-osoite (Internet Protocol-osoite) (Mikä on henkilötieto? n.d). Pseudonymisoidut tiedot ovat Suomessa luettu myös henkilötiedoiksi, vaikka niistä suoraa ei pysty henkilöä tunnistamaan (Pseudonymisoidut ja anonymisoidut tiedot n.d). Jos taas tiedot ovat anonymisoitu, se tarkoittaa, että henkilöä, jota tieto koskee, ei voi enää tunnistaa, eikä dataa voi palauttaa sellaiseen muotoon, josta pystyisi tunnistamaan, ja näitä tietoja ei enää katsota henkilötiedoiksi (mt).

Yksityistä dataa eivät välttämättä ole pelkät henkilötiedot, vaan mikä tahansa tieto, jonka henkilö haluaa pitää vain omassa ja niiden tiedossa, joille hän on päättänyt tämän tiedon jakaa. Ihmiset saattavat sanoa, että he eivät niin välitä yksityisyytensä invaasiosta, koska heillä ei ole mitään salattavaa. Glenn Greenwald (2014) kertoi Tedtalk-puheessaan, että kun hän on kohdannut ihmisiä, jotka näin väittävät, hän antaa heille sähköpostiosoitteen ja pyytää henkilöä lähettämään hänelle kaikki sähköpostitilit ja salasanat, jotta Greenwald voi tutkia, mitä hän tekee ja julkaista tietoja, jotka haluaa. Yksikään ihminen ei ollut lähettänyt tietoja Greenwaldille, kun Tedtalk kuvattiin ja tuskin tulee lähettämäänkään, koska vaikka ihminen ei olisikaan paha, jokaisella on jotain salattavaa.

Yksityisyydessä ei kuitenkaan ole kyse siitä, että välttämättä haluaisi salata jotain, vaan valinnasta. Ihmisten täytyy olla vapaita valitsemaan, mitä tietoja he haluavat jakaa, milloin ja kenelle. Usein kun ihmisen asioista kerrotaan esimerkiksi toimittajan kautta lehdessä tai verkossa ilman kyseisen henkilön lupaa tai suostumusta, tuntee henkilö, että hänen yksityisyyttään on loukattu. Yksityisyys vaikuttaa merkitsevästi demokratiaan, talouteen ja henkilökohtaiseen hyvinvointiin. Mitä enemmän tietoja kerätään ilman lupaa, sitä enemmän vapaus rajoittuu, ja mitä vähemmän on valtaa omaan yksityiseen dataan, sitä vähemmän on valinnanvapautta. (Claypoole & Payton 2014, 1–3.)

Tietosuoja-asetus henkilötietojen suojana

25.5.2018 EU:n alueella astui voimaan tietosuoja-asetus GDPR (General Data Protection Regulation), jolla pyritään mahdollistamaan parempi suoja henkilötiedoille ja niiden käsittelylle. GDPR:n myötä ihmisillä, jotka asuvat EU maassa, on oikeus pyytää yrityksiltä ja organisaatioilta (myös niiltä, jotka ovat EU:n ulkopuolella), mitä henkilötietoja heistä on tallennettu ja mihin tarkoitukseen niitä käsitellään. Tarvittaessa väärät tai epätarkat tiedot täytyy korjata ja kaikki tiedot pitää olla mahdollista poistaa pyydettäessä silloin, kun niiden käsittelylle ei ole laillista perustetta. Tietosuoja-asetus antaa oikeuden myös vastustaa henkilötietojen käsittelyä, ja niiden käsittelyn ra-

joittamista voi pyytää. Omat henkilötiedot on myös mahdollista siirtää toiselle organisaatioille. GDPR soveltuu kuitenkin käytettäväksi ainoastaan silloin, kun kyseessä on henkilötiedot. (Usein kysyttyä EU:n tietosuoja-asetuksesta n.d.)

3.2 Digitaalinen jalanjälki

Joka kerta kun käyttäjä avaa selaimen, hän jättää vierailusta jäljen verkkoon. Joka kerta kun verkkosivu kysyy, hyväksytäänkö evästeet, ja käyttäjä painaa kyllä tai ei, hän jättää itsestään jäljen. Jokainen verkossa tehty klikkaus ja toiminta jättää jälkeensä kirjauksen, ja näistä digitaalisista jäljistä ja henkilökohtaisesta datasta saatua kokoelmaa kutsutaan *digitaaliseksi jalanjäljeksi*. Digitaalinen jalanjälki voidaan jakaa kahteen eri ryhmään: aktiivinen ja passiivinen. Aktiivinen jalanjälki syntyy, kun käyttäjä itse jakaa jotain, kuten valokuvia, ja on tietoinen siitä, että jättää verkkoon jotain. Passiivinen jalanjälki taas syntyy ilman, että käyttäjä on välttämättä tietoinen siitä, että jotain on jäänyt johonkin muistiin. (Gencoglu, Honko, Isomursu & Similä 2015.)

Aktiivinen digitaalinen jalanjälki on käyttäjän omassa hallinnassa hyvinkin helposti; ei julkaise kuvia verkossa, ei jaa tykkäyksiä Facebookissa tai kommentoi foorumeilla niin paljon. Passiivinen jalanjälki koostuu tiedoista, joita kerätään ”huomaamatta”, kuten IP-osoite, sijainti, selaimen tiedot, mikä käyttöjärjestelmä tietokoneessa on ja mitkä ovat käyttäjän ostostottumukset verkossa (Holland & Jones 2019). Kuviossa 1 on esimerkki BrowserLeaks-sivustolta, jossa pystyy näkemään, mitä tietoja jokainen verkkosivu pystyy keräämään ilman lupaa. Passiivisen digitaalisen jalanjäljen pienentämiseksi täytyy tehdä vähän enemmän töitä ja näitä keinoja avataan luvussa 4.2.

What Is My IP Address

My IP Address :

IP address	195.148.26.18
Hostname	nat18.labranet.jamk.fi

IP Address Location :

Country	Finland (FI)
State/Region	Central Finland
City	Jyväskylä
ISP	CSC - Tieteen tietotekniikan keskus Oy
Organization	Jyvaskyla University of Applied Sciences
ASN	AS1741 Tieteen tietotekniikan keskus Oy
Connection Type	Cable/DSL
Timezone	Europe/Helsinki
Local Time	Thu, 07 Nov 2019 10:14:45 +0200
Latitude/Longitude	62.2294,25.7191

IPv6 Leak Test :

IPv6 Address	n/a

WebRTC Leak Test :

Local IP address	192.168.44.37
Public IP address	n/a

Flash Leak Test :

Flash IP address	

TCP/IP Fingerprint :

Passive, SYN	Windows NT kernel | Language: Unknown | Link: Ethernet or modem | MTU: 1500 | Distance: 16 Hops

DNS Leak Test :

	IP Address :	ISP :	Location :
Your DNS Servers	195.148.26.4	CSC - Tieteen tietotekniikan keskus Oy	Finland, Jyväskylä

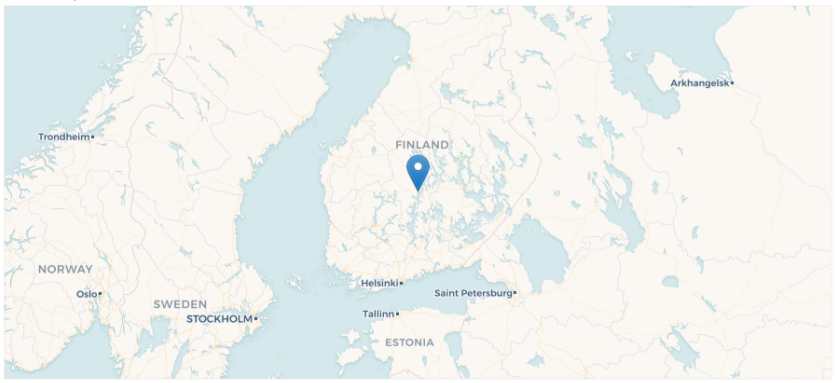
HTTP Headers :

Request	GET /ip HTTP/2.0
Host	browserleaks.com
User-Agent	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:68.0) Gecko/20100101 Firefox/68.0
Accept	text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language	en-US,en;q=0.5
Accept-Encoding	gzip, deflate, br
Referer	https://browserleaks.com/
Upgrade-Insecure-Requests	1
TE	trailers
Cookie	_ga=GA1.2.1328585326.1573114483; _gid=GA1.2.726092166.1573114483; _gat=1

Tor Relay Details :

Relays	This IP is not identified to be a Tor Relay

Where is My IP :



IP Address Whois :

Source Registry	RIPE NCC
Net Range	195.148.26.0 - 195.148.27.255
Name	JYPOLY-NET2B
Handle	195.148.26.0 - 195.148.27.255
Parent Handle	195.148.0.0 - 195.148.255.255
Net Type	ASSIGNED PA
Country	FI
Last Changed	Mon, 23 Feb 2015 12:25:55 GMT
Description	Jyvaskyla University of Applied Sciences Rajakatu 35 40200 Jyvaskyla

Full Name	Abuse-C Role
Handle	AR31397-RIPE
Entity Roles	Abuse
Email	abuse@jamk.fi
Organization	ORG-JAMK1-RIPE
Address	Rajakatu 35 40200 Jyvaskyla, FI

Kuvio 1. BrowserLeaks-sivuston keräämät IP-tiedot

Digitaalinen jalanjälki syntyy käyttäjän omien toimintojen perusteella ja sisältää tietoa hänestä itsestään. Analysoimalla tätä tietoa voi oppia paljonkin käyttäjästä, mistä hän pitää, mitä hän harrastaa, katsoo, kuuntelee, mitä hän on ostanut, milloin ja mistä, onko hän uskonnollinen ja kuinka aktiivisesti hän seuraa politiikkaa. Näitä tietoja voidaan käyttää mainosten ja palveluiden kohdentamiseen ja räätälöintiin käyttäjäystävällisemmäksi. Tietoja myös myydään eteenpäin kolmansille osapuolille, mikä mahdollistaa muun muassa sen, että verkkopalvelut ovat käyttäjille ilmaisia, mutta myös altistaa käyttäjät profiloinnille. (Mt.)

Barfin Bakir (2019) tutki loppukäyttäjien suhtautumista digitaaliseen jalanjälkeen verkkokyselyn avulla. Tulokset osoittivat, että vaikka käyttäjät ovat tietoisia, mistä digitaalinen jalanjälki koostuu, 34,8 % vastaajista ei osannut sanoa, onko huolissaan digitaalisesta jalanjäljestään. Bakirin totesi tutkimuksen perusteella, että käyttäjät eivät halua, että heistä kerätään tietoja, koska se nähdään yksityisyyden loukkaamisena, mutta eivät kuitenkaan ole valmiita muuttamaan toimintaansa verkossa. (Bakir 2019, 29–34.)

3.3 Behavioral Targeting – kohdennettu mainonta

Ihmiset saattavat olla joskus varmoja siitä, että heidän älypuhelimensa kuuntelee keskusteluja, joita he käyvät muiden kanssa, koska sovelluksessa tulee vastaan mainos asiasta, josta he juuri puhuivat, mutta jota he eivät ole kuitenkaan esimerkiksi hakeneet Googlesta. Mainokset kuitenkin saattavat ilmestyä sen takia, että joku toinen käyttäjä samassa lähiverkossa on etsinyt asiaa Googlesta, ja siksi mainos näkyy myös toisten laitteilla. Tämä kuitenkin kertoo myös siitä, kuinka hyvin nykypäivän algoritmit osaavat ennustaa ihmisen käyttäytymistä. (Rautanen 2019.)

Behavioral targeting eli käyttäytymiseen perustuva kohdentaminen rakentuu käyttäjien digitaalisen jalanjäljen perusteella. Se voidaan jakaa kahteen tyyppiin: yhdellä sivustolla ja muualla verkossa tapahtuvaan kohdentamiseen. Sivustolla tapahtuva kohdentaminen on yleensä implementoitu osaksi sivuston personointia, jotta se pystyy näyttämään mainoksia, jotka perustuvat dataan, jotka on saatu käyttäjän toimista samalla sivustolla. Esimerkiksi verkkokaupassa, jossa on katsellut kuulokkeita, tulee

yleensä mainoksia ”sinua saattaisi kiinnostaa nämäkin kuulokkeet”, joka on kohdentunut sivustolla tapahtuvien käyttäjätoimien kautta. Verkossa tapahtuva kohdentaminen tarkoittaa laajempaa datan keräystä, eli eri sivustoilta saatu tieto ajetaan algoritmiin, joka datan perusteella pyrkii määrittämään muun muassa henkilön iän, sukupuolen ja mitä tuotteita hän on mahdollisesti kiinnostunut ostamaan ja lajittelee käyttäjät näiden tietojen perusteella erilaisiin ryhmiin. (Wlosik n.d.)

Tämän päivän trendi on, että käyttäjän ei tarvitse maksaa verkkopalvelusta tai sovelluksesta, mutta koska yritys kuitenkin tarvitsee tuloja, ne hankitaan usein mainonnan kautta. Facebook on suosittu mainostajien keskuudessa, koska se tarjoaa muun muassa analyysi- ja analytiikkapalveluita kolmansille osapuolille, tilastojen ja tietojen avulla, jotka se on kerännyt Facebook-käyttäjien datan perusteella (Tietokäytäntö 2018).

Eräässä Harvardin yliopiston tutkimuksessa esitettiin, että mitä enemmän käyttäjät tietävät, mitä kautta mainokset ovat heille valikoituneet, sitä vähemmän ihmisiä kiinnosti ostaa kyseinen tuote. Tutkimuksessa näytettiin kohdehenkilöille mainoksia ja samalla kerrottiin, miksi kyseinen mainos oli heille valikoitunut. Mainoksista, joista paljastui, että sen kohdentamiseksi on käytetty dataa, joka oli saatu muilta käyttäjän vierailemilta sivuilta, väheni tehokkuus huomattavasti enemmän, kuin taas jos kerrottiin, että mainos on kohdentunut käyttäjän toimista samalla sivustolla tai ei kerrottu ollenkaan, miksi kyseinen mainos näkyi. Näin ollen mainostajien ei olisi järkevää paljastaa, kuinka markkinointi tapahtuu, ja harvemmin verkkosivuilla näkeekään selvää informaatioita siitä, mitä kautta mainos on valikoitunut. Muun muassa Facebookia on kuitenkin kritisoitu siitä, että se ei kerro tarkemmin, kuinka sen mainontapalvelut toimivat. (Barasz, John, & Kim 2018.)

GDPR on hankaloittanut mainonnan kohdentamista ja toinen vaihtoehto on contextual targetin, joka ei vaadi käyttäjäseurantaa.

3.4 Big data

3.4.1 Mitä on big data?

Big data eli niin kutsuttu isodata on vaikea määritellä, mutta yleisesti se tarkoittaa suurten tietomassojen keräämistä ja analysointia, johon ei pystytä enää normaaleja data-analysointityökaluja käyttäen. Big datan ominaisuuksia kuvataan yleisesti kolmen V:n avulla:

- Volyymi (Volume) kuvaa datan määrää. Esimerkiksi Facebookista saatava käyttäjädata kuten kuvat, videot ja tykkäykset, mutta myös auton tietokoneen sensoreiden keräämä data, älykellon keräämä tieto käyttäjän sykkeestä ja puhelinsovellukset luovat eksponentiaalisesti uutta dataa.
- Vauhdilla (Velocity) kuvataan datan kertymisen nopeutta tietojärjestelmiin sekä aikaa, joka kerätyn datan prosessointiin tarvitaan, jotta saadaan käyttökelpoista tietoa ulos.
- Vaihtelevuudella (Variety) kuvataan datan monimuotoisuutta. Data tulee erilähteistä ja eri formaateissa aina audio-formaatista json-dataan. (Salo 2013, 20 - 22.)

Hajautetun datan prosessointia varten on tiettyjä ohjelmistokehyksiä kuten Hadoop, joka on avoimenlähdekoodin ohjelmistokehys Apache Software Foundationilta (Jantunen 2017, 9).

Big datan avulla organisaatiot pystyvät keräämään dataa eri kohteista valtavia määriä, mutta pelkästä datasta ei ole mitään hyötyä, vaan sitä pitää osata myös jäsennellä ja löytää oleellinen tieto. Tähän löytämiseen tarvitaan vielä usein ihmistä, mutta tähänkin on tulossa muutos kehittyvän tekoälyn eli AI:n (Artificial Intellegence) avulla. Kun data on valmis, se voidaan analysoida kehittyneempien analyysiprosessien tavoin, joka sisältää työkaluja

- tiedonlouhintaan
- ennustavaan analytiikkaan

- koneoppimiseen (machine learning)
- syväoppimiseen (deep learning). (Rouse 2019.)

3.4.2 Mihin big dataa hyödynnetään?

Big data-analysoinnin avulla yritykset pyrkivät löytämään ennakoivia malleja muun muassa mahdollisten asiakkaiden selainhistorian ja sosiaalisen media datan avulla, jotta pystytään tarjoamaan sellaisia tuotteita, joista asiakas on valmis maksamaan. Analytiikan avulla on mahdollista tunnistaa mahdollisia uusia asiakkaita sekä tehdä havaintoja omasta henkilökunnasta ja keitä yritykseen kannattaa palkata (Michalowicz 2018). Yrityksille data, johon heillä on niin sanotusti yksinoikeus, tuottaa valtavasti kilpailuetua (Salo 2013, 32).

Big datan avulla on mahdollista tehdä tulevaisuuden arvioita, ei vain yritysten asiakkaista, mutta myös yhteiskuntatasolla esimerkiksi kaupunkisuunnittelussa, rikosten ehkäisyssä ja terroriuhkien havaitsemisessa (Khojaye & Shamsi 2018, 73).

Big datan avulla pystytään myös kehittämään tekoälyä koneoppimisen eli machine learningin kautta. Esimerkiksi sen sijaan että itseohjautuvalle autolle koodataan kaikki liikennesäännöt, auton tietokoneelle syötetään dataa ja annetaan koneen itse opetella ne datan perusteella. (Cukier 2014.)

Valtiot eri puolilla maailmaa käyttävät eri valvontaohjelmia eri tarkoituksiin aina liikennerikkomuksien havaitsemisesta, kansallisen turvallisuuden parantamiseksi valtion vastaisten toimintojen tunnistamisessa. Tämäkin toiminta hyödyntää big dataa. Vaikka tarkoituksena on parantaa kansalaisten turvallisuutta, se sisältää riskit yksityisyyden vaarantumisesta. Palveluiden tarjoajien keräämä tieto sisältää samat riskit. Vaikka käyttäjiltä pyydetäänkin useilla sivuilla evästelupa, se ei kuitenkaan tarkoita sitä, että käyttäjät ymmärtäisivät täysin, mihin antavat suostumuksensa. (Khojaye & Shamsi 2018, 74.)

3.4.3 Data brokers – “datan välittäjät”

Data brokers eli ”datan välittäjät” on multimiljoona teollisuus, jossa yritykset keräävät ja ostavat, pakkaavat ja myyvät yksityiskohtaisia tietoja verkossa toimivista käyttäjistä. Datatiedot koostuvat käyttäjien sosiaalisesta mediasta, selaushistoriasta, verkko-ostoshistoriasta, luottokorttitapahtumista sekä henkilötiedoista. Yritykset operoivat eri tavoin, mutta yleisesti ne myyvät eteenpäin kategorisoituja listoja ihmisten tiedoista, esimerkiksi ”fitness intoilijat” tai ”uudet vanhemmat”, mutta myös kyseenalaisempia listoja on tullut julki, jotka käsittelevät muun muassa ihmisten terveystietoja kuten ”HIV:stä kärsivät”. Jossain tapauksissa datavälittäjät, kuten Datacoup, maksavat käyttäjille, jotta ne saavat pääsyn käyttäjän sosiaalisen median kanaville ja tiedot maksukorttiostoksista. (Wlosik 2019.)

Datan välittäjiä käytetään markkinoinnin ja mainonnan lisäksi esimerkiksi arvioimaan, onko lainanhakijan antamat tiedot paikkansa pitäviä vai onko pankki myöntämässä lainaa huijarille. Myös vakuutusyhtiöt voivat käyttää data brokereita arvioidakseen, millä hinnalla kannattaa tietylle asiakkaalle myydä vakuutus. (Mt.)

3.5 Riskit käyttäjille datan keruussa

3.5.1 Tietoturvaloukkaukset

Tietoturvaloukkaus tarkoittaa sitä, kun luvaton taho pääsee käsiksi arkaan, suojattuun tai luottamukselliseen dataan tai tapahtumaa, jonka seurauksena henkilötietoja tuhoutuu, häviää tai muuttuu. Tietoturvaloukkauksella on yleensä negatiivinen vaikutus organisaatioon tai yritykseen, jota se koskee kuin myös sen asiakkaisiin. GDPR:n myötä yritysten ja organisaatioiden täytyy informoida henkilöitä, joita tietovuoto koskee sekä valvontaviranomaisia 72 tunnin kuluessa siitä, kun tietoturvaloukkaus on huomattu. (Tietoturvaloukkaukset n.d.)

3.5.2 Kiristys

Jos henkilö on julkaissut esimerkiksi anonyymissa blogissa tekstin, jossa hän kertoo omasta uskonnostaan, hänellä on syy siihen, miksi hän haluaa pitää tekstin anonyymina. Kuitenkin blogin ylläpitäjä, tässä tapauksessa se ei ole kirjoittaja itse, saattaa tietää kirjoittajan henkilöllisyyden esimerkiksi rekisteröidyn sähköpostiosoitteen perusteella tai IP-tiedoilla. Mikäli sivusto hakkeroidaan ja tapahtuu tietovuoto, saattaa hakkeri kiristää käyttäjältä rahaa tai muuten hän julkaisee tekstin kirjoittajan identiteetin, mikä saattaa aiheuttaa psyykkistä ahdistusta, nöyryytystä tai maineen menetyksen. Nämä kaikki seuraukset katsotaan olevan vakavan tietoturvaloukkauksen aiheuttamia, kuten myös identiteettivarkaus ja petos. (Tietoturvaloukkaukset n.d.)

3.5.3 Identiteettivarkaus

Identiteettivarkaus tarkoittaa, että esiinnytään toisen henkilöllisyydellä ja tarkoituksella erehdytetään kolmatta osapuolta. Mikäli palvelin, jossa henkilötietoja säilytetään, hakkeroidaan ja tapahtuu tietovuoto, on mahdollista, että vuotaneita tietoja käyttää joko hakkeri itse tai levittää muille. Rikollinen saattaa esiintyä toisena ihmisenä sosiaalisessa mediassa ja julkaista sisältöä, josta on haittaa uhrille tai aiheuttaa taloudellista vahinkoa tekemällä verkko-ostoksia uhrin nimiin. Rikollinen saattaa käyttää varastettuja tietoja vielä vuosia sen jälkeen, kun tietovuoto on tapahtunut. Identiteettivarkaus tuli rangaistavaksi 4.9.2015 ja on rangaistava vain siinä tapauksessa, että siitä aiheutuu taloudellista vahinkoa tai vähäistä suurempaa haittaa uhrille. (Identiteettivarkaudessa esiinnytään toisen henkilöllisyydellä n.d)

3.5.4 Profilointi

Mitä enemmän teknologian avulla kerätään käyttäjädataa, sitä helpommin yritykset, valtiot ja rikolliset pystyvät sitä analysoimaan ja luomaan kuvaa käyttäjästä. Mikäli tähän tarkoitukseen on käytetty henkilötietoja automaattisen käsittelyn kautta, on kyse profiloinnista. Profiloinnissa henkilöt sijoitetaan johonkin tiettyyn kategoriaan tai ryhmään henkilökohtaisten ominaisuuksien perusteella ja tarkoituksena on usein

analysoida tai ennakoida ihmisten kykyä suoriutua jostain tehtävästä, mielenkiinnonkohteita ja todennäköistä käyttäytymistä. (Automaattinen päätöksenteko ja profilointi n.d.)

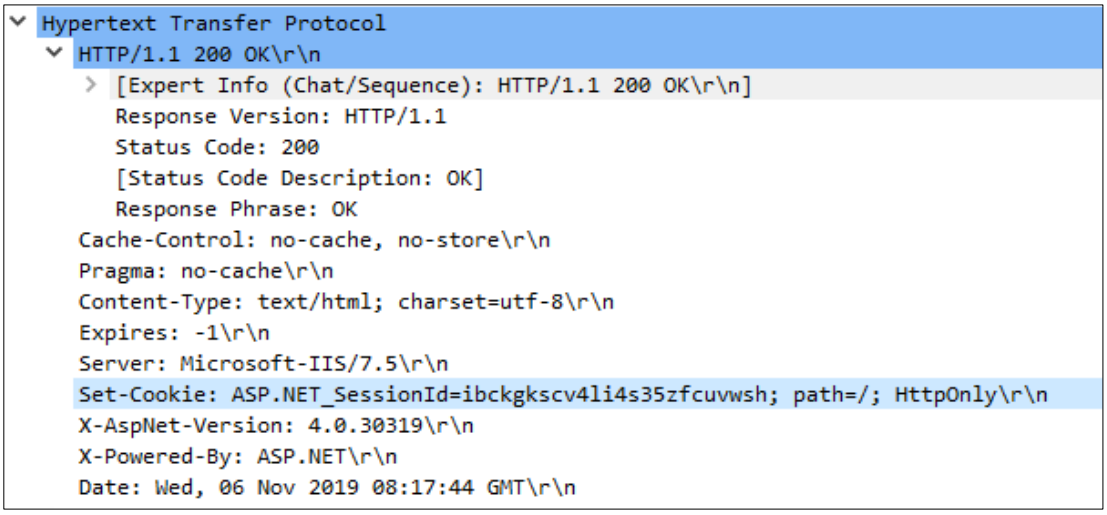
Data-analysoinnin avulla ei vaikuteta pelkästään ihmisten ostostapoihin, vaan sen avulla voidaan vaikuttaa jopa yhteiskuntatasolla esimerkiksi vaaleissa. Vuoden 2016 Yhdysvaltojen presidentinvaalien aikana, laittomasti Facebookista louhitun datan avulla, voitiin profiloida äänestäjiä ja suunnata heille sosiaalisessa mediassa räätälöityä sisältöä, jolla pyrittiin kääntämään kannatus Trumpin puolelle (Cadwalladr & Graham-Harrison 2018).

4 Passiivisen digitaalisen jalanjäljen seuranta ja peittäminen

4.1 Kuinka käyttäjädataa kerätään?

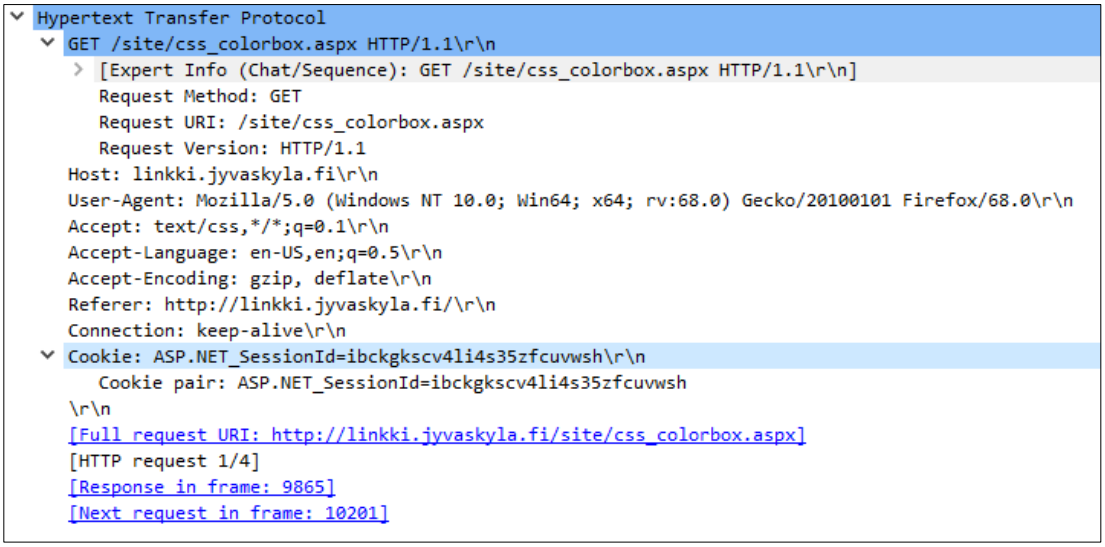
4.1.1 Evästeet

Eväste (cookie) on pieni tekstitiedosto, jonka verkkosivu, jolla käyttäjä vierailee lähettää käyttäjän tietokoneelle ja tallennetaan selaimeen. Käyttäjän vieraillessa sivustolla, joka käyttää evästeitä, sivuston HTTP-vastauksessa (Hypertext Trasport Protocol) on otsikkotieto (header) *Set Cookie* (ks. kuvio 2). (HTTP cookies 2019.)



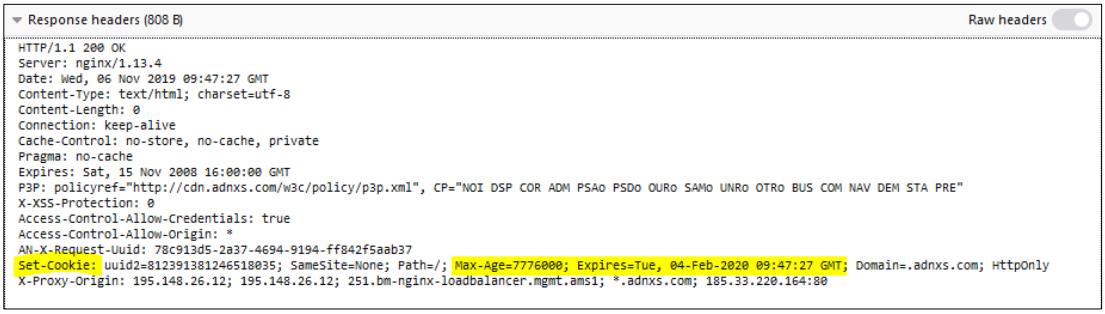
Kuvio 2. Wiresharkilla kaapattu palvelimen vastaus

HTTP-pyyntö, jonka selain lähettää takaisin palvelimelle sisältää evästeet *Cookie*-otsikossa (ks. kuvio 3). Evästeisiin tallennetaan muun muassa käyttäjän IP-osoite, kellonaika, käytetyt sivut, selaintyyppi, mistä osoitteesta, miltä palvelimelta ja mistä verkkotunnuksesta käyttäjä on tullut verkkosivulle.



Kuvio 3. Client-pyyntö selaimelta, sisältää evästeet

Kuvioissa 3 esitetty eväste on niin kutsuttu istuntoeväste (session cookie), joka tallentuu selaimeen siihen asti, että istunto suljetaan. Kuviossa 4 taas on niin kutsuttu pysyvä eväste (persistent/stored cookie), koska sille on asetettu sekä *Max-Age* eli maksimi-ikä ja *Expires* eli vanhenemisarvo. Eväste on tallennettu selaimeen, niin pitkäksi aikaa, kunnes *Max-Age*-arvo täyttyy tai *Expires*-päivämäärä saavutetaan tai käyttäjä poistaa evästeet käsin.

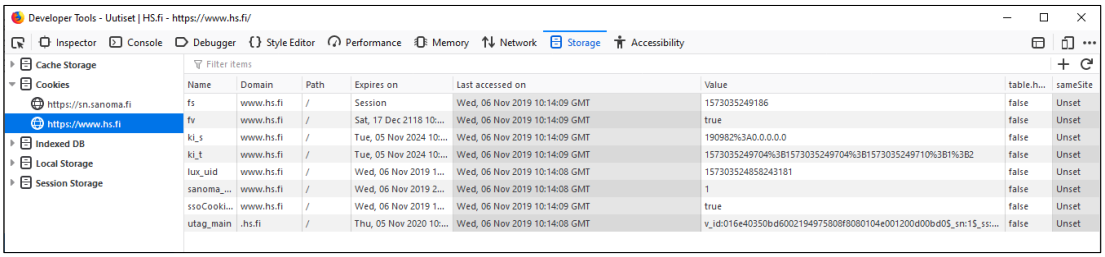


Kuvio 4. Set-Cookie-otsikko, jossa määritelty Max-Age ja Expires-aika

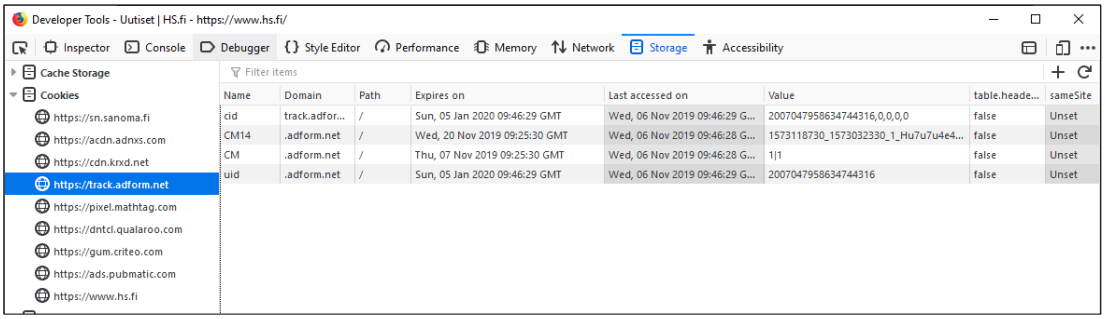
Koska HTTP on yhteydetön protokolla, ilman evästeitä sivustoilla kirjautuneena pysyminen ei onnistuisi, ostoskorissa olevat tuotteet häviäisivät ja käyttäjän tekemät muutokset esimerkiksi sivuston ulkoasuun unohtuisivat joka istunnon jälkeen. Yksityisyysongelmaksi evästeissä nouseekin kolmannen osapuolen evästeet, jotka pystyvät seuraamaan käyttäjää eri sivustoilla. Seuranta evästeet (tracking cookies) jaetaan yhden tai useamman sivuston tai palvelun kanssa, usein markkinoinnin ja mainonnan kohdentamiseksi tai analytiikkayhtiöitä varten. Kolmannen osapuolen eväste lähetetään samalla lailla HTTP-otsikossa, mutta sen mukana tulee aina käyttäjälle yksilöity id. Kun käyttäjä vierailee sivustolla, joka käyttää samaa kolmannen osapuolen evästettä, lähetetään sivustolle käyttäjän id, ja näin pystytään keräämään tietoja siitä, millä sivustoilla sama käyttäjä vierailee. (Jokinen 2018, 23.)

Ensimmäisen osapuolen evästeet (first-party cookies) tunnistaa siitä, että niillä on sama domain-nimi kuin sivustolla, jossa vierailee. Kuviossa 5 on hs.fi-sivuston eväs-

teet yksityisen istunnon kautta, jolloin kaikki evästeet ovat vain hs.fi domainille. Kuvi-ossa 6 taas on lista evästeistä, jotka tulevat, kun kaikki hyväksytään. Tällöin tietoja lähtee myös kolmansille osapuolille.



Kuvio 5. Helsingin Sanomien sivuilla ensimmäisen osapuolen evästeet yksityinen selaus-tilassa

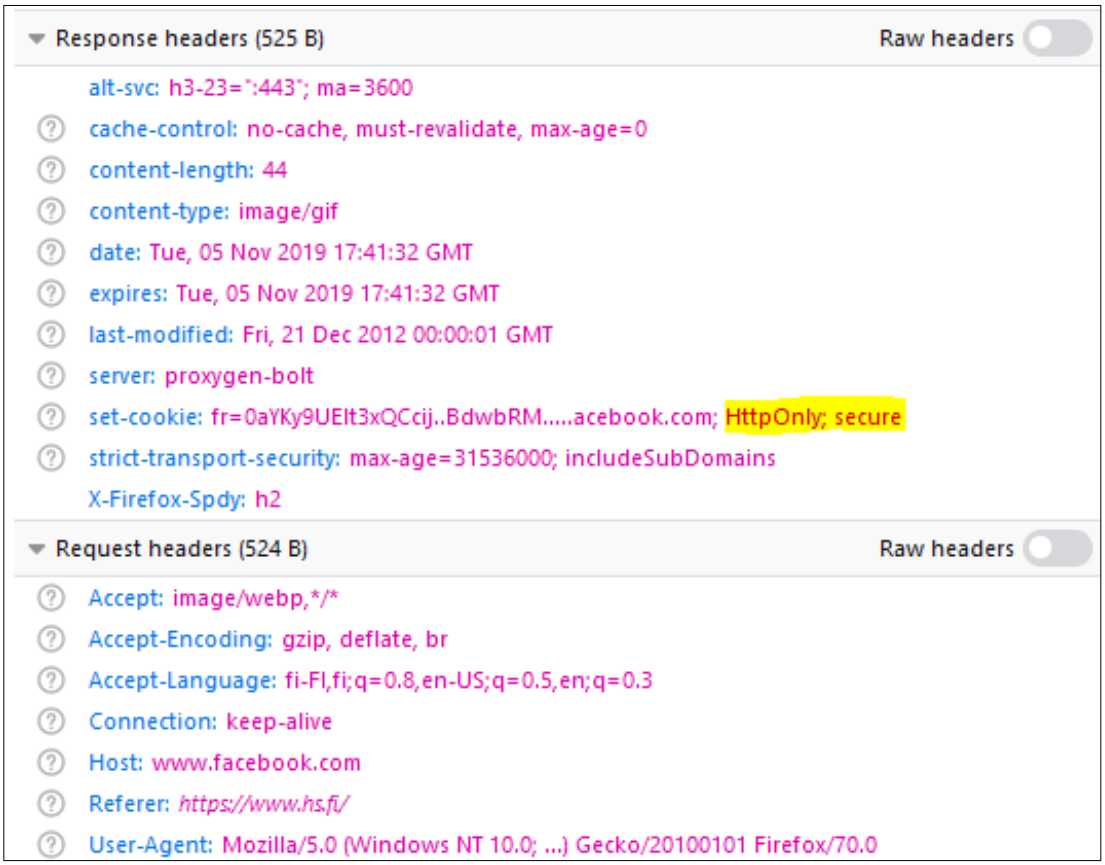


Kuvio 6. Helsingin Sanomien sivujen evästeet, jossa myös kolmannen osapuolen evästeet

Koska kolmannen osapuolen evästeet sisältävät tietoja käyttäjästä, niitä voidaan käyttää käyttäjän seurantaan myös negatiivisessa mielessä ja käyttäjän on vaikea myös GDPR aikaan tietää, kuinka paljon hänen dataansa jaetaan ja kenellä se on.

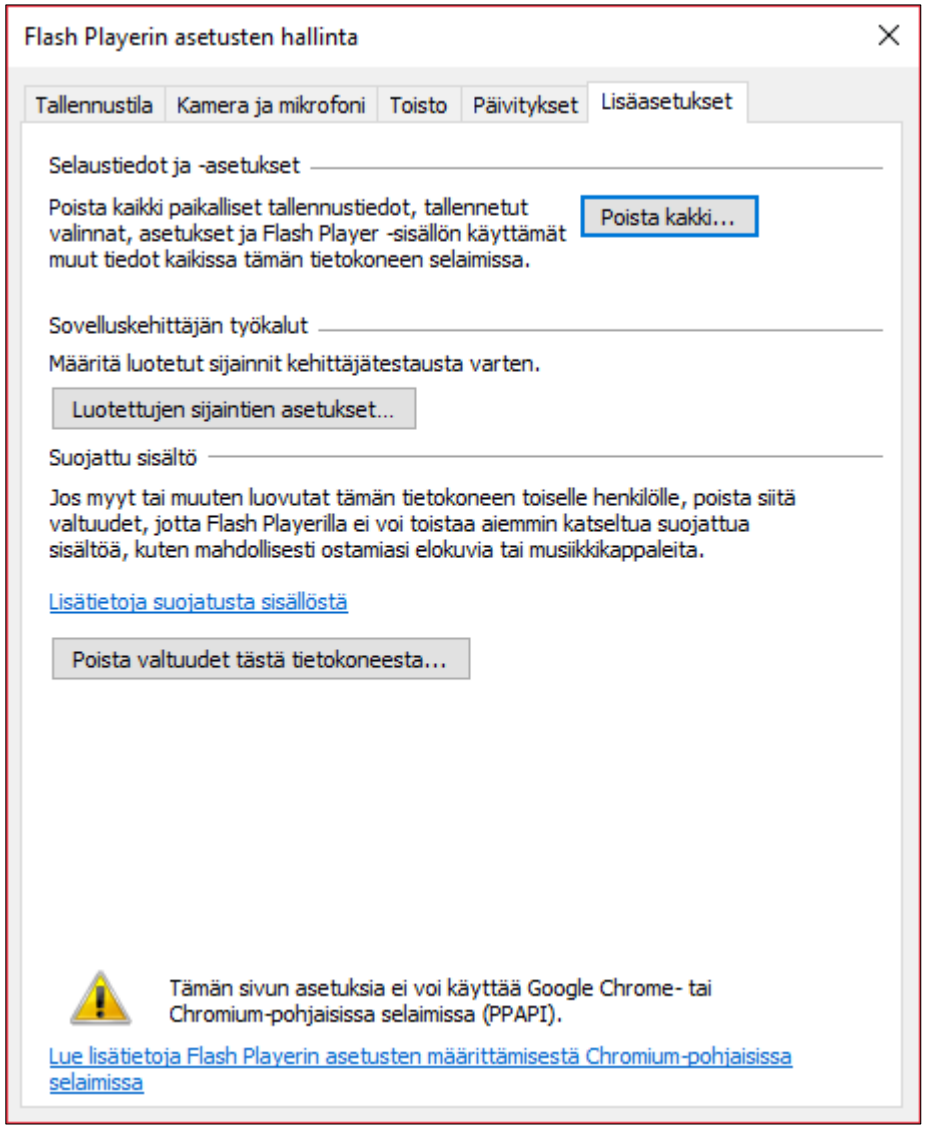
HTTPS (Hypertext Transport Protocol Secure) on TLS salattu HTTP-protokolla, joten tiedot, jotka yhteyden kautta kulkee, lukuun ottamatta kohde osoitetta ja porttia, on kryptattu, eivätkä tästä syystä näy Wiresharkissa, mutta ne näkyvät selaimessa. Kun on kyse HTTPS yhteyden evästeestä, *Set-Cookie*-otsikkoon lisätään *Secure* (ks. kuvio

7). Vaikka eväste olisi suojattu, ei se saisi ikinä sisältää arkaluontoista tietoa. *HttpOnly*-attribuutti estää evästeen avaamisen käyttäjäpuolella, koska tällä tavalla voidaan ehkäistä cross-site scripting (XSS) hyökkäyksiä. (HTTP cookies 2019.)



Kuvio 7. Set-Cookie-otsikossa on määritelty Http-Only ja Secure

Flash eväste (Flash cookie, local shared object) on tekstitiedosto, jonka palvelin lähettää samalla lailla, kuin HTTP-evästeet ja niitä hyödynnetään Adobe Flash-lisäosan kanssa. Esimerkiksi videoita toistettaessa flash-evästeisiin tallennetaan, mihin kohtaan on jääty, kun istunto-lopetetaan. Flash-evästeet tallentuvat erilliseen Adobe-kansioon, joka täytyy muistaa tyhjentää erikseen, jos haluaa tuhota kaikki evästeet (ks. kuvio 8).



Kuvio 8. Windows 10:ssä saa poistettua Flash Playerin tallentamat tiedot ohjauspaneelin kautta

Evästeiden käyttö vaatii käyttäjän suostumuksen ja niiden käyttötarkoituksesta on kerrottava selkeästi ja kattavasti. Suomessa tätä varten ei kuitenkaan vaadita erillistä ponnahdusikkunaa. (Luottamuksellinen viestintä 2019.)

Leppäsen ja Ruohosen (2017) tekemässä tutkimuksessa tutkittiin pysyviä evästeitä suosituilla suomalaisilla verkkosivuilla. Tutkimuksessa havaittiin, että 91.4% evästeistä oli kolmannen osapuolen evästeitä ja kahdella suositulla sivulla, joita ei paljastettu, tallennetaan session id ja kirjautumistiedot selkokielisenä, minkä voidaan todeta olevan vähintäänkin häiritsevää. Tutkimuksessa myös todettiin, että suurin osa

suomalaisista verkkosivuista on mediayhtiöiden omistuksessa, jotka toimivat tiiviissä yhteistyössä mainostajien kanssa. Esimerkiksi Alma Media omistaa yli 30 verkkosivua ja käyttää yli 120:tä kolmannen osapuolen evästettä. (Leppänen & Ruohonen 2017.)

4.1.2 Jäljite – web bug

Beacon eli jäljite tai web bug on evästettä huomaamattomampi keino seurata käyttäjiä. Jäljitteitä käytetään ajastamaan epäsynkronisia ja estottomia pyyntöjä palvelimelle. Pyynnöt lähtevät ennen kuin sivusto on ladannut ja käyttävät HTTP POST metodia, eivätkä vaadi palvelimelta vastausta. Jäljitteet ovat sivustolle mahdollisesti kokonaan piilotettuja pikselin kokoisia GIFejä (Graphics Interchange Format), PNG (Portable Network Graphics) kuvia, scriptejä tai elementtejä, jotka näkyvät vain sivuston lähdekoodissa (ks. kuvio 9). Jäljitteitä voidaan käyttää myös yhdessä evästeiden kanssa, jolloin pystytään lähettämään enemmän dataa. Myös sosiaalisen median kuvakkeet, ovat jäljitteitä; Facebook kerää dataa kaikkien Tykkää- ja Jaa-nappien kautta, vaikka käyttäjä ei koskaan painaisi niitä. (Jokinen 2018, 26.)

Kuvio 9. hm.com-sivuilla piilotettu jäljite, joka on ad.service.google-sivustolta

Web bugit keräävät muun muassa IP-osoitteen, URL:n missä jäljite on ja mistä osoitteesta se on ladattu, evästetiedot, mikäli se on asetettu toimimaan yhdessä evästeen kanssa, sekä selain tiedot. Web bugeja käytetään myös sähköposteissa; niiden kautta voidaan lähettää tieto, onko vastaanottaja avannut viestin. (What is a Web Bug/Beacon? n.d.)

4.1.3 Superevästeet

Superevästeillä (supercookie) tarkoitetaan evästeitä, jotka eivät tallennu käyttäjän selaimeen samalla tavalla kuin HTTP-evästeet.

Evercookie on JavaScript kirjasto, joka luo evästeitä, jotka tunnistavat käyttäjät senkin jälkeen, kun käyttäjä on poistanut evästeet selaimeltaan. Puhutaan niin sanotusti pysyvistä evästeistä tai zombie-evästeistä. JavaScriptin avulla evästeiden dataa tallennetaan eri puolelle selainta, ja kun havaitaan, että joku tietty eväste on poistettu, se luodaan uudestaan hakemalla data muualta muistista. Dataa voidaan tallentaa esimerkiksi HTML5 paikallis- ja istuntomuistiin, WebSQL:ään ja ETageihin (EntityTag). Käyttäjä ei saa tietoa, että eväste on uudelleen luotu. Flash evästeet saatetaan uudelleen luoda jopa selaimen vaihdon tai uudelleenasennuksen jälkeen, koska data on tallennettu erikseen Flash Player-lisäosan välimuistiin. (Matuszewska 2019; Jokinen 2018, 29.)

4.1.4 HSTS

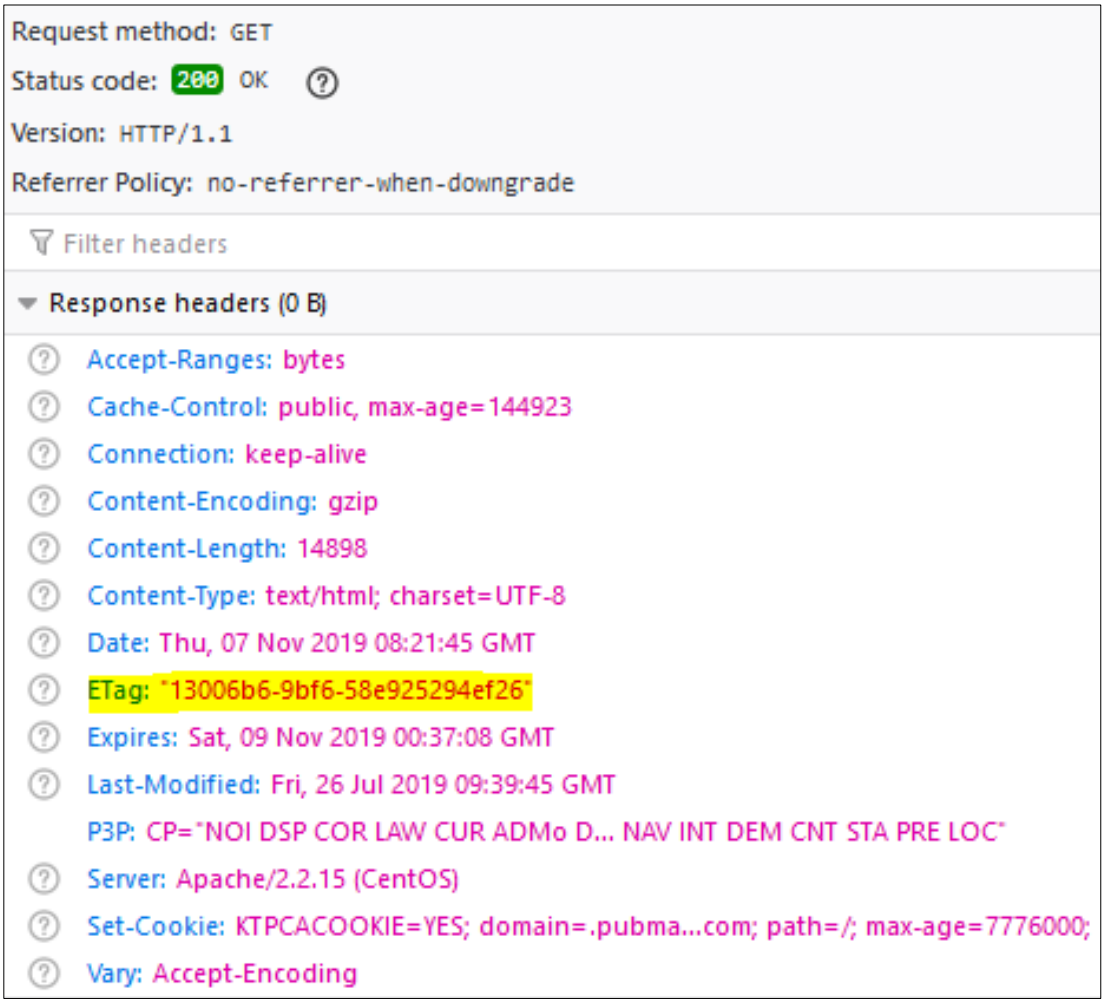
HSTS (HTTP Strict Transport Security) on tietoturvaominaisuus, jonka ylläpitäjä voi laittaa verkkosivulle. Kun käyttäjä selaa hs.fi- tai www.hs.fi-sivulle, pyynnöt lähtevät http://www.hs.fi muodossa, mutta koska Helsingin Sanomien verkkosivuilla on käytössä HSTS, palvelin vastaa pyyntöihin, että vain HTTPS-yhteys on mahdollinen, jolloin selain vaihtaa pyynnön https://hs.fi-osoitteeseen jolloin yhteys on suojattu. Selain myös muistaa tämän tiedon myöhemmin. HSTS:n avulla voi estää mahdolliset SSL stripping- ja man-in-the-middle (MiTM)-hyökkäykset. (Stockley 2015.)

Käyttäjäseurantaan HSTS:a voi käyttää luomalla väärennettyjä kansiota, joita käyttäjä pyytää palvelimelta aina tietyssä järjestyksessä. Palvelin voi esimerkiksi lähettää tiedon useammasta kuvakansiosta, joista osalle on asetettu HSTS. Selain tallentaa tiedon kansioita, jotka vaativat HTTPS-yhteyden ja seuraavalla kerralla sivustolle tullessa, se pyytää nämä kansiot automaattisesti HTTPS-yhteydellä. Mikäli palvelin on asettanut kansioille binaaritiedot, että kansiot HTTP-yhteydellä pyydettynä on nollia ja HTTPS-yhteydellä on ykkösiä, voidaan käyttäjälle luoda yksilöllinen tunnus, josta tämän tunnistaa, kun käyttäjä seuraavan kerran tulee sivustolle. Mikäli väärennettyjä

kansioita on kymmenen, on mahdollista tunnistaa yksi käyttäjä 1024 joukosta, kahdellakymmenellä kansiolla yli miljoonan joukosta. (Stockley 2015.)

4.1.5 ETag

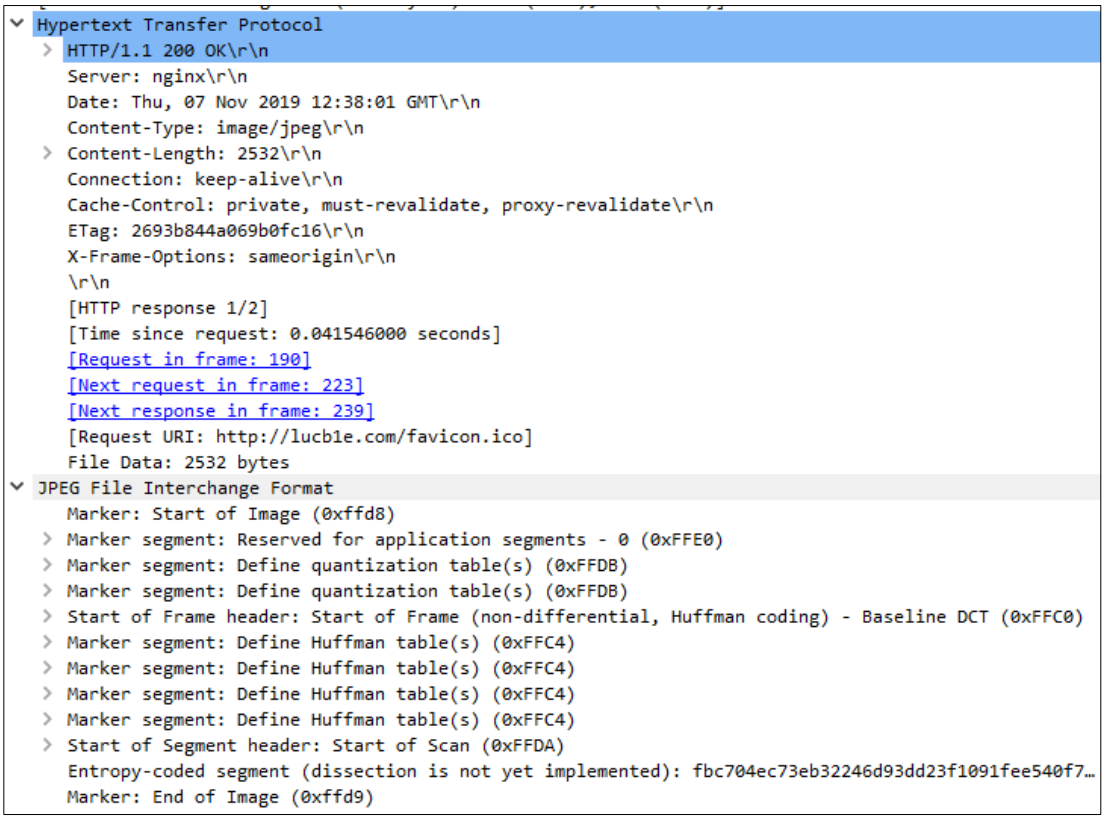
EntityTag (ETag) on välimuistin validointi metodi, jota web palvelimet käyttävät tunnistaakseen lähteitä ja säästääkseen kaistanopeutta ja välttääkseen yhteen törmäykset (mid-air collision). ETag on yksilöllinen arvo, joka annetaan jokaiselle välimuistin elementille. Käyttäjän vieraillessa samalla sivustolla useamman kerran, palvelin vertaa ETageja ja mikäli se tunnistaa, että lähteellä on sama ETag se lähettää HTTP 304 Not Modified vastauksen, jolloin elementti latautuu käyttäjän välimuistista. ETagit on tallennettu käyttäjän selaimeen ja kulkevat HTTP otsikkokentässä (ks. kuvio 10). (Brinkmann 2014; ETag 2019.)



Kuvio 10. Helsingin Sanomien sivuilla oleva ETag

ETageja käytetään myös seurantaan, jolloin ei tarvita evästeitä, JavaScriptia tai paikallismuistia. ETagien avulla on mahdollista seurata käyttäjien vierailuja sivustolla, vaikka heillä olisi käytössä VPN (Virtual Private Network) ja/tai IP-osoite vaihtuisi, mikäli selain on sama, eikä välimuistia ole tyhjennetty.

Osoitteessa *http://lucb1e.com/rp/cookielesscookies/* on demonstroitu, miten seurantaan tarkoitetut ETagit toimivat. Sivusto sisältää kuvan, jossa on ETag ja tämä tallentaa sivustovierailut ja tekstin, jos on kirjoittanut sivustolle jotain (ks. kuvio 11). Sivusto vaikuttaa toimivan samalla lailla kuin, jos se käyttäisi evästeitä.



Kuvio 11. Cookieless cookie-sivuston jpeg-kuva sisältää ETagin, joka kerää tiedot

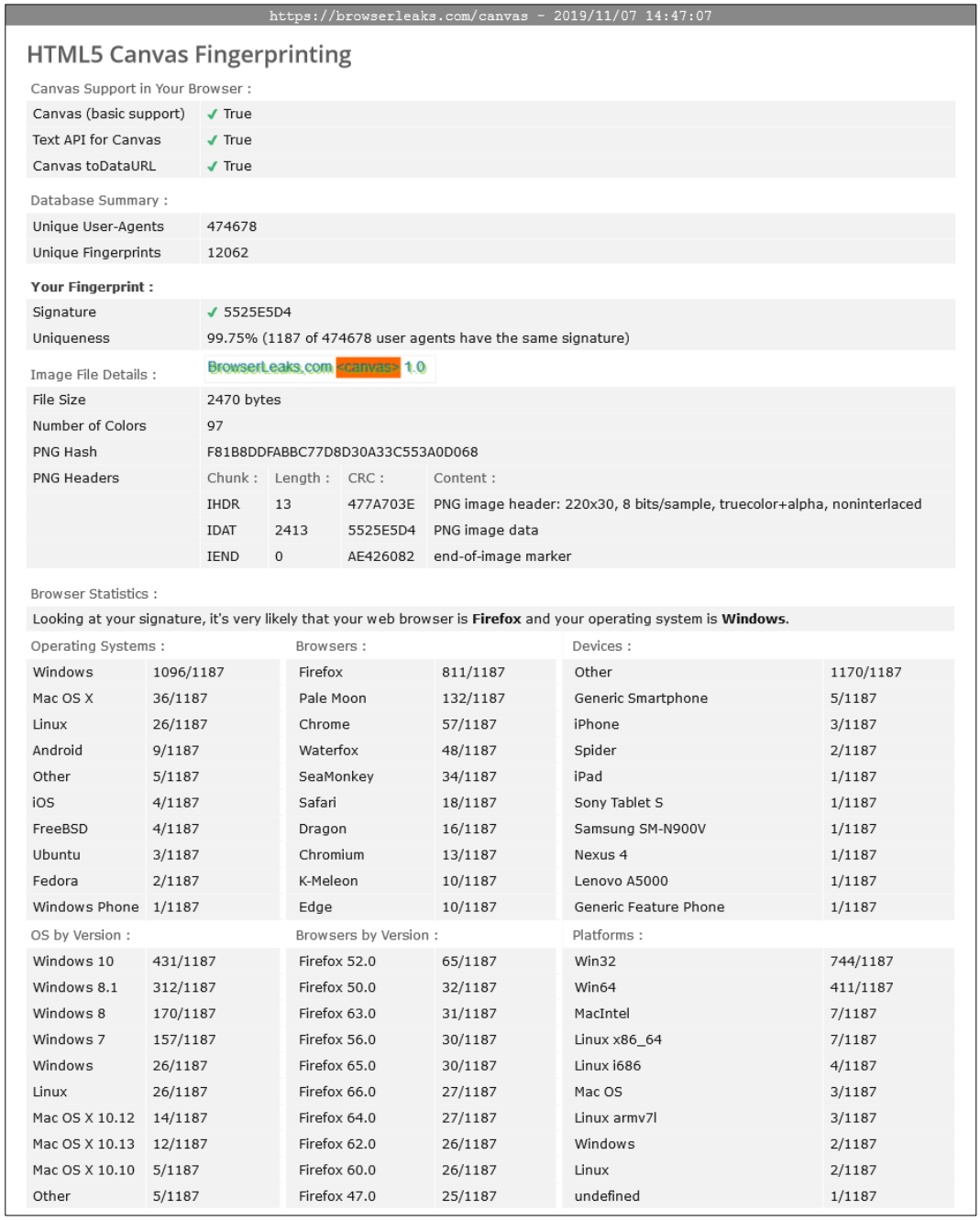
ETagit poistuu, kun tyhjentää selaimen välimuistin.

4.1.6 Canvas fingerprinting

Web Fingerprinting (suomeksi verkon sormenjälkitunnistus) viittaa fyysiseen sormenjälkitunnistukseen, jossa henkilön sormen ihon kohoumista muodostuu yksi numeerinen malli, samoin verkossa voidaan kerätä tietoja ja näistä muodostaa yksilöllinen tunniste, kuten hash-arvo. Tähän ei tarvita evästeitä. Canvas fingerprinting on yksi tällainen keino.

Canvas on HTML5 API (Application Programming Interface, ohjelmointirajapinta), jolla verkkosivut piirtävät grafiikat ja animaatiot käyttäen JavaScriptiä. Kun halutaan käyttää canvasia käyttäjäseurantaan, palvelin ohjaa selaimen piirtämään piilotetun tekstin tai 3D grafiikan, joka esitetään yhtenä digitaalisena merkkinä. Koska tietokoneilla on erilaisia näytönohjaimia, käyttöjärjestelmiä, käyttää eri selainta ja sisältää erilaisia fontteja, teksti/kuva, jonka sivusto piirtää on yksilöllinen. Tämä uniikkimerkki voidaan jakaa muiden sivustojen kesken, jolloin pystytään seuraamaan sen käyttäjän toimia verkossa. (Acar, Diaz, Englehardt, Eubank, Juarez & Narayanan 2014.)

BrowserLeaks-sivustolla on esimerkki siitä, kuinka canvas fingerprinting toimii. Testatessa kolmella eri tietokoneella sivusto kuitenkin antoi aina samat tiedot, kun vain käytti samaa selainta. Kuviossa 12 on tiedot, jotka sai joka tietokoneella Firefox-selainta käyttäessä.



https://browserleaks.com/canvas - 2019/11/07 14:47:07

HTML5 Canvas Fingerprinting

Canvas Support in Your Browser :

Canvas (basic support)	✓ True
Text API for Canvas	✓ True
Canvas toDataURL	✓ True

Database Summary :

Unique User-Agents	474678
Unique Fingerprints	12062

Your Fingerprint :

Signature	✓ 5525E5D4
Uniqueness	99.75% (1187 of 474678 user agents have the same signature)

Image File Details : BrowserLeaks.com <canvas> 1.0

File Size	2470 bytes			
Number of Colors	97			
PNG Hash	F81B8DDFABBC77D8D30A33C553A0D068			
PNG Headers	Chunk :	Length :	CRC :	Content :
	IHDR	13	477A703E	PNG image header: 220x30, 8 bits/sample, truecolor+alpha, noninterlaced
	IDAT	2413	5525E5D4	PNG image data
	IEND	0	AE426082	end-of-image marker

Browser Statistics :

Looking at your signature, it's very likely that your web browser is **Firefox** and your operating system is **Windows**.

Operating Systems :		Browsers :		Devices :	
Windows	1096/1187	Firefox	811/1187	Other	1170/1187
Mac OS X	36/1187	Pale Moon	132/1187	Generic Smartphone	5/1187
Linux	26/1187	Chrome	57/1187	iPhone	3/1187
Android	9/1187	Waterfox	48/1187	Spider	2/1187
Other	5/1187	SeaMonkey	34/1187	iPad	1/1187
iOS	4/1187	Safari	18/1187	Sony Tablet S	1/1187
FreeBSD	4/1187	Dragon	16/1187	Samsung SM-N900V	1/1187
Ubuntu	3/1187	Chromium	13/1187	Nexus 4	1/1187
Fedora	2/1187	K-Meleon	10/1187	Lenovo A5000	1/1187
Windows Phone	1/1187	Edge	10/1187	Generic Feature Phone	1/1187
OS by Version :		**Browsers by Version :**		**Platforms :**	
Windows 10	431/1187	Firefox 52.0	65/1187	Win32	744/1187
Windows 8.1	312/1187	Firefox 50.0	32/1187	Win64	411/1187
Windows 8	170/1187	Firefox 63.0	31/1187	MacIntel	7/1187
Windows 7	157/1187	Firefox 56.0	30/1187	Linux x86_64	7/1187
Windows	26/1187	Firefox 65.0	30/1187	Linux i686	4/1187
Linux	26/1187	Firefox 66.0	27/1187	Mac OS	3/1187
Mac OS X 10.12	14/1187	Firefox 64.0	27/1187	Linux armv7l	3/1187
Mac OS X 10.13	12/1187	Firefox 62.0	26/1187	Windows	2/1187
Mac OS X 10.10	5/1187	Firefox 60.0	26/1187	Linux	2/1187
Other	5/1187	Firefox 47.0	25/1187	undefined	1/1187

Kuvio 12. BrowserLeaks-sivulla näkee esimerkin siitä, kuinka sivustot rakentaa yksilöllisen merkin

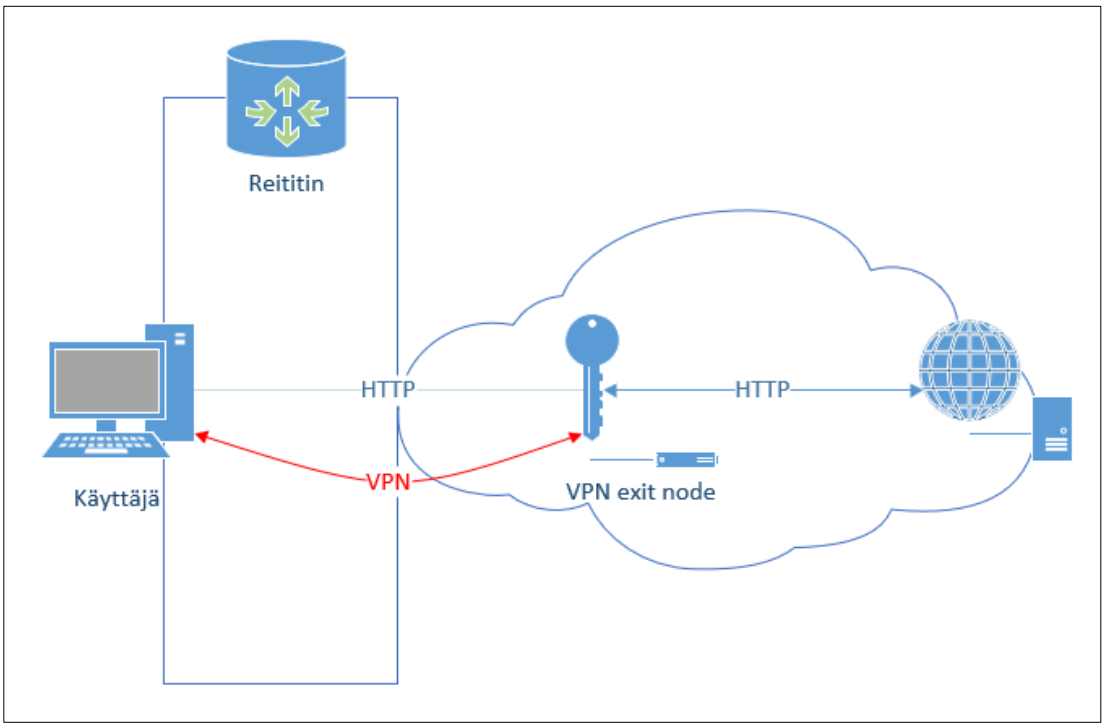
Muun muassa verkkolehdet käyttävät canvas-elementtiä seuratakseen vierailijoita, joille tarjotaan esimerkiksi viisi ilmaista lukukertaa. Jos sivusto käyttäisi evästeitä, käyttäjä voisi vain poistaa ne ja jatkaa ilmaisten artikkelien lukemista seuraavat viisi ja taas poistaa evästeet. Canvas fingerprinting on käyttäjälle täysin huomaamaton. Jos haluaa varmistua, että canvas-elementtiä ei käytetä seurantaan, sen voi poistaa

kokonaan käytöstä, mutta tällöin myös menettää sivuston grafiikat, jotka hyödyntävät elementtiä näkyvän grafiikan piirtämiseen.

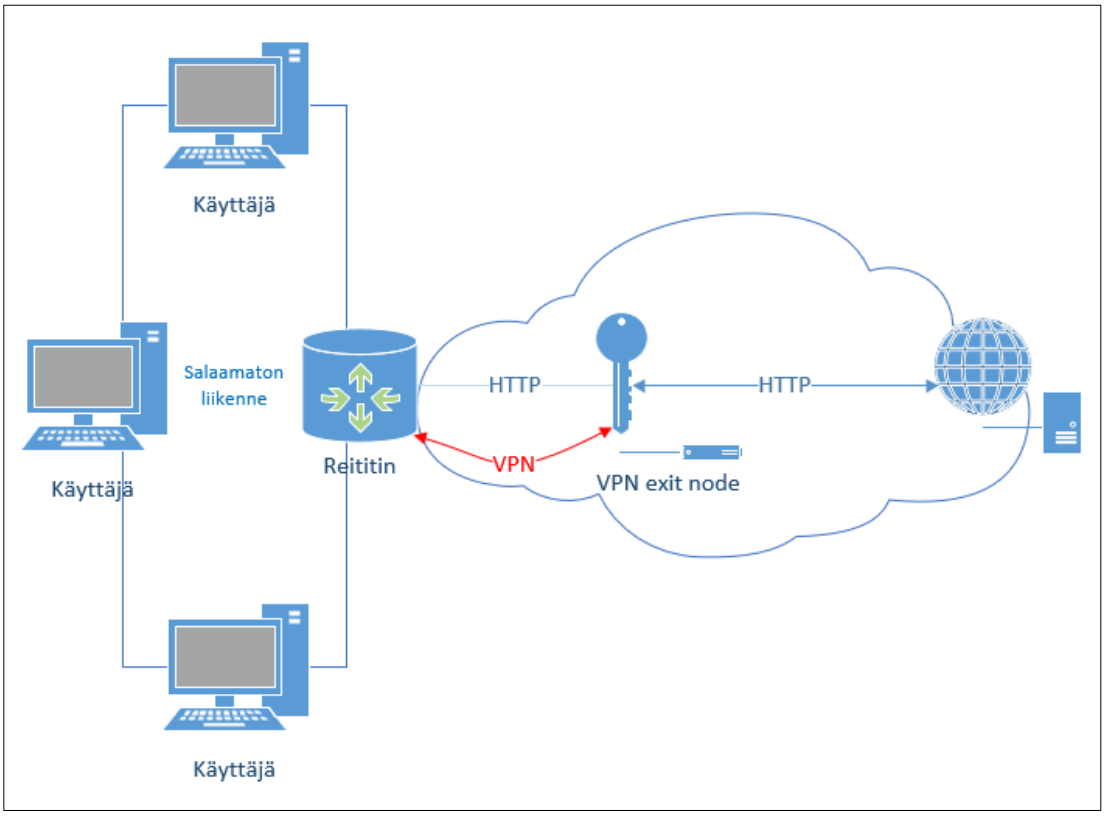
4.2 Keinoja digitaalisen jalanjäljen pienentämiseksi

4.2.1 VPN

Virtuaalinen yksityinen verkko eli VPN (Virtual Private Network) on keino, jolla voidaan peittää käyttäjän todellinen IP-osoite ja estää muita, mukaan lukien verkko-operaattoria näkemästä, mitä käyttäjä verkossa selaa. VPN:n avulla luodaan tunneli käyttäjän tietokoneen (ks. kuvio 13), reitittimen (ks. kuvio 14) tai virtuaalikoneen ja VPN palvelimen välille. Vain tämä yhteys on suojattu, mutta kyselyt, jotka menevät lopulliselle palvelimelle, sisältävät VPN:n luoman IP-osoitteen. Koska käyttäjän todellinen IP ei näy verkossa, ei käyttäjän sijaintia voi tietää, eikä dataa kerätä. Operaattori näkee vain, että yhdistetään VPN palvelimeen. Passiivisen valvonnan avulla valtio näkee, mitä verkossa tapahtuu, mutta täytyy tehdä aktiivista valvontaa, jotta voi tietää kuka verkossa toimii.



Kuvio 13. VPN:n avulla salataan liikenne käyttäjän koneelta vain VPN-palvelimelle asti



Kuvio 14. VPN-yhteys voidaan luoda myös suoraan reitittimeltä VPN-palvelimelle

VPN teknologioita on muun muassa:

- Point-to-Point Tunneling Protocol (PPTP)
- Layer Two Tunneling Protocol (L2TP)
- Internet Protocol Security (IPsec)
- OpenVPN
- Secure Socket Tunneling Protocol (SSTP)
- Internet Key Exchange (IKEv2)

VPN palvelimen voi rakentaa itse, mutta on olemassa myös kaupallisia VPN palveluita kuten NordVPN (Kaikki mitä tarvitset parempaan verkkokokemukseen n.d.) tai F-Securen Freedome (F-Secure FREEDOME VPN n.d.). VPN:n käyttö hidastaa liikennettä ja vaihtelee riippuen siitä, missä VPN palvelin sijaitsee. Esimerkiksi Suomesta yhdistettäessä Yhdysvaltoihin, jotta voi katsoa Yhdysvaltojen suoratoistopalveluita, hidastuu liikenne huomattavasti enemmän, kuin jos yhdistää Suomessa sijaitsevaan palvelimeen. VPN suojaa yksityisyyttä ja lisää tietoturvaa tiettyjä hyökkäyksiä kuten MiTM vastaan, mutta se ei suojaa esimerkiksi canvas fingerprintingiltä, joten seurantaa vastaan se ei ole täysin aukoton.

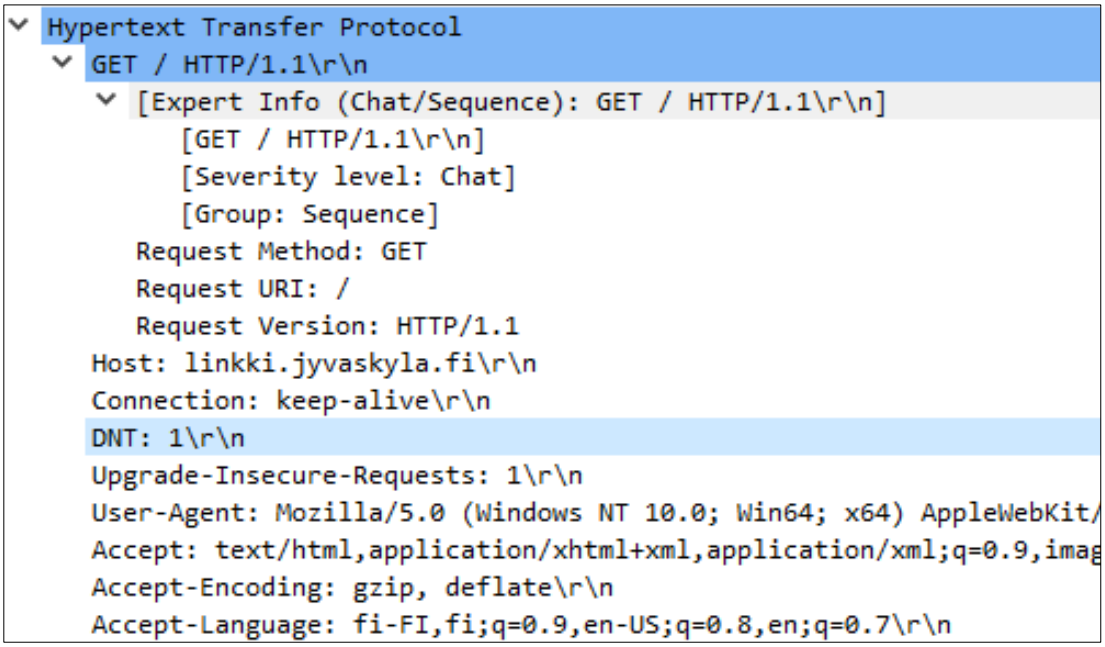
Jotkut VPN-palveluntarjoajat tallentavat käyttäjien selainhistorian, joten palvelua hankkiessa kannattaa tarkistaa, että tällaista tietojen keruuta ei sallita ja jos sallitaan, niin mihin tarkoitukseen. Ilmaisia VPN palveluita käyttäessä kannattaa olla erityisen tarkka, koska jotkut palvelut eivät ainoastaan tallenna selaustietoja, vaan saattavat jopa myydä ne eteenpäin.

4.2.2 Selaimen vahvistaminen

Selaimille on kehitetty erilaisia liitännäisiä, joilla voi rajoittaa seurantaa.

Do Not Track

Useista selaimista löytyy asetuksista Do Not Track – Älä seuraa-ominaisuus. Tämän ollessa käytössä, käyttäjän vieraillessa jollain sivustolla, palvelimelle lähetetään HTTP-pyynnössä DNT-otsikko, jossa kerrotaan, halutaanko sivuston seuraavan häntä arvolla 0 tai 1 (ks. kuvio 15).

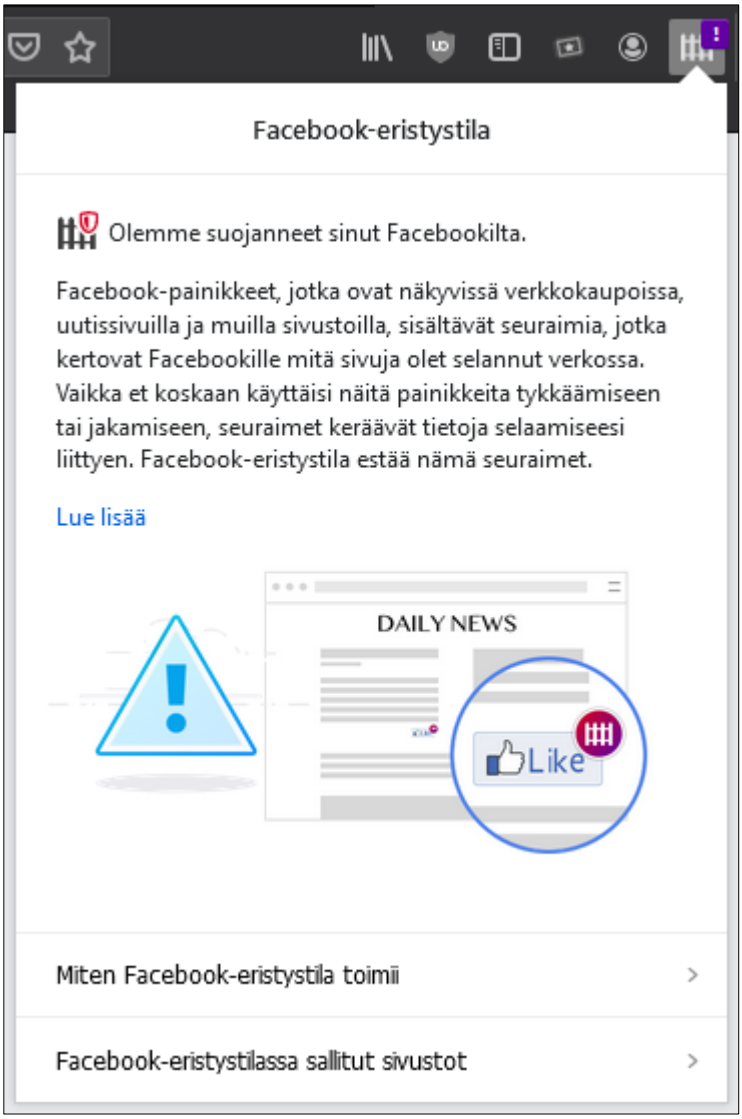


Kuvio 15. DNT-arvo 1 kertoo, että käyttäjä ei halua sivuston seuraavan häntä

Älä seuraa-toiminto on verrattain turha, koska se ei millään tavalla varmista huomioiko palvelin pyyntöä. Sivusto voi hyvinkin jättää otsikon täysin huomioimatta ja käyttäjä ei saa tästä mitään tietoa. Muun muassa Google myöntää, ettei sen toiminta muutu millään lailla Do Not Track-pyynnöstä (Do Not Track -asetuksen käyttöön ottaminen ja käytöstä poistaminen n.d).

Facebook Container

Facebook Container on Mozillan kehittämä lisäosa Firefox-selaimelle, jonka avulla huomaa, milloin sivusto, joka ei ole Facebookin omistama, käyttää Facebookin seurantaevästeitä tai jäljitteitä, ja estää niiden käytön (ks. kuvio 16). Facebook-henkilöllisyys on eristetty omaan välilehteen, jolloin muita sivustovierailuja ei voi yhdistää käyttäjän Facebook-tiliin. (Facebook Container n.d.)



Kuvio 16. Aidan yllä oleva huutomerkki kertoo jos sivustolla on Facebook-seuraimia

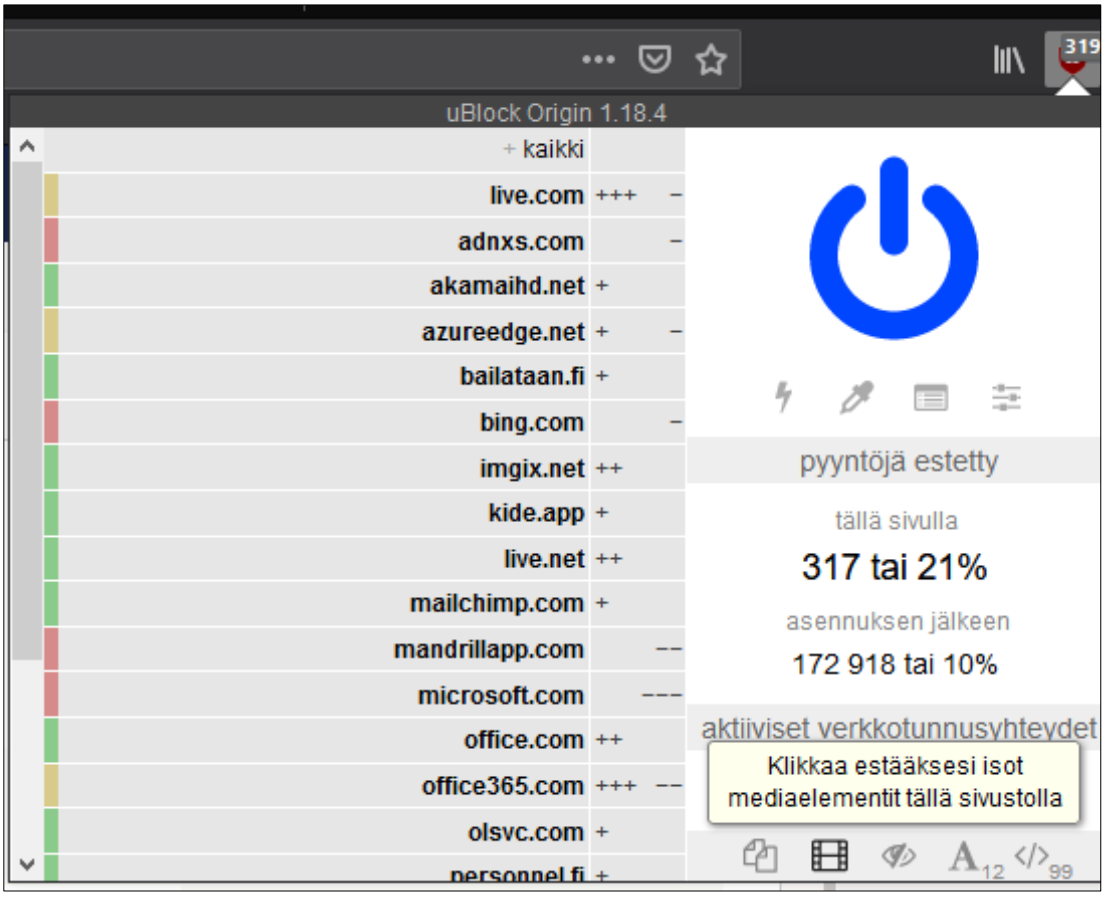
Firefox versio 70:ssä on automaattisesti käytössä *Enhanced Tracking Protection in Firefox,* joka automaattisesti estää seurantayritykset tunnetuilta kolmannen osapuolen seurantaevästeiltä, kryptolouhinnalta (*cryptomining*, jolloin käyttäjän tietokonetta käytetään kryptovaluutan louhintaan), sekä sormenjälkitunnistuksilta. Myös sosiaalisen median kuten Twitterin, LinkedInin ja Facebookin seurantayritykset estetään, mutta se ei ole niin kokonaisvaltaista kuin Facebook Containerin avulla. (Trackers and scripts Firefox blocks in Enhanced Tracking Protection n.d.)

Privacy Badger

Privacy Badger on Electronic Frontier Foundationin (EEF) kehittämä lisäosa Firefox-, Chrome- ja Opera-selaimille, joka automaattisesti estää näkymättömät seurantayritykset. Privacy Badger lähettää HTTP otsikossa DNT 1 pyynnön, ja mikäli palvelin silti lähettää seurantayrityksen, Privacy Badger torjuu seurannan sen jälkeen, kun se on, nähnyt saman seuraimen kolmella eri sivustolla. Tällä hetkellä se myös estää ulospäinlähtevät linkkien seuraamisen Facebookilta, Googlelta ja Twitteriltä. (Privacy Badger n.d.)

uBlock Origin

uBlock Origin on Raymond Hillin kehittämä avoimenlähdekoodin lisäosa Firefox-, Chrome-, Chromium-, Edge-, Opera- ja Safari-selaimille. Sen avulla voi estää mainokset, seurantayritykset ja haitalliset sivustot. Suodatinlistat *EasyList*, *Peter Lowe's Ad-servers*, *EasyPrivacy* ja *Malware domains* ovat automaattisesti käytössä. Näiden lisäksi on mahdollista lisätä muitakin suodatinlistoja, mutta liian monen listan käyttäminen kuluttaa tietokoneen keskusmuistia. Lisäosa on hyvin helppokäyttöinen ja tehokas. Mikäli tulee eteen sivusto, joka vaatii, että mainostenesto on otettava pois käytöstä, jotta voi nähdä sivuston sisällön, uBlockin voi ottaa helposti pois käytöstä vain yhdellä sivustolla ilman, että se vaikuttaa esimerkiksi muilla välilehdillä oleviin sivustoihin (ks. kuvio 17).



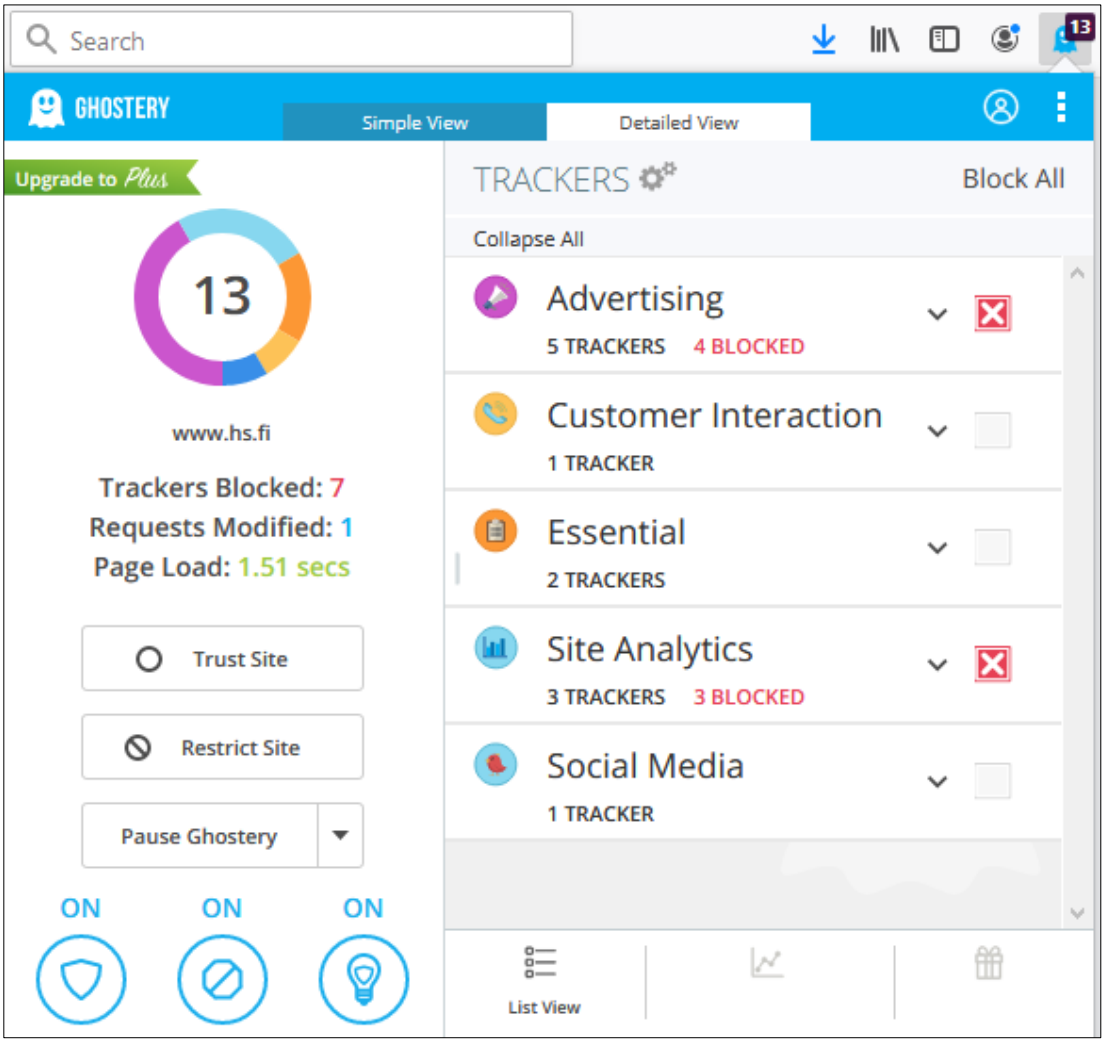
Kuvio 17. Painamalla sinistä Play-nappia voi ottaa uBlockin pois päältä

uMatrix

UMatrix on myös Raymond Hillin kehittämä lisäosa Firefox-, Chrome- ja Opera-selaimille. Lisäosan avulla käyttäjä pystyy muokkaamaan selaimen elementtejä, ottamalla niitä pois käytöstä. Lisäosa vaatii huomattavasti enemmän perehtymistä kuin uBlock, koska sivusto saattaa lakata täysin toimimasta, mikäli poistaa käytöstä elementin, joka on sivuston toiminnalle pakollinen. uMatrixin avulla voi muokata muun muassa evästeitä, kuvia, mediaa, scriptejä ja frameja.

Ghostery – Privacy Ad Blocker

Ghostery on lisäosa, joka estää mainokset ja seurannan, sekä nopeuttaa selainta. Lisäosa on saatavilla Cliqz-, Firefox-, Chrome-, Opera- ja Edge-selaimille (ks. kuvio 18).



Kuvio 18. Ghosteryn avulla näkee helposti, mitä sivustolla olevat seuraimet ovat

Ghosterya on syytetty siitä, että se myy käyttäjädataa eteenpäin kolmansille osapuolille, minkä takia sen lähdekoodi on nykyään julkinen.

Cookie Quick Manager

Cookie Quick Manager on lisäosa evästeiden hallintaan. Sen avulla voi katsoa domain-kohtaisesti evästeet, evästeiden attribuutteja voi muuttaa, varmuuskopioida evästeet tai poistaa ne kokonaan. (Cookie Quick Manager n.d.)

HTTPS Everywhere

EEF:n luoma lisäosa Firefox-, Chrome- ja Opera-selaimille, HTTPS Everywhere luo automaattisesti HTTPS yhteyden sivustoille, jotka tukevat sitä (HTTPS Everywhere n.d.). Sen lisäksi, että HTTPS-yhteys on turvallisempi kuin HTTP-yhteys, tämä mahdollistaa myös eston HSTS-seurannalle.

4.2.3 Hide signal in noise

Selaimille on kehitetty lisäosia, jotka luovat ylimääräistä liikennettä hämätäkseen mainostajia ja seuraajia.

Ad Nauseam

Ad Nauseam on avoimen lähdekoodin projekti ja rakennettu uBlock Origin pohjalta. Lisäosa piilottaa mainokset ja estää haittaohjelmia, mutta taustalla Ad Nauseam lisää jokaisen mainokset AdVaultiin, joita käyttäjä voi interaktiivisesti selata. Ad Nauseamin saa myös konfiguroitua klikkaamaan jokaista estettyä mainosta, käyttäjän normaalin selaamisen taustalla. Näin mainostajat ja muut datan kerääjät saavat seurantaevästeiden avulla käyttäjästä tiedon, että hän olisi kiinnostunut tuotteesta, vaikka ei oikeasti ole, eikä todellista profiilia ei voi rakentaa.

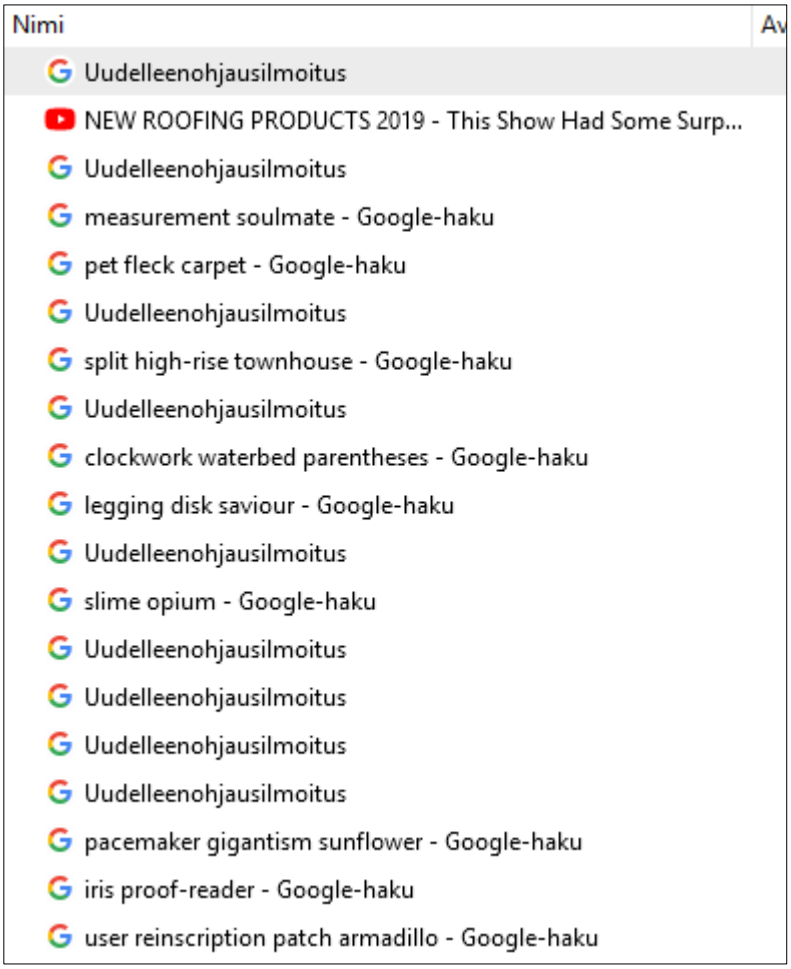
Ad Nauseam on ladattavissa Firefox- ja Opera-selaimille. Google Chrome on poistanut AdNauseamin Chrome Web Storesta 2017 sillä perusteella, että se olisi haittaohjelma, mutta se on silti mahdollista saada asennettua myös Chrome-selaimeen lataamalla sen manuaalisesti tietokoneelle. Vastaavanlainen lisäosa on myös TrackMeNot, joka luo sattumanvaraisia hakuja suosituille hakusivustoille.

Huijausklikkauksia vastaan on myös tehty ohjelmistoja mainostajille. Muun muassa PPC Protect on automatisoitu ohjelmisto, joka valvoo mainoksia, analysoi joka klikkauksen vertaamalla sitä muihin klikkauksiin ja tekee päätöksen, onko klikkaus käyttäjältä, joka voisi oikeasti olla kiinnostunut mainostettavasta tuotteesta tai palvelusta. PPC Protect kerää käyttäjädataa klikkauksesta, muun muassa IP-osoitteen, laite

id:n, selaimen ja sijainnin ja analysoimalla näitä tietoja algoritmi tekee päätöksen, onko klikkaus kohteesta, joka on aiemmin käynyt sivulla. Mikäli algoritmi havaitsee, että klikkaus on vilpillinen, käyttäjä laitetaan listalle, jolle ei enää näytetä mainoksia. (How We Prevent Click Fraud n.d.)

Internet Noise

Dan Schultzin kehittämä verkkosivu Internet Noise, toimii kuin selaimenlisäosa olematta sitä. Sivusto avaa välilehtiä automaattisesti joka kymmenes sekunti perustuen Google-hakuihin, jotka ovat täysin sattumanvaraisia. Samalla tavoin kuin Ad Nauseam, se luo nimensä mukaisesti ylimääräistä ”melua” internet liikenteeseen, jolloin tarkkojen käyttäjäprofiilien luominen on vaikeampaa. Kehittäessään sivua Schultz haki Googlesta ”Top 4000 substantiivia” ja lisäsi listan koodiinsa, joten joka kerta, kun sivulla painaa ”Make some noise”-nappia se käyttää Googlen ”Kokeilen onneani”-nappia listassa oleviin sanoihin. Kuviossa 19 on joitain hakusanoja, joita sivusto loi. (Dreyfuss 2017.)



Kuvio 19. Internet Noisen googlehakujen avaamista sivuista suurinosa estettiin

Sivustolla kerrotaan selvästi, että se ei lisää turvallisuutta, vaan on ”digitaalinen protesti” mainostajia vastaan. Sivusto vierailut ovat lyhyitä ja niin sattumanvaraisia, että ne voi päätellä koneen tekemäksi, eikä estä analytiikkayhtiöitä, jotka hakevat toistuvia ja pidempiä klikkauksia.

4.2.4 TOR-selain

The Onion Router (TOR) -selain mahdollistaa anonyymin selaamisen verkossa, niin että edes valtiot, yritykset tai operaattorit eivät tiedä, millä sivuilla käyttäjä vierailee. Tor-selainta käyttäessä jokainen pyyntö reititetään useamman pisteen kautta kryptattuna. Tor-selain sai alkunsa tor-verkko projektista 2008 ja on täysin ilmainen, vapaan lähdekoodin ohjelmisto.

Tor-suojaa yksityisyyttä, lähinnä peittämällä IP-osoitteen, salaamalla liikenteen vähintään kolmella kerroksella ja reitittämällä sen kolmen vapaaehtoisen tietokoneen läpi, jotka on valittu tuhansien koneiden joukosta ympäri maailmaa. Nämä kolme tietokonetta poistaa yhden salauksen, ennen kuin lähettää datan eteenpäin seuraavalle koneelle. Näin on vaikea seurata, mistä pyyntö on tullut ja kenelle se on menossa. Reitityksen viimeinen kone näkee kuitenkin liikenteen, joka loppupalvelimen välillä kulkee, vaikkei se näekään, mistä pyyntö on tullut. Tätä “heikkoutta” voi käyttää hyväkseen ja esimerkiksi verkkorikolliset, valtiot ja tiedustelupalvelut voivat asentaa omia koneita, jotka toimivat viimeisenä pisteenä reitityksessä ja näin seurata verkkoliikennettä. Tämän takia kannattaa vierailla vain sivustoilla, jotka käyttävät HTTPS-yhteyttä, jos haluaa välttää mahdollisen seurannan. (Greenberg 2017.)

4.2.5 Muita keinoja

Selaimen asetuksista voi asentaa sen unohtamaan kaikki historiatiedot heti, kun selain suljetaan, jolloin muun muassa evästeitä ja ETageja ei tallenneta. Jos haluaa, että tietyt sivut muistavat käyttäjätunnuksen, sen voi tallentaa selaimeen erikseen. Yksityinen selaustila on käytännössä sama, kuin jos selain on asennettu unohtamaan historiatiedot.

ffprofile.com-sivulla pystyy helposti muuttamaan Firefox-selaimen asetuksia ja tietoturvaaparannettua, ilman selaimen konfiguroinnin muuttamista käsin. Sivustolla on esitelty englanniksi selkokielellä, mitkä asetukset voi asentaa halutessaan ja lopuksi sivu antaa JavaScript tiedoston, jossa on valinnanmukaiset konfiguroinnit. Myös GitHubista löytyy erilaisia profiileja, jotka voi asentaa Firefox-selaimelle tietoturvan parantamiseksi.

Hakukoneen käytöllä on myös vaikutusta yksityisyyteen verkossa. Esimerkiksi Google kerää paljon käyttäjädataa ja vaihtoehto sille on DuckDuckGo, joka lupaa olla tallentamatta mitään käyttäjädataa, hakutietoja ja estää samalla mainoksia. DuckDuckGo lisäosan voi asentaa selaimelle, jolloin sen voi asettaa oletushakukoneeksi, mutta hakukoneen löytää myös osoitteesta https://duckduckgo.com.

Tietosuojalausunnot on tärkeä lukea, koska niiden ehtoihin on pakko suostua, jos haluaa palvelua käyttää. Usein ne ovat kuitenkin pitkiä ja sisältää lakitekstiä, jota on vaikea ymmärtää. Terms of Service; Didn't Read (ToS;DR)-lisäosa tarjoaa apua tietosuojalausuntoihin, tiivistämällä niistä tärkeimmät kohdat ja ilmoittamalla A:sta D:hen arvion siitä, kuinka luotettavasta sivustosta on kyse. Lisäosa ilmoittaa tiedot heti sivustolle selattaessa, mutta samat tiedot löytävät myös tosdr.org-sivulta, jolloin lisäosaa ei ole pakko ladata, nähdäkseen arvioita tietosuojalausunnoista. Projekti on vielä alussa ja vain pieni osa tietosuojalausunnoista on arvosteltu, mutta usean sivun lausuntojen pääkohtia on listattu sivustolle paljon. ToS;DR ylläpitää myös sivustoa tosback.org, joka on EFF:n ja Internet Securityn kanssa yhteistyössä aloitettu sivusto, joka tarkistaa joka päivä, onko muutoksia tehty tietosuojalausuntoihin, jotka sivustolla on listattu. Tätä opinnäytetyötä kirjoittaessa projekti oli vielä testausvaiheessa. (TOSBack n.d.)

Tietokoneen ja puhelimen ohjelmien ja sovelluksien oikeudet kannattaa tarkistaa ja rajoittaa niitä jos mahdollista. Ohjelmat ja sovellukset joita ei käytä, kannattaa poistaa kokonaan.

Verkossa ei ole aina pakko kertoa oikeita tietoja. Esimerkiksi Apple tarjoaa puhelimissaan mahdollisuuden luoda, niin sanotusti kertakäyttöisen sähköpostiosoitteen, jonka voi antaa sovellukselle, jonka haluaa ladata. Osoite on voimassa, niin kauan kuin on tarvetta ja tuhoutuu sen jälkeen. Vastaavaa käytäntöä voi harjoittaa myös internetissä sivustoilla, jotka vaativat sähköpostiosoitteen, mutta eivät kuitenkaan lähetä käyttäjälle mitään oleellista. Esimerkiksi 10MinuteMail-sivusto tarjoaa sähköpostiosoitteen kymmeneksi minuutiksi, minä aikana varmistusviestin (jonka sivusto, joka sähköpostiosoitteen vaatii, yleensä lähettää) ehtii lukea ja sen jälkeen sähköpostiosoitetta, ei enää ole olemassa. Eri sivustoille kannattaa aina luoda omat profiilit, eikä kannata kirjautua Facebook-tunnuksilla muihin palveluihin, koska data sivustolta kerääntyy suoraan henkilökohtaiseen Facebook-profiiliin, jolloin käyttäjästä saadaan tarkempi kuva rakennettua.

5 Tulokset

Käyttäjätietoja kerätään verkossa yritysten, organisaatioiden, viranomaisten ja valtioiden toimesta. Käyttäjätietoja kerätään palveluiden ja mainosten kohdentamiseksi sekä käyttökokemuksen parantamiseksi, räätälöimällä sivustoja käyttäjäystävällisemmiksi datan perusteella. Big data analysoinnin avulla valtavasta määrästä dataa pystytään ennustamaan, paitsi ihmisten käyttäytymistä, myös tulevaisuuden tarpeita, ja tätä tietoa voidaan hyödyntää kaupunkisuunnittelussa, rikosten ehkäisyssä ja tekoälyn kehityksessä. Käyttäjädataa kerätään muun muassa evästeiden, ETagien, jäljitteiden ja erilaisten sormenjälkitunnistusten kautta. Muitakin keinoja varmasti on, kuten cross-device tracking, jossa saman henkilön eri laitteet pyritään tunnistamaan ja keräämään näistä kaikista saatu data yhteen. Yritykset eivät kuitenkaan halua jakaa tietoa siitä, kuinka he käyttäjiä seuraavat.

Jos keskiverto ihminen käytti vuonna 2018 aikaa sosiaalisessa mediassa 144 minuuttia päivässä, kuinka paljon sisällöstä oli sellaista, joka oli algoritmien perusteella suunnattu juuri hänelle (Average Time Spent Daily on Social Media (with 2019 Data))? Millaisia valintoja hän teki, jotka perustuivat sisältöön, jonka hän oli nähnyt verkossa? Lähdekriittisyyttä täytyy harjoittaa verkossa koko ajan, koska profiloinnin avulla pyritään vaikuttamaan juuri tietyn käyttäjän mielipiteisiin. Koska dataa kerätään jopa joka klikkauksen tasolla, käyttäjästä saa hyvinkin tarkan kuvan pelkän verkkoselaamisen avulla. Palvelimille kertyy valtavasti tietoja käyttäjästä, ja tietojen varastaminen tietovuoden seurauksena, voi johtaa identiteettivarkauteen tai kiristykseen, mikä saattaa vaikuttaa käyttäjään vielä vuosienkin päästä.

Mainonnan takia kerätään käyttäjätietoja ja dataa, mutta kuka haluaisi, että heidän jokaista klikkausta seurataan? Mainostajia kuitenkin kiinnostaa se, mikä käyttäjää kiinnostaa, ei se mitä hän tekee. Mainostajien ainoa tavoite on saada tuotteita myydyksi, ja vaikka käyttäjiä saattaa häiritä se, että heidän tietojaan kerätään, tuskin kukaan voi valittaa siitä, että mainokset, joita he näkevät, ovat aiheista, jotka heitä kiinnostaa. Siskoni totesi yksi päivä, että on se hyvä, että Instagramissa tuli mainos siitä, että eräs artisti on tulossa Suomeen, koska hän ei sitä muuten tietäisi ja voisi ostaa lippua. Jotta mainonta ei vaikuttaisi yksityisyyden loukkaamiselta, yritykset voisivat

kertoa, kuten muun muassa Facebook jossain määrin tekee, ”miksi sinä näet tämän mainoksen”. Näin ihmiset, joita häiritsee se, että heidän tietojaan kerätään, saavat tietoa, mistä mainostajat ovat tiedot saaneet ja käyttäjä voi tehdä tietyt toimet sen estämiseksi. Toisaalta käyttäjät, joita ei itselleen sopivat mainokset haittaa, saa rauhan tietäessään, että esimerkiksi puhelin ei kuunnellut salaa heidän keskusteluaan aiheesta.

Selaimille on olemassa monia erilaisia lisäosia estämään käyttäjädatan keräämistä, eikä näiden asennus vie käyttäjältä juuri aikaa tai vaadi osaamista. Kun lisäosan on kerran asentanut, kannattaa laittaa automaattiset päivitykset päälle, niin on saanut jo jonkin verran suojausta seurannalta. VPN:n avulla on mahdollista suojata identiteettiä ja jos haluaa käyttää palveluntarjoajaa, luotettavan palvelun saamisesta täytyy maksaa. Tor-verkko on tarkoitettu anonyymille selaamiselle, minkä takia se onkin saanut kyseenalaisen maineen. Vaikka tor onkin hyödyllinen ilmiantajille ja toimittajille, sekä käytettäväksi maissa, joissa sensuuri on suurta, ei se peruskäyttäjälle ole tarpeellinen ja toimii huomattavasti hitaammin kuin perusselain. Yksityisyys on hyvästä, mutta toisaalta on lohdullista tietää, että tarvittaessa lähde-IP:t pystytään selvittämään, vaikka käytössä olisi ollut tor-selain. Täysin anonyymi internet päästää valloilleen pedofiilit ja muut rikolliset, terroristit, nettikiusaajat ja trollit.

Vaikka selaimet ja lisäosat kuinka lupaisivat yksityisyyttä, on verkossa pakko muistaa, että se minkä sinne kerran laittaa, pysyy siellä, eikä voi luottaa siihen, että seurantaa ei tapahtuisi. Jos haluaa estää Facebook seurannan, on luotettava siihen, että lisäosa, jonka sen estämiseksi lisää, ei kerää käyttäjädataa, koska sille on kuitenkin asennuksen yhteydessä pakko antaa lupa seurata käyttäjää, jotta se osaa etsiä Facebook-lähteitä.

Käyttäjädataa kerätään valtavia määriä ja käyttäjän on vaikea, myös GDPR aikaan tietää, kuinka paljon hänen dataansa jaetaan ja kenellä sitä on. GDPR lisäsi tietoisuutta datan keräämisestä, mutta se ei millään lailla kieltänyt sitä. Käyttäjädatan keräämistä vastaan on kehitetty lisäosia ja keinoja, mutta se ei poista ongelmaa niin kauan kuin datan kerääminen on laillista. Verkossa on myös mietittävä, kuka datan omistaa. Jos

lataa kuvan sosiaaliseen mediaan käyttäjä kokee, että hän omistaa kuvan, se on ladattu hänen profiiliinsa ja hän on mahdollisesti sen myös ottanut. Mutta koska se sijaitsee jonkun toisen omistamalla palvelimella, saattaa palvelun tarjoaja myöskin ajatella omistavansa kuvan. Esimerkiksi tästä syystä tietosuojalausunnot olisi tärkeä lukea, jotta tietää, mitä oikeuksia antaa palvelulle ja siksi tietosuojalausuntojen pitäisi olla mahdollisimman helposti luettavissa.

Ihmisen käytös muuttuu silloin, kun hän tietää, että häntä seurataan. Verkossa on helppo unohtaa, että joku saattaa seurata ja tallentaa tietoja, silloin on oma itsensä esittämättä ja varomatta, että paljastaa itsestään tietoja, joita ei halua ulkopuolisten tietävän. Sen takia on tärkeää kasvattaa tietoisuutta siitä, että käyttäjädataa kerätään, analysoidaan ja tallennetaan jatkuvasti. Mitä enemmän ihmiset ymmärtävät datankeruusta, sitä vastuuntuntoisempia heistä tulee sen jakamisessa, ja useammat alkavat vaatia oikeuksia omiin tietoihinsa ja estää niiden jakamista kolmansille osapuolille.

6 Pohdinta

Tämän tutkimuksen tutkimusongelmana oli käyttäjädatan kerääminen internetissä. Tutkimuskysymykseen, kuinka käyttäjädataa seurataan verkossa, etsittiin ratkaisua kvalitatiivisen tutkimusmetodologin kautta. Tietoa aiheesta haettiin kirjoista, verkkojulkaisuista ja aiemmista tutkimuksista ja näiden pohjalta ilmiötä tarkasteltiin eri puolilta. Koska aihe oli hyvin laaja, tutkimus rajattiin pelkästään internetissä selaimen kautta tapahtuvaan seurantaan.

Kvalitatiivinen tutkimusmenetelmä sopi työhön hyvin, koska tarkoituksena oli saada ilmiöstä syvällinen näkemys ja hyvä kuva. Laadullisen aineiston tulkinnan ainoa keino on lukeminen. Suurin osa tämän opinnäytetyön aineistosta on peräisin verkosta, ja koska tekniikka kehittyy koko ajan, lähteiksi pyrittiin valitsemaan mahdollisimman uusia julkaisuja. Työtä tehdessä huomasi, miten laajasta ilmiöstä on kyse ja tässä työssä siitä saatiin avattua vain pieni osa, mutta mielestäni käyttäjädatan seuraamisen laajuus saatiin tuotua esille. Kvalitatiivinen tutkimus sopi työhön myöskin sen

puolesta, että ei lähdetty tuottamaan ratkaisua käytännössä käyttäjäseurannan tuomiin ongelmiin, vaan kuvattiin, kuinka se tapahtuu ja kuinka käyttäjän on mahdollista minimoida riskejä. Tutkimuksia siitä, kuinka käyttäjäseuranta olisi jo aiheuttanut haittaa, ei löytynyt, mutta esimerkiksi tutkivan journalismin kautta on paljastettu, kuinka käyttäjädataa ja käyttäjien manipulointia on jo hyödynnetty demokraattisissa vaaleissa. Yritykset ja organisaatiot eivät myöskään kerro julkisesti, kuinka he käyttäjädataa keräävät, joten tässä on luotettava tietoon, joka saadaan tietotekniikka-alan asiantuntijoilta, ei yrityksiltä itseltään.

Aihetta olisi voinut rajata vielä tarkemmin esimerkiksi pelkän sosiaalisen median seurantaan, jota tässä työssä tarkasteltiin lähinnä vain muutamien esimerkkien kautta. Muita tutkimuksia aiheeseen liittyen, voisi olla puhelinten ja sovellusten kautta tapahtuva seuranta tai kuinka eri laitteiden välinen seuranta (cross-device tracking) tapahtuu.

Toimeksiantajanalle Jyväskylän ammattikorkeakoulun IT-instituutin JYVSECTEC kyberturvallisuuden tutkimus-, kehitys- ja koulutuskeskuksen ja POLAMKin yhteishankkeelle CYBERDI:lle työ toi tietoa siitä, miten käyttäjätietoja kerätään verkossa ja mitä datalla tehdään, ilman että asiaa tarkasteltiin liian teknisestä näkökulmasta, joten myös muun alan ammattilaiset voivat ymmärtää, mistä tutkimuksessa on kyse. CYBERDI:n yhtenä osa-alueena on tietoisuuden ja osaamiseen kasvattaminen, jossa työtä voidaan hyödyntää.

Lähteet

Acar, G., Diaz, C., Englehardt, S., Eubank, C., Juarez, M. & Narayanan, A. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. KU Leuven, ESAT/COSIC and iMinds,Leuven, Belgium and Princeton University.

Automaattinen päätöksenteko ja profilointi. N.d. Tietosuojavaltuutetun toimiston verkkosivujen informaatiosivu. Viitattu 29.10.2019. https://tietosuoja.fi/automaattinen-paatoksenteko-profilointi

Average Time Spent Daily on Social Media (with 2019 Data). N.d. Broadband Serchin verkkosivuilla oleva artikkeli ihmisten sosiaalisen verkon käytöstä. Viitattu 30.9.2019. https://www.broadbandsearch.net/blog/average-daily-time-on-social-media

Bakir, B. 2019. Loppukäyttäjien suhtautuminen digitaaliseen jalanjälkeen. Opinnäytetyö, AMK. Haaga-Helia ammattikorkeakoulu, tietojenkäsittelyn koulutusohjelma. Viitattu 13.10.2019. https://www.theseus.fi/bitstream/handle/10024/165935/OPINN%C3%84YTETY%C3%96.pdf?sequence=2.

Barasz, K., John, L.K. & Kim, T. 2018. Why Am I Seeing This Ad? The Effect of Ad Transparency on Ad Effectiveness. Oxford University Press on behalf of Journal of Consumer Research, Inc. 2018. Viitattu 1.11.2019. https://www.hbs.edu/faculty/Publication%20Files/Kim%20et%20al%202019%20-%20Why%20Am%20I%20Seeing%20This%20Ad_c30f31d3-1f7f-4fe3-a167-73e2d35e40b8.pdf

Brinkmann, M. 2014. How to make sure that ETags are not used to track you on the Internet. Viitattu 7.11.2019. https://www.ghacks.net/2014/03/11/make-sure-etags-used-track-internet/

Cadwalladr, C. & Graham-Harrison, E. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian 17.3.2018. Viitattu 11.10.2019. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Claypoole, T. & Payton, T.M. 2014. Privacy in the Age of Big Data : Recognizing threats, defending your rights, and protecting yout family. United States of America: Rowman & Littlefield.

Cookie Quick Manager. N.d. Firefox Browser Add-Ons. Viitattu 15.11.2019. https://addons.mozilla.org/en-US/firefox/addon/cookie-quick-manager/?src=search

Cukier, K. 2014. Big data is better data. TED Talk-konferenssipuhe, kesäkuussa 2014 Berliinissä. Lataaja TED. https://www.youtube.com/watch?v=8pHzROP1D-w

CYBERDI. 2019. CYBERDI-projektin esittelysivusto. Viitattu 26.8.2019. https://www.jamk.fi/fi/Tutkimus-ja-kehitys/projektit/CYBERDI/etusivu/.

Do Not Track -asetuksen käyttöön ottaminen ja käytöstä poistaminen. N.d. Google Chrome Ohjeet. Viitattu 13.11.2019. https://support.google.com/chrome/answer/2790761?visit_id=637092675481246372-2235289403&p=settings_do_not_track&rd=1

Dreyfuss, E. 2017. Wired-verkkoartikkeli 29.3.2017. https://www.wired.com/2017/03/wanna-protect-online-privacy-open-tab-make-noise/

ETag. 2019. MDN web docs, päivitetty 28.8.2019. Viitattu 7.11.2019. https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/ETag

Facebook Container. N.d. Firefox Browser Add-ons-sivulta. Viitattu 14.11.2019. https://addons.mozilla.org/fi/firefox/addon/facebook-container/

F-Secure FREEDOME VPN. N.d. Tuotetietoja F-Securen sivuilta Freedomesta. Viitattu 20.11.2019. https://www.f-secure.com/fi/home/products/freedome

Gencoglu, O., Honko, H., Isomursu, M. & Similä, H. 2015. Collecting a citizen's digital footprint for health data mining. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Viitattu 9.10.2019. https://ieeexplore.ieee.org/document/7320158.

Greenberg, A. 2017. The Grand Tor: How to Go Anonymous Online. Wired-verkkoartikkeli. Viitattu 12.11.2019. https://www.wired.com/story/the-grand-tor/

Greenwald, G. 2014. Why privacy matters. TED Talk-konferenssipuhe, lokakuussa 2014 RioDeJaneirossa. Lataaja TED. Viitattu 15.10.2019. https://www.youtube.com/watch?v=pcSlowAhvUk

Hirsjärvi, S., Remes, P. & Sajavaara, P. 2009. Tutki ja kirjoita. 15. uud. p. Hämeenlinna: Tammi.

Holland, M. & Jones, C. 2019. What is your digital footprint? ITPro 30.9.2019. Viitattu 12.10.2019. https://www.itpro.co.uk/strategy/29259/what-is-your-digital-footprint

How We Prevent Click Fraud. N.d. PPC Protect-verkkosivuilta tieto, kuinka PPC Protect toimii. Viitattu 13.11.2019. https://ppcprotect.com/how-it-works/

HTTP cookies. 2019. MDN web docs, päivitetty 5.8.2019. Viitattu 6.11.2019. https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies

HTTPS Everywhere. EFF Electronic Frontier Foundation-sivustolta. Viitattu 16.11.2019. https://www.eff.org/https-everywhere

Identiteettivarkaudessa esiinnytään toisen henkilöllisyydellä. N.d. Tietoa identiteetti-varkaudesta rikosuhripäivystyksen verkkosivuilla. Viitattu 27.10.2019. https://www.riku.fi/erilaisia-rikoksia/identiteettivarkaus-2/

Jantunen, M. 2017. Big datan analysointi. Opinnäytetyö, AMK. Lahden ammattikorkeakoulu, tekniikan ala, tietotekniikan koulutusohjelma, ohjelmistotekniikka. Viitattu 14.10.2019. https://www.theseus.fi/bitstream/handle/10024/128517/Jantunen_Mikael.pdf?sequence=1

Jokinen, J. 2018. Personal Internet Privacy and Surveillance : Implementation and evasion of user tracking. Opinnäytetyö, YAMK. Jyväskylän ammattikorkeakoulu, tietotekniikan koulutusohjelma. Viitattu 5.11.2019. https://www.theseus.fi/bitstream/handle/10024/146658/Jokinen_Juha.pdf?sequence=1&isAllowed=y

Kaikki mitä tarvitset parempaan verkkokokemukseen. N.d. Nord VPN verkkosivulta. Viitattu 20.11.2019. https://nordvpn.com/fi/features/

Kananen, J. 2015. Opinnäytetyön kirjoittajan opas : näin kirjoitan opinnäytetyön tai pro gradun alusta loppuun. Suomen Yliopistopaino Oy – Juvenes Print.

Khojaye, M. A. & Shamsi, J. 2018. Understanding Privacy Violations in Big Data Systems. IT Professional 20, 3, 73 - 81. Viitattu 10.10.2019. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8378964

Leppänen, V. & Ruohonen, J. 2017. Whose hands are in the Finnish cookie jar? 2017 European Intelligence and Security Informatics Conference. Viitattu 6.1.2019. https://ieeexplore-ieee-org/document/8240779

Luottamuksellinen viestintä. 2019. Traficom liikenne- ja viestintäviraston Kyberturvallisuuskeskuksen verkkosivuilta, päivitetty 30.10.2019. Viitattu 6.11.2019. https://www.kyberturvallisuuskeskus.fi/fi/toimintamme/saantely-ja-valvonta/luottamuksellinen-viestinta

Matuszewska, K. 2019. What Is Evercookie and Why You Should Avoid It for Privacy's Sake. Piwik Pro verkkosivuilta, julkaistu 19.7.2018, päivitetty 24.6.2019. Viitattu 15.11.2019. https://piwik.pro/blog/what-is-evercookie-and-why-you-should-avoid-it-for-privacys-sake/

Michalowicz, M. 2018. Big Data Business Benefits: The Perks of Predictive Analysis. 2018. American Express verkkoartikkeli. Viitattu 8.10.2019.

https://www.americanexpress.com/en-us/business/trends-and-insights/articles/big-data-business-benefits-the-perks-of-predictive-analysis/

Mikä on henkilötieto? N.d. Tietosuojavaltuutetun toimiston verkkosivujen informaatiosivu henkilötiedoista. Viitattu 30.9.2019. https://tietosuoja.fi/mika-on-henkilotieto

Privacy Badger. N.d. Firefox Browser Add-Ons. Viitattu 15.11.2019. https://addons.mozilla.org/fi/firefox/addon/privacy-badger17/?src=search

Pseudonymisoidut ja anonymisoidut tiedot. N.d. Pseudonymisoidut ja anonymisoidut tiedot. Viitattu 30.10.2019. https://tietosuoja.fi/pseudonymisointi-anonymisointi

Rautanen, S. 2019. Epäiletkö, että puhelimesi salakuuntelee keskusteluja? Teimme yksinkertaisen testin ja saimme hämmentävän tuloksen. Aamulehden verkkoartikkeli tilaajille, 16.9.2019, päivitetty 14.10.2019. Viitattu 17.11.2019. https://www.aamulehti.fi/a/703b3cec-b26b-40ac-91da-9e69313f812d

Rouse, M. 2019. Big data analytics. Big datan analysoinnista kertova artikkeli TechTarget-sivustolla. Viitattu 20.10.2019. https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics

Salo, I. 2013. Big Data, tiedon vallankumous. Jyväskylä: Docendo

Stockley, M. 2015. Anatomy of a browser dilemma – how HSTS 'supercookies' make you choose between privacy or security. Naked security by Sophos-verkkosivulta. Viitattu 16.11.2019. https://nakedsecurity.sophos.com/2015/02/02/anatomy-of-a-browser-dilemma-how-hsts-supercookies-make-you-choose-between-privacy-or-security/

Suomen kyberturvallisuusstrategia 2019. 2019. Valtioneuvoston periaatepäätös 3.10.2019. Viitattu 27.10.2019. https://turvallisuuskomitea.fi/wp-content/uploads/2019/10/Kyberturvallisuusstrategia_A4_SUOMI_WEB_300919.pdf

Tietokäytäntö. 2018. Facebookin tietokäytäntö, jota päivitetty viimeksi 19.4.2018. Viitattu 31.10.2019. https://www.facebook.com/full_data_use_policy

Tietoturvaloukkaukset. N.d. Tietosuojavaltuutetun toimiston verkkosivujen informaatiosivu tietoturvaloukkauksista. Viitattu 20.10.2019. https://tietosuoja.fi/tietoturvaloukkaukset

TOSBack. N.d. The terms-of-service tracker Terms of Service Didn't Read-verkkosivulta. Viitattu 16.11.2019. https://tosback.org/

Trackers and scripts Firefox blocks in Enhanced Tracking Protection. N.d. Mozilla Firefoxin ylläpitosivu. Viitattu 13.11.2019. https://support.mozilla.org/fi/kb/trackers-

and-scripts-firefox-blocks-enhanced-track?as=u&utm_source=inproduct#w_social-media-trackers

Usein kysyttyä EU:n tietosuoja-asetuksesta. N.d. Tietosuojavaltuutetun verkkosivuilta usein kysyttyä vastaussivusto. Viitattu 15.10.2019. https://tietosuoja.fi/gdpr

What is a Web Bug/Beacon? N.d. What is my IP address-verkkosivulta. Viitattu 7.11.2019. https://whatismyipaddress.com/web-beacon

Wlosik, M. 2019. What is a data broker and how does it work? Blogiteksti Clearcode-sivustolla. Viitattu 20.10.2019. https://clearcode.cc/blog/what-is-data-broker/

Wlosik, M. N.d. What is Behavioral Targeting and How Does It Work? Blogiteksti Clearcode- sivustolla. Viitattu 01.11.2019. https://clearcode.cc/blog/behavioral-targeting/