YKSITYISYYDEN SUOJAA PARANTAVAT VERKKOPROFIILIT

Kandidaatintyö

Lappeenrannan–Lahden teknillinen yliopisto LUT

Tietotekniikan kandidaatintutkielma

2023

Veikko Moilanen

Tarkastaja: Tutkijaopettaja Jouni Ikonen

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT Teknis-luonnontieteellinen

Tietotekniikka

Veikko Moilanen

Yksityisyyden suojaa parantavat verkkoprofiilit

Tietotekniikan kandidaatintyö

2023

32 sivua ja 23 kuvaa.

Tarkastaja: Tutkijaopettaja Jouni Ikonen

Avainsanat: Yksityisyys, internet, Pi-hole, Raspberry Pi, DNS-palvelin, mainokset, mainosten esto, seurantatyökalut, seuranta.

Työn tavoitteena on tutkia internetin käyttäjiin kohdistuvaa seurantaa erilaisten seurantatapojen toiminnan ja seurannanestotyökalujen osalta. Työssä kartoitetaan ja pohjustetaan tietoa seuraimiin ja verkkoselailun yksityisyyteen liittyen tieteellisiä tutkimuksia hyödyntämällä. Löydetyistä tutkimuksista käsitellään niitä erityisesti seurainten yleisyyden ja toimintatapojen osalta, käsitellen tietoa, joka antaa kattavan kuvan seurainten toiminnasta internetselailussa. Yleisimmät seurannan estotavat sekä niiden toimintatavat käydään läpi ja niiden eroavaisuudet sekä käyttötarkoitukset tuodaan ilmi.

Työssä testataan kotiverkon laajuista DNS-suodatukseen perustuvaa laitetta. Laite eroaa toimintatyypiltään läpikäydyistä muista yleisistä käyttäjäystävällisistä tavoista parantaa yksityisyyttä verkossa sekä on myös laajempikäyttöinen. Kun laite konfiguroidaan olemaan koko kotiverkon laajuinen DNS-palvelin, saadaan seurannanestohyödyt koko verkon laitteille käyttöön. Laitteen vaikutusta käyttäjän verkkoselauskokemukseen testataan yhden laitteen tapauksessa useilla esimerkeillä.

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

School of Engineering Science

Software Engineering

Veikko Moilanen

Privacy enabled network profiles

Bachelor's thesis

2023

32 pages and 23 figures.

Examiner: Associate professor Jouni Ikonen

Keywords: Privacy, Internet, Pi-hole, Raspberry Pi, DNS-server, ads, blocking ads, blocking trackers, anti-tracking tools, tracking, tracking on the web, tracking on the internet.

The aim of the work is to gather information on tracking of users in terms of the operation of different tracking methods and anti-tracking tools. The work maps and primes information related to trackers and the privacy of web browsing privacy by utilizing scientific studies. Of the studies found, they are discussed in particular regard to the prevalence and methods of operation of trackers, processing information that gives a comprehensive picture of the influence of trackers on private browsing. The most common tracking prevention methods and their operating methods are reviewed and their differences and intended uses are revealed.

In the work a home network-wide device that is based on DNS filtering is tested. The device differs from the other general user-friendly ways to improving privacy online and also works more broadly. When the device is configured to be a DNS server for the entire home network, the anti-tracking benefits are applied for the devices of the entire network. The effect of the device on the user's online browsing experience is tested in the case of one device with multiple examples.

Sisällysluettelo

1 Johdanto

1.1 Tausta

Nykymaailma on digitalisoitunut, mikä on johtanut siihen, että lähes jokainen ihminen on päivittäin jollain tavalla yhteydessä internetiin. Harvempi kuitenkaan tietää, mitä kaikkea itsestään jakaa ja kenelle selatessaan verkkoa, sosiaalisen median alustoja tai esimerkiksi käyttäessä puhelinsovelluksia. Lerner ym. [1] kertoo tutkimuksessaan, että yleisimmät 20 seurainta (engl. tracker) kattaa jopa 70 prosenttia eniten vierailtujen verkkosivujen listasta. Seuraimet ovat joko esimerkiksi tiedostoja kuten evästeitä, jotka sisältävät dataa käyttäjästä ja jota jaetaan kolmansille osapuolille, tai esimeriksi seurantatapoja, joiden avulla käyttäjä voidaan yksilöidä käyttämättä erinäisiä tiedostoja. Käyttäjien seuranta verkossa erinäisin tavoin on hyvin yleistä, ja vasta viime vuosina on seurantaan tullut muutoksia. Esimerkiksi Euroopassa vuonna 2018 voimaan tullut Euroopan GDPR on tietosuojalakipaketti, joka velvoittaa yrityksiä eri tavoin, parantaen käyttäjän yksityisyyttä sekä eurooppalaisten datan turvallisuutta [14].

Myös isot yritykset kuten Apple ovat parantamassa käyttäjien yksityisyyttä erinäisten toimien myötä. Applen luoma selain, Safari, käyttää koneoppimista seurainten tunnistamisessa ja estämisessä. Tämän lisäksi sormenjälkiteknologiaa käyttävät yritykset saavat täsmällisemmän personoidun datan sijaan yksinkertaisemman profiilin, joka jaetaan verkkosivuille kaiken datan sijasta. Applella on myös monia muita yksityisyyttä parantavia työkaluja, kuten ”App Tracking Transparency”. Käyttäjiltä on tämän myötä pakollista kysyä lupa, jos seurantaa halutaan jatkaa sovellusten ja verkkosivujen yli. Tämän avulla käyttäjä voi evätä lupia sovellusten kysyessä sovellusten välisen seurannan käytöstä. [8]

Yksityisyys on käsitteenä laaja. Eri konteksteissa voi yksityisyydelle löytää erilaisia määritelmiä. Yksityisyyteen liitetään esimerkiksi ihmisen oikeus ”suojautua ulkopuoliselta puuttumiselta” [15]. Esimerkiksi kirjesalaisuus, joka takaa, että ainoastaan saaja saa avata kirjeen, koskee usein myös digitaalisia viestejä. Yksityisyyteen liittyy yleensä ihmisen oikeus määrätä itseensä liittyvistä asioista, kuten itseensä liittyvien tietojen käsittelystä.

Internet on mahdollistanut täysin uudenlaisen yksityisen tai osittain yksityisen tiedon keräämisen ja hyödyntämisen erilaisiin tarkoituksiin, mistä esimerkiksi verkkosivujen selaaja ei välttämättä ole tietoinen. Seuraimet ja niiden esto liittyvät siis suoraan käyttäjän yksityisyyteen. Käyttäjien seuranta on nykypäivänä todella yleistä, minkä takia on tärkeää, että käyttäjät tietävät, kuinka heitä seurataan ja miten siltä voi yrittää välttyä.

Tämä kandidaatintyö käsittelee käyttäjien seurantaa verkossa. Pääasioihin kuuluvat seuraimien toiminta yleisesti, yleisimmät seurannan estokeinot ja työkalut seurannan estämiseen sekä näiden estotyökalujen vaikutus verkkosivujen toimintaan.

Tämän kandidaatintyön tutkimuskysymykset ovat seuraavat:

1. Miten käyttäjien seurantaa verkossa on tutkittu akateemisesti?
2. Mitä yleisiä keinoja tai työkaluja seuraamisen estämiksi on olemassa?
3. Miten käyttäjä voi parantaa yksityisyyttään verkossa laajemmin?

1.2 Tavoitteet ja rajaukset

Tämän työn tarkoituksena on parantaa tietoisuutta käyttäjien seurannasta, kertoa yleisistä työkaluista seurannan vähentämiseksi sekä antaa yksi ratkaisu lisää seurannan vähentämistä varten.

Kirjallisuuskatsaus on rajattu niin, että vain akateemiset tutkimukset käyttäjien seurantaan liittyen huomioidaan. Kun työssä pohjustetaan, miten seuranta toimii, mitä työkaluja on olemassa ja miten työkalut vaikuttavat verkkosivujen toimivuuteen, ei lähteitä voida rajata vain akateemisiksi tai työ ei olisi yhtä kattava.

1.3 Työn rakenne

Työn rakenne jakaantuu kirjalliseen osioon sekä laitteen suunnitteluosioon. Työn kirjallisen osuuden tarkoituksena on selvittää, miten seurantaa on tutkittu akateemisesti ja kertoa yleisimmistä seurantatavoista sekä mitkä ovat yleisimmät työkalut seuraamisen estämiseksi. Akateemisista lähteistä odotetaan löytyvän tilastotietoa siitä, kuinka tehokkaasti erilaiset seurannan estotyökalut toimivat ja kuinka seuranta on kehittynyt vuosien varrella.

Akateemisten tutkimusten lisäksi työssä annetaan pohjatietoa seuraimista ja niiden estotavoista. Lähteinä käytetään muitakin kuin akateemisia lähteitä. Muista lähteistä - esimerkiksi Wikipediasta ja eri verkkosivuilta - odotetaan löytyvän neuvoja seurannan estoon sekä tietoa yleisistä työkaluista tehokkaammin kuin täysin akateemisista lähteistä.

Työn toisessa osassa rakennetaan soveltuvuusselvitys (engl. proof of concept) -laite, joka estää seurantaa ”DNS-sinkhole” -teknologialla. DNS on lyhenne Dynamic Name Systemistä, ja sitä käytetään, jotta verkkosivujen tunnukset voidaan muuntaa IP-osoitteiksi. Tässä vaihtoehdossa laite lataa mainossivun sijaan tyhjän sivun tilalle, jotta mainosta tai seuranta ei ladata ollenkaan.

2 Tutkimukset seuraimista sekä käyttäjien yksityisyydestä

Työtä varten etsittiin tutkimuksia, jotka liittyisivät jollain tavalla työn aihealueeseen. Tietoa etsittiin esimerkiksi käyttäjien seurantatavoista, seurannan yleisyydestä, seurannan estotavoista, yksityisyydestä sekä sen parantamisesta. Hyvät hakusanat auttoivat löytämään aihetta läheltä sivuavia artikkeleita. Osa tutkimuksista oli hyvin samankaltaisia, jonka takia eri tutkimuksia käydään läpi vain vähän.

2.1 Hakuprosessi

Tieteellisiä tutkimuksia lähdettiin etsimään aluksi Google Scholarista. Hakukoneen tuloksia selattiin eri hakusanoilla, jotka ovat listattu alla. Hakusanoja käytettiin erikseen. Kun sopivan näköinen tutkimus oli löytynyt, lähdettiin sitä etsimään LUT Primon kautta, joko suoraan tai siirtymällä ensin ACM (Association for Computing Machinery) tai IEEE Xplore (Institute of Electrical and Electronics Engineers) kirjaston kautta. Sopivaksi tutkimukseksi katsottiin sellaiset, joissa alla olevat hakusanat olivat vahvasti läsnä, ja jotka käsittelivät seurantaa tai käyttäjien yksityisyyttä verkkoselailun tai datanjaon osalta. Löydetyistä tieteellisistä artikkeleista karsittiin ne, jotka vähiten sivusivat tai käsittelivät aihetta. Parhaimmat artikkelit selattiin läpi, selvitettiin tutkimuksen tulokset ja tiivistettiin seuraavissa kappaleissa. Parhaimmissa artikkeleissa tehtiin tilastotietoa esimerkiksi seurainten yleisyyteen liittyen. Tämän lisäksi käyttäjien tiedonjaon halukkuutta käsittelevä tutkimus valittiin, sillä sen tuloksia voidaan verrata siihen, mitä käyttäjistä oikeasti kerätään.

Käytetyt hakusanat:

- online tracking
- users tracked online
- block trackers
- trackers
- cookies

- web-tracking
- privacy of users

Haettaessa yllä olevilla hakusanoilla, löytyi erityyppisiä tutkimuksia. Joissakin testattiin selainlisäosien toimintaa eri verkkosivuilla, joissakin kartoitettiin käyttäjien halukkuutta erilaisen datan jakamiseen ja jotkin olivat yleisemmänpuoleisia, kertoen esimerkiksi historiaa seuraimista ja niiden määrästä verkossa. Jokaisesta tutkimuksesta löytyi jotain hyödyllistä tietoa aiheeseen liittyen.

2.2 Hakuprosessissa löytyneet tutkimukset

Merzdovnik kumppaneineen [2] tutki seuraimien estotyökalujen toimivuutta. Työssä kerrottiin lyhyesti yleisimmistä seurantatavoista, yleisimmistä työkaluista sekä mitattiin, kuinka tehokkaita nämä työkalut ovat. Tutkimus löysi eroavaisuuksia myös täysin samankaltaisilta vaikuttavilta työkaluilta, esimerkiksi AdBlock Plus -selainlisäosan sekä muiden mainosten sekä seurainten estäjien väliltä. AdBlock Plus antaa joidenkin sallittujen luetteloon (engl. whitelisted) lisättyjen mainosten näkyä käyttäjälle, ellei käyttäjä tätä erikseen estä lisäosan asetuksista. Ublock Origin taas esti seuraimia tehokkaammin kuin muut samankaltaiset työkalut. Tutkimus antoi tärkeää tietoa siitä, ettei kaikki seuraimien estotyökalut ole yhtä tehokkaita tai laadukkaita kuin toiset.

Leonin ja kumppaneiden tutkimuksessa [3] tutkittiin käyttäjien halukkuutta jakaa erilaisia tietoja itsestään verkkosivuille. Tutkimuksessa pyydettiin ensin käymään terveyssivulla, lukemaan, mihin heille annettua dataa käytettäisiin, ja sitten vastata kolmenkymmenen kysymyksen kyselyyn. Monet olivat valmiina antamaan tietoa esimerkiksi selaimestaan, missä maassa tai osavaltiossa asuu ja minkä sukupuolinen henkilö on kyseessä. Käytännössä kukaan ei ollut valmis jakamaan sosiaaliturvatunnustaan, luottokorttinsa numeroaan, oikeaa osoitettaan tai IP-osoitettaan. Tutkimuksen mukaan se, mitä dataa kerätään ja kuinka kerätyn datan käyttöä rajattiin, vaikutti paljolti siihen, miten halukkaita käyttäjät olivat jakamaan dataansa.

Lernerin tutkimuksessa [1] tutkittiin seuraimien historiaa. Tutkimuksessa selvisi, että seuraimien määrä ja kompleksisuus on vain vuosien myötä, aikavälillä 1996-2016, lisääntynyt.

Tutkimus myös kertoo seuraimien yleisyydestä. Kuten johdannossa jo mainittiin, suosituimmista 20 seuraimesta on käytössä vähintään yksi yli 70 prosentilla eniten käydyillä sivuilla. Vielä vuonna 2006 sama prosenttiluku oli 30. Vuonna 2016 kaikkein suosituin seurain kattoi jopa 30 prosenttia viidestäsadasta suosituimmasta sivusta yksinään.

2.3 Tutkimusten ohella löytyneet havainnot

Monissa yllä mainituissa sekä monissa aihetta käsittelevissä tutkimuksissa tulee esille nettisivujen toimimattomuus joillakin seurannan estokeinoilla. Esimerkiksi mainostenesto-tyyppistä selainlisäosaa käyttämällä voi käyttäjä joutua laittamaan lisäosan pois päältä, jotta pääsee selaamaan esimerkiksi uutissivua [9]. Jotkin sivut voivat näyttää käyttäjän silmään oudoilta, kun jotkut sivun osat eivät lataudu. Tästä näemme esimerkin kandidaatintyön viidennessä luvussa.

Sivustojen oheisvaikutukset seurannanestotyökalua kuten selanlisäosaa käyttämällä voi vaikuttaa käyttäjän päätökseen olla käyttämättä seurannan estotyökalua joillakin sivuilla, joka on yksityisyyden kannalta huono asia. Tämän vuoksi työssä myös käsitellään näitä tapauksia hieman.

3 Miten käyttäjiä seurataan verkossa

Selatessasi verkkoa, verkkosivut saavat monenlaista dataa sinusta ja käyttämästäsi laitteesta. Jotkin tiedot kuten istuntoevästeet voivat olla hyödyllisiä, ja tietynlaiset palvelut kuten verkkokaupat voivat toimia järkevästi käyttämällä istuntoevästeitä. Tällöin käyttäjän toimet, esimerkiksi ostoskoriin valitut esineet, voidaan muistaa siirtyessä verkkosivulla pidemmälle. Myös diagnostiikkatiedot esimerkiksi selaimen kaatuessa voivat olla hyödyllisiä selaimen toiminnan kannalta. Kaikki data ei kuitenkaan ole sellaista dataa, joka olisi tarpeellista kerätä, ja monet entiteetit keräävätkin dataa erilaisiin tarkoituksiin, joista käyttäjä ei välttämättä ole tietoinen. Suurin osa kerätystä datasta käytetään mainostarkoituksiin. Tapoja, miten seurantaa ja datankeräystä tehdään, on monia. Seuraavaksi käsittelemme erilaisia seurantatapoja sekä yleisiä tapoja estää seurantaa.

3.1 Erilaisia seurantatapoja ja -palveluita

Erilaisia seuranta- ja datankeräystapoja on useita. Joillakin tavoilla on tarkoitus yksilöidä käyttäjiä esimerkiksi parempaa mainostenkohdentamista varten. On myös yleistä, että monet seurantatavat ovat samanaikaisesti käytössä. Yleisimpiä tapoja ovat:

- Pysyvät evästeet (engl. persistent cookies)
- HTTP E-tagit (engl. HTTP ETags)
- Selaimen sormenjälki (engl. Browser fingerprinting)
- Analyysipalvelut (engl. Analytics services)
- Mainospalvelut (engl. Advertisement services)
- Social widget -napit ja palvelut

Jotkin seurantatavat, kuten pysyvät evästeet, ovat tallessa käyttäjän laitteella, eivätkä käytännössä ikinä poistu itsestään. Ne lähettävät talletetun tiedon takaisin verkkosivulle, kun käyt samalla verkkosivulla uudestaan. Seuraavaksi avaamme yleisimmät seurantatavat pääpiirteittäin.

3.1.1 Pysyvät evästeet

Evästeet (engl. cookies) ovat yleensä pieniä, lyhyitä tekstitiedostoja, joita käytetään sekä erinäisten palveluiden tarjoamisessa käyttäjälle sekä käyttäjän seuraamisessa verkossa. Evästeiden avulla voidaan esimerkiksi pitää yllä verkkokaupassa käyttäjän ostoskorissa olevia tavaroita. Tällaisia evästeitä kutsutaan istuntokohtaisiksi evästeiksi (engl. session cookie).

Pysyvät evästeet (engl. persistent cookies) ovat kolmannen osapuolen luomia ja niiden tarkoituksena on seurata käyttäjää sivulta toiselle. Pysyvät evästeet voivat kerätä dataa aina, kun käyttäjä käy sivulla, joka käyttää tämän kolmannen osapuolen työkaluja. Esimerkiksi mainokset, sosiaalisten medioiden napit kuten Facebookin tykkäysnappi tai Twitterin jakonappi ja web-analytiikkatyökalut ovat kolmannen osapuolen työkaluja.

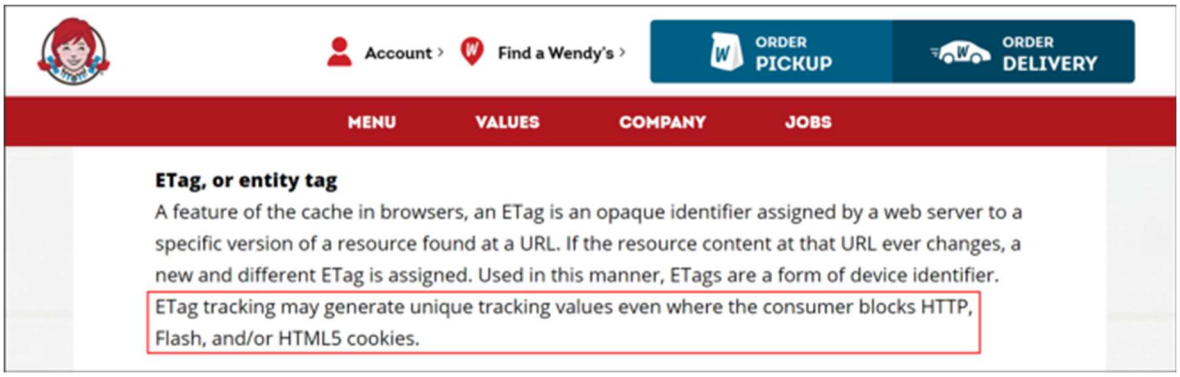
Jos esimerkiksi käyt verkkosivulla katsomassa tennismailaa ja kyseinen verkkosivu käyttää kolmannen osapuolen mainontaa, lähettää verkkosivu evästeesi mainostajalle. Tämän jälkeen vieraillessasi muilla sivuilla, jotka käyttävät samaa kolmannen osapuolen palvelua, voit nähdä mainoksia esimerkiksi tennismailoista tai tennispalloista. Mainokset siis voidaan kohdentaa evästeiden avulla todella tarkasti.

3.1.2 HTTP E-tagit

HTTP E-tagit ovat tapa tunnistaa saman resurssin, kuten verkkosivun elementin, eri versioita. Aina kun resurssin versio muuttuu, lähetetään uusi E-tagi. Jos E-tagi on sama kuin aikaisemmin, ei resurssia ladata uudestaan. Kun avaat ensimmäisen kerran verkkosivun, lataa selain koko sivun. Seuraavalla kerralla selain voi ladata sivun välimuistista, ja ladata vain ne sivun osat, jotka ovat muuttuneet viimekäynnin jälkeen, verkosta.

E-tagit ovat uniikkeja, joten niitä voidaan käyttää seuraamisessakin. E-tagi voidaan luoda esimerkiksi jokaiselle verkkosivukäyttäjälle erikseen. Kun henkilö palaa sivulle, tarkistetaan, täsmääkö välimuistin E-tagi verkkosivun esimerkiksi näkymättömään elementtiin, joka E-tagin käyttäjälle antoi. Jos ei, käyttäjä on uusi, ja hänelle luodaan E-tagi. Jos täsmää, käyttäjä on ollut verkkosivulla aiemmin. [13]

Vaikka E-tagi ei ole tehty seurantatarkoituksiin, se sopii niihin käytännössä yhtä hyvin, ainakin käyttäjän tunnistamisen osalta. E-tagit eroaa evästeistä esimerkiksi siinä, ettei niihin voi suoraan liittää dataa käyttäjään liittyen toisinkuin evästeeseen, joka voi sisältää monenlaista dataa. Kuvassa 1 näemme verkkosivun tiedoista E-tagin käytöstä seurantatarkoituksessa.



Kuva 1: Wendy's ravintolakonsernin E-tagista kertova osio eväste- ja seurantateknologiaa koskevasta käytännöstä. [13]

3.1.3 Selaimen sormenjälki

Selaimen sormenjälki on seuraustapa, joka ei perustu evästeisiin lainkaan. Käyttäjän selaimesta muodostetaan joko uniikki tai melkein uniikki sormenjälki, jonka perusteella käyttäjä voidaan tunnistaa ympäri internetiä. Sormenjälkiseurantatapa ei perustu evästeisiin.

Selaimesta voi yleensä kerätä eri tietoja, jotka yhdistämällä voidaan käyttäjä tunnistaa melko tarkasti, ellei todella tarkasti. Näitä tietoja ovat esimerkiksi selainversiot, selainlisäosat, laitteen tiedot, selainhistoria sekä käytössä oleva fontti ja fontin tiedot. Kuvassa 2 on Laperdrixin ym. tutkimuksessa ollut taulukko eri parametreista sekä niiden arvoista, mikä selventää, miten ja kuinka tarkasti eri tietoja katsomalla voidaan käyttäjän yksilöintiä tehdä.

TABLE I
BROWSER MEASUREMENTS OF AMIUNIQUE FINGERPRINTS WITH AN EXAMPLE

Attribute	Source	Distinct values	Unique values	Example
User agent	HTTP header	11,237	6,559	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2272.118 Safari/537.36
Accept	HTTP header	131	62	text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Content encoding	HTTP header	42	11	gzip, deflate, sdch
Content language	HTTP header	4,694	2,887	en-us,en;q=0.5
List of plugins	JavaScript	47,057	39,797	Plugin 1: Chrome PDF Viewer. Plugin 2: Chrome Remote Desktop Viewer. Plugin 3: Native Client. Plugin 4: Shockwave Flash...
Cookies enabled	JavaScript	2	0	yes
Use of local/session storage	JavaScript	2	0	yes
Timezone	JavaScript	55	6	-60 (UTC+1)
Screen resolution and color depth	JavaScript	2,689	1,666	1920x1200x24
List of fonts	Flash plugin	36,202	31,007	Abyssinica SIL,Aharoni CLM,AR PL UMing CN,AR PL UMing HK,AR PL UMing TW...
List of HTTP headers	HTTP headers	1,182	525	Referer X-Forwarded-For Connection Accept Cookie Accept-Language Accept-Encoding User-Agent Host
Platform	JavaScript	187	99	Linux x86_64
Do Not Track	JavaScript	7	0	yes
Canvas	JavaScript	8,375	5,533	Cwm fjordbank glyphs vext quiz, 😃
WebGL Vendor	JavaScript	26	2	NVIDIA Corporation
WebGL Renderer	JavaScript	1,732	649	GeForce GTX 650 Ti/PCIe/SSE2
Use of an ad blocker	JavaScript	2	0	no

Kuva 2: Minkälaista dataa voidaan kerätä uniikin sormenjäljen tekemiseksi [4]

Laperdrixin, Rudametkinin, ja Baudryn tekemässä tutkimuksessa [4] heidän reilusta sadasta tuhannesta selainsormenjäljestä melkein 90 prosenttia oli uniikkeja. Puhelimien osalta tämä luku oli 81 prosenttia, joka on pienempi, mutta silti todella suuri.

Acarin ym. tutkimuksessa [5] tutkittiin, kuinka yleistä selaimen sormenjäljen keräys käyttäjältä on suosituimmalta miljoonalta sivulta käyttäjän käydessä näiden sivujen etusivulla. Tutkimuksessa löydettiin 404 sivua miljoonasta sivusta, jotka keräsivät sormenjälkidataa käyttäjistään, eli noin 0.04 prosenttia testatuista sivuista. Tutkimus ei ole täysin tarkka, sillä sivut, joissa sivun tiedot on piilotettu esimerkiksi CAPTCHA-lomakkeen taakse, ei pystytty testaamaan.

Näistä neljästäsadasta neljästä sivusta, jotka sormenjälkitekniikkaa käyttävät, käyttivät vain kolmentoista eri yrityksen teknologiaa, joista vain osa oli tunnistettu aiemmissa tutkimuksissa.

Vaikka sormenjäljen kerääminen ja hyödyntäminen on aika tarkka tapa yksilöidä käyttäjiä, on kyseisen teknologian hyödyntäminen silti vain pieni osa seurantaa verkossa, ainakin tällä hetkellä.

3.1.4 Analyysipalvelut

Monilla verkkosivuilla on käytössään jonkin sortin analyysipalvelu, kuten Google Analytics. Analyysipalveluiden tarkoituksena on saada tietoa, millaiset ihmiset käyvät verkkosivulla sekä mitä he siellä tekevät.

Analyysipalvelut ovat eräänlainen osa seurantaa, vaikka kerätystä datasta ei välttämättä saada yksilöityä käyttäjiä, ainakaan kovin tarkalla tasolla. Jokainen käyttäjä, jonka dataa on kerätty, on käytännössä ”pseudoanonyymi”. Jokaisella on tunniste, josta ei suoraan saada selville, kuka henkilö on, mutta esimerkiksi IP-tietoja keräämällä voidaan tämä saada selville.

Analyysipalvelut voidaan luokitella seurantatavaksi, vaikka seuranta ei välttämättä ole yksilöllistä. Analyysipalveluiden tarkoituksena on yleensä parantaa sivujen toimivuutta sekä seurata mainosten toimivuutta esimerkiksi mainosten klikkausmäärällä verrattuna sivun kävijämäärään. Analyysipalvelut käyttävät monia mittareita saadakseen selville, miten käyttäjät heidän verkkosivuaan käyttävät. Näihin kuuluu esimerkiksi vierailuaika sivulla, avainsanat, jolla käyttäjä löytää kyseisen verkkosivun ja onko käyttäjä uusi, vanha, uudestaan käyvä vai yksittäinen kertakävijä.

3.1.5 Mainospalvelut

Mainospalvelut ovat palveluita, kuten yrityksiä, jotka suorittavat mainontaa verkkosivuilla. Mainospalvelut voivat käyttää esimerkiksi pysyviä evästeitä mainonnan kohdentamiseen. Mainospalvelut ei itsessään ole tekninen seurantatapa, mutta ne voidaan kategorisoida omaksi, sillä ne sisältävät eri seurantatapoja esimerkiksi kohdentaessaan mainontaa.

Mainoksia klikkaamalla voi päätyä jopa haitalliselle verkkosivulle, jossa käyttäjää voidaan houkutella antamaan tilitietojaan erinäisiä asioita varten. Mainokset voivat myös olla täysin

legitiimejä. Yleensä tämä riippuu alkuperäisen verkkosivun, jossa mainoksia näytetään, luotettavuudesta.

3.1.6 Social widget -napit ja palvelut

Social widget -napit ja palvelut ovat isojen sosiaalisten media-alustojen tapoja integroida palveluitaan muihin verkkosivuihin, kuten uutissivuihin. Esimerkiksi Facebookin jakonappi on yleinen näky uutissivustoilla, sillä ihmiset haluavat jakaa mielenkiintoisia uutisia ystävilleen. Toinen yleinen integroitu palvelu on Twitterin upotteet, jotka voivat olla esimeriksi videoita tai tekstiä. Jos verkkosivu käyttää jompaakumpaa yllä olevista, voivat yritykset jatkaa seurantaa omien verkkosivujensa yli myös kyseisille palveluja käyttäville verkkosivuille. Seuraavassa luvussa on esimerkki uutissivustosta, jossa Twitterin upote ei kolmannen osapuolen evästeet estämällä käyttäjälle näy.

4 Seurannan estotavoista

Seurannan estämiseen on monia eri keinoja. Selaimissa näitä ovat selainlisäosat. Selaimissa itsessään on myös joitakin työkaluja lisätty, jotka saattavat auttaa seuraamisen estossa. Seurannan estotapoihin sisältyy esimerkiksi mainostenestäjät, sisällönestäjät, DNS-sinkholet ja ”Do not track” -pyynnöt. Seuraavaksi avaamme yleisimpiä keinoja estää seurantaa verkossa.

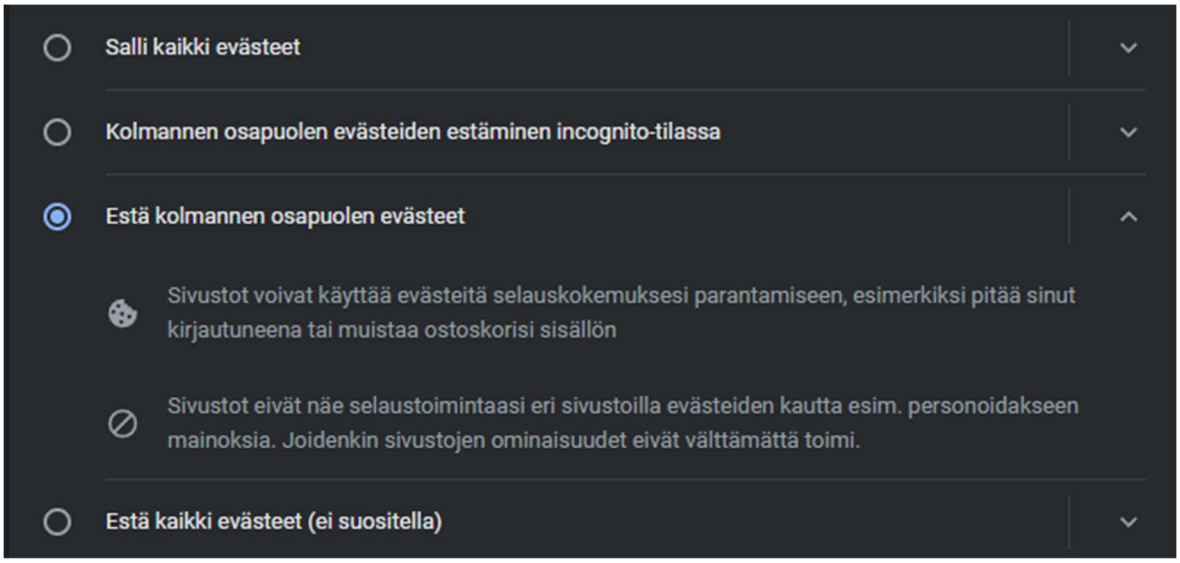
4.1 Selaimen asetukset

Nykyaikaisissa selaimissa on joitakin asetuksia, jotka voivat auttaa käyttäjää parantamaan yksityisyyttään seurannan estossa. Nämä eivät kuitenkaan välttämättä muuta verkkosivun seuraamistoimintaa mitenkään. Seuraavaksi tarkastelemme Google Chrome -selaimen seuraimiin liittyviä asetuksia.

Selain voi lähettää selaustietojen mukana ”Do Not Track” -pyynnön, joka ilmoittaa verkkosivulle, ettet halua, että sinua seurataan. Pyyntö lähetetään HTTP-tunnisteen mukana. Tämä pyyntö ei velvoita verkkosivua tekemään yhtään mitään, etkä todennäköisesti saa tietoa siitä, salliiko sivu seurannan tai seuraako sivu sinua silti vai ei. Selainten sormenjälkitutkimuksessa huomattiin, ettei sormenjälkeä luovat ja keräävät sivut huomioineet tätä pyyntöä laisinkaan. [5] Kuvassa 3 näemme suositun Google Chrome selaimen kuvauksen ”Do Not Track” -pyynnöstä.

Kuva 3: Kuvankaappaus Google Chrome-selaimen ”Do not track” -napin lisätiedoista kertovalta verkkosivulta.

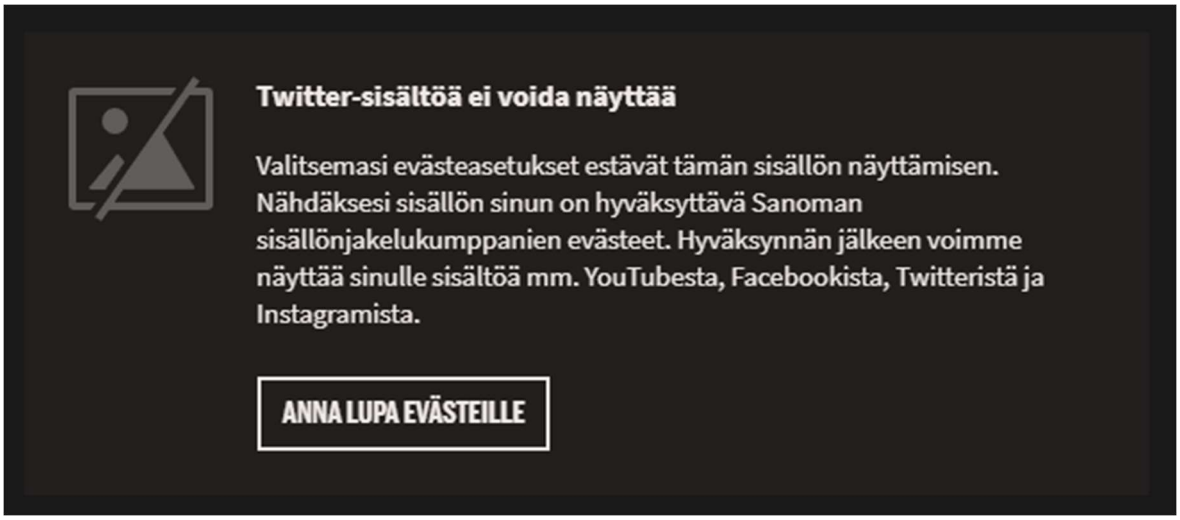
Kuvassa 4 näkyy Google Chromen asetuksista löytyvä asetus, jonka avulla käyttäjä voi estää kaikki kolmannen osapuolen evästeet. Tämän lisäksi käyttäjä voi myös estää kaikki evästeet, jota ei suositella, koska joidenkin verkkosivujen jotkin ominaisuudet voivat lakata toimimasta, jos toiminnallisia evästeitä kuten istuntoevästeitä ei sallita.



Kuva 4: Kuvankaappaus Google Chrome -selaimen evästeasetuksista.

Evästeiden estäminen ei itsessään estä kaikkea seurantaa, sillä evästeiden lisäksi seurantaan on muitakin jo aikaisemmin mainittuja keinoja. Mainokset, E-tagit ja selaimen sormenjälkitekniikka toimivat edelleen, vaikka evästeet olisi estetty. ”Social widget” -napit todennäköisesti lakkaavat toimivasta, ellei käyttäjä erikseen anna lupaa kolmannen osapuolen sisällön näyttämiseen ja evästeiden käyttämiseen.

Esimerkiksi Ilta-Sanoman hetkittäin päivittyvässä artikkelissa käytetään monessa kohtaa Twitterin upotettuja julkaisuja. Näitä varten on käyttäjän hyväksyttävä ”Sanoman sisällönjakelukumppanit” -evästeponnahdusikkunasta, jossa kerrotaan, että nämä kolmannen osapuolen tarjoajat voivat kerätä dataasi omiin tarkoituksiinsa hyväksyttyäsi evästeen. Muutoin upote ei lataudu. Kuvassa 5 ja 6 näemme tiedotteen sekä evästeiden seurantatiedosta tarkempaa tietoa liittyen isoihin sosiaalisten medioiden palveluihin, kuten Facebookiin, Twitteriin ja YouTubeen.



Kuva 5: Evästeiden estämisen takia sosiaalisen median palvelun sisältö ei näy Ilta-Sanomien sivulla (is.fi).

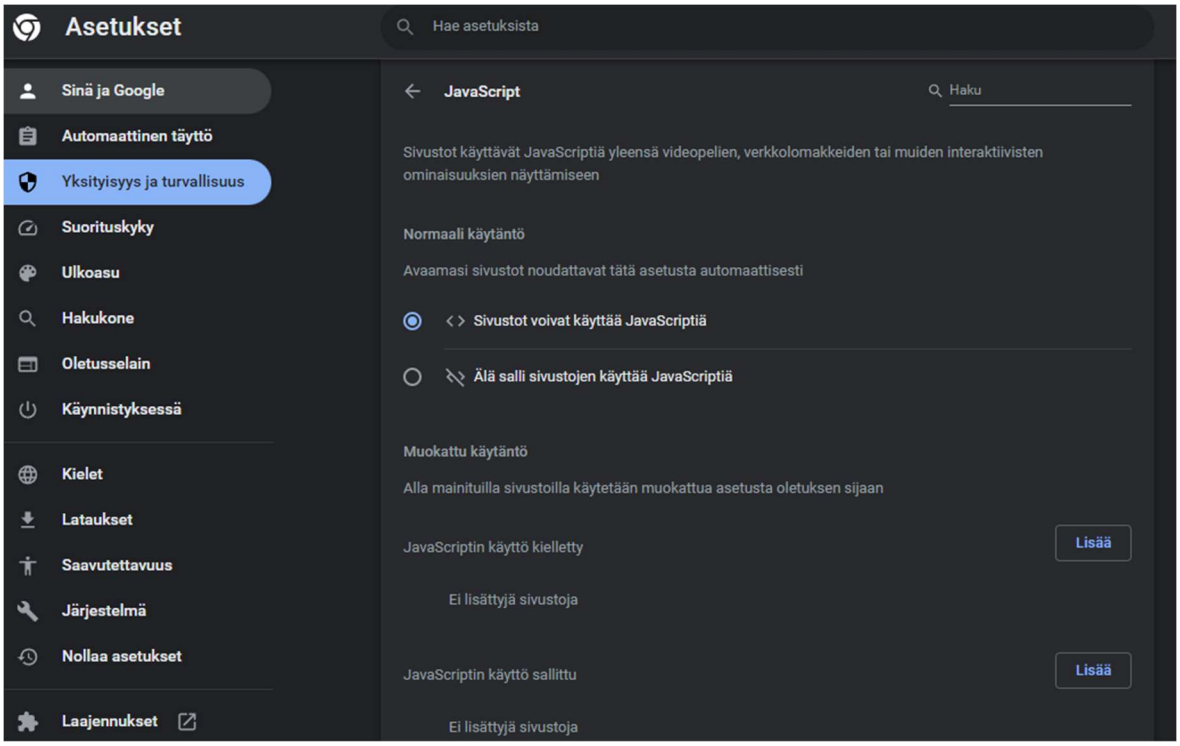
Kuva 6: Sanoman evästevalikon laajennettu kohta ”Sanoman sisällönjakelukumppanit”

Selainten asetuksissa on myös mahdollista estää JavaScriptin käyttö. JavaScript on yleisesti käytössä oleva ohjelmointikieli, jonka avulla voidaan tehdä monenlaista, myös tavallisen käyttäjän kokemusta parantavaa, toimintaa.

JavaScriptin avulla voidaan tehdä seurantaa hyödyntäen selaimen sormenjälkiteknologiaa. JavaScript sisältää monia objekteja ja ominaisuuksia, joiden kautta voidaan saada käyttäjästä tietoa. Esimerkiksi ”navigator” -objekti [6] sisältää tietoja käyttäjän geolokaatiosta, selaimen kielen, tiedon siitä, onko evästeet päällä sekä tietoa käyttöjärjestelmästä ja arkkitehtuurista. JavaScriptin estämisellä voi käyttäjä välttyä joiltain sormenjälkiteknologiaa käyttäviltä seuraimilta.

JavaScriptin estämisellä voi olla isoja verkkosivujen toimintaan vaikuttavia vaikutuksia. Tästä syystä seuraintapoja, yksityisyyttä ja seuraimia käsittelevissä tutkimuksissa ei kehuta tai suositella JavaScriptin estämistä hyvänä seurannanestokeinona, koska sillä on isoja haittavaikutuksia verkkosivujen toimintaan [6].

Google Chrome -selaimen asetuksissa JavaScript on myös mahdollista estää tai sallia tietyillä verkkosivuilla käyttäjän mieltymysten mukaan. Kuvassa 7 näemme esimerkkinä Chrome-selaimen asetuksissa mitä vaihtoehtoja käyttäjällä on JavaScriptin suhteen.



Kuva 7: JavaScript asetusten muuttaminen Chrome-selaimen asetuksissa.

4.2 Selainlisäosat

Seuraamisen estoon on tarjolla erilaisia selainlisäosia, joiden tarkoituksena on yleensä estää mainoksia sekä evästeitä. Jotkut selainlisäosat, kuten uBlock Origin, mainostaa itseään sisällönestotyökaluna (engl. content-blocking tool) eikä pelkästään mainostenestotyökaluna. Jotkin selainlisäosat, kuten AdBlock Plus, puolestaan sallii jotkut mainokset, vaikka mainostaa itseään mainosten estäjänä ja on suosituin selainlisäosa mainosten estoon. [2]

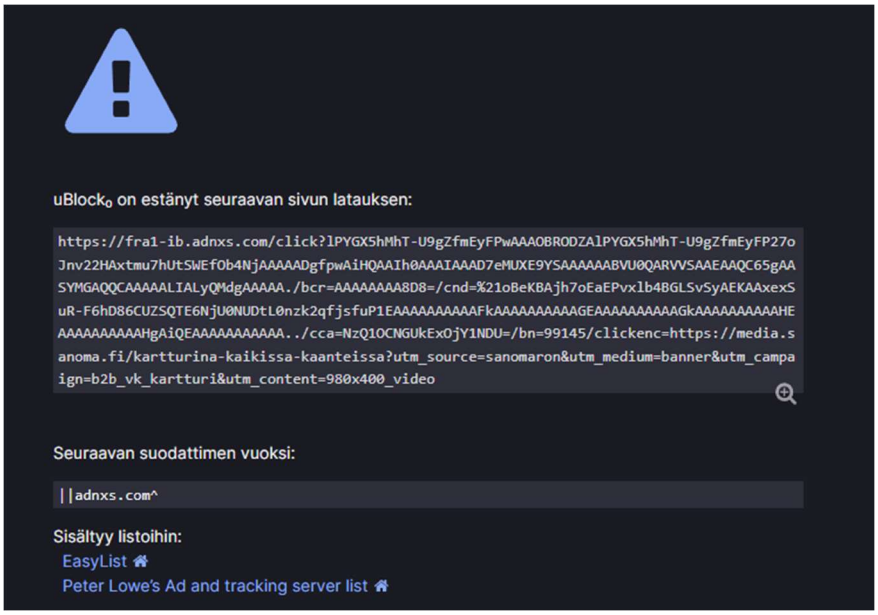
Suosituimpiin selainlisäosiin kuuluu esimerkiksi:

- AdBlock
- AdBlock Plus
- uBlock Origin
- Ghostery
- Privacy Badger

AdBlock sekä AdBlock Plus ovat nimensä mukaisesti työkaluja, jotka estävät mainontaa. Ne yrittävät estää mainoksia kokonaan, myös piilottamalla näiden elementit sivulta, jottei turhia tyhjiä elementtejä näytettäisi.

AdBlock Plus sekä AdBlock kuuluvat kummatkin ”Acceptable Ads” -ohjelmaan, joka käytännössä sallii tiettyjen mainoksien näyttämisen käyttäjällä, ellei tämä tätä erikseen estä. Acceptable Ads -ohjelma ylläpitää listaa mainostajista ja mainoksista, jota voi käyttäjlle näyttää, vaikka heillä olisi jompikumpi selainlisäosa mainostenestotarkoitukseen käytössä. [18] [19]

UBlock Origin haluaa erottua AdBlocker -lisäosista olemalla laajan skaalan sisällönestäjä. Mainosten lisäksi, uBlock Origin kertoo estävänsä myös seuraimia ja haittaohjelmasivustoja. UBlock Originissa käyttäjä voi paremmin itse muokata, mitä sivuja estetään. UBlock ei myöskään pyydä lahjoituksia sen käyttäjiltä, toisin kuin AdBlock ja AdBlock Plus. [10] Se ei myöskään näytä joitain aikaisemmin mainittuja sallittuja (engl. whitelisted) mainoksia, toisinkuin AdBlocker lisäosat tekevät. Kuvassa 8 näemme uBlock originin toimimassa käytännössä.



Kuva 8: Mainoksen klikkaaminen aiheuttaa popup-ikkunan avautumisen, joka estyy uBlock Originin toimesta

Monet selainlisäosat voivat käyttää pitkälti samoja listoja mainosten, seuraimien sekä haittaohjelmaa levittävien sivujen estossa. Näihin kuuluu esimerkiksi:

- EasyList
- EasyPrivacy
- Peter Lowe's Blocklist
- Online Malicious URL Blocklist

Esimerkiksi AdBlock, AdBlock Plus sekä uBlock Origin käyttävät EasyList ja EasyPrivacy listoja. Näiden lisäksi Ublock Origin käyttää myös Peter Lowe's Blocklistiä sekä Online Malicious URL Blocklistia.

Jotkin selainlisäosat taas voivat keskittyä enimmäkseen seuraimiin mainosten sijaan. Esimerkiksi EFF's Privacy Badger -selainlisäosa käyttää heuristiikkoja oppiakseen kolmannen osapuolen seuraimet [2]. Ghostery on lisäosa, jonka takana on yritys, joka ylläpitää listoja, joihin seurannanesto perustuu [2]. Tämänkaltainen listan ylläpito on yleensä heikompaa kuin yhteisön ylläpitämien listojen käyttäminen.

5 Kotiverkon laajuinen seurannanestotyökalu

Tässä luvussa käsitellään kandidaatintyön osana testattua laitetta, jonka tarkoituksena on estää mainoksia sekä edistää käyttäjän yksityisyyttä. Laite on Raspberry Pi, jolle on asennettu Pi-hole -sovellus.

Pi-hole sovelluksen voi laittaa koko verkon laajuiseksi konfiguroimalla reititin käyttämään DNS-palvelimenaan Pi-holea, jolloin kaikki verkossa käytävä liikenne menee Pi-holen läpi. Tällöin myös muutkin laitteet kuin esimerkiksi selainlisäosalla varustettu tietokone pääsee mainostenesto- sekä seuraintenestotyökalun piiriin. Tässä luvussa käsittelemme yhtä Windows 11 -laitetta, joka on yhdistetty Pi-holeen sekä verrataan verkkosivukokemusta Pi-holeen yhdistettynä ja ilman.

5.1 Raspberry Pi

Raspberry Pi on Raspberry Pi -säätiön luoma kokoelma pieniä tietokoneita. Raspberry Pi:sta on olemassa erilaisia ja kokoisia malleja eri tarkoituksiin sekä hintaluokkiin. Tämän lisäksi Pi:lle on saatavilla erilaisia lisäosia, kuten kosketusnäyttö, kamera, sekä erilaisia sensoreita. Näiden lisäksi voi minitietokoneelle ostaa esimerkiksi erillisen kotelon sekä virtalähteen.

Raspberry Pi laitteita voi käyttää monella tavalla. Esimerkkejä näistä ovat esimerkiksi mediapalvelin, retrokonsoli, sääasema, pelipalvelin, pöytätietokone sekä monet muut.

Tässä työssä käytämme Raspberry Pi -laitekokoelman tehokkaammasta päästä olevaa laitetta. Käyttämämme sovellus toimii myös esimerkiksi Raspberry Pi Zero -sarjan laitteilla, sillä vaatimukset sovelluksen käyttöön eivät ole kovin isot [11].

5.2 Pi-hole

Pi-hole on avoimen lähdekoodin sovellus, joka on luotu vuonna 2014 AdTrap-sovelluksen korvaamiseksi. Sovellusta voi käyttää melkein millä vain Linux-laitteella. [16] Pi-holea voi myös ajaa virtuaalitietokoneella tai esimerkiksi docker-kontissa. Tässä työssä asennamme Pi-holen Raspberry Pi 4 model B+ -laitteelle, ja katsomme, miten kyseinen sovellus toimii.

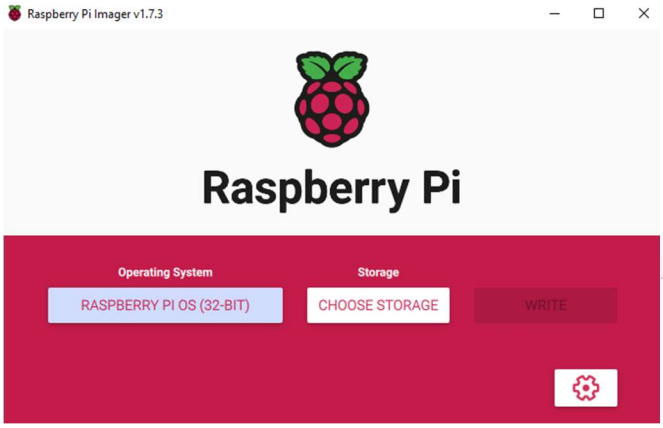
Pi-holen toiminta perustuu ”DNS-sinkhole” mekaniikkaan, sillä Pi-hole toimii itsessään DNS-palvelimena. Kun Pi-hole on asetettu jonkin samassa verkossa olevan laitteen DNS-palvelimeksi, saadessaan DNS-kyselyn, Pi-hole vertaa onko sen tiedoissa olevissa, ennalta määrätyissä listoissa kyseistä kohdetta ja jos on, se on mainos tai seurainkohde, joka pitää estää. Tällöin Pi-hole vastaa kyselyä pyytäneelle laitteelle osoitteella, joka ei ole alkuperäinen osoite mistä ladattaisiin jotain, vaan osoite, josta ei löydy mitään ladattavaa, kuten ”0.0.0.0” [20]. Tällöin mainosta tai seurainta ei ladata verkkosivulle.

Koska Pi-hole toimii DNS-palvelimena, voi sen käyttöä hyödyntää myös esimerkiksi Smart TV:ssa, puhelimessa sekä muissa laitteissa, jossa selaimista tuttuja seurannanestotyökaluja ei ole helposti saatavilla. Näin käyttäjän yksityisyyden parantaminen ei jää pelkästään esimerkiksi käyttäjän tietokoneen selainlisäosien varaan, vaan sama estotyökalu on käytössä kaikilla siihen kytketyillä laitteilla, tai laitteilla, jotka on kytketty verkkoon, joka käyttää DNS-palvelimenaan Pi-holea ajavaa laitetta.

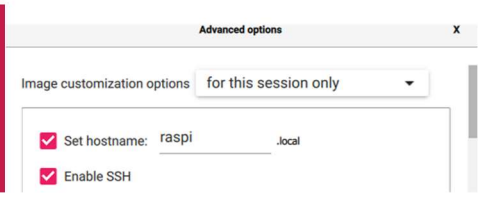
5.3 Laitteen asennus sekä käyttöönotto

Sovelluksen asentamiseksi tarvitsemme käyttämämme laitteen (Raspberry Pi model 4 B+), muistikortin, muistikortin lukijan sekä SSH-yhteyden laitteeseen, sillä asennamme sovelluksen laitteelle komentokehotteen avulla (ns. ”headless install”). Näiden lisäksi erillinen virtalähde sekä toinen verkkoon kytketty laite, jolla Pi-hole DNS-palvelinta kokeillaan, ovat tarpeellisia.

Ensimmäiseksi asennamme muistikortille Raspberry Pi OS -käyttöjärjestelmän kuvassa 9 näkyvän Raspberry Pi -säätiön ”Raspberry Pi Imager” -sovelluksen avulla. Sovelluksen voi ladata Raspberry Pi -säätiön verkkosivuilta. Ennen asennusta pistämme SSH-yhteyden päälle, annamme Pi:lle Wi-Fi-verkon tiedot - SSID:n ja salasanan - internet-yhteyttä varten sekä luomme käyttäjänimen ja salasanan. Nämä voi luoda ja laittaa päälle painamalla rataskuvaketta ennen asennusta, jolloin pääsee lisäasetuksiin. Kuvassa 10 näemme lisäasetusten ylimmän osion, jossa asetamme isäntänimen ja otamme käyttöön SSH-yhteyden.



Kuva 9: Asennusohjelma Raspberry Pi OS -käyttöjärjestelmälle.

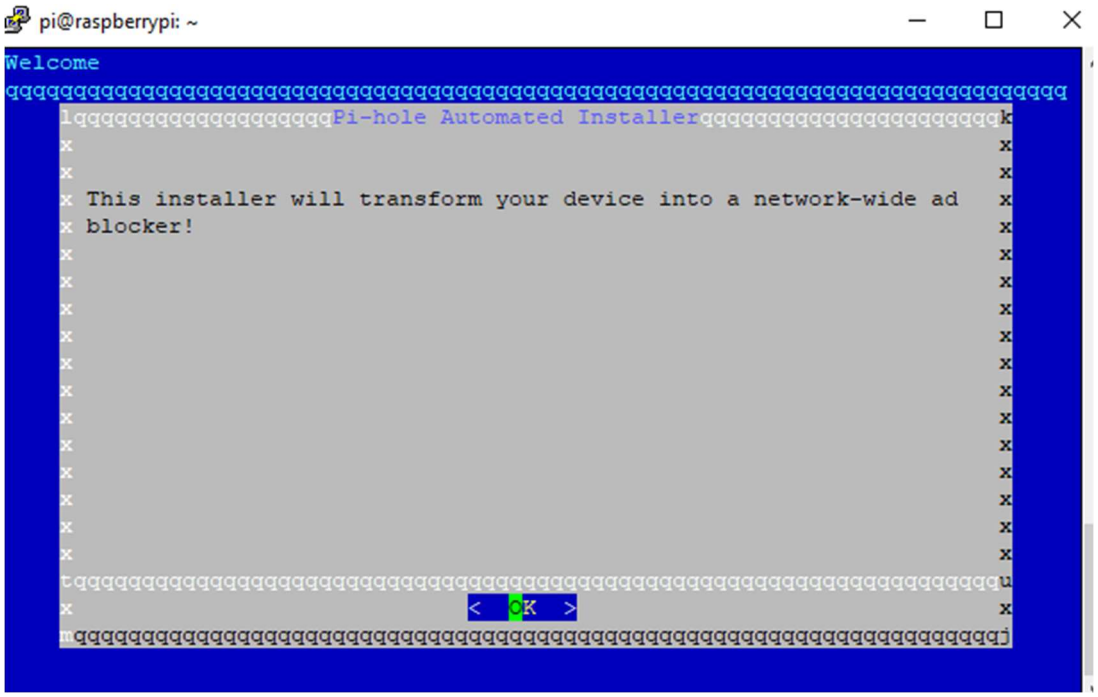


Kuva 10: Asennusohjelman lisäasetusten olennaisimmat osiot.

Seuraavaksi asennamme ohjelman, jolla SSH-yhteyden luominen on helppoa. Tätä varten asennamme PuTTY-ohjelman, joka on SSH sekä Telnet-asiakaskone. Luodaksemme yhteyden, pitää meidän tietää laitteen IP-osoite tai isäntänimi. On olemassa skannereita, jolla IP-osoitteen voi saada selville. Reitittimen asetuksia katsomalla voi myös saada selville laitteen IP-osoitteen, katsomalla yhteydessä olevien laitteiden listaa. Koska asennusohjelmassa on mahdollista asettaa laitteen isäntänimi, voimme käyttää tätä yhdistäessämme laitteeseen PuTTy-ohjelman avulla.

Tämän jälkeen päivitämme Raspberry Pi:n päivitykset, ja lataamme Pi-holen nettisivuilta olevan kuvassa 11 näkyvän komennon avulla Pi-hole sovelluksen.

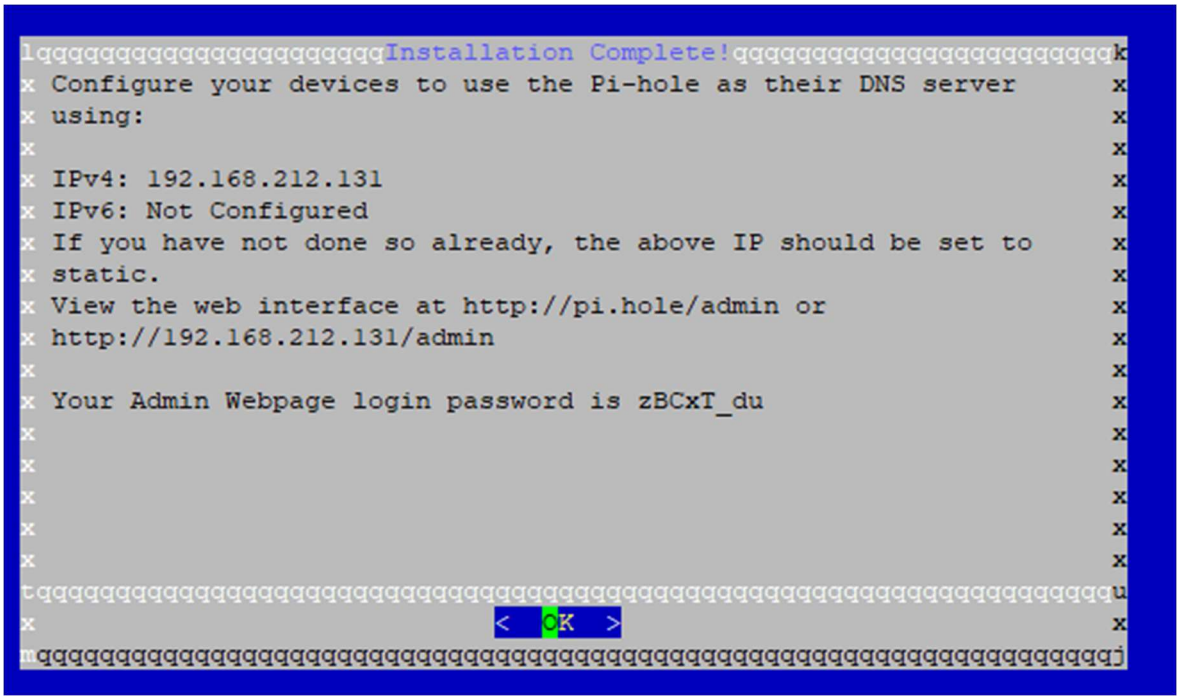
Kuva 11: Pi-holen asennuskomento Raspberry Pi -tietokoneen terminaalissa.



Kuva 12: Asennusohjelma Pi-hole sovellukselle

Tämän jälkeen pääsemme kuvassa 12 näkyvään asennusohjelmaan, jossa Pi-hole asennetaan. Asennuksen aikana valitsemme, haluammeko lisätä valmiin listan mainossivuista ja seuraimista, vai lisätä manuaalisesti kaikki käyttämämme listat. Valitsimme sovelluksen tarjoaman esiasennetun listan. Tämän jälkeen asennusohjelma kysyy mieltymyskysymyksiä, joita voimme halutessamme muokata myöhemmin, joten valinnalla ei ole juurikaan väliä. Valitaan web-käyttöliittymän asennus, jotta pääsemme tutkimaan laitteen toimintaa tarkemmin verkossa.

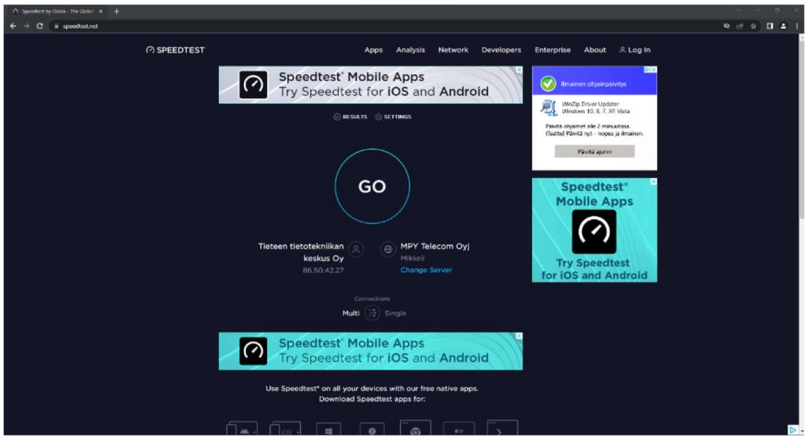
Lopuksi varmistetaan tai tarpeen vaatiessa määritetään laitteelle staattinen IP-osoite. Staattisen IP-osoitteen voi yleensä määrittää reitittimen asetuksista. Tämän jälkeen asennus on valmis. Kuvassa 13 näemme valmiin asennuksen loppuruudun.



Kuva 13: Valmistunut asennus Pi-hole ohjelmasta.

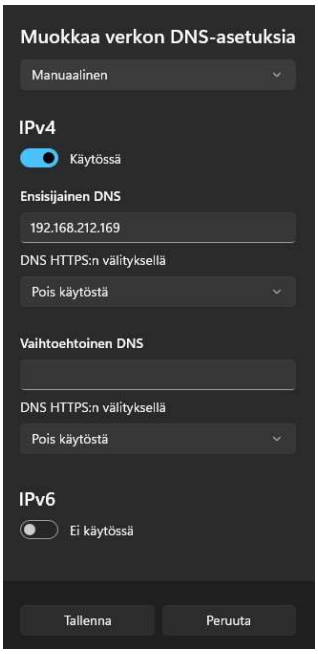
Tässä tapauksessa Pi-hole on automaattisesti aina päällä, kun Raspberry Pi on päällä, ellei asetuksia muuta toisin jälkikäteen.

Tässä vaiheessa Pi-hole on nyt käynnissä. Esimerkin vuoksi, kokeilemme muutamia verkkosivuja – *speedtest.net, foreca.fi* sekä *is.fi*. Kuvassa 14 kokeilemme mennä *speedtest.net* -verkkosivulle jotta näemme, miten mainokset muuttuvat, kun olemme yhdistäneet Pi-holeen.



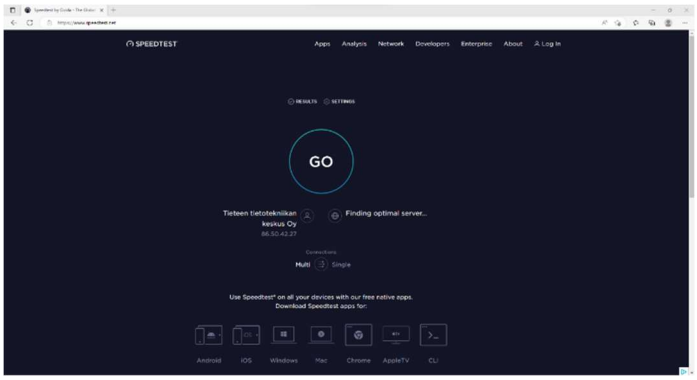
Kuva 14: speedtest.net -verkkosivu ennen yhdistämistä Pi-holeen.

Seuraavaksi konfiguroimme laitteemme käyttämään DNS-palvelimenaan Raspberry Pi:ta. Laitteemme käyttää Windows 11 -käyttöjärjestelmää. Verkkoasetuksista, valitsemme käyttämämme verkon ja DNS kohdasta painamme ”muokkaa”. Valitsemme automaattisen sijaan manuaalinen, pelkkä IPv4, ja syötämme ensimmäiseen kohtaan Pi-holen IP-osoitteen. Vaihtoehtoisen DNS-kohdan jätämme tyhjäksi, sillä muutoin olisi riski, että mainoksia latautuisi toisen DNS-osoitteen kautta. Kuvassa 15 näemme asetusten tilan muutoksien jälkeen ennen tallentamista.



Kuva 15: Windows 11 -käyttöjärjestelmän DNS-asetukset.

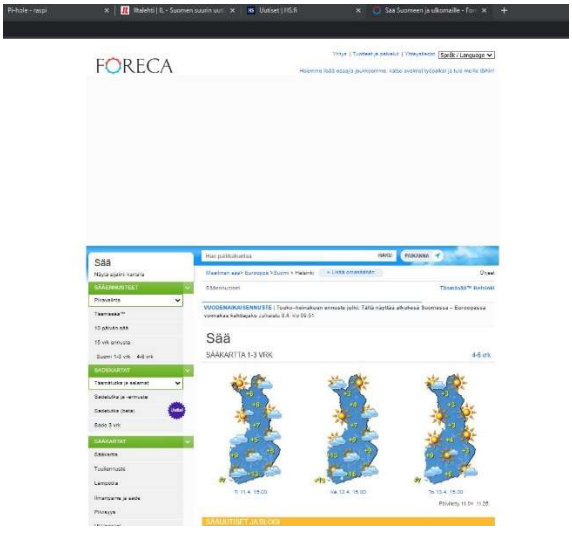
Yhdistettyämme Raspberry Pi:hin, kokeillaan, kuinka mainosten määrä muuttuu käyttämällä samaa esimerkkisivua *speedtest.net* ja kuinka mainoselementeille käy, kun mainostenestotyökalu Pi-hole on käytössä. Kuvassa 16 näemme uuden lopputuloksen verkkosivua ladatessa.



Kuva 16: Mainokset ovat kadonneet yhdistettyämme Raspberry Pi:hin.

Tässä tapauksessa mainokset ovat hävinneet kokonaan, eikä sivulle ole jäänyt tyhjiä elementtejä siitä, että jotain on jäänyt lataamatta, niille paikoille, missä mainokset olisivat olleet latautuessaan. Näin voi käydä, jos esimerkiksi käyttäjä on käynyt sivulla aikaisemmin ja elementti on ladattu, mutta Pi-holeen yhdistettyä ei elementin dataa enää saatu. Tällöin selain ei ole hakenut verkkotunnusta sekä verkkosivun tietoja uudestaan, vaan käyttänyt välimuistissa olevaa versiota.

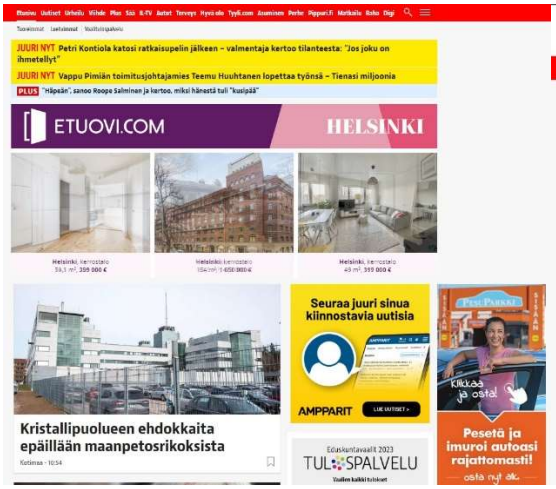
Esimerkiksi kuvassa 17 *foreca.fi* -sivulla näkyy esimerkki käyttäjäkokemukseen vaikuttavasta seurantenestotyökalujen vaikutuksesta. Missä normaalisti olisi mainos, onkin tilalla vain tyhjää. Kuvassa 18 on esimerkki, jossa verkkosivulle on menty ilman Pi-holeen yhdistämistä.

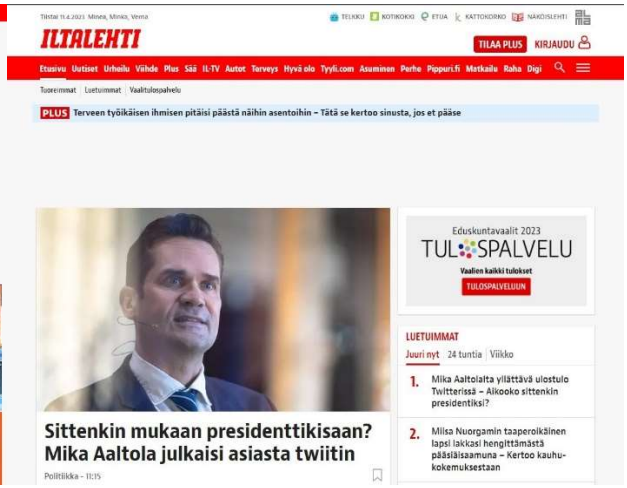


Kuva 17: Forecan nettisivu Pi-holen kanssa.

Kuva 18: Forecan nettsivu ilman Pi-holeen yhdistämistä.

Suurinpiirtein samoin käy myös *is.fi* -verkkosivun kanssa. Kuvassa 19 näemme *is.fi* -verkkosivun ilman Pi-holea ja kuvassa 20 Pi-holen kanssa.





Kuva 19: is.fi ilman yhdistämistä Pi-holeen. Kuva 20: is.fi Pi-holen kanssa.

Kuva 21: Pi-holen hallintapaneeli verkkosivuselailun jälkeen.

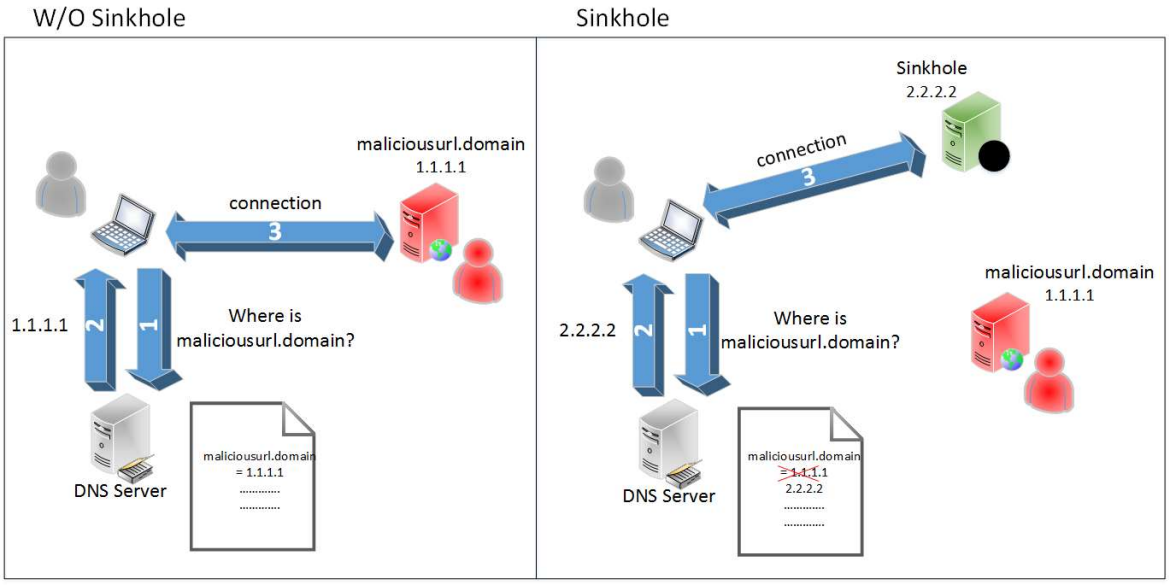
Muutaman tunnetun verkkosivun sekä muutaman satunnaisesti valitun sivun kokeilemisen jälkeen palasin Pi-holen hallintapaneeliin katsomaan, miten estotyökalu on toiminut. Kuvassa 21 näemme tehtyjen pyyntöjen määrän, estettyjen pyyntöjen määrän, estettyjen pyyntöjen suhteellisen osuuden kaikista pyynnöistä, sekä estolistalla olevien verkkotunnusten määrän.

5.4 Pi-holen toimintatapa

Pi-holen toimintatapa perustuu ”DNS-Sinkhole” toimintatapaan. Pi-hole toimii kuten DNS palvelin, jos nimipalvelupyynnön osoite ei ole estetty jollain tapaa. Pi-holeen voi ladata valmiita listoja verkkotunnuksista, lisätä yksittäisiä verkkotunnuksia suoraan tai käyttää regex (Regular Expression) toimintoa, jolla voi etsiä tiettyyn kaavaan sopivia verkkotunnuksia. Kuvassa 22 on kuvankaappaus esimerkkiregex-listasta.

```
21  ^adim(age|g)s?[0-9]*[_.-]
22  ^adtrack(er|ing)?[0-9]*[_.-]
23  ^advert(s|is(ing|ements?))?[0-9]*[_.-]
24  ^aff(iliat(es?|ion))?[_.-]
25  ^analytics?[_.-]
26  ^banners?[_.-]
```

Kuva 22: Esimerkki regex-listasta, joka etsii tiettyjä sanoja verkkotunnuksista.



Kuva 23: DNS-toiminta kuvattuna DNS-sinkhole kanssa sekä ilman. [12]

Kuvataan Pi-holen toimintaa yllä olevan kuvan 23 numeroitujen kohtien avulla. Aloitetaan kuvan vasemmasta puolesta, jossa käyttäjä ei käytä sinkholea.

Kun käyttäjä yrittää yhdistää verkkosivulle *maliciousurl.domain*, lähettää hän nimipalvelupyynnön, jotta saa tietoonsa tarvitsemansa verkkotunnuksen IP-osoitteen. Kuvassa 23 tämä

on ensimmäinen kohta. Saatuaan vastauksen DNS-palvelimelta kohdassa 2, käyttäjä siirtyy palvelimelta saamaansa IP-osoitteeseen kohdassa 3. Kuvan tapauksessa tämä verkkotunnus on haitallinen, mutta koska mitään yrityksiä haitallisen verkkosivun estämiseksi ei välivaiheissa ollut tehty, on käyttäjä nyt mahdollisesti vaarallisella sivulla.

Oikeanpuoleisessa kuvassa on käytössä DNS-sinkhole, jossa tunnetut pahaenteiset sivut on listattu. Kun käyttäjä kysyy mahdollisesti vaarallisen verkkosivun tietoja kohdassa 1, DNS-palvelin on tunnistanut sivun haitalliseksi, sillä sivu on löytynyt ennalta määritetyltä listalta. Oikean IP-osoitteen sijaan DNS-palvelin vastaa käyttäjälle tyhjällä osoitteella, tai varta-vasten esimerkiksi tutkimista varten tehdyllä IP-osoitteella kohdassa 2 [12]. Näin käyttäjä ei yhdistäessään saadulle verkkosivulle kohdassa 3 pääse oikealle mahdollisesti vaaralliselle sivulle.

Pi-holen tapauksessa toimitaan yllä kuvatulla tavalla. Jos pyydettyä verkkotunnusta ei löydy estolistoilta eikä osoite täsmää regex-listoja käyttämällä, käyttäjän nimipalvelukysely sallitaan menevän eteenpäin.

Koska Pi-hole ei esimerkiksi monen selainlisäosan tavoin lataa mainoksia ja seuraimia ja sen jälkeen piilota niitä, vaan ei suorita koko latausta verkkotunnusten perusteella, voi Pi-holen käyttäminen nopeuttaa verkon toimintaa hieman. Jos Pi-holeen on lisätty monia estolistoja sisältäen esimerkiksi satoja tuhansia tai jopa miljoonia verkkotunnuksia, voi nopeusero normaaliin verkkoon verrattuna hiipua.

5.5 Pi-holen hyötyjä

Muihin aiemmin mainittuihin seurantatapoihin verrattuna, voi Pi-hole estää esimerkiksi Windows-käyttöjärjestelmän eri ohjelmien lähettämää telemetriaa. Windows on listannut omilla verkkosivuillaan verkkotunnuksia, samalla listaten niiden toimintaa [17]. Täten käyttäjä voi DNS-suodatusta hyödyntämällä estää myös tämänkaltaiset ei-niin-persoonalliset datankeruutavat.

Pi-hole on ilmainen, avoimen lähdekoodin, lahjoitusten sekä käyttäjien korjausten varassa toimiva sovellus. Koska sovellus ei ole yrityksen varassa ja lähdekoodi on avointa, ja Pi-holella on aktiivinen yhteisö, voidaan tämä katsoa hyödyksi. Jos sovellus olisi kolmannen osapuolen tekemä, suljettu sovellus, olisi se todennäköisesti myös maksullinen, eikä

sovelluksen toiminnasta voisi olla täysin varma, sillä lähdekoodi ei olisi saatavilla. Mahdollisten korjausten sekä lisäominaisuuksien saaminenkin jäisi yrityksen päätöksen varan. Pi-hole ei myöskään joihinkin ilmaisin selainlisäosiin verrattuna erikseen salli mainoksia rahaa vastaan.

Pi-holessa on laajat mahdollisuudet estämisen ja sallimisen räätälöintiin käyttäjän tarpeen mukaan. Laajan, aktiivisen yhteisön avulla voi käyttäjä saada vastauksensa esimerkiksi Reddit -palvelussa olevan Pi-hole yhteisön avulla tai Pi-holen nettisivuilta löytyvän perinteisemmän yhteisösivun kautta.

5.6 Muut harkitut vaihtoehdot

Pi-hole ei ole ainoa tapa millä käyttäjä voi parantaa yksityisyyttään. Toinen samankaltainen vaihtoehto on AdGuard Home, joka on myös avoimen lähdekoodin version, joka hyödyntää DNS-sinkhole toimintatapaa. Ominaisuuksiltaan Pi-hole ja AdGuard Home ei käytännössä eroa toisistaan paljoa.

Pi-holen, kuten myös AdGuard Homen, tapauksessa käyttäjällä on paljon valtaa sen määrittämisessä, mitkä verkkosivut ovat saatavilla ja mitkä eivät. Sovelluksen toimiessa koko verkon laajuisesti, eroaa se merkittävästi aikaisemmin mainituista seurannan estotavoista, kuten selainlisäosista, jotka auttavat vain kyseisen selaimen tai korkeintaan yhden laitteen yksityisyyden parantamisessa.

Käyttäjän yksityisyyttä voi myös parantaa VPN-palvelua (Virtual Private Network) käyttämällä, jotta käyttäjän oikea IP-osoite pysyisi piilossa käyttäjän vierailemilta verkkosivuilta. Koska tämä kandidaatintyö keskittyy enemmänkin mainoksiin sekä seurantatapoihin kuten evästeisiin, johon VPN ei suoraan vaikuta, ei sitä tässä työssä käsitelty.

5.7 Pi-holen ongelmia

Käyttäjällä voi olla tarve saada monta eri profiilia saman verkon eri aliverkoille käyttöön. Jos käyttäjä esimerkiksi tekee Pi-holeen asetettujen estolistojen määrästä ison, voi käyttäjän internet-selauskokemus huonontua ja ongelmia ilmaantua, jos käytössä ei ole raskaamman tason laitteistoa. Jotkin estolistat päivittyvät usein, ja jotkut yleisemmätkin verkkosivut

voivat joutua estolistalle vahingossa. Jos verkossa on muitakin käyttäjiä, joilla ei ole tarvetta seurannan estolle, voi ongelmia aiheutua lisää. Olisi kätevää, jos Pi-hole voisi antaa eri verkoille omat profiilit, joille voisi asettaa eri estolistat käyttäjän tarpeen mukaan.

Etsiessäni tietoa kahden Pi-holen käyttämisestä Pi-holen yhteisöfoorumeilta sekä verkosta yleisesti, oli käyttäjien toisen Pi-holen ajaminen käytössä ainoastaan siltä varalta, että ensimmäinen lakkaa toimimasta.

Yksi vaihtoehto on, että käyttäjä ajaa useaa Pi-holea eri laitteilla, mutta tavalliselle käyttäjälle tämä on turhan vaivalloista sekä mahdollisesti haastavaa. Yksi kehityskohta Pi-holelle voisi olla monen eri profiilin salliminen ja niiden ajo samanaikaisesti. Tällöin eri laitteet voisivat käyttää eri profiileja tarpeidensa mukaan ilman, että käytössä olisi enemmän kuin yksi laite.

6 Yhteenveto

Tässä työssä tutkittiin käyttäjien seurantaa verkossa sekä erilaisten seurantatapojen että seurannanestotapojen osalta. Vaikka nykyajan selaimet sisältävät jo joitain keinoja seurannan vähentämiseksi ja estämiseksi, on seuranta edelleen yleistä, joka näkyy myös erilaisten seurannanestotapojen määrässä. Työssä käytiin läpi seurannanestotyökalujen sekä seurantatapojen eroavaisuuksia. Työn ensimmäisen osan tutkimusosuus tuo yleiskatsauksen verkkoseurantaan liittyen.

Työn toisessa osassa testattiin kotiverkon laajuista laitetta, joka toimii useammalle laitteelle eikä sen toiminta-ala rajoitu pelkästään esimerkiksi yhteen selaimeen. Laitteen toimintatapa eroaa yleisemmin käytetyistä tavoista estää seurantaa. Testatun laitteen avulla voi jokainen käyttäjä halutessaan parantaa omaa yksityisyyttään entisestään samankaltaisella laitteella.

Tutkimuskysymyksiin onnistuttiin vastaamaan hyvin. Akateemisista lähteistä löydettiin tietoa käyttäjien seurantaan liittyen. Yleiset seurantatavat sekä keinot käytiin lävitse, ja tuotiin esille yksi tapa lisää estää seurantaa testaamalla olemassa olevaa, aktiivisen yhteisön omaavaa sovellusta laitteella, joka on käyttäjän saatavilla yleisissä elektroniikkakaupoissa.

Lähdeluettelo:

[1] A. Lerner, A. K. Simpson, T. Kohno, ja F. Roesner, ”Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016”, s. 17.

[2] G. Merzdovnik *ym.*, ”Block Me If You Can: A Large-Scale Study of Tracker-Blocking Tools”, teoksessa *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*, Paris: IEEE, huhti 2017, ss. 319–333. doi: 10.1109/EuroSP.2017.26.

[3] P. G. Leon *ym.*, ”What matters to users?: factors that affect users’ willingness to share information with online advertisers”, teoksessa *Proceedings of the Ninth Symposium on Usable Privacy and Security - SOUPS ’13*, Newcastle, United Kingdom: ACM Press, 2013, s. 1. doi: 10.1145/2501604.2501611.

[4] P. Laperdrix, W. Rudametkin, ja B. Baudry, ”Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints”, teoksessa *2016 IEEE Symposium on Security and Privacy (SP)*, San Jose, CA: IEEE, touko 2016, ss. 878–894. doi: 10.1109/SP.2016.57.

[5] G. Acar *ym.*, ”FPDetective: dusting the web for fingerprinters”, teoksessa *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security - CCS ’13*, Berlin, Germany: ACM Press, 2013, ss. 1129–1140. doi: 10.1145/2508859.2516674.

[6] I. Sanchez-Rola, X. Ugarte-Pedrero, I. Santos, ja P. G. Bringas, ”The web is watching you: A comprehensive review of web-tracking techniques and countermeasures”, *Logic Jnl IGPL*, vsk. 25, nro 1, ss. 18–29, helmi 2017, doi: 10.1093/jigpal/jzw041.

[7] W3S. [verkkosivu] Saatavilla: https://www.w3schools.com/jsref/obj_navigator.asp

[8] Apple. [verkkosivu] Saatavilla: https://www.apple.com/privacy/

[9] Wired. [verkkosivu] Saatavilla: https://www.wired.com/how-wired-is-going-to-handle-ad-blocking/

[10] uBlock Origin, GitHub. [verkkosivu] Saatavilla: https://github.com/gorhill/uBlock

[11] Pi-hole documentation. [verkkosivu] Saatavilla: https://docs.pi-hole.net/main/prerequisites/

[12] enisa, EU. [verkkosivu] Saatavilla: https://www.enisa.europa.eu/topics/incident-response/glossary/dns-sinkhole?v2=1&tab=details

[13] Hinternesch Nicolas, Medium. 2020. [verkkosivu] Saatavilla: https://levelup.gitconnected.com/no-cookies-no-problem-using-etags-for-user-tracking-3e745544176b

[14] Yleinen tietosuoja-asetus, Euroopan unionin virallinen verkkosivusto. [verkkosivu] Saatavilla: https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_fi.htm

[15] Yksityisyys, Wikipedia. [verkkosivu] Saatavilla: https://fi.wikipedia.org/wiki/Yksityisyys

[16] Pi-hole, Wikipedia. [verkkosivu] Saatavilla: https://en.wikipedia.org/wiki/Pi-hole

[17] Manage connection endpoints for Windows 10 Enterprise, version 2004. Microsoft. 02/2023. [verkkosivu] Saatavilla: https://learn.microsoft.com/en-us/windows/privacy/manage-windows-2004-endpoints

[18] About the Acceptable Ads program and “non-intrusive” ads, AdBlock. 03/2023. [blogi] Saatavilla: https://helpcenter.getadblock.com/hc/en-us/articles/9738480686483-About-the-Acceptable-Ads-program-and-non-intrusive-ads

[19] Acceptable Ads. [verkkosivu] Saatavilla: https://acceptableads.com/

[20] 0.0.0.0, Wikipedia. [verkkosivu] Saatavilla: https://en.wikipedia.org/wiki/0.0.0.0