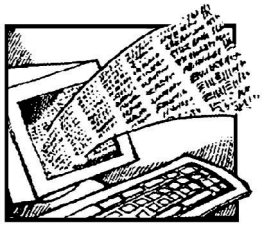


Solmu päivitti tietosuojaselosteensa

Jouni Seppänen

Toukokuun lopulla jokainen Internetiä yhtään pidempään käyttänyt on saanut suuren määrän viestejä, joissa vedotaan EU:n uuteen tietosuoja-asetus GDPR:ään ja joko todetaan, että jokin palvelu on päivittänyt käyttöehtonsa, tai sitten pyydetään lupaa jatkaa sähköpostien lähettämistä. Mistä on kyse ja liittyykö tämä jotenkin Solmuun? Onko ilmiöllä yhtymäkohtia matematiikkaan?

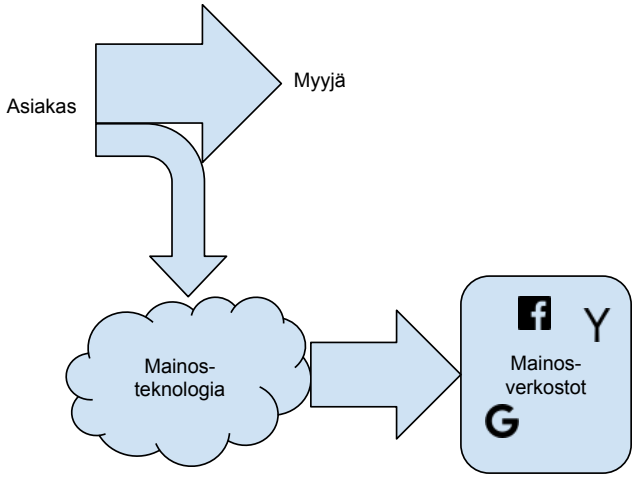
Kun Internetissä tarjotaan ilmaiseksi jotain, kuten sosiaalisen median palveluita, viihdyttäviä pelejä tai vaikka matematiikkalehtiä, on perusteltua kysyä, miten se on mahdollista, koska jostain näiden palveluiden ylläpitämiseen täytyy saada rahaa. Usein raha tulee mainoksista tai maksetusta tuotesijoittelusta, mutta näitä Solmussa ei ole. Monien sivustojen on väitetty myyvän käyttäjien henkilötietoja, joskaan sivustojen omistajat eivät mielellään kerro tällaisesta kaupankäynnistä, vaan se saattaa olla verhottu käyttäjäkokemuksen parantamiseen tai mainosten kohdentamiseen. Yleensä käyttäjältä saadut tai hänestä päätellyt henkilötiedot kytkeytyvät mainoksiin siten, että mainosten näyttämisen kautta kerätään tietoja käyttäjistä, ja toisaalta mainokset kohdennetaan käyttäjille reaaliajassa käyttäen kaikkea saatavilla olevaa tietoa. Moni on varmaan huomannut, että kun jossain nettikaupassa katselee vaikka housuja, monella muullakin sivulla alkaa näkyä housumainoksia. Saattaa tuntua uskomattomalta, että mainosvaroin ja käyttäjätietoja kauppaamalla voitaisiin rahoittaa kovin merkittäviä sivustoja. Jos tilannetta vertaa vaikka paperisiin julkaisuihin, niin postiluukusta tulee ilmaiseksi lehdyköitä, joiden sisältö on melkein kokonaan mainoksia, ja hyvää journalismia lukeakseen täytyy maksaa oikean lehden tilausmaksu (olkoonkin, että mainoksia pääsee lukemaan sittenkin).

Erään matemaattisen mallin tästä kaupasta esitti Maciej Cegłowski esitelmässään *The Website Obesity Crisis*[1], jonka varsinainen aihe oli se, miksi www-sivut latautuvat niin hitaasti. Osaselitykseksi hän tarjosi mainosteknologian yrityksiä, joiden määrä oli kasvanut niin suureksi ja joiden väliset kytkennät olivat tulleet niin monimutkaisiksi jo vuonna 2015, että kuluttajan oli tullut mahdottomaksi edes ymmärtää, ketkä kaikki kilpailevat hänen lukemistaan rahoittavien mainosdollarien maksamisesta.

Sankey-diagrammi A esittää yksinkertaistettuja rahavirtoja, kun kuluttaja ostaa jotain kauppiaalta. Suurin osa hinnasta menee kauppiaan taskuun. Osa siitä toki jatkaa matkaansa mm. verottajalle ja työntekijöille, mutta tässä tapauksessa olemme erotelleet rahavirrasta vain mainoksiin kuluvan rahan. Diagrammissa nuolen paksuus esittää rahavirran suuruutta ja pilvi esittää salaperäisiä mainosteknologian yrityksiä. Mainosteknologia tarkoittaa tässä sellaisia toimialoja kuin mainospaikkojen huutokauppaa, kuluttajakäyttäytymisen tilastollista mallinnusta ja käyttäytymistä kuvaavan datan kauppaamista. Näihin liittyy tietysti matemaatikkoa houkuttelevia työpaikkoja mutta myös yksityisyysnäkökohtia. Housumainosten näkeminen ei ole kovin vakavaa, mutta kun ahkerat mallintajat tutkivat osto-

[1] `http://idlewords.com/talks/website_obesity.htm`

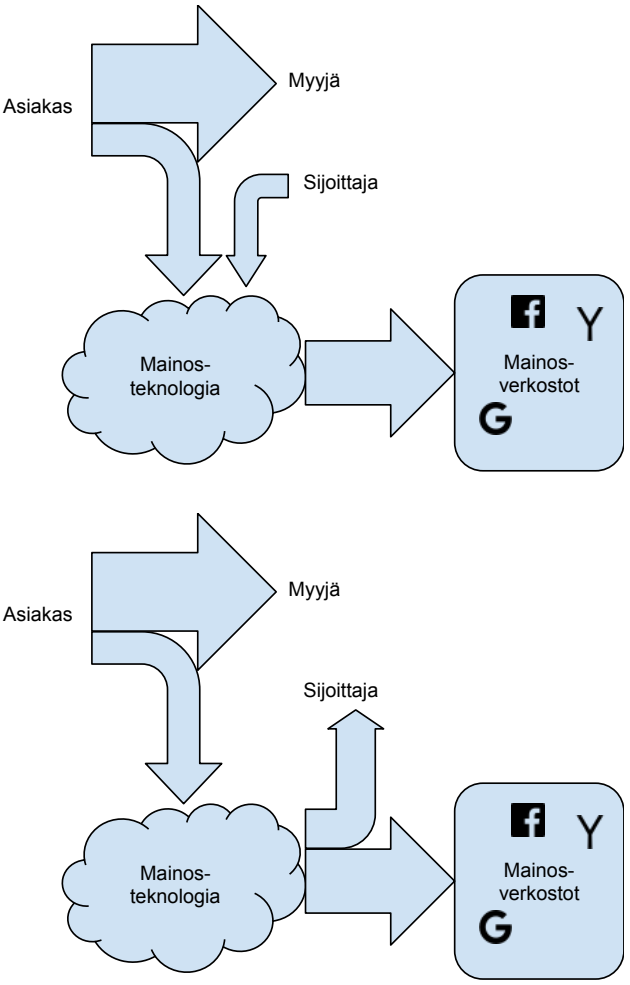
käyttäytymistämme, heitä kiinnostaa myös esimerkiksi asuinalueemme, poliittinen suuntautumisemme ja tiedot terveydentilastamme. On väitetty, että jossain tapauksessa mainostajat arvasivat perheen teinitytön olevan raskaana ja alkoivat lähettää elämäntilanteeseen sopivia mainoksia ennen kuin tyttö oli kertonut asiasta perheelleen[2]. On myös väitetty, että ulkopuolinen taho olisi vaikuttanut Yhdysvaltain vaalien tuloksiin Internet-mainonnalla[3].



Sankey-diagrammi A.

Palataan rahavirtamalliin. Emme varmastikaan edes tiedä, mitä kaikkea mainosteknologiapilvessä tapahtuu, mutta tässä mallissa tällä ei ole väliä: jokin määrä rahaa menee sisään ja jokin määrä tulee ulos. Tämän rahan kohde ovat kannattavia mainosverkostoja pyörittävät yritykset, joita ovat mm. Facebook ja Google. Näistä taas kulkee jonkin verran rahaa ns. sisällöntuottajille, kuten lehdille. Jos rahaa koskee jonkinlainen säilymislaki, niin sisään tulevan ja ulos poistuvan rahan määrä on sama, mutta tässä diagrammissa näin ei ole. Mainosteknologiapilvessä rahaa näyttää syntyvän tyhjästä. Jonkun mielestä mainosteknologia varmaan tuottaa lisäarvoa, mutta rahaa se ei sentään voi painaa. Diagrammi B näyttää, mistä raha tulee: sijoittajat pumppaavat koneistoon rahaa.

Mainosteknologiaan sijoittaneet kapitalistit siis rahoittavat lehtiä. Jonain päivänä nämä sijoittajat kyllä haluavat siirtyä saamapuolelle ja saada jopa jotain tuottoa sijoitukselleen, mitä esittää diagrammi C. Se on vielä pahemmin epätasapainossa kuin diagrammi A, joten mitä silloin tapahtuu? Cegłowskin mukaan mainosteknologiayritykset käyvät entistä kovempaan kilpailuun, joka uhkaa yksityisyyttämme vielä paljon pahemmin kuin tällä hetkellä. Cegłowski esitti, että mainosteknologiaa pitäisi säännellä tiukemmin, ja ehkä EU kuunteli.



Sankey-diagrammit B ja C.

Solmu ei kerää eikä kauppaa henkilötietoja, vaan lehteä rahoitetaan apurahoin ja lahjoituksin ja nämä rahat saadaan riittämään tiukalla kulukurilla. Pitkäaikainen lukija on saattanut huomata tämän esimerkiksi siitä, että lehden ulkoasua ei uudisteta ihan yhtä usein kuin kaupallisten tuotteiden. Solmu ei siis ainakaan mainosteknologian kautta ole osallistunut minkään maan vaalien horjuttamiseen.

Vaikka Solmua eivät varsinaiset henkilötiedot kiinnosta, tiukan laintulkinnan mukaan Solmulla kuitenkin on lukijoitaan koskeva henkilörekisteri: Internetissä viestinnästä jää nimittäin aina jälkiä. Kun selain ottaa yhteyden mihin tahansa palvelimeen, tiedonsiirto tapahtuu lähettämällä viestejä edestakaisin. Jokaisessa viestissä täytyy olla paluuosoite, jotta vastapuoli pystyy lähettämään vastauksen. Näihin osoitteisiin liittyy joitakin matemaattisia tai ainakin numeerisia näkökohtia, jotka sopinevat Solmussa esitettäviksi.

Puheena olevat osoitteet tunnetaan IP-osoitteina (lyhennys tarkoittaa *Internet Protocol*), jotka ovat perinteisesti olleet 32-bittisiä lukuja. Ne kirjoitetaan ylei-

[2] `https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html`
[3] `https://reut.rs/2ppSGBk`

sesti neljänä pisteillä erotettuna oktettina eli kahdeksan bitin rykelmänä, jotka voidaan esittää desimaalijärjestelmän lukuina nollasta 255:een. Esimerkiksi osoite 11000000101010000000000000001010 kirjoitetaan 192.168.0.10. Mahdollisia osoitteita on 2^{32} eli vähän yli neljä miljardia, mikä näytti aikanaan valtavalta määrältä, joten niitä jaettiin eri käyttäjille aluksi melko avokätisesti. Apple sai kaikki 2^{24} osoitetta, joiden ensimmäinen oktetti on 17, Stanfordin yliopisto ne, joiden ensimmäinen oktetti on 36, jne. Nykyään osoitteet alkavat olla lopussa, koska huomattava osa maailman väestöstä haluaa päästä Internetiin, ja monilla on jopa useampi erillinen laite, joille pitäisi saada osoitteet. Siksi näiden rinnalla on käytössä 128-bittisiä osoitteita. Nämä IPv6-osoitteet kirjoitetaan yleensä kahdeksana kaksoispisteillä erotettuna 16-järjestelmän lukuna, mutta peräkkäisten nollien jonot korvataan kahdella kaksoispisteellä. Esimerkiksi Facebookilla on osoite 2a03:2880:2110:df07:face:b00c::1, jonka numeroihin on piilotettu pieni sanaleikki. Osoitteiden määrä on niin valtava, että yleensä loppukäyttäjälle annetaan suurpiirteisesti 64-bittinen osoiteavaruus, jossa on 2^{64} osoitetta, yhtä paljon kuin vanhojen osoitteiden määrän neliö. Siinä mahtuu toteuttamaan vaikka numeerisia sanaleikkejä. Näitä avaruuksia on samoin 2^{64}, mistä riittää käyttäjille vielä aika pitkään. Kaikki laitteet eivät vielä näitä pidempiä osoitteita ymmärrä.

Kun siis selaat Solmun sivuja, selaimesi ottaa yhteyden Solmun palvelimeen, ja palvelimen on muistettava koneesi IP-osoite vastauksen lähettämistä varten. Kotikäyttäjien IP-osoitteet vaihtuvat usein, osin siksi, että näin saadaan neljä miljardia osoitetta riittämään. Siksi tänään käyttämäsi osoite voi olla huomenna jollakulla muulla. Osoitteita on käytännössä mahdoton yhdistää sinuun henkilönä, paitsi että oikeus voi velvoittaa teleoperaattorin luovuttamaan tiedon osoitteen tilaajasta, ja Facebookin kaltainen palvelu, jolla on lonkeronsa monella sivustolla ympäri maailmaa, voi yhdistää osoitetiedot omiin kirjautumistietoihinsa ja siten seurata, millä sivuilla kukin käyttäjä vierailee. Tämä tapahtuu niin, että sivuston ylläpitäjä haluaa olla osallisena mainoksiin liittyvissä verkostoissa ja sijoittaa sivuille Facebookin jakokuvakkeen. Selain hakee kuvakkeen Facebookin palvelimelta, joka kirjaa käyttäjän IP-osoitteen ja tiedon sivusta, jolle kuvake on upotettu. Ehkä tästä syystä on katsottu, että IP-osoite on laskettava henkilötiedoksi[4], ja tässä mielessä jokainen www-sivusto muodostaa ainakin lyhytaikaisesti henkilörekisterin.

Solmunkin joillakin sivuilla on mm. Facebookin jakokuvake, mutta se on toteutettu niin, että kuva haetaan Solmun palvelimelta. Siksi Facebook ei saa tietoja ennen kuin käyttäjä nimenomaisesti klikkaa kuvaketta.

Solmu käyttää oman palvelimensa edustalla Cloudflare-nimisen yrityksen cdn-palvelua (*content delivery network*, sisällönjakeluverkosto). Cloudflarella on ympäri maailmaa välityspalvelimia, joista sivut on nopeampi ladata ja jotka suodattavat automaattisesti useita tietoturvauhkia. Solmun kannalta merkityksellistä on myös, että käyttäjän IP-osoitteet näkyvät vain Cloudflarelle, ja Solmun palvelin tietää vain Cloudflaren eikä loppukäyttäjän IP-osoitteen. Solmulla on Cloudflaren kanssa sopimus tietojen huolellisesta säilyttämisestä ja tarkoituksenmukaisesta käytöstä. Virhetilanteissa on kuitenkin mahdollista, että koneesi ottaa yhteyttä suoraan Solmun palvelimeen Cloudflaren ohi. Todennäköisemmin tällaisia yhteyksiä ottavat ovat kuitenkin hyökkääjiä tai koputtelijoita, jotka etsivät huonosti suojattuja palvelimia. Koputtelijoiden IP-osoitteet tallentuvat palvelimen lokiin vianmääritystä ja hyökkäysten analysointia varten, mutta harvinaisissa tilanteissa sinne voi tosiaan päätyä myös loppukäyttäjän osoite, joten lokia on ehkä pidettävä henkilörekisterinä. Siksi loki kryptataan ns. julkisen avaimen salauksella, joka lienee 1900-luvun lukuteorian tunnetuin saavutus.

Perusidea on, että muodostetaan kaksi erillistä avainta, jotka toimivat parina. Salainen avain koostuu esimerkiksi kahdesta suuresta alkuluvusta p ja q, julkinen avain on näiden alkulukujen tulo pq. Koska suurten lukujen tekijöihinjako on matemaatikoiden parhaan tiedon mukaan vaikea tehtävä, julkisesta avaimesta ei osata päätellä salaista avainta. Palvelimella on julkinen avain ja vain Solmun ylläpitäjillä on salainen avain. Jos lokissa on viesti m, palvelin laskee luvun

$$m' \equiv m^e \pmod{pq},$$

missä e on sopiva eksponentti (usein erinäisistä syistä valitaan $e = 65537$). Ylläpitäjät osaavat ratkaista tästä luvun

$$m \equiv (m')^d = m^{de} \pmod{pq},$$

missä d on yhtälön $de \equiv 1 \pmod{(p-1)(q-1)}$ ratkaisu, mutta jos joku ulkopuolinen pääsee murtautumaan palvelimelle, tehtävä on hänelle hyvin vaikea, kun luvut p ja q eivät ole tiedossa.

Kaikkiaan Solmu on siis varautunut GDPR-aikaan niin, että se ei kerää henkilötietoja juuri ollenkaan, mutta koska sille välttämättä päätyy henkilötiedoiksi tiukan tulkinnan mukaan luettavia tietoja, se käyttää parasta tunnettua matematiikkaa näiden tietojen salaamiseen. Tilanne on hankalampi mainosteknologian yrityksille tai muille yrityksille, joiden toiminta perustuu henkilötietojen kauppaamiseen. Näiden pitäisi kertoa toiminnastaan avoimesti ja antaa kuluttajalle mahdollisuus tarkastaa tiedot ja kieltää niiden käyttö[5]. Muillekin yrityksille tilanne voi olla vaikea, koska ison yrityksen voi olla vaikea kartoittaa kaikki henkilötietojen käsittely. Siksi esimerkiksi suosittu Instapaper-palvelu lakkasi ainakin väliaikaisesti toimimasta EU:n

[4] http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_fi.pdf#page=16

[5] https://yle.fi/uutiset/3-10222442

alueella, ja eräät Internetiin kytketyt sähkölamput ja tietokoneiden oheislaitteet lakkasivat toimimasta, ellei käyttäjä hyväksynyt tietojen keruuta. Ainakin jälkimmäiset tapaukset lienevät asetuksen vastaisia, ellei yrityksillä ole esittää perustelua sille, että henkilötietojen kerääminen on välttämätöntä tietokoneen hiiren toiminnalle. Ilmaiset EU:n ulkopuolella tuotetut Internetpalvelut saattavat kuitenkin päätyä siihen, että on helpompi sulkea palvelu eurooppalaisilta kuin noudattaa asetusta. Siten sääntelyllä on hintansa, mutta ehkä sillä saavutetaan entistä läpinäkyvämpi yhteiskunta.

Suurempi ongelma on ehkä Solmun kaltaisilla, heikosti rahoitetuilla mutta ehkä vähän isomman mittakaavan kansalaisyhteiskunnan toimijoilla, kuten urheiluseuroilla, taloyhtiöillä ja kulttuurintuottajilla. Niiden toiminnassa henkilörekisterien syntyminen on ehkä väistämätöntä, eikä sitä voi helposti kutistaa ihan niin pieneksi kuin Solmun palvelimen salatut IP-osoitteet. GDPR-asetus on pitkä ja vaikeaselkoinen, joten toimijan pitää ehkä palkata lakimies hoitamaan asiaa, mikä voi lohkaista ison osan toimintaan tarkoitetusta budjetista. Facebookin ja Googlen kaltaisilla suuryrityksillä on lakiosastot valmiina, eikä niille varmasti ole minkäänlainen ongelma tuottaa vaikeasti ymmärrettäviä sopimuksia käyttäjien hyväksyttäviksi ja selittää tietosuojaviranomaisille, miksi niiden toiminta pitäisi sallia. Toivoa sopii, että tämä kustannus ei käy kansalaisyhteiskunnalle liian kalliiksi ja aja loppuakin kansalaistoimintaa suurten sosiaalisten verkostojen huomaan.

Verkko-Solmun oppimateriaalit

Osoitteesta `matematiikkalehtisolmu.fi/oppimateriaalit.html` löytyvät oppimateriaalit:

- Sata lukion matematiikan tehtävää (Markku Halmetoja)
- Suppeaa suhteellisuusteoriaa alusta alkaen (Lasse Pantsar)
- Lukion matemaattisen analyysin mestarikurssi (Markku Halmetoja ja Jorma Merikoski)
- Ensiaskeleet Einsteinin avaruusaikaan, osa 1: Kinematiikka: aika, paikka ja liike (Teuvo Laurinolli)
- Ensiaskeleet Einsteinin avaruusaikaan, osa 2: Dynamiikka: liikelait, liikemäärä ja energia (Teuvo Laurinolli)
- Kilpailumatematiikan opas (Matti Lehtinen)
- Geometrian perusteita (Matti Lehtinen)
- Geometria (K. Väisälä)
- Lukualueiden laajentamisesta (Tuomas Korppi)
- Jaksolliset desimaaliesitykset algebrallisesta näkökulmasta (Jaska Poranen ja Pentti Haukkanen)
- Algebra (Tauno Metsänkylä ja Marjatta Näätänen)
- Algebra (K. Väisälä)
- Matemaattista fysiikkaa lukiolaiselle 1: Mekaniikkaa (Markku Halmetoja ja Jorma Merikoski)
- Matemaattista fysiikkaa lukiolaiselle 2: Sähköoppia (Markku Halmetoja ja Jorma Merikoski)
- Lukuteorian helmiä lukiolaisille (Jukka Pihko)
- Matematiikan peruskäsitteiden historia (Erkki Luoma-aho)
- Matematiikan historia (Matti Lehtinen)
- Reaalianalyysiä englanniksi (William Trench)