Markus Koskela

Metadata ja digitaalinen forensiikka

Tietotekniikan kandidaatintutkielma

30. huhtikuuta 2020

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Markus Koskela

Yhteystiedot: markus.j.koskela@student.jyu.fi

Ohjaaja: Sanna Juutinen

Työn nimi: Metadata ja digitaalinen forensiikka

Title in English: Metadata and digital forensics

Työ: Kandidaatintutkielma

Sivumäärä: 31+0

Tiivistelmä: Digitaalinen forensiikka on kehittyvä ala kuten digitaalisten rikosten tutkiminen sekä yleinen järjestyksen ja turvallisuuden ylläpito digitaalisessa maailmassa edellyttää. Tässä tutkielmassa tehdään katsaus digitaalisen forensiikan yleiseen kehykseen tarkastellen siinä metadatan roolia ja lisäksi metadatan sekä suuren mittakaavan Big Datan tuomaa yksityisyysproblematiikkaa valvonnan yhteydessä.

Avainsanat: Big Data, digitaalinen forensiikka, metadata, valvonta, yksityisyys

Abstract: Digital forensics is an evolving field in digital crime investigation as general public order and safety requires in the digital world. This thesis reviews the framework of digital forensics examining the role of metadata in it and also the privacy concerns that arise from using metadata as well large-scale Big Data surveillance programs.

Keywords: Big Data, digital forensics, metadata, privacy, surveillance

Termiluettelo

Big Data	Nopeasti kasvavaa useasta tietolähteestä kerättyä dataa.
DNS	DNS eli Domain Name System on internetin nimipalvelujärjestelmä, joka muuntaa verkkotunnuksia IP-osoitteiksi.
I/O-laite	Tarkoittaa laitetta, jolla siirretään tietoa tietokoneen komponenttien välillä. Esimerkiksi tietokoneen kiintolevy.
MAC-ajat	MAC-ajat ovat tiedostojärjestelmän metadatan osia. MAC-ajat tallentuvat, kun tietyt tapahtumat tiedostoon ovat tapahtuneet.
Metadata	Dataa datasta. Liitännäistätietoa, mikä kuvailee tai määrittelee tietolähdettä.
OSI-malli	Open Systems Interconnection Reference Model kuvaa tiedonsiirtoprotokollien yhdistelmän seitsemässä kerroksessa: 1) fyysinen kerros, 2) siirtoyhteyskerros, 3) verkkokerros, 4) kuljetuskerros, 5) istuntokerros, 6) esityspaikkakerros, 7) sovelluskerros.
RAID	Redundant Array of Independent Disks. Tekniikka, jolla tietokoneiden vikasietoisuutta ja/tai nopeutta kasvatetaan käyttämällä useita erillisiä kiintolevyjä, jotka yhdistetään yhdeksi loogiseksi levyksi.
Steganografia	Tiedon piilottaminen.
Tiedonlouhinta	Joukko menetelmiä, joilla pyritään oleellisen tiedon löytämiseen suurista tietomassoista.

Kuviot

Sisältö

1 Johdanto

Tässä tutkielmassa rajoitutaan tarkastelemaan kahta digitaalisen forensiikan haaraa: Tietokoneforensiikkaa ja tietoverkkoforensiikkaa. Alussa tehdään katsaus yleisemmin digitaalisen forensiikan kehykseen, sitten tietokoneforensiikan kehykseen sekä samoin tietoverkkoforensiikan kehykseen. Lisäksi tehdään suppea katsaus digitaaliseen anti-forensiikkaan.

Tutkielma tarkastelee metadatan roolia niin tietokoneforensiikassa kuin tietoverkkoforensiikassa. Metadataa esiintyy niin tiedostojärjestelmissä kuin tietoverkoissa sekä joskus pelkkä metadata voi antaa perustaviin rikostutkinnallisiin kysymyksiin vastauksen.

Digitaalisen forensiikan on pysyttävä tietotekniikan kehityksen mukana, jotta ajantasaisesti voidaan tutkia digitaalisia rikoksia sekä lisäksi ennaltaehkäistä niin digitaalisia kuin muunlaisia rikoksia. Joskus rikos itsessään välttämättä ei ole digitaalinen, mutta rikokseen saattaa liittyä taustatekijöinä digitaalisia aineksia tai erilaiset digitaaliset ainekset eri lähteistä, kuten tietoverkoista voivat olla rikostutkinassa tukena.

Niin rikostutkinnassa kuin erilaisten rikosten ennaltaehkäisemisessä erilainen tietoverkkojen ja -liikenteen valvonta on oleellisessa osassa. Tähän liittyen tarkastellaan yksityisyyden roolia digitaalisessa forensiikassa. Tässä metadatalla on oma tärkeä roolinsa Big Datan sisällä. Big Data onkin Das, Shaw ja Medhi (2017) mukaan äärimmäinen haaste digitaaliselle forensiikalle turvallisuuden ylläpitämisessä. Mitä suuremmaksi tietomäärä kasvaa, sitä suuremmaksi kasvaa haaste jollain olennaisella tavalla hallita tämä tieto.

2 Digitaalinen forensiikka

Digitaalisen forensiikan on pysyttävä informaatioteknologian kehityksen mukana, jotta nykyaikainen mielekäs rikostutkimus on mahdollista. Garfinkel (2010) mukaan ilman selkeää strategiaa, joka rakentuu erilaisiin tieteellisiin tutkimustuloksiin, forensinen tutkimus ei pysy ajantasaisena ja työkalut vanhenevat yhä enemmän ja lainvalvontaviranomaiset, armeija ja muut tietokoneiden forensisten tuotteiden käyttäjät eivät voi luottaa forensisen tutkimuksen tuloksiin.

Digitaalinen forensiikka koostuu useista haaroista, jotka kukin muodostavat oman forensisen tutkimusalueen. Wazid ym. (2013) mainitsevat seuraavat digitaalisen forensiikan haarat: Tietokoneforensiikka, tietoverkkoforensiikka, mobiililaiteforensiikka, tietokantaforensiikka, muistiforensiikka sekä sähköpostiforensiikka. Das, Shaw ja Medhi (2017) mainitsevat lisäksi Big Data-forensiikan. Tämän ollessa suhteellisen uusi digitaalisen forensiikan haara, ovat myös sen asettamat haasteet suuria, kuten Zawoad ja Hasan (2015) toteavat. He toisaalta tuovat esiin, että rikostutkimuksessa Big Data-forensiikan myötä on mahdollisuus saada laaja-alaisesti todisteita.

Digitaalisen forensiikan jaottelu eri haaroihin vaihtelee eri tutkijoiden keskuudessa. Esimerkiksi monien tutkijoiden mainitsemaa tietokoneforensiikkaa, eivät kaikki mainitse ollenkaan, vaan sen aihepiiriä lähestytään tiedostojärjestelmäforensiikalla, kuten tulee esiin esim. Martini ja Choo (2014) artikkelista.

2.1 Määritelmä

Käsitteelle digitaalinen forensiikka Reith, Carr ja Gunsch (2002) mainitsevat seuraavan määritelmän: ”Tieteellisesti johdettujen ja todistettujen metodien käyttö digitaalisten todisteiden säilyttämiseen, validiointiin, identifiointiin, analysointiin, tulkintaan, dokumentointiin ja esittämiseen digitaalisista lähteistä tarkoituksena rikollisiksi todettujen tapahtumien rekonstruoiminen tai laittomien toimien ennakoiminen.”

2.2 Prosessimalli

Digitaalinen forensiikka käsittää monivaiheisen prosessin. Raghavan (2013) kertoo ensimmäisen vaiheen olevan relevantin todistusaineiston *tunnistamisen*, mikä käsittää yhden tai useamman digitaaliseen varastoimisen mahdollistavan laitteen tiedon tallentamisen liittyen käsillä olevaan rikostutkimukseen. Reith, Carr ja Gunsch (2002) mukaan kyse on tapahtuman tunnistaminen indikaattoreista ja sen tyypin määrittäminen. Esimerkkeinä laitteista jotka voivat mahdollistaa digitaalisen todistusaineiston hankkimisen Raghavan (2013) mainitsee mm. tietokoneiden kiintolevyasemat ja ulkoiset massamuistit kuten USB-tikut. Kun nämä ovat tunnistettu, todistusaineisto hankitaan laitteista ja säilötään rikosteknisesti.

Hankkimisella Raghavan (2013) viittaa binäärisen bittitason kopion hankkimisprosessiin kaikesta sisällöstä, mikä tunnistetussa digitaalisessa mediassa on. Näin hankittu todistusaineisto on säilytetty muuttumattomana ja standardeja hajautusallekirjoituksia kuten MD5 tai SHA1 voidaan käyttää digitaalisen todistusaineiston ehjyyden varmistamiseen. MD5- ja SHA1-algoritmeja pidetään alkuperäisessä muodossaan epäturvallisina, kuten tulee esiin Rasjid ym. (2017) artikkelista. Molempia algoritmeja on kuitenkin kehitetty niiden ollessa yhä käytössä, mikä tulee esiin samasta artikkelista. Algoritmien parantelu onkin paikallaan niiden heikkouksien vuoksi, jotta edelleen voidaan uskottavasti osoittaa rikostutkimuksessa, että tutkimuksen aikana todistusaineisto on säilynyt muuttumattomana.

Todistusaineiston hankkimisen yhteydessä Raghavan (2013) mukaan rikostutkijat ovat tekemisissä erilaisten digitaalisten tietojen kanssa tutkimista varten. Tietojen kerrotaan voivan olla erilaisia muodoltaan ja tyypiltään. Esimerkiksi digitaalistet tiedot voivat olla dokumentteja tietokoneella, puhelimen yhteystietoluettelo, lista tehdyistä puhelinsoitoista, matkapuhelimen tukiaseman jäljityssignaalin voimakkuus, tallennettuja ääni- ja videotiedostoja, sähköpostikeskusteluja, tietoverkkojen liikennekuvioita sekä virustunkeutumisia ja -havaintoja.

Lyhyesti ottaen digitaalinen todistusaineisto kattaa Raghavan (2013) mukaan:

- käyttäjädatan
- käyttäjädataan liittyvän metadatan
- aktiviteettilokit; ja mahdollisesti
- järjestelmälokit.

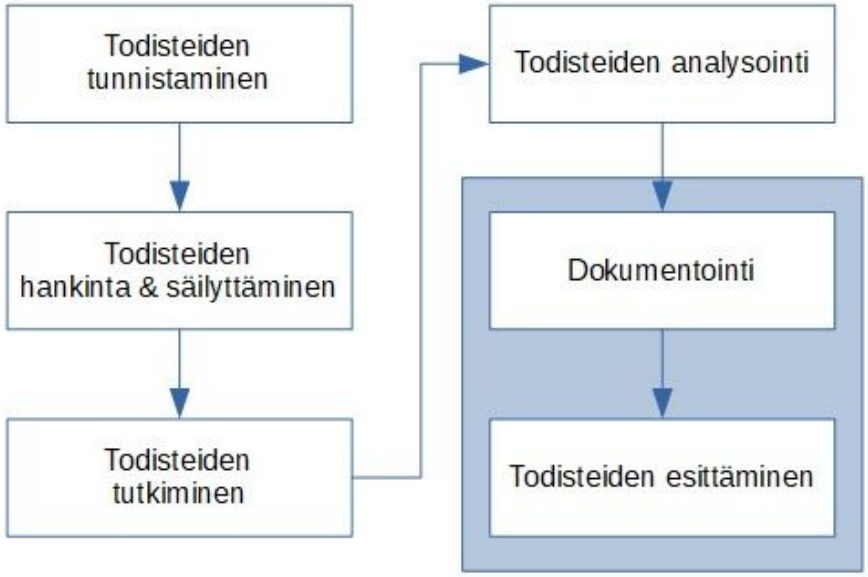
Käyttäjädata koskee suoraan yhden tai useamman rikostutkimukseen liittyvän käyttäjän luomaa, muokkaamaa tai käyttämää dataa. Metadata liittyy datan kontekstiin kysymyksillä miten, milloin, kuka ja missä muodossa käyttäjä oli luonut, muokannut tai käyttänyt dataa. Aktiviteettilokit ovat tietoja käyttäjän tekemistä aktiviteeteista järjestelmässä tai sovelluksessa tai molemmissa antaen yksityiskohtaista tietoa tietyistä toimista, joita yksi tai useampi käyttäjä on tehnyt. Järjestelmälokit liittyvät variaatioihin järjestelmän käyttäytymisestä verrattuna normaaliin käyttäen vertailukohtana yhtä tai useampaa käyttäjää.

Raghavan (2013) kertoo, että kun digitaalinen todistusaineisto on hankittu, on aina tarpeen tehdä siitä kopioita ja tehdä kaikki tekniset testit tällaisiin vain luettavissa oleviin (engl. read-only) kopioihin, jotta mikään toiminta ei muuta dataa alkuperäisistä lähteistä. Näin pystytään *säilyttämään* todistusaineisto muodossa, jossa se alunperin löydettiin. Tämä onkin tärkeää, sillä monet sovellukset muokkaavat tiedostojen sisältöä, erityisesti tiedoston metadataa, kun tiedosto avataan, vaikka mitään muutoksia ei tehtäisi.

Digitaalinen todistusaineisto tutkitaan käyttäen forensisia työkaluja. Nämä forensiset työkalut yleensä ottaen antavat jonkinlaisen tiedostojärjestelmän abstraktion digitaalisista todisteista siten, että niiden sisällöstä voidaan tutkia sisältävätkö ne jälkiä todistusaineistosta. Tätä vaihetta kutsutaan todisteiden *tutkimiseksi*, missä digitaaliset lähteet tutkitaan ja niiden sisältö mahdollisesti indeksoidaan hakujen suorittamiseksi. Castiglione, Santis ja Soriente (2007) määrittelevät teknisen tutkinnan prosessiksi todistusaineiston erottamiseksi digitaalisesta todistusaineistosta sekä sen saattamiseen analysoitavaksi. Joissakin tapauksissa digitaalisen todistusaineiston tutkiminen voi paljastaa joitakin piilotettuja tai muutoin ei-eksplisiittistä informaatiota, mikä täytyy poimia ja analysoida myöhemmin. Tällaista informaation tunnistamista kutsutaan todistusaineiston *löytämiseksi*.

Raghavan (2013) mukaan todistusaineiston tutkimisen ja löytämisen jälkeen tekninen *analyysi* alkaa siitä, että todistelähteet ja löydetyt tiedot analysoidaan tapahtumien järjestyksen määrittämiseksi, jotka johtivat tutkittuun rikokseen. Tässä on kyse Reith, Carr ja Gunsch (2002) mukaan datan tärkeyden määrittämisestä, datan segmenttien rekonstruoimisesta sekä johtopäätösten tekemisestä löydetyn näytön perusteella. Raghavan (2013) kertoo, että nämä kaikki yksittäiset forensiset vaiheet *dokumentoidaan* perusteellisesti. Dokumentaatio *esitetään* oikeudessa, jossa usein kuullaan digitaalisen todistusaineiston asiantuntijatodista-

jaa. Kuvio 1 havainnolistaa edellä esitettyjä vaiheita.



Kuvio 1. Digitaalisen forensiikan prosessi mukaillen (Raghavan 2014).

Uusimmissa ehdotuksissa digitaalisen forensiikan kehykselle tekoäly on keskiössä. Esimerkiksi Rughani (2017) on ehdottanut digitaalisen forensiikan kehystä, jossa todisteiden analyysi, hankinta ja esittäminen perustuisivat tekoälyyn. Tämä edustaisi hänen mukaansa koko digitaalisen forensiikan yleistä kehystä.

2.3 Analyysin tyypit

Digitaalinen rikostutkimus käsittää useita erilaisia datan muotoja, joille kullekin on oma analyysityyppinsä. Seuraavassa esitetään Carrier ja Spafford (2004) mainitsemat analyysityypit:

Media-analyysi: Tallennuslaitteen datan analyysi. Tämä analyysi ei ota huomioon mitään partitioita tai muita käyttöjärjestelmäspesifisiä tietorakenteita. Jos tallennuslaite käyttää kiinteäkokoista yksikköä, kuten sektoria, laitetta voidaan käyttää tässä analyysissä.

Medianhallinnan analyysi: Hallintajärjestelmän analyysi median organisoimiseen. Tämä tyypillisesti käsittää partitioita sekä saattaa sisältää RAID-järjestelmiä, jotka yhdistävät mediaa useilta tallennuslaitteilta yhdelle virtuaaliselle tallennuslaitteelle.

Tiedostojärjestelmäanalyysi: Partition tai levyn tiedostojärjestelmän analyysi. Tämä käsittää tyypillisesti datan prosessointia tiedoston sisällön eriyttämiseksi tai poistetun tiedoston sisällön palauttamiseksi.

Sovellusanalyysi: Tiedoston sisältämän datan analyysi. Tiedostot ovat käyttäjien ja sovellusten luomia sekä sisällön formaatti on sovelluskohtainen.

Tietoverkkoanalyysi: Tietoverkon tietoliikenteen analyysi. Tietoverkkopaketit voidaan tutkia käyttämällä OSI-mallia raakadatan tulkitsemiseksi sovellustason virraksi. Sovellusanalyysi käsittää suuren joukon analysoimistekniikoita useiden sovellustyyppien vuoksi. Osa yleisimmistä listataan seuraavassa:

Käyttöjärjestelmäanalyysi: Käyttöjärjestelmä itsessään on sovellusohjelma, vaikka se on erityinen sovellus, koska se ajetaan ensimmäisenä tietokoneen käynnistyessä. Tällä analyysillä tutkitaan konfiguraatiotiedostoja sekä käyttöjärjestelmän tulostamaa dataa tapahtumien määrittämiseksi.

Ajettavien tiedostojen analyysi: Ajettavat tiedostot ovat digitaalisia objekteja, jotka voivat aiheuttaa tapahtumia sekä niitä tutkitaan usein tunkeutumistapausten yhteydessä, koska rikostutkijan on selvitettävä, mitä tapahtumia on ilmentynyt ajettavien tiedostojen käytöstä.

Kuva-analyysi: Digitaaliset kuvat ovat monien digitaalisten rikostutkimusten kohteena, koska kuvat voivat olla eritavoin lainvastaisia. Tässä analyysissä tarkastellaan, missä kuva on otettu sekä kuka tai mitä kuvassa on. Kuva-analyysi käsittää myös kuvien tutkimisen steganografiaan liittyvien todisteiden saamiseksi.

Videoanalyysi: Digitaalisia videoita käytetään turvallisuuskameroissa sekä henkikökohtaisissa kameroissa mukaanlukien webbikamerat. Verkon rikosten tutkimuksessa voi joskus olla videota webbikameroista. Tämä analyysi tutkii videota tunnistaakseen kohteet videossa sekä paikantaakseen, missä video on kuvattu.

2.4 Tietokoneforensiikka

Tiedotusvälineiden kertoessa taloudellisten petosten, terrorismiepäilyjen ja muiden nykyaikaisten rikosten tutkimuksista, ne mainitsevat yhä useammin tietokoneista kerätyn näytön merkityksen Allen (2005) mukaan. Hän jatkaa, että tietokoneiden ja erilaisen digitaalisen datan forensisesta tutkimuksesta on tullut välttämätöntä. Rogers ym. (2006) mukaan käsitys todisteista ja mahdollisista todisteiden lähteistä onkin muuttunut dramaattisesti.

2.4.1 Määritelmä

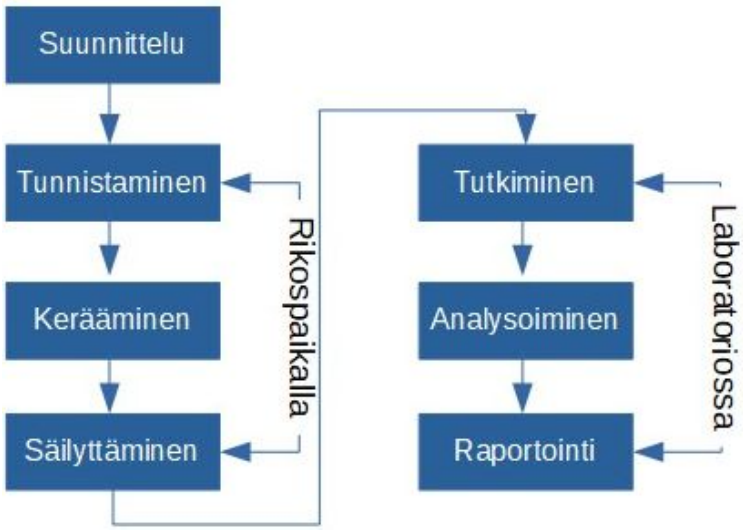
Allen (2005) sanoo erään määritelmän tietokoneforensiikalle olevan: ”Hankkia, säilyttää, hakea ja esittää tietoja, jotka on käsitelty sähköisesti ja tallennettu tietokonemedialle.” Hän jatkaa, että monien asiantuntijoiden mukaan tarkan määritelmän antaminen ei ole kuitenkaan vielä mahdollista johtuen siitä, että digitaalinen todistusaineisto kasvavassa määrin hankitaan laitteilta, joita ei perinteisesti ole käsitetty tietokoneiksi.

Tietokoneforensiikan välttämätön prosessimalli voidaan Rogers ym. (2006) mukaan määritellä seuraavasti: ”Ne tutkintaprosessit, jotka suoritetaan tutkinnan ensimmäisten tuntien aikana ja jotka tarjoavat tietoja epäillyn kuulustelun ja etsinnän suorittamisvaiheessa. Koska tarve saada tietoja koskee suhteellisen lyhyttä aikaa, malli käsittää yleensä kyseessä olevan tietokonejärjestelmän tai järjestelmien paikan päällä tehtävän analyysin.”

2.4.2 Vaiheet

Tietokoneforensiikaa varten on kehitetty useita tutkintamalleja. Rogers ym. (2006) mukaan useissa malleissa on oletettu, että koko rikostutkintaprosessi tietokoneforensiikkaa varten olisi suoritettu. Tämä voi olla erittäin aikaa vievää tutkittavan tietomäärän vuoksi ja useimmissa tapauksissa siihen liittyy järjestelmän (järjestelmien) tai forensisten jäljennösten (tietojen) siirtäminen tallennusvälineellä sijaitsevista tiedoista laboratorioympäristöön perusteellista tutkimusta ja analysoimista varten (Rogers ym. 2006). Kuviossa 2 on esitettynä tällaisen perinteisen tietokoneforensiikan vaiheet.

Edellisessä alaluvussa esitetyn välttämättömän tietokoneforensiikan fokus käsittää Rogers



Kuvio 2. Digitaalisen tietokoneforensiikan perinteiset vaiheet mukaillen (Rogers ym. 2006).

ym. (2006) mukaan seuraavat vaiheet:

1. löydä käyttökelpoinen näyttö heti;
2. tunnista akuutin riskin uhrit;
3. ohjaa meneillään olevaa tutkimusta;
4. tunnista mahdolliset kulut; ja
5. arvioi oikein rikoksentekijän vaara yhteiskunnalle.

Rogers ym. (2006) jatkaa, että samanaikaisesti suojataan todisteiden ja/tai mahdollisten todisteiden eheyttä lisätutkimuksia ja analysointia varten. Allen (2005) mukaan tarkkojen, hyväksyttävien todisteiden keräämisen vaikeus tietokoneista ja muista laitteista kasvaa rikollisten omaksuessa uusia menetelmiä.

2.5 Tietoverkkoforensiikka

Internet on nykypäivään asti tarkasteluna tehokkain väline, joka tarjoaa monipuolisia palveluja useille käyttäjille. Samalla internetistä on tullut myös tietoverkkojen sodankäynnin ympäristö, jossa käynnistetään monenlaisia hyökkäyksiä, kuten taloudellisia, ideologisia tai kostohyökkäyksiä. Verkossa suoritettavat sähköiset kaupat ovat kiinnostavia kyberrikollisille. Internet on suojattava näiltä hyökkäyksiltä, ja niiden käsittelemiseksi on luotava asian-

mukaisia keinoja haitallisten vaikutusten vähentämiseksi. Tietoverkkoforensiikka on tiede, joka käsittelee verkkoliikenteen kaappaamista, tallentamista ja analysointia rikostutkmistarkoituksia varten ja tapahtumien ratkaisemiseksi. (Emmanuel S Pilli, RC Joshi ja Niyogi 2010b)

2.5.1 Määritelmä

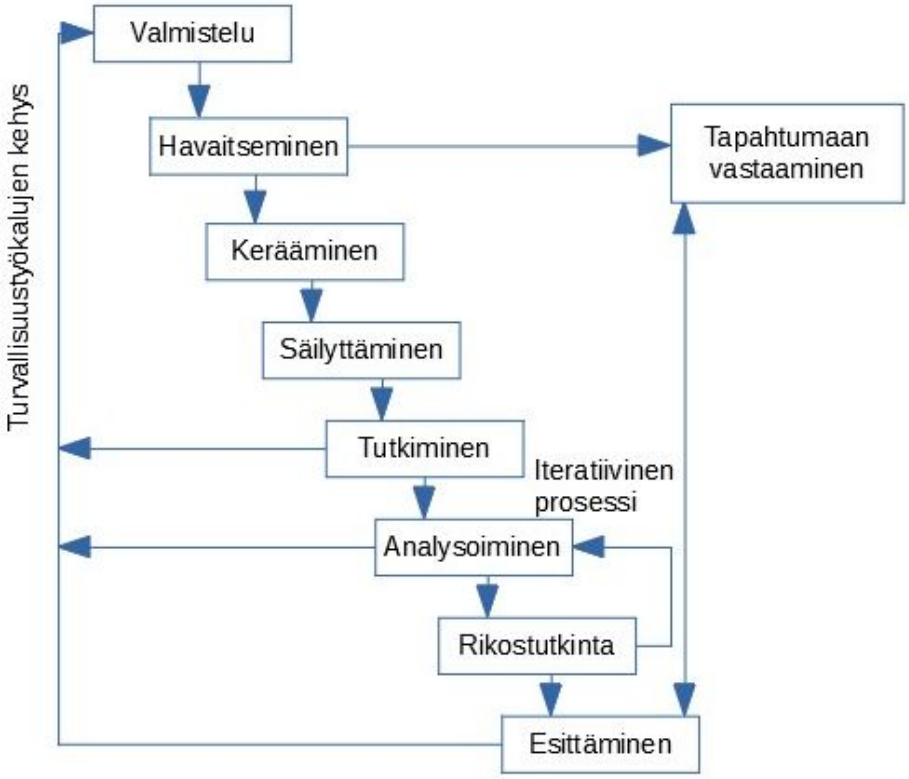
Emmanuel S Pilli, RC Joshi ja Niyogi (2010b) mainitsevat seuraavan määritelmän tietoverkkoforensiikalle: ”Tieteellisesti todistettujen tekniikoiden käyttäminen digitaalisen näytön keräämiseen, sulauttamiseen, tunnistamiseen, tutkimiseen, korreloimiseen, analysointiin sekä digitaalisen todistusaineiston dokumentoimiseen useista aktiivisesti prosessoivista ja lähettävistä lähteistä suunniteltujen rikosten faktojen paljastamiseksi tai luvattomien toimien mittaamiseen, joiden tarkoitus on häiritä, vioittaa ja/tai vaarantaa järjestelmän osia, samoin kuin tietojen tarjoaminen näihin toimiin vastaamiseksi tai näistä toimista toipumiseksi.”

2.5.2 Vaiheet

Emmanuel S. Pilli, R.C. Joshi ja Niyogi (2010a) kertovat, että Wei Ren ja Hai Jin (2005) esittivät ensimmäisenä tietoverkkoforensiikan yleistä prosessimallia seuraavilla vaiheilla: verkkoliikenteen sieppaaminen, kopiointi, siirtäminen, analysointi, tutkimus ja esittäminen. Emmanuel S. Pilli, R.C. Joshi ja Niyogi (2010a) ovat esittäneet hieman erilaisen mallin vaiheista, joita on havainnollistettu kuviossa 3. Seuraavassa käydään lyhyesti näiden vaiheiden sisältöä läpi.

Valmistelu ja valtuutus: Tietoverkkoforensiikkaa voidaan soveltaa vain ympäristöihin, joissa verkon eri strategisissa kohdissa käytetään verkon tietoturvatyökaluja (antureita), kuten tunkeutumisen havaitsemisjärjestelmiä, pakettianalysaattoreita, palomuureja ja liikennevirranmittausohjelmistoja. Vaadittavat valtuudet ja lailliset optiot toimeenpannaan, jotta yksityisyyttä ei loukata.

Havaitseminen: Tarkkaillaan eri tietoturvatyökalujen generoimia hälytyksiä, jotka ilmaisevat tietoturvaloukkauksia tai käytäntöjen rikkomuksia. Luvattomat tapahtumat ja havaitut poikkeamat analysoidaan.



Kuvio 3. Tietoverkkoforensiikan prosessimalli mukaillen (Emmanuel S Pilli, RC Joshi ja Niyogi 2010b).

Tapahtumaan vastaaminen: Vastaus havaittuun rikokseen tai tunkeutumiseen käynnistetään tapahtuman validointia ja arviointia varten kerättyjen tietojen perusteella.

Kerääminen: Tiedot kerätään antureilta, joita käytetään liikennetietojen keräämiseen. Antureiden on oltava turvattuja, vikasietoisia rajoitetun pääsyn antureita, jotta kompromisseilta vältytään.

Suojaaminen ja säilyttäminen: Alkuperäiset tiedot jotka on saatu jälkien ja lokien muodossa, tallennetaan varmuuskopiolaitteeseen. Kaikkien jäljitetietojen tiiviste (engl. hash) ja sitä myöten myös tietorakenne säilyy. Kopio tiedoista analysoidaan ja alkuperäinen kerätty verkkoliikenne säilyy.

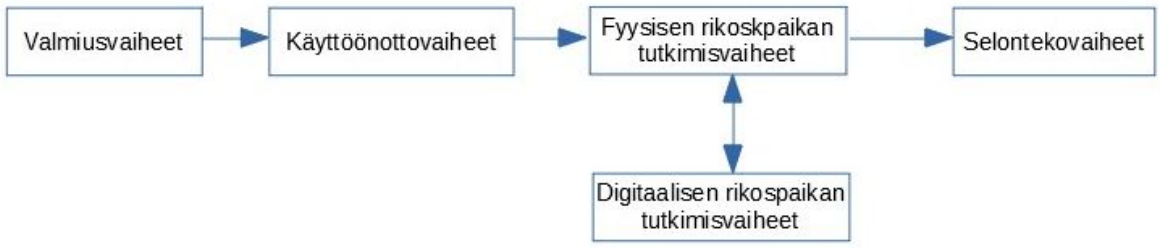
Tutkiminen: Eri turva-antureista saadut jäljet integroidaan ja sulautetaan yhdeksi suureksi tietojoukoksi, jolla analyysi voidaan suorittaa.

Analysoiminen: Indikaattorit luokitellaan ja korreloidaan tärkeiden havaintojen johtamiseksi olemassaolevia hyökkäysmalleja käyttämällä. Tilastollista ja ohjelmallista laskentaa sekä tiedon louhintamenetelmiä käytetään tietojen hakuun ja hyökkäyskuvioiden yhteensovittamiseen. Jotkut tärkeistä parametreista liittyvät verkkoyhteyden luomiseen, DNS-kyselyihin, pakettien hajanaisuuteen, protokollaan ja käyttöjärjestelmän sormenjälkiin.

Rikostutkinta: Tavoitteena on määrittää polku uhriverkosta tai -järjestelmästä kaikkien välijärjestelmien ja viestintäreittien kautta takaisin hyökkäyksen lähtöpisteeseen. Pakettien sieppauksia ja saatuja tilastoja käytetään hyökkäyksen määrittämiseen.

Esittäminen ja tulosten tarkastelu: Havainnot esitetään juridiselle henkilöstölle ymmärrettävällä kielellä samalla, kun selitetään erilaisia päätelmiin pääsemiseksi käytettyjä menettelyjä. Järjestelmällinen dokumentointi sisältyy myös lakien vaatimusten täyttämiseen.

Tietoverkkoforensiikan tutkimisvaiheita Almulhem (2009) kuvaa yleisesti digitaalisen forensiikan integroituna mallina kuvion 4 mukaisesti.



Kuvio 4. Digitaalisen forensiikan integroitu malli mukaillen (Almulhem 2009).

Tietoverkkojen forensisen tutkimisen onnistumiseksi tietoverkkojen itsessään on Almulhem (2009) mukaan oltava varustettuna infrastruktuurilla, joka tukee forensista tutkimusta täysin.

2.6 Metadatasta lyhyesti todisteessa ja todisteena

Digitaalisen forensiikan tutkinnan aikana on usein välttämätöntä löytää digitaaliseen todistusaineistoon jollain tavalla liittyviä tiedostoja tai lokitietoja analysointia varten. Tällä Raghavan (2014) viittaa digitaalisten todisteiden määrän kasvuun ja teknologian monimuotoisuuteen tietojen tallentamisessa eri tiedostomuodoissa ja esityksissä. Molemmat haasteet

vaativat skaalautuvaa lähestymistapaa liittyvien tiedostojen ja lokien määrittämiseksi yhdestä tai useammasta digitaalisen näytön lähteestä.

Raghavan (2014) kertoo, että metadata sisältää tietoja, jotka kuvaavat digitaalisiin kohteisiin tai esineisiin liittyviä aspekteja. Metadata antaa kontekstitietoja, jotka mahdollistavat vastaavan tiedon helpon käsittelyn ja hallinnan. On monentyyppistä metadataa: Järjestelmän metadataa, tiedostojärjestelmän metadataa, sovellusten metadataa, asiakirjan metadataa, sähköpostin metadataa, liiketoiminnan metadataa, maantieteellistä metadataa sekä monia muitakin metadatan tyyppejä (Raghavan 2014).

Tiedostojärjestelmän metadata kuvaa tiedostojärjestelmän tallentamia ominaisuuksia, kuten tiedostojen sijaintia, MAC-aikaleimaa, tiedoston kokoa, tiedoston omistajaa sekä käyttöoikeuksia. Sovelluksen metadata kuvaa määritteitä, kuten tiedostojen tekijöitä, tiedostomuotoja, sisältötyyppejä ja sovelluksen tallentamaa tietoa tallennetusta koodauksesta. Täten termi metadata kattaa kaiken tällaisen erityyppisen metadatan erilaisista digitaalisista kohteista. (Raghavan 2014)

3 Digitaalinen anti-forensiikka

Digitaalisella anti-forensiikalla tarkoitetaan Kessler (2007) mukaan joukkoa taktiikoita ja toimenpiteitä, joilla pyritään estämään digitaalista rikostutkintaprosessia. Varsinaista yhtenäismääritelmää digitaaliselle anti-forensiikalle ei kuitenkaan Harris (2006) mukaan ole, vaan määritemät vaihtelevat. Tämä johtuu Harris (2006) mukaan siitä, että digitaalinen anti-forensiikka on varsin tutkimaton kenttä.

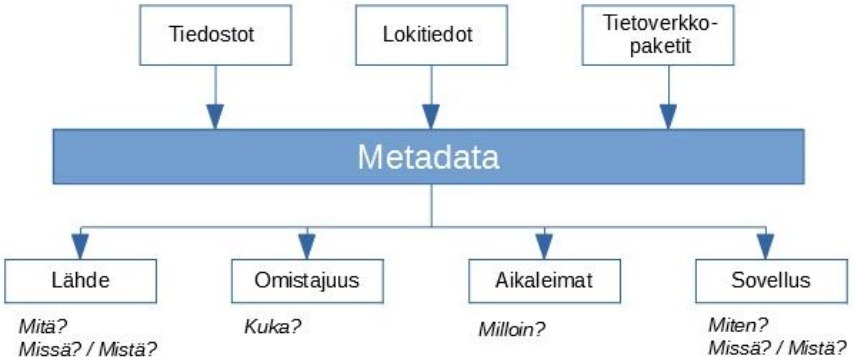
Uusin ja uhkaavin digitaalisen anti-forensiikan tyyppi on Kessler (2007) mukaan haitata eri digitaalisen forensiikan prosessin vaiheita. Tälle digitaalisen anti-forensiikan tyypille on Kessler (2007) mukaan ominaista hyökätä rikostutkimuksen työkaluja vastaan. Harris (2006) jakaa artikkelissaan digitaalisen anti-forensiikan neljään kategoriaan: Todisteiden tuhoamiseen, piilottamiseen, manipulointiin sekä todisteiden luomisen estämiseen.

Todisteiden tuhoamisen Harris (2006) kertoo tarkoittavan esimerkiksi todisteiden tekemistä jollain tavoin käyttökelvottomiksi. Kuitenkin digitaalisessa kontekstissa hänen mukaansa esimerkiksi tiedoston päällekirjoittamisen yhteydessä jokin ohjelma voi tehdä jonkin jäljen, joka antaa viitettä tällaisen toimenpiteen tekemisestä. Todisteiden piilottamisen yhteydessä puolestaan Harris (2006) kertoo, että todisteita ei tuhota, vaan ne esimerkiksi sijoitetaan siten, että niitä on vaikea sisällyttää rikostutkinnan yhteyteen. Kessler (2007) viittaa eksplisiittisesti steganografiaan todisteiden piilottamisessa. Todisteita voidaan lisäksi Harris (2006) mukaan manipuloida siten, että niiden lähde neutralisoidaan. Hän jatkaa, että kuitenkin todisteiden lähteiden eliminoiminen itsessään voi luoda todisteita. Todisteiden luomisen estämistä voi tehdä väärentämällä todisteita. Harris (2006) kertoo, että todisteiden väärentämisen tarkoituksena on saada jokin näyttämään todisteelta, mikä kuitenkaan ei ole todellinen todiste. Tällainen johtaisi viranomaisia harhaan.

4 Metadatan roolista digitaalisessa forensiikassa

Tietokoneiden I/O-laitteiden tallennuskapasiteetin kasvaessa erilaisen kuvailevan tiedon lisääminen järjestelmään on tullut mahdolliseksi. Tällaista tietoa sanotaan metatiedoksi. Metadata ja metatieto ovat määritelmällisesti eri asioita, mutta ne usein samaistuvat tietotekniikan eri konteksteissa. Datan ja tiedon täsmällinen määritelmä sivuutetaan, oleellista on ymmärtää metadatan -ja tiedon kuvaileva olemus. Ne edustavat kuvailevaa dataa tai tietoa kohteestaan. Toisinsanoen metadata on data datasta sekä puolestaan metatieto on tietoa tiedosta.

Raghavan (2014) mukaan metadata sisältää erilaisia tietoja dokumenteista ja muista tiedostoista tietokoneen tai digitaalisen laitteen tallentamana tiedoston tallentamisen ja haun helpottamiseksi. Raghavan (2014) jatkaa, että tiedot voivat olla hyödyllisiä myös järjestelmänhallinnassa, koska ne antavat tietoja asiakirjan tai tiedoston luomisesta, käsittelystä, siirrosta ja varastoinnista tietokoneessa tai digitaalisessa laitteessa. Kuviossa 5 on esitettynä metadatan roolia tietokonejärjestelmässä.



Kuvio 5. Forensiset metadataperheet mukaillen (Raghavan 2014).

4.1 Metadata tiedostoissa ja tiedostojärjestelmissä

Rikostutkinnallisesti metadata voi toimia digitaalisessa ympäristössä eräänlaisena sormenjälkenä tai ylipäätään todisteena, sen voidessa antaa viitettä eri rikostutkinnallisesti keskeisiin kysymyksiin. Samoin metadata voi auttaa rikostutkijaa rekonstruoimaan ja selvittämään tapahtuneeseen rikokseen liittyviä taustoja ja tapahtumia. Seuraavassa alaluvussa tarkastellaan

tarkemmin kuviossa esitettyjen rikostutkinnallisesti keskeisten kysymysten osuutta metadatassa.

4.1.1 Metadata ja rikostutkinnalliset kysymykset

Kysymykset *kuka*, *mistä*, *milloin*, *miten*, *mitä* ja *miksi* ovat rikostutkinnallisesti keskeisiä. Tässä luvussa tehdään katsaus, kuinka erityisesti Unix (tai Linux) -tyyppisessä ympäristössä metadata voi avata näitä kysymyksiä.

Unix-järjestelmien tiedostojärjestelmän metadata on kuitenkin kokonaisuudessaan suppea, paljon suppeampi kuin uusimpien Linux-käyttöjärjestelmien tiedostojärjestelmien, sillä vanhojen Unix-järjestelmien tiedostojärjestelmä kattaa lähinnä seuraavat tiedot Wikipedia (2020) mukaan:

- tiedoston nimi sekä hakemisto, missä tiedosto sijaitsee,
- käyttöoikeudet (omistaja, omistajaryhmä ja oikeusasetukset) sekä
- ajat, jolloin tiedostoa on viimeksi muutettu, luettu tai metadata on muuttunut.

Buchholz ja Spafford (2004) kertovat, että Unix-käyttöjärjestelmien yhteydessä pohdittaessa kysymystä *kuka*, kysymys tiedoston ”omistajasta” on epärelevantti; relevanttia on tiedoston suhteen tietää, kuka on luonut, muokannut, käyttänyt ja poistanut tiedoston. Usein käyttäjätunnus vastaa tähän kysymykseen. Buchholz ja Spafford (2004) mukaan näin ei kuitenkaan aina ole: Esimerkiksi käyttäjä A voi luoda tiedoston, muuttaa sitten käyttäjätunnuksen käyttäjälle B käyttämällä Unixin `chown`-komentoa, minkä jälkeen tiedoston alkuperäinen luoja häviää järjestelmästä. Jäljelle jäävät operaatiot, joita tiedostolle voidaan tehdä (muokkaa, käytä ja tuhoa), voidaan Buchholz ja Spafford (2004) yhdistää henkilöön sillä perusteella, keillä on pääsymahdollisuus tiedostoon. Kuitenkaan täysin yleisessä tapauksessa ei ole mahdollista saada tiedostojärjestelmästä aina korrektia vastausta kysymykseen *kuka*, mutta approksimaatiot voivat olla kuitenkin riittävän hyviä useimmissa tapauksissa (Buchholz ja Spafford 2004). Juola (2007) huomauttaakin, että avoimiin toimistoihin voi periaatteessa kuka hyvänsä tulla käyttämään tietokonetta. Haettaessa vastausta kysymkseen ”kuka” onkin olennaista tietää tai selvittää, onko esimerkiksi tietyn käyttäjätunnuksen haltija todella tehnyt havaitut asiat.

Edellä Buchholz ja Spafford (2004) kertoivat, että kysymys tiedoston omistajasta ei ole relevantti; esitetyn kuvion yhteydessä voikin omistajuutta tulkita laveasti myös käyttöoikeutena tiedostoon. Raghavan (2014) kertookin tekstissään kysymyksen ”kuka” osalta tarkastelevansa, kenellä on oikeus esim. muuttaa tiedostoa.

Pohdittaessa *mistä* tiedosto on tullut järjestelmään, näkökulmaksi voidaan löytää järjestelmän sisäinen tai ulkoinen alkuperä. Tiedosto voi olla käyttäjän itsensä järjestelmän sisällä järjestelmään tekemä, tiedoston alkuperä voi olla ulkoiselta siirrettävältä laitteelta tai tiedosto voi olla verkosta ladattu (Buchholz ja Spafford 2004).

Kaikki yleisesti käytetyt tiedostojärjestelmät yhdistävät aikaleiman tiedostoihin ja hakemistoihin (Buchholz ja Spafford 2004), mikä voi auttaa vastaamaan kysymykseen *milloin*. Aikaleimaa voi kuitenkin muuttaa useissa järjestelmissä. Esimerkiksi Buchholz ja Spafford (2004) mainitsevat Unix-järjestelmän `touch`-komennon, jolla voi antaa mielivaltaisen aikaleiman tiedostolle.

Kysymys miten jokin asia tehtiin järjestelmässä voi olla hyvin tärkeä rikostutkijalle. Kysymyksellä ”miten” Buchholz ja Spafford (2004) viittaavat siihen, mitä kontrollointiagenttia tai mitä ajettavaa ohjelmaa käytettiin toimintojen suorittamiseen rikoksen yhteydessä. Pääsy käyttäjän koko käyttäjäagenttiketjuun voi olla arvokasta tietoa tutkimuksessa järjestelmän tapahtumien rekonstruoimiseksi. Tällaisen järjestelmän agenttiketjun datan tallentaminen ei kuitenkaan ole mahdollista monissa kotitietokonejärjestelmissä, kuten Buchholz ja Spafford (2004) toteavat.

Kysymystä *mitä* pohdittaessa ovat muokkaukset tiedostoon tiedoston luomishetkestä lukien tärkeitä Buchholz ja Spafford (2004) mukaan. He jatkavat, että ideaalissa tapauksessa koko muokkausketju ja nykyinen tila ovat saatavilla.

Miksi-kysymys koskee rikoksen taustalla olevia motiiveja, joihin tietokonejärjestelmä sinällään tuskin voi antaa perustavia vastauksia. Kuitenkin ”miksi”-kysymystä avattaessa, voidaan muista kysymyksistä, joihin tietokonejärjestelmä voi antaa vastauksen, päätellä jotain ”miksi”-kysymykseen liittyen (Buchholz ja Spafford 2004).

4.1.2 Metadata ja steganografia

Raghavan (2013) mukaan todisteiden tutkinnan aikana kaikki tiedot eivät ole välttämättä helposti saatavilla, sillä tietoja on voitu esimerkiksi piilottaa. Tätä kutsutaan steganografiaksi. Se on tiedettä piilotettujen viestien kirjoittamiseen siten, että kukaan lähettäjän ja aiotun vastaanottajan lisäksi ei epäile viestin olemassaoloa Raghavan (2013) mukaan. Digitaalinen steganografia voi sisältää tietojen piilottamista ohjelmiin tai protokolliin sekä asiakirja- ja kuvatiedostoihin. Mediatiedostot ovat suuren koon vuoksi ihanteellisia steganografiseen tiedon lähettämiseen (Raghavan 2013).

Tietoja piilotettaessa salaiset viestit upotetaan muihin viesteihin – näennäisesti viattomiin. Upotusoperaatio suoritetaan apuavaimella: Ilman avaimen tuntemusta, kolmannen osapuolen on vaikea havaita ja käyttää salaisia viestejä (Anderson ja Petitcolas 1998). Kaikki dokumentit, jotka ovat jäsennelty metadatan avulla, mahdollistavat tietojen piilottamisen tähän metadataan. Tästä käy esimerkkinä Microsoftin Office-dokumentit, kuten käy ilmi Castiglione, Santis ja Soriente (2007) artikkelista. Artikkelissa kerrotaan, että jäsentelyyn liittyvän metadatan lisäksi dokumenttien metadata, joka pitää sisällään aiemmat muokkaukset dokumentista, mahdollistaa salatun informaation kätkemisen.

4.2 Metadata tietoverkkopaketeissa

Tietoverkoissa liikkuvat tietoverkkopaketit ovat Hunt ja Zeadally (2012) mukaan eräs digitaalinen artefakti, jota käytetään tietoverkkoforensiikassa. Näihin paketteihin liityvät aikaleimat, lähde- ja kohde-IP-osoitteet sekä yhteys- ja hyötykuorman kokoa koskevat tiedot ovat Raghavan (2013) mukaan metadataa. Yksittäisistä tietoverkkopakettien metadatasta IP-osoite-tietoja sekä protokollatietoa käytetään Raghavan (2013) mukaan tietoverkkopakettien organisoimiseen verkkoon liittyvien tunkeutusmistapausten tutkimisen yhteydessä.

Eräs suurimmista haasteista liittyen tietoverkkopakettien aikaleimoihin on Raghavan (2014) mukaan luoda yhtenäinen aikajana tapahtumista, jotka liittyvät tutkittavaan rikokseen, koska hänen mukaansa mm. aikaleimassa oleva aika voi olla väärennetty. Lisäksi hänen mukaansa aikavyöhyke ja siihen liittyvän aikaleiman tulkinta voi luoda oman ongelmansa kuten myös aikaleimaan liittyvät syntaksiset näkökulmat.

4.3 Metadata matkapuhelimissa

Käyttämällä ihmisjoukkolähteisiä puhelinlokeja ja sosiaalisten verkostojen tietoja, voidaan havaita, että eri puhelimien metatiedot ovat tiheästi toisiinsa kytkettyjä ja altistavat uudelleen tunnistamisen sekä erittäin herkkäluonteisia päätelmiä ihmisistä. Matkapuhelimien puhelujen metadata kertoo osapuolien puhelinnumerot, jotka ovat puhelun käyneet, lisäksi puhelun ajankohdan sekä puhelun keston ja tekstiviestin pituuden merkeissä. (Mayer, Mutchler ja Mitchell 2016)

Mayer, Mutchler ja Mitchell (2016) kertovat kokeestaan, jossa he jäljittelivät tiedustelupalvelujen harjoittamaa metadataan pohjautuvaa tiedustelua matkapuhelimien metadataan pohjaten. Tuloksista selvisi, että metadatavalvontaohjelmilla on voitu tunnistaa henkikökohtaisesti ihmisiä. Lisäksi puhelu- ja tekstiviesti metadatan pohjalta pystyttiin tekemään sijaintiin perustuvia tarkkoja päätelmiä. Esimerkiksi kotipaikkasijainnin pystyi päättelemään epätarkasta ja niukasta puhelimen metadatasta.

Kokeesta selvisi myös, että puhelimien metadatasta pystyttiin päättelemään ihmisuhteiden erilaisia ryhmiä. Kokeessa keskityttiin romanttisten ihmissuhteiden tunnistamiseen. Tässä käytettiin alustavasti hyväksi Facebookin profiilien tietoja sivilisäädystä, jolla saatiin käyttökelpoisia päätelmiä pohjaten puhelimien puhelutietojen metadataan.

Ehkä suurimmaksi huolenaiheeksi kokeen yhteydessä mainittiin mahdollisuus tehdä arkaluontoisia päätelmiä ihmisistä puhelimien sisältäämän metadataan pohjaten. Tällä viitattiin päätelmiin ihmisten perhekeskeisiin, poliittisiin, ammatillisiin, uskonnollisiin sekä seksuaalisiin yhteyksiin.

5 Verkonvalvonnan yksityisyysproblematiikasta

Sirkkunen ja Haara (2017) kutsuvat raportissaan verkonkäyttäjien seuraamista sekä tiedon keräämistä heistä teknisillä keinoilla verkkovalvonnaksi. He kertovat, että viranomaisten harjoittaman verkonvalvonnan tarkoitus on taata ihmisten yleinen turvallisuus ehkäisemällä vakavia rikoksia, erityisesti terrorismia.

Termi Big Data-analytiikka turvallisuustiedusteluun viittaa Gahi, Guennoun ja Mouftah (2016) mukaan prosessiin, jossa analysoidaan ja louhitaan suuria määriä tietoja eri lähteistä, kuten Zawoad ja Hasan (2015) mukaan verkkosivuilta, tietoverkko- ja prosessilokeista, postauksista sosiaaliseen mediaan, sähköposteista ja erilaisista antureista. Yksityisyyshuolet, jotka nousevat Big Data analytiikasta turvallisuustiedustelussa ovat Gahi, Guennoun ja Mouftah (2016) mukaan mm. siihen liittyvien erilaisten vaatimusten noudattaminen sekä ylipäätään tällaista analyytiikkaa käytettäessä suuren datamäärän vuoksi ihmisten anonymisaatio voi käydä mahdottomaksi. Erilaista dataa verkonkäyttäjistä liikkuukin epäilemättä paljon. Yksityisyyden puolustajat yrittävätkin Lyon (2014) mukaan edistää uusia poliittisia lähestymistapoja Big Datan kaltaiseen ilmiöön.

Joukkovalvonnalla Sirkkunen ja Haara (2017) viittavat valtioiden usein salaa harjoittamaan läpitunkevaan valvontaan, joka koskettaa kokonaista väestöä tai merkittävää osaa siitä. Valtioiden salaa harjoittaman valvonnan kohteeksi voi joutua tietämättään, vaikka olisi lainkuuliainen ihminen. Tästä käy esimerkkinä NSA:n vuonna 2013 paljastettu PRISM-vakoiluohjelma, jonka myötä NSA:lla oli pääsy Lyon (2014) mukaan mm. Googlen ja Microsoftin palvelimille, joten riippumatta asuinpaikastaan, on ihminen voinut altistua tälle valvonnalle, jos on käyttänyt em. yritysten palveluilta. Van Dijck (2014) kirjoittamasta artikkelista käy ilmi, että NSA olisi PRISM:n kautta kerännyt Big Datasta metadataa, ei itse sisältöä.

Kuitenkin pelkän metadatan pohjalta voidaan Van Dijck (2014) mukaan automatisoitujen lokien pohjalta tietää kenen kanssa kukin kommunikoi, mistä paikasta ja kuinka kauan. Samassa artikkelissa tulee esiin, että samoin kuin jotkut yritykset, lisäksi valtioiden virastot harjoittaisivat datafikaatiota: Ihmisten elämä muuttuu verkossa olevaksi tiedoksi, mikä mahdollistaa reaaliaikaisen jäljittämisen ja ennakoivan analyysin. Toimintamallien tunnistami-

nen tietämättä jätetyistä (meta)datajäljistä sosiaalisten verkostojen sivustoilla palvelee Van Dijck (2014) mukaan yhä enemmän tulevaisuuden käyttäytymisen ennustamista. Tätä eksponentiaalisesti kasvavaa metadatamäärää valtion virastot louhivat Van Dijck (2014) mukaan kommunikaatioalustoilta kuten Facebook, Twitter, LinkedIn, WhatsApp, Skype, YouTube ja GMail ihmisten käyttäytymisen seuraamiseksi. Ehkä hieman yllättäenkin on paljastunut lisäksi Van Dijck (2014) mukaan, että yritykset jakavat omista sosiaalisen median verkostoistaan joko mielellään tai vastahakoisesti tietojaan tiedustelupalveluiden käyttöön.

Suomessa vuoden 2015 hallitus esitti vuonna 2017 eduskunnalle säädettävästä tietoliikennekommunikaation siviilitiedustelulaista liittyen mm. metadatan keräämiseen esityksen sisältäen mm. seuraavaa: ”Tietoliikennetiedusteluun liitettävien oikeusturvajärjestelyjen tulisi olla sekä ennakollisia että jälkikäteisiä. Ennakollinen oikeusturva toteutuisi säätämällä tuomioistuin tietoliikennetiedustelun käytön päätöksentekijäksi. Tuomioistuimen edellytettäisiin hyväksyvän suodatuksessa käytettävien viestin sisältöä kuvaavien hakuehtojen käytön. Tietoliikennetiedustelun yhteydessä kertyvä metadata tallennettaisiin tarkoitusta varten luotavaan tietovarantoon, johon kohdistuvat haut tuomioistuin myös hyväksyisi (Finlex 2015).”

Esityksestä käy ilmi, että Suomessa tiedustelua harjoittaisi suojelupoliisi, jonka toimenkuvaan lisääntyneiden toimintavalmiuksien puitteissa kuuluisi nimenomaan rikostorjunta. Puolustusvoimat analysoisivat tilastollisesti tietoliikennevirtoja. Uusi siviilitiedustelulaki tuli voimaan kesäkuun 1. päivänä 2019.

Ihminen voi valvoa myös itse itseään. Tätä Sirkkunen ja Haara (2017) kutsuvat notkeaksi valvonnaksi. Tämän he kertovat perustuvan siihen, että ihmisen ollessa tietoinen olevansa jonkinlaisen valvonnan alaisena tietämättä kuitenkaan, kuinka hänestä kerättyjä tietoja todella käytetään, hän kontrolloisi itse omaa toimintaansa verkossa yksityisyytensä suojelemikseksi. Yksityisyys itsessään on Solove (2008) mukaan laaja käsite, joka käsittää mm. ajatuksenvapauden, henkilökohtaisten tietojen hallinnan, valvonnan vapauden, maineen suojaamisen sekä suojan etsinnöiltä ja kuulusteluilta. Voi katsoa, että nykyajan ihminen maksaa turvallisuudestaan jonkinasteisella yksityisyyden menettämisellä. Yhdysvaltain presidentti Barack Obama sanoikin PRISM-vakoiluohjelmaan liittyen Van Dijck (2014) mukaan, että kansalaiset eivät voi odottaa sekä 100 prosenttista turvallisuutta että 100 prosenttista yksityisyyttä.

6 Yhteenveto

Digitaalinen forensiikka on kehittyvä ala, joka koostuu useasta haarasta siten, että kullakin haaralla on itselleen sekä spesifinen prosessimalli että spesifiset vaiheet. Ajansaatossa tutkijat ovat ehdottaneet kehittämismielessä vaihtoehtoisia malleja digitaalisen forensiikan eri haaroille. Kuitenkin voidaan, kuten luvussa 2 on esitetty, esittää yleinen digitaalisen forensiikan prosessimallin konsepti. Tällekin Valjarevic ja Venter (2012) ovat esittäneet ns. harmonisoidun prosessimallin, jossa vaiheita on lisätty ja linkitetty toisiinsa. Big Data-forensiikan ollessa suhteellisen uusi digitaalisen forensiikan haara ovat sen haasteet myös erityisesti Big Datan itsensä luonteen vuoksi suuria, mutta samalla myös sen luomat mahdollisuudet melkoisia.

Tässä tutkielmassa ovat varsinaiset kyberrikollisuuden muodot sivuutettu tutkielman ottaessa kantaa vain digitaaliseen anti-forensiikkaan tavalla, mitä se erään määritelmän mukaan on yhtenäismääritelmän kuitenkin puuttuessa. Näin kyseeseen rikollisuuden ilmentymistä on tullut digitaalisiin todisteisiin liittyvät rikolliset toimet sekä uusimpana ja suurimpana uhkana itse rikostutkintaprosessia vastaan hyökkääminen.

Useassa tutkimusartikkelissa tulee metadatan rooli jollain tavoin esiin liittyen digitaaliseen forensiikkaan. Aikaleimojen mahdollinen väärentäminen luo kuitenkin esimerkiksi Raghavan (2014) mukaan suuren haasteen tietokone- ja tietoverkkoforensiikassa rekonstruoida rikollisten tapahtumien aikajana. Lisäksi rikostutkinnallista kysymystä *kuka* päätellessä tiedot tiedostojen omistajuudesta ja käyttöoikeuksista eivät välttämättä anna varmaa tietoa siitä, ketkä rikollisiksi todetut toimet ovat tehneet.

Verkonvalvonnan yksityisyysproblematiikkaa tarkastellessa on tullut esiin ihmisen yksityisyyttä vaarantavana ilmiönä joukkovalvonta ja tästä ilmentymänä vuonna 2013 paljastettu PRISM-vakoiluohjelma. Pelkästään Yhdysvalloissa on tuollut tietoon useita muitakin suuren mittakaavan tiedusteluohjelmia, kuten käy ilmi Reddick, Chatfield ja Jaramillo (2015) artikkelista. Näiden ohjelmien tarkoitus on erityisesti terrorismin ennaltaehkäisy.

Myös Suomessa suojelupoliisin lisääntyneiden toimintavalmiuksien puitteissa tapahtuva tietoliikennevalvonnan tarkoitus on siihen liittyvän lain mukaan mm. Suomeen mahdollisesti

kohdistuvan terrorismin ennaltaehkäisy sekä lisäksi mm. Suomeen kohdistuvan ulkomaisen tiedustelutoiminnan valvonta. Kuitenkaan Suomessa siviilitiedustelu ei saa lain mukaan olla yleistä ja kohdentamatonta. Meidän kaikkien tiedostaessa, että olemme verkossa periaatteessa aina jonkinlaisen valvonnan alaisuudessa, ehkä kontrolloimme omaa toimintaamme siellä vastuuntuntoisina – tai sitten ehkä vain varovaisina – verkonkäyttäjinä, kuten Sirkkunen ja Haara (2017) raportissaan notkean valvonnan käsitteenä ovat tuoneet esiin. Ihmisten kantaessa vastuuta itsestään ja kanssaihmisistä digitaalisessa maailmassa, on myös digitaalisen forensiikan ammattilaisilla vähemmän työtä. Ainakin he voivat keskittyä tuolloin enemmän työnsä olennaisimpaan asiaan: Meidän kaikkien elämämme turvaamiseen.

Lähteet

Allen, W. H. 2005. “Computer forensics”. *IEEE Security Privacy* 3, numero 4 (heinäkuu): 59–62. ISSN: 1558-4046. doi:10.1109/MSP.2005.95.

Almulhem, A. 2009. “Network forensics: Notions and challenges”. Teoksessa *2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT),* 463–466. Joulukuu. doi:10.1109/ISSPIT.2009.5407485.

Anderson, R. J., ja F. A. P. Petitcolas. 1998. “On the limits of steganography”. *IEEE Journal on Selected Areas in Communications* 16, numero 4 (toukokuu): 474–481. ISSN: 1558-0008. doi:10.1109/49.668971.

Buchholz, Florian, ja Eugene Spafford. 2004. “On the role of file system metadata in digital forensics”. *Journal of Digital Investigation* 1 (4): 297–308.

Carrier, Brian, ja Eugene H Spafford. 2004. “An event-based digital forensic investigation framework”. Teoksessa *Digital forensic research workshop,* 11–13.

Castiglione, A., A. De Santis ja C. Soriente. 2007. “Taking advantages of a disadvantage: Digital forensics and steganography using document metadata”. Component-Based Software Engineering of Trustworthy Embedded Systems, *Journal of Systems and Software* 80 (5): 750–764. ISSN: 0164-1212. doi:https://doi.org/10.1016/j.jss.2006.07.006. http://www.sciencedirect.com/science/article/pii/S0164121206001981.

Das, Dolly, Urjashee Shaw ja Smriti Priya Medhi. 2017. “REALIZING DIGITAL FORENSICS AS A BIG DATA CHALLENGE”. Teoksessa *4th International Conference on “Computing for Sustainable Global Development”, New Delhi (INDIA) 01st-03rd March.*

Finlex. 2015. “Hallituksen esitys eduskunnalle siviilitiedustelua koskevaksi lainsäädännöksi”. https://www.finlex.fi/sv/esitykset/he/2017/20170202.

Gahi, Y., M. Guennoun ja H. T. Mouftah. 2016. “Big Data Analytics: Security and privacy challenges”. Teoksessa *2016 IEEE Symposium on Computers and Communication (ISCC),* 952–957.

Garfinkel, Simson L. 2010. “Digital forensics research: The next 10 years”. *digital investigation* 7:S64–S73.

Harris, Ryan. 2006. “Arriving at an anti-forensics consensus: Examining how to define and control the anti-forensics problem”. The Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS ’06), *Digital Investigation* 3:44–49. ISSN: 1742-2876. doi:https://doi.org/10.1016/j.diin.2006.06.005. http://www.sciencedirect.com/science/article/pii/S1742287606000673.

Hunt, R., ja S. Zeadally. 2012. “Network Forensics: An Analysis of Techniques, Tools, and Trends”. *Computer* 45, numero 12 (joulukuu): 36–43. ISSN: 1558-0814. doi:10.1109/MC.2012.252.

Juola, Patrick. 2007. “Future Trends in Authorship Attribution”. Teoksessa *Advances in Digital Forensics III,* toimittanut Philip Craiger ja Sujeet Shenoi, 119–132. New York, NY: Springer New York. ISBN: 978-0-387-73742-3.

Kessler, Gary C. 2007. “Anti-forensics and the digital investigator”.

Lyon, David. 2014. “Surveillance, Snowden, and Big Data: Capacities, consequences, critique”. *Big Data & Society* 1 (2): 2053951714541861. doi:10.1177/2053951714541861. eprint: https://doi.org/10.1177/2053951714541861. https://doi.org/10.1177/2053951714541861.

Martini, Ben, ja Kim-Kwang Raymond Choo. 2014. “Distributed filesystem forensics: XtreemFS as a case study”. *Digital Investigation* 11 (4): 295–313. ISSN: 1742-2876. doi:https://doi.org/10.1016/j.diin.2014.08.002. http://www.sciencedirect.com/science/article/pii/S1742287614000942.

Mayer, Jonathan, Patrick Mutchler ja John C Mitchell. 2016. “Evaluating the privacy properties of telephone metadata”. *Proceedings of the National Academy of Sciences* 113 (20): 5536–5541.

Pilli, Emmanuel S., R.C. Joshi ja Rajdeep Niyogi. 2010a. “Network forensic frameworks: Survey and research challenges”. *Digital Investigation* 7 (1): 14–27. ISSN: 1742-2876. doi:https://doi.org/10.1016/j.diin.2010.02.003. http://www.sciencedirect.com/science/article/pii/S1742287610000113.

———. 2010b. “A generic framework for network forensics”. *International Journal of Computer Applications* 1 (11): 1–6.

Raghavan, Sriram. 2013. “Digital forensic research: current state of the art”. *CSI Transactions on ICT* 1 (1): 91–114.

———. 2014. “A framework for identifying associations in digital evidence using metadata”. Tohtorinväitöskirja, Queensland University of Technology. https://eprints.qut.edu.au/72659/.

Rasjid, Zulfany Erlisa, Benfano Soewito, Gunawan Witjaksono ja Edi Abdurachman. 2017. “A review of collisions in cryptographic hash function used in digital forensic tools”. *Procedia computer science* 116:381–392.

Reddick, Christopher G., Akemi Takeoka Chatfield ja Patricia A. Jaramillo. 2015. “Public opinion on National Security Agency surveillance programs: A multi-method approach”. *Government Information Quarterly* 32 (2): 129–141. ISSN: 0740-624X. doi:https://doi.org/10.1016/j.giq.2015.01.003. http://www.sciencedirect.com/science/article/pii/S0740624X15000246.

Reith, Mark, Clint Carr ja Gregg Gunsch. 2002. “An examination of digital forensic models”. *International Journal of Digital Evidence* 1 (3): 1–12.

Rogers, Marcus K, James Goldman, Rick Mislan, Timothy Wedge ja Steve Debrota. 2006. “Computer forensics field triage process model”. *Journal of Digital Forensics, Security and Law* 1 (2): 2.

Rughani, Parag H. 2017. “ARTIFICIAL INTELLIGENCE BASED DIGITAL FORENSICS FRAMEWORK.” *International Journal of Advanced Research in Computer Science* 8 (8).

Sirkkunen, Esa, ja Paula Haara. 2017. *Yksityisyys ja notkea valvonta: Yksityisyys ja anonymiteetti verkkoviestinnässä-projektin loppuraportti.*

Solove, Daniel J. 2008. “Understanding privacy”.

Valjarevic, A., ja H. S. Venter. 2012. “Harmonised digital forensic investigation process model”. Teoksessa *2012 Information Security for South Africa,* 1–10.

Van Dijck, José. 2014. “Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology”. *Surveillance & society* 12 (2): 197–208.

Wazid, M., A. Katal, R. H. Goudar ja S. Rao. 2013. “Hacktivism trends, digital forensic tools and challenges: A survey”. Teoksessa *2013 IEEE Conference on Information Communication Technologies,* 138–144. Huhtikuu. doi:10.1109/CICT.2013.6558078.

Wei Ren ja Hai Jin. 2005. “Modeling the network forensics behaviors”. Teoksessa *Workshop of the 1st International Conference on Security and Privacy for Emerging Areas in Communication Networks, 2005.* 1–8.

Wikipedia. 2020. “Comparison of file systems”. https://en.wikipedia.org/wiki/Comparison_of_file_systems.

Zawoad, Shams, ja Ragib Hasan. 2015. “Digital Forensics in the Age of Big Data: Challenges, Approaches, and Opportunities”. Teoksessa *Proceedings of the 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conf on Embedded Software and Systems,* 1320–1325. HPCC-CSS-ICESS ’15. USA: IEEE Computer Society. ISBN: 9781479989379. doi:10.1109/HPCC-CSS-ICESS.2015.305. https://doi.org/10.1109/HPCC-CSS-ICESS.2015.305.