Jenny Hornborg

VALTUUTETUN KÄYTTÄJÄN VALTUUTTAMATTOMIEN HAKUJEN TUNNISTAMINEN LOKIANALYTIIKAN KEINOIN

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2021

TIIVISTELMÄ

Hornborg, Jenny
Valtuutetun käyttäjän valtuuttamattomien hakujen tunnistaminen lokianalytiikan keinoin
Jyväskylä: Jyväskylän yliopisto, 2021, 82 s.
Tietojärjestelmätiede, pro gradu-tutkielma
Ohjaaja: Seppänen, Ville

Digitalisoituvan maailman myötä organisaatioiden toimintatavat ja työkalut ovat muuttuneet huomattavasti tietoteknisempään suuntaan. Muutos on koskettanut myös yksilöitä, sillä heistä kerätään nykyisin enemmän dataa, kuin koskaan ennen. Euroopan Unionin Yleinen tietosuoja-asetus (eng. EU General Data Protection Regulation, GDPR) pyrkii tuomaan yksilöiden datan käsittelyyn sen kaipaamaa turvaa. Organisaatioille GDPR toi mukanaan lisää vastuuta ja tarvetta varmistua työntekijöiden valtuutetusta toiminnasta. Valtuutetut käyttäjät voivat asettaa organisaatiolle tietoturvauhan, esimerkiksi väärinkäyttämällä oikeuksiaan tietojärjestelmiin hakemalla tietoja, joihin heillä ei ole valtuutusta. GDPR:n näkökulmasta onkin keskeistä huomioida, että käyttöoikeus järjestelmään ei tarkoita oikeutta käyttää järjestelmistä löytyvää dataa, vaan siihen vaaditaan aina peruste. Yhdeksi keskeiseksi keinoksi tällaisten valtuuttamattomien hakujen tunnistamiseen ja monitorointiin on ehdotettu lokianalytiikkaa. Kuitenkin tietojärjestelmien suuren käyttöasteen takia myös lokitietoa tallentuu paljon, jolloin valtuuttamattoman haun tunnistaminen voi olla haastavaa. Suurten lokitapahtumien määrän lisäksi myös lokien laatu voi asettaa organisaatioille suuria haasteita toteuttaa lokianalytiikkaa. Lokimassan keskeiseksi rajauskriteeriksi valtuuttamattomien hakujen tunnistamisessa havaittiin työntekijöiden poikkeava toiminta. Tämä Pro gradu-tutkielma toteutettiin toimeksiantona yhdelle Suomen suurimmista teleoperaattoreista. Tutkielma seuraa suunnittelutieteellistä tutkimusmenetelmää ja luo sen metodein artefaktina suodatinmallin. Suodatinmalli koostuu 12 suodattimesta, joiden avulla organisaatioiden on mahdollista tehdä poikkeavia hakuja nostavia suodatuksia analysoitavaan lokimassaan, joka edelleen mahdollistaisi valtuuttamattomien hakujen tunnistamisen. Tutkielma esittää myös perustelut lokianalytiikan suorittamiselle, sekä sille miksi lokien laatuun huomion kiinnittämistä voidaan pitää keskeisenä lokianalytiikan näkökulmasta.

Asiasanat: Lokianalytiikka, GDPR, Tietoturva, Valtuutettu käyttäjä, Valtuuttamaton haku

ABSTRACT

Hornborg, Jenny
Detecting authorized user's unauthorized search with log analytics
Jyväskylä: University of Jyväskylä, 2021, 82 pp.
Information Systems Science, Master's Thesis
Supervisor: Seppänen, Ville

The way of working has changed in the past decade due to the digitalization that is happening across the world. The change has also affected individuals, since their data is being processed more than ever before. European Union's General Data Protection Regulation has brought the much-needed security to individual's data processing. To organizations GDPR has given more responsibility and a bigger need to ensure that their employees are acting based on authorizations. However, it has been detected that users can create an information security threat to organizations for example by misusing their accesses to information systems and searching for data that they do not have an authorization for. From GDPR's perspective it is crucial to understand that having an authorized access to information systems does not equal to authorized usage of the data. One of the most significant tools for detecting and monitoring this type of unauthorized searches is log analytics. However, information systems are widely used in organizations, and this leads to significant amount of log events being generated, which can set a challenge to detecting unauthorized searches. Besides significant amounts of log events also log quality can set a challenge on organization's log analytics. It was acknowledged that filtering log events of employee's abnormal behavior can help detect unauthorized events. This Master thesis was carried out as an assignment for one of the biggest telecom operators acting in Finland. This thesis follows the design science method and creates a filtering model as an artefact. The filtering model includes 12 filters and offers filters that will help to raise abnormal events from the log events that are being analyzed, helping to detect unauthorized searches. This thesis also justifies the need for log analytics and explains why log quality is a significant factor for log analytics' effectiveness.

Keywords: Log analytics, GDPR, Information security, Authorized user, Unauthorized search

KUVIOT

KUVIO 1 Tahallinen ja tahoton tiedon vaarantaminen työntekijän toimesta 13
KUVIO 2 Esimerkki Audit Trailistä 26
KUVIO 3 CITD työkalun kuvaus 33
KUVIO 4 Suodatinmallin ensimmäinen versio 56
KUVIO 5 Suodatinmallin lopullinen versio 62

TAULUKOT

TAULUKKO 1 GDPR keskeiset profiilit 15
TAULUKKO 2 Lokityypit tarkoitusperän mukaan 19
TAULUKKO 3 Ominaisuuksien erittely ryhmittely-lajittelu-jalostus viitekehyksessä 30
TAULUKKO 4 Observointiin perustuva lajittelu 32
TAULUKKO 5 Viitekehyksen soveltaminen PoSeID-On alustaan 38
TAULUKKO 6 Haastateltavien organisaatiolliset taustatiedot 43
TAULUKKO 7 Suodatinmallin ensimmäisen version selitteet 57
TAULUKKO 8 Lopullisen suodatinmallin selitteet 63

SISÄLLYS

TIIVISTELMÄ
ABSTRACT
KUVIOT JA TAULUKOT

1 JOHDANTO 7

2 TARVE SISÄISEN TIETOTURVAN TARKISTELULLE 10
2.1 Tietoturvallisuuden merkitys organisaatiolle 10
2.2 Organisaation sisäiset uhat tietoturvallisuudelle 11
2.3 GDPR – Yleinen tietosuoja-asetus 15
2.3.1 Yksilö ja data 16
2.3.2 GDPR ja organisaation sisäiset uhat 17

3 KATSAUS LOKITIETOIHIN 18
3.1 Lokien käyttötarkoitukset 18
3.1.1 Lokitietojen laadun merkitys 21
3.1.2 Lokitiedot ja lakivelvoitteet 23
3.2 Lokianalytiikka 24
3.2.1 Audit Trail lokitiedoissa 25
3.2.2 Lokianalytiikan merkitys organisaatiolle 27

4 LOKIANALYYTIIKKA VÄÄRINKÄYTÖSTEN TUNNISTAMISESSA 29
4.1 Poikkeavuuksien tunnistaminen lokitiedoista 29
4.2 Käyttäjä- ja roolipohjainen analytiikka 31
4.3 GDPR ja valtuuttamattomat haut 35
4.3.1 GDPR ja lokitiedot 36
4.3.2 PoSeID-on projekti 37

5 TUTKIMUSMENETELMÄT 40
5.1 Toimeksiantajayrityksen esittely 40
5.2 Suunnittelutieteellinen tutkimusmenetelmä 41
5.3 Datan keräystavat ja analysointi 42

6 SUODATINMALLIN LUOMINEN 45
6.1 Ongelma, motivaatio ja ratkaisun objektiivit 45
6.2 Lokianalytiikkaa ja sen merkitys 46
6.2.1 GDPR lokianalytiikan tarpeen vahvistajana 47
6.2.2 Lokianalytiikka työntekijöiden turvana 48
6.2.3 Prosessipuutteiden havaitseminen 49
6.2.4 Muunlaisten valtuuttamattomien hakujen lokianalytiikka 50
6.3 Lokien laadun vaikutus analytiikkaan 51
6.4 Suodatinmallin ensimmäisen version luominen 52
6.5 Suodatinmallin arviointi ja edelleen kehitys 58

6.5.1 Esiteltyjen suodattimien kehitys 58
6.5.2 Uudet suodattimet ja mallin jatkokehitys 60
6.5.3 Suodatinmallin lopullinen versio 61

7 JOHTOPÄÄTÖKSET JA POHDINTA 65
7.1 Tutkielman tulokset 65
7.1.1 Lokianalytiikan keskeisyys 66
7.1.2 Lokien laadun merkitys lokianalytiikassa 67
7.1.3 Suodatinmalli lokianalytiikan tukena 67
7.2 Tutkielman tulosten merkitys 68
7.3 Tutkielman rajoitteet ja ehdotukset jatkotutkimusaiheiksi 69

8 YHTEENVETO 71

LÄHTEET 73

LIITE 1 HAASTATTELUN KYSYMYSPATTERISTO 79

LIITE 2 SAATEKIRJE HAASTATELTAVILLE ITEROINTI 80

LIITE 3 SUODATINMALLIN ESITTELY ITEROINTIA VARTEN 81

1 JOHDANTO

Maailma ja sen mukana organisaatioiden toimintatavat ovat kehittyneet vahvasti teknologian kehityksen mukana. Yritysten toimintatavat keskittyvät yhä enemmissä määrin toimiin, joissa hyödynnetään teknologiaa. Tämä kehitys on tuonut mukanaan organisaatioille monia etuja ja tehostanut toimintatapoja. Teknologian kehitys on kuitenkin tuonut mukanaan myös negatiivisia ilmiöitä ja uusia uhkakuvia. (Safa, Maple, Watson, & Von Solms, 2018.) Tietoa voidaan pitää yhtenä yrityksen tärkeimmistä resursseista, ja tietoturvaa uhkaavat tekijät ovatkin nykyisin yrityksille asia, joka tulee huomioida jokapäiväisessä toiminnassa (Nieles, Dempsey & Pillitteri, 2017). Useat organisaatiot ovatkin tunnistaneet keskeisen tarpeen tietoturvallisuuden varmistamiseen. Tarpeeseen on vastattu tietoturvallisuutta parantavilla toimilla ja investoinneilla. Investointeihin voi kuulua esimerkiksi tietoturvaa suojaavia ohjelmistohankintoja. Nämä ennaltaehkäisevät toimet ovat kuitenkin usein kohdistettuja organisaation ulkoisia uhkia kohtaan. (Mulligan & Schneider, 2011.) On kuitenkin tunnistettu, että yrityksille merkittävä ja todennäköisin tietoturvariski on kuitenkin organisaation sisäiset uhkatekijät (Safa, yms., 2018). Organisaatiota, tai sen omaavia tietoja vastaan sisältä päin tulevat hyökkäykset ovat saaneet merkittävästi huomiota viime vuosikymmenen aikana, esimerkiksi WikiLeaksin myötä (Wall, 2013).

Osa organisaatioiden tietoturvan sisäisistä uhkakuvista on tiedostettu myös laissa. EU:n yleinen tietosuoja-asetus 2016/679 General Data Protection Regulation (GDPR) velvoittaa organisaatiot varmistamaan, että heidän työntekijänsä käsittelevät heidän asiakkaidensa henkilötietoja vaaditulla tavalla noudattaen lakia ja tietosuojaperiaatteita (GDPR, 2016). Useiden organisaatioiden työntekijöiden tulee luonnollisesti käsitellä henkilötietoja työssään, mutta on tärkeää huomioida, että näille käsittelytapauksille tulee olla oikeutettu peruste. On tärkeää tunnistaa, että vaikka työntekijällä on pääsy asiakkaiden henkilötietoihin, ei hänellä silti välttämättä ole oikeutta tietoja tarkastella. Sisäisiä uhkia vastaan puolustautuminen on haasteellista, sillä organisaation sisäisille toimijoille on annettu pääsy tietoihin, tietojen tarkastelu voi sisältyä heidän työtehtäviinsä ja työntekijöiden toimintaa kohtaan on organisaatiolla olemassa lähtökohtaisesti luottamus. Yhdeksi keskeiseksi ja tehokkaaksi tavaksi tunnistaa valtuutetun käyttäjän

valtuuttamattomia henkilötietohakuja on suorittaa organisaation sisäistä lokianalytiikkaa. (Myers, Grimaila & Mills, 2009.)

Kyberturvallisuuskeskus (2020) on määritellyt lokien tarkoittavan aikajärjestyksessä kerättyjä tapahtumia ja niiden aiheuttajia. Esimerkiksi tapahtumat tietojärjestelmissä, sovelluksissa, tietoverkoissa ja tietosisällöissä voidaan kirjata lokiin. Näiden tietojen analysoinnin kautta, voidaan havaita organisaation sisäisen toimijan toteuttamia hakuja, joille ei kuitenkaan löydy oikeutettua perustetta. Valtuuttamattomien hakujen tunnistamisen ohella lokitietojen analysointi tarjoaa arvokasta tietoa organisaatiolle siitä, miten sisäiset toimijat käyttäytyvät yrityksen tietoverkoissa ylipäätään. Lokianalytiikka kohtaa kuitenkin haasteen siinä, miten erottaa valtuuttamattomat haut valtuutetuista hauista. (Myers, Grimaila & Mills, 2009.) Sisäiset toimijat toteuttavat valtuutetusti useita toimia, joiden joukkoon voi helposti sulautua myös valtuuttamattomia hakuja. Näiden tunnistaminen valtuutettujen hakujen joukosta asettaa organisaatioille suuria haasteita. (Liu, Qin, Guan, Jiang & Wang, 2018.)

Tämä tutkielma toteutetaan toimeksiantona yhdelle Suomen suurimmista teleoperaattoreista. Teleoperaattorien tulee luonnollisesti noudattaa GDPR:n mukaisia toimintatapoja ja taata asiakkaidensa tietoturva. Kuitenkin haasteet valtuutettujen käyttäjien valtuuttamattomien hakujen tunnistamisessa koskevat myös operaattoritoimintaa. Lokitietojen analysointia toteutetaan kohdeorganisaatiossa tutkielman aloitushetkellä. Tutkielma on kuitenkin oleellinen tähän hetkeen toteutettavaksi, sillä henkilötietohakuihin kohdistuvan lokianalytiikan kehitysprosessi on käynnissä. Tutkielman tavoitteena on luoda organisaatioihin soveltuva suodatinmalli. Suodatinmallissa kuvataan mitä rajauksia tulisi tehdä suureen lokimassaan, jotta henkilötietohakujen lokianalytiikka tehostuu. Tehostamisen on tarkoitus mahdollistaa poikkeavien, ja näin ollen mahdollisesti valtuuttamattomien hakujen tunnistaminen suuresta lokitietojen datamassasta, jossa suurin osa tapahtumista on valtuutetusti toteutettuja. Lisäksi tutkimuksessa tullaan perustelemaan, miksi lokianalytiikan toteuttaminen on yrityksessä merkityksellistä ja miksi lokitietojen laadulla on merkitystä lokianalytiikan toteutuksen tehokkuuteen. Tutkimus rakentuu seuraavien tutkimuskysymysten ympärille:

- Miksi sisäiseen toimintaan kohdistuvaa lokianalytiikka tulisi tehdä?
- Miksi lokien laadulla on vaikutus lokianalytiikan toteuttamiseen?
- Millaisilla suodattimilla valtuuttamattomien henkilötietohakujen tunnistamista lokianalytiikassa voidaan tehostaa?

Tutkimuksen keskeisenä tutkimuskysymyksenä, jonka avulla tutkielman artefakti luodaan, toimii viimeisenä esitetty kysymys lokianalytiikan tehostamisesta. Tukevina tutkimuskysymyksinä toimivat syyt lokianalytiikan toteuttamisen takana, sekä lokien laadun ja johdonmukaisuuden merkitys lokianalytiikan toteutuksen kannalta. Tutkimuksen tuotoksena syntyvän suodatinmallin odotetaan täyttävän valtuuttamattomien hakujen lokianalytiikassa esiintyneet tarpeet, tarjoten tukea myös muille organisaatioille heidän sisäiseen toimintaansa kohdistuvan lokianalytiikkansa suunnittelussa. Tutkimuksen ulkopuolelle rajataan

GDPR:n vaikutus lokitietojen tallennukseen ja käsittelyyn lokittajan vastuun näkökulmasta. Lisäksi tutkimuksen ulkopuolelle rajataan lokituksen, lokianalytiikan, sekä esitettävien suodattimien tekninen toteutus. Lisäksi on merkityksellistä tunnistaa, että ratkaisu tehdään toimeksiantajaorganisaation ympäristöön, eikä se välttämättä ole suoraan sovellettavissa muihin ympäristöihin.

Tutkielman teoreettinen osuus on toteutettu kirjallisuuskatsauksena. Tutkielman teoreettinen taustatieto on haettu pääosin Google Scholar ja JYKDOK hakemistoista. Teoriaosuudessa hyödynnetään pääasiallisesti JUFO julkaisufoorumilla luokiteltua tieteellistä aineistoa, mutta tutkimuksen aihepiirin vuoksi myös niin sanottua harmaata lähdetietoa on hyödynnetty. Esimerkiksi erilaisten virastojen ja tutkimushankkeiden data on tutkielman aiheen kannalta arvokasta.

Tutkielman rakenne on rakentunut seuraavasti. Johdanto esitellään tutkimuksen aihealue lyhyesti, sekä tutkimusongelma ja tutkimuskysymykset. Tutkielman alkuosa on teoreettinen osuus, jossa käsitellään tutkimuskysymysten kannalta oleellista aiempaa tutkimusta ja tietoutta kirjallisuuskatsauksen muodossa. Toinen ja kolmas luku pyrkivät määrittämään tutkimuksen kannalta oleellista aihealuetta ja käsitteitä, sekä tarjoamaan näkökulmaa tutkimusongelman tärkeydestä. Neljännessä luvussa pyritään avaamaan lokianalytiikan merkitystä valtuuttamattomien toimien tunnistamisessa syvemmin, myös GDPR:n näkökulmasta, sekä nostamaan esiin mahdollisia suodattimia ja keinoja, joilla valtuuttamattomat haut voidaan tunnistaa. Tämän jälkeen viides luku esittelee tutkimusmenetelmän. Tutkielman empiirinen osuus on toteutettu suunnittelutieteellisin laadullisin metodein. Kuudennessa luvussa esitellään tutkimuksen tulokset. Luodaan ensimmäinen versio artefaktista. Sitten suoritetaan suodatinmallin ensimmäisen version iterointi. Kuudennessa luvussa esitellään myös iteroinnin kautta toteutettu tutkielman puitteissa lopullinen versio artefaktista. Seitsemäs luku kokoaa yhteen johtopäätökset ja pohdinnan tutkielmasta, johon sisältyy artefaktin tavoitteiden onnistuminen, sekä tutkimustulosten merkityksen pohtiminen. Lisäksi luvussa esitellään tutkielman rajoitteet ja mahdollisia jatkotutkimusaiheita. Tutkielman viimeinen luku keskittyy tutkielman keskeisten havaintojen ja toteutuksen yhteenvetoon.

2 TARVE SISÄISEN TIETOTURVAN TARKISTELULLE

Informaatioteknologian kehityksen myötä organisaatioiden toiminta on kehittynyt huomattavaa vauhtia. Dataa käsitellään enemmän, kuin koskaan ennen. Valitettavasti myös datan vaarantavat uhkakuvat ovat lisääntyneet ja kehittyneet teknologian kehityksen myötä. Moderni liiketoimintamaailma asettaa uudenlaisia uhkakuvia erityisesti dataa kohtaan. Tietoa voidaan pitää organisaation yhtenä keskeisenä resurssina ja siksi tietoturvan varmistaminen on organisaatioille keskeistä. Datan määrän kasvamisen myötä myös yksilöistä tallennettavan datan määrä on kasvanut merkittävästi. Tarve yksilöiden tietoturvan vahvistamiselle on yleisesti tunnistettu. Tämän puolesta puhuu myös Euroopan Unionin asettama tietosuoja-asetus, GDPR. Tässä luvussa käsitellään organisaatioiden kokemia sisäisiä tietoturvauhkia, sekä sitä, millaisen tarpeen Euroopan Unionin tietosuoja-asetus on luonut sisäisten tietoturvauhkien tarkistelulle.

2.1 Tietoturvallisuuden merkitys organisaatiolle

Organisaatioiden siirtyminen teknisempiin toimintatapoihin ja työvälineisiin on kehittänyt liiketoimintamaailmaa, mutta se on aiheuttanut myös haasteita organisaatioiden toimintaympäristöihin. Modernissa liiketoimintaympäristössä organisaatiot kohtaavat tietoturvauhkia, jotka kohdistuvat yrityksen dataan, informaatioteknologian infrastruktuuriin, sekä henkilökohtaisiin laitteistoihin. (Johnston & Warkentin, 2010.) Tietoa voidaankin pitää yhtenä organisaation tärkeimmistä varoista (Rickett, 2015). Tietoturvallisuuden tulisi taata tiedon saatavuus, eheys, sekä luotettavuus kaikissa datan toimintaa koskevissa vaiheissa. Nämä vaiheet sisältävät niin datan säilytyksen, käsittelyn, kuin siirron. (Whitman & Mattord, 2011.) Tietoturvallisuuden tulisi myös kattaa sen ”pehmeä” puoli, kuten ihmiset, organisaatiot, kulttuurin, etiikan, käytänteet ja lain, sekä myös sen ”kova” ja teknologinen puoli. Onkin keskeistä huomioida, että

tietoturvallisuuden on taattava tiedon, kuin tietojärjestelmien turvallisuus ollakseen kokonaisvaltaista. (Lundgren & Möller, 2019.) Erityisesti organisaatiot, jotka toimivat informaatioteknologian alalla joutuvat nykyisin kohdentamaan entistä enemmän resurssejaan tietoturvan toteuttamiseen. Informaatioteknologian alalla toimivat yritykset käsittelevät todella merkittäviä määriä informaatiota, ja tietoturvallisuutta voidaankin pitää yhtenä kriittisimpänä tekijänä organisaation toiminnassa. (Foroughi, 2008.)

Tietoturvarikkomukset ja tietomurrot voivat aiheuttaa usealla tavalla organisaatioille merkittäviä kustannuksia. Tietomurtoja voidaan pitää yhtenä organisaatioiden todennäköisimmistä ja kalliimmista uhista. Menetetyn tiedon lisäksi, tietomurrot voivat myös johtaa oikeustoimenpiteisiin, sekä häiriöaikaan organisaatiossa, jotka molemmat voivat aiheuttaa suuria kustannuksia. Tietomurron kustannukset voivat nousta jopa miljooniin dollareihin. Organisaatioiden menetykset eivät välttämättä tietomurron sattuessa rajaudu pelkästään taloudellisiin menetyksiin, vaan myös organisaation maine, sekä brändi voivat kärsiä merkittävistä vahingoista. (Rickett, 2015.) Liikekumppanit ja mahdolliset tulevat liikekumppanit edellyttävät, että tiedot ovat suojattu ja turvassa organisaation hallinnassa. Mikäli liikekumppaniyritys kokee, että organisaatio ei ole kykeneväinen tietoturvallisuuden takaamiseen, voi myös yhteistyö olla vaarassa. (Von Solms, 1998.)

Lisäksi merkittävä syy sille, miksi tietoturvallisuuteen tulisi organisaation sisällä keskittyä, on aihepiiriä koskevat lakivelvoitteet. Organisaatioiden voidaan katsoa olevan vastuussa omasta datastaan. Tämän lisäksi heidän toimintansa on oltava linjassa regulaatioiden kanssa, jotka määräävät asiakkaiden tietojen säilytyksestä. (Bulpett, 2020.) Paikallisesti, kansallisesti ja kansainvälisesti tarkisteltuna, voidaan todeta, että useilla mailla on säädöksiä, jotka vaikuttavat organisaation tietoturvallisuuden velvoitteisiin. Viimeisten vuosikymmenten aikana on havaittu, että muuttunut liiketoimintaympäristö jopa vaatii tietosuojalainsäädäntöä. Säädökset voivat vaikuttaa esimerkiksi tiedon säilytykseen, suojaukseen, kuin myös tietojen käyttöön. (Harkins, 2013.) Yhtenä keskeisimmistä säädöksistä voidaan pitää vuonna 2016 julkaistua yleistä tietosuoja-asetusta, GDPR:ää, joka asettaa organisaatioille tarkat raamit henkilötietojen käsittelyyn. GDPR:n asetuksen vastaiset toimet organisaatiossa voivat aiheuttaa yritykselle merkittävät sakot, jopa 20 miljoonaan euroon tai 4 prosenttiin yrityksen vuosittaisesta liikevaihdosta asti. GDPR voidaankin nähdä velvoittavan organisaatiot varmistamaan teknologiset, sekä organisaatiolliset turvallisuustoimet, jotka ovat riittävät riskit ja uhat huomioiden. (Chabinsky, 2018.)

2.2 Organisaation sisäiset uhat tietoturvallisuudelle

Arvokkaan tiedon ja teknologian yhdistäminen on tuonut mukanaan myös täysin uudenlaisia ulkoisia, kuin myös sisäisiä tietoturvauhkia, joihin organisaatioiden on pystyttävä vastaamaan. Ulkoisia uhkia organisaation tietoturvalle voivat olla esimerkiksi virukset, hakkerit, kuin myös rikolliset (Dorethy & Fulford, 2005).

Usein organisaatioiden keskittyminen tietoturvauhkien torjumisessa fokusoituukin juuri ulkoisiin uhkiin. Onkin arvioitu, että jopa 90 prosenttia turvallisuustoimista kohdistetaan ulkoisten uhkien torjumiseen. (Colwill, 2009.) Fokuksen kiinnittyminen ulkoisiin uhkiin johtuu useasta tekijästä. Ensiksi on huomioitava, että suuri osa tarjottavista tietoturvapalveluista, ovat keskittyneet juuri ulkoisten tietoturvauhkien torjumiseen. Tällaisia tietoturvapalveluita ovat esimerkiksi tunkeutumisen tunnistavat järjestelmät, palomuurit ja virustentorjuntaohjelmistot (Warkentin & Willison, 2009). Lisäksi usein tietoturvan toimenpiteitä toteuttavien henkilöiden suojautuminen on kohdennettu ulkoisia uhkia vastaan. Esimerkiksi on tutkittu, että median uutisointi ulkoisista tekijöistä johtuvista turvallisuusmurroista motivoi käyttäjiä luomaan itselleen vahvempia salasanoja (Mamonov & Benbunan-Fich, 2018). Organisaatiot voivat myös tehdä oletuksen, että yleisesti ulkoiset uhat ovat vakavampia sillä hyökkääjinä ovat usein tahot, jotka haluavat aiheuttaa organisaatiolle merkityksellistä vahinkoa, kuten kilpailijat, vihamiehet ja rikolliset. Kuitenkin on tunnistettava, että ulkoisilla uhilla on usein hyvin rajalliset mahdollisuudet uhkien toteuttamisessa. (Walton, 2006.)

On tunnistettu, että ulkoisten uhkien lisäksi myös sisäiset toimijat ja haasteet voivat luoda merkittävän uhan organisaation tietoturvalle. Tietoturvallisuus on asia, joka organisaation tulee itse rakentaa. Duncan ja Whittington (2016) esittivät, että vaikka osa uhkakuvista johtuvat ulkoisista tekijöistä, turvallisuuteen liittyvät haasteet tulevat usein organisaation sisäpuolelta. He tunnistivat, että pilviturvallisuuteen liittyvät avainhaasteet ovat seuraavat:

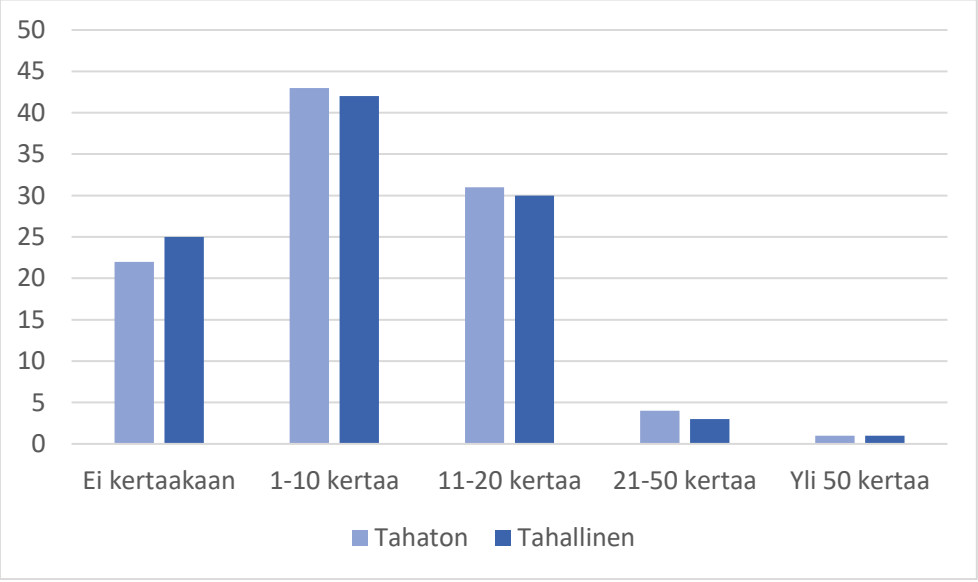
- Turvallisuustavoitteiden määrittely
- Standardien noudattaminen
- Auditoinnin ongelmat
- Johdon lähestymistavat
- Pilvipalveluiden monimutkaisuus
- Vastuun ja vastuullisuuden puute
- Mittaaminen ja monitorointi
- Johdon suhtautuminen turvallisuuteen
- Turvallisuuskulttuuri organisaatiossa
- Uhkaympäristö

Voidaan siis todeta, että merkittävä osa tietoturvallisuuden haasteista kumpuaa organisaation sisältä. Osa haasteista liittyy suoraan työntekijöiden käyttäytymiseen tietoturvaohjeistuksia kohtaan, kuten standardien noudattaminen. Tämän lisäksi sisäiset toimijat voivat myös itse muodostaa uhan organisaation tietoturvallisuudelle. Sisäiset toimijat tuntevat organisaation toimintatavat ja heillä on pääsy niin fyysisiin tiloihin, kuin dataan. Siksi sisäisillä uhilla onkin lähtökohtaisesti parempi mahdollisuus saavuttaa merkittävää tietoa, jättämällä mahdollisimman vähän todisteita, verrattuna ulkoisiin uhkiin. (Colwill, 2009.) Organisaation turvallisuustoimien tuntemus antaa sisäisille uhille myös paremmat mahdollisuudet ohittaa asetetut turvallisuustoimet. On kuitenkin tutkittu, että suurin

osa sisäisistä uhista ei ole pitkälti ennalta suunniteltuja hyökkäyksiä, vaan ennemmin tietoturvauhka muodostuu opportunistisesti. (Harkins, 2013.)

Sisäiset uhat itsessään ovat jo monimutkainen käsite. Elifoglu, Abel ja Taşseven (2018) määrittelivät sisäisen toimijan olevan henkilö, jolla on oikeutetusti pääsy organisaation tietoteknisiin resursseihin. Tämä saattaa sisältää niin nykyiset, kuin entiset työntekijät, alihankkijat, asiakkaat, sekä liiketoimintapartnerit. Sisäisen uhan he puolestaan määrittelivät olevan yksilö, jolla on oikeutettu pääsy organisaation resursseihin ja jota pääsyä hän käyttää tahallisesti tai tahattomasti väärin ja näin vaikuttaa negatiivisesti organisaation toimintaan. Bishop ja Gates (2008) määrittelivät sisäisen uhan sen omaaman pääsyn perusteella, esimerkiksi pääsy tietoihin tai fyysisiin tiloihin voi tehdä jostakin sisäisen uhan. Dodge JR., Ferguson ja Cappelli (2013) puolestaan määrittelivät sisäisen uhan olevan sellainen uhka, joka kohdistuu organisaatioon luotetulta taholta. Ulkoisia uhkia vastaan organisaatiot voivat suojautua esimerkiksi hyödyntämällä erilaisia teknisiä toteutuksia, ja laitteita. Kuitenkin sisäisten tietoturvauhkien tunnistaminen onkin usein haastavampaa. (Ambre & Shekokar, 2015.) Sisäisillä toimijoilla on usein oikeutetusti pääsy tietoon, heille on luotu käyttäjätunnukset ja salasanat erinäisiin järjestelmiin ja sisäiset toimijat voivat olla työssään oikeutetusti tiedon kanssa tekemisissä. Työntekijät voivat kuitenkin tahattomilla tai tahallisilla haitallisilla toimillaan vaarantaa tiedon luotettavuuden, eheyden ja saatavuuden. Työntekijät voivat asettaa tietoturvallisuuden vaaraan myös tietoturvallisuusselosteen toimien kiertämiselle, laiskuudella, huolimattomuudella, huonolla koulutuksella tai organisaation, kumppaneiden ja asiakkaiden tiedon suojamisen motivaation puutteella. (Warkentin & Willison, 2009.)

Egress (2020) antoi toimeksiantona riippumattomalle Opinon Mattersille toteuttaa tutkimuksen organisaatioiden sisäisistä tietomurroista. Tutkimuksessa selvisi, että IT-johtajien mukaan sekä tahattomat, että tahalliset tapaukset datan vaarantamisessa ovat organisaatioissa yleisiä (Kuvio 1).



KUVIO 1 Tahallinen ja tahoton tiedon vaarantaminen työntekijän toimesta (Egress, 2020)

Kuviosta voimme havaita, että työntekijät vaarantavat datan niin tahallisesti, kuin tahattomasti. Kuviossa on esitetty datan vaarantaminen tutkimusta edeltävän 12 kuukauden aikana. Tavallisempaa on, että data vaarannetaan, kuin että data olisi täysin turvassa. Tahaton datan vaarantaminen on tutkimuksen mukaan kuitenkin yleisempää, kuin tahallinen. Tahaton datan vaarantaminen johtua esimerkiksi työntekijöiden kiirehtimisestä, koulutuksen puutteesta, ymmärryksen puutteesta tai turvallisuusjärjestelmien tehokkuuden puutteesta. Tahallinen datan vaarantaminen voi olla esimerkiksi datan vieminen uuteen työpaikkaan luvattomasti, datan vuotaminen kyberrikollisille tai datan jakaminen omiin henkilökohtaisiin järjestelmiinsä. Tutkimus osoitti lisäksi haasteita datan omistajien määrittelyssä ja turvallisuuden vastuiden määrittelemisessä. IT-johtajista jopa 97 prosenttia tiedostivat, että sisäiset uhat voivat olla riski heidän organisaationsa tietoturvalle. Ylipäätään sisäisten uhkien esiintyminen organisaatioissa näyttää olevan kasvava trendi. (Egress, 2020.)

Sisäiset toimijat voivat muodostaa organisaatioille jopa suuremman riskin, kuin kaikki ulkoiset uhat yhdistettynä. Sisäiset uhat voivat myös helposti aiheuttaa organisaatioille merkittävää taloudellista tappiota. Onkin todettu, että keskiverto vaurio sisäisestä tietoturvatapahtumasta on tiettävästi suurempi, kuin ulkoisen uhan aiheuttama vaurio. Joissain tapauksissa sisäinen uhka voi aiheuttaa jopa miljoonien dollareiden haitan organisaatioille, esimerkiksi petoksen, sabotaasin tai varastettujen liikesalaisuuksien myötä. (Elifoglu, ym., 2018.) Vauriota voi syntyä myös myymällä tai luovuttamalla merkityksellistä tietoa kilpailijalle. On huomioitava, että sisäiset uhat saattavat olla organisaatioille haitallisia myös muista näkökulmista tarkasteltuna. Esimerkiksi yrityksen maine voi olla vaakalaudalla sisäisten uhkien paljastuttua. (Harkins, 2013.)

Kuten todettu, ulkoisten uhkien tunnistamiseen on kehitetty useita erilaisia työkaluja. Sisäisten uhkien tunnistaminen on kuitenkin osoittautunut haastavammaksi. (Elifoglu, ym., 2018.) Yksi keino sisäisten uhkien minimoimisessa on rajata tunnushallinnalla ja oikeuksienhallinnalla työntekijöiden pääsy erilaisiin järjestelmiin ja näin ollen järjestelmän sisältämiin tietoihin. Oikeuksienhallinnan kautta varmistetaan se, että jokaisella työntekijällä on pääsy vain järjestelmiin, joita hän tarvitsee työssään. (Ferreira & Alonso, 2013.) Kuitenkin suurissa organisaatioissa kaikkien oikeuksien pyyntöjen tarkka läpikäynti voi olla haastavaa. Lisäksi on todettu, että yksinään oikeuksien rajaaminen tietoresursseihin ei ole riittävä toimenpide sisäisten uhkien rajaamisessa. Kuitenkin oikeushallinta voidaan nähdä keskeisenä esimerkiksi tunnistettaessa valtuuttamaton resurssien käyttö. (Kandias, Mulonas, Virvilis, Theoharidou & Gritzalis, 2010.) Sisäisten uhkien tunnistamisen hankaluus johtuu suurelta osin siitä, että sisäiset uhat ovat usein saaneet oikeutetusti pääsyn tietoihin, esimerkiksi työtehtävien johdosta. Järjestelmissä tapahtuvien tapahtumien monitoroinnin on havaittu olevan yksi keskeinen toimi sisäisten uhkien havaitsemisessa. Lokitietojen on puolestaan havaittu olevan yksi tehokkaimmista työvälineistä järjestelmien monitoroinnissa. (Ambre & Shekokar, 2015.)

2.3 GDPR – Yleinen tietosuoja-asetus

Henkilötiedot ovat käsite, joka on ollut viime vuosikymmenen aikana paljon esillä, muun muassa Euroopan Unionin (EU) yleisen tietosuoja-asetuksen, GDPR:n (engl. General Data Protection Regulation) ja erilaisten tietovuotojen ja tietoloukkausten johdosta. Henkilötiedot ovat monitahoinen käsite, joka sisältää tietoa niin perushenkilötiedoista, kuin myös arkaluotoisemmasta tiedosta. Riippuen asiayhteydestä henkilötiedolla voidaan tarkoittaa monenlaista tietoa. GDPR määrittelee henkilötiedon olevan tietoa, josta yksilö voidaan joko suorasti, tai epäsuorasti tunnistaa. Luonnollisesti henkilötiedot käsittävät siis yksilöiden nimet ja esimerkiksi sähköpostiosoitteet, mutta on huomioitava, että esimerkiksi myös paikkatieto tai biometrinentieto voivat myös olla henkilötietoa. Lisäksi esimerkiksi käyttäjätunnusta, tai nimimerkkiä voidaan pitää henkilötietona. (GDPR, 2016.) Asetuksen, ja sen vaikutusten laajuuden johdosta tämän tutkielman puitteissa ei ole mahdollista käsitellä koko lainsäädäntöä, vaan tämän luvun tarkoitus tutkimuksen osana on tarjota yleinen näkemys siihen, miksi GDPR on yksi keskeisistä motivaatioista henkilötietohakujen analysoinnille.

Vuonna 2016 julkaistu Yleinen tietosuoja-asetus, GDPR toi tarkennuksia luonnollisten henkilöiden tietojen käsittelyyn. GDPR:n soveltaminen sellaisenaan alkoi Euroopan Unionin jäsenvaltioissa vuoden 2018 toukokuussa. Säädös koskee myös yrityksiä, jotka käsittelevät tietoja henkilöistä EU:ssa, riippumatta yrityksen tai tietojen sijainnista. (GDPR, 2016.) Säädöksen keskeisten tavoitteiden voidaan määrittää olevan henkilötietojen suojan ja tietosuojaoikeuksien parantaminen, vastaaminen globalisaation ja digitalisaation haasteisiin, tietosuojasäädöksien yhtenäistäminen koko Euroopan Unionissa ja edelleen edistää digitaalisten sisämarkkinoiden kehitystä. (Tietosuojavaltuutetun toimisto, 2019.) GDPR (2016) myös selkeyttää eri toimijoiden rooleja tietosuojan käsittelyssä. Säädös esittelee kaksi keskeistä profiilia (Taulukko 1) henkilötietojen hallinnassa: rekisterinpitäjän, sekä tietojenkäsittelijän.

TAULUKKO 1 GDPR keskeiset profiilit

Profiili	Selite
Rekisterinpitäjä	Päättää miten henkilötietoja käsitellään ja mitä varten niitä käsitellään
Tietojenkäsittelijä	Säilyttää henkilötiedot ja käsittelee tietoa rekisterinpitäjän puolesta

Taulukossa esitettyihin kahteen keskeiseen profiiliin kohdistuu paljon velvollisuuksia. GDPR on hyvin laaja säädös, jossa on määritelty muun muassa, että minkälaisissa tilanteissa henkilötietoja saa tarkastella ja miten organisaation tulee toimia vastaanottaessaan yksilöltä tietopyynnön koskien hänen tietojaan. Asetuksessa on myös määritelty, mikä on kunkin toimijan rooli tietojen säilytyksessä. (GDPR, 2016.) GDPR:n osalta on keskeistä huomioida, että säädös tosiaan

koskee kaikkien EU kansalaisten tietojen käsittelyä ja prosessointia, riippumatta siitä missä yritys toimii tai missä henkilötietoja säilytetään, mukaan lukien myös pilvipalveluiden hyödyntämisen datan säilytyksessä. Tämä olikin yksi oleellisimmista muutoksista, joita säädöksen voimaan tuleminen toi mukanaan. (Tankard, 2016.) Tietosuoja-asetusta ei kuitenkaan sovelleta, mikäli rekisteröity on menehtynyt tai mikäli rekisteröity on oikeushenkilö. Lisäksi asetusta ei sovelleta, mikäli tietoja käsittelevä henkilö toimii sellaisessa tarkoituksessa, joka ei kuulu hänen alaansa, liiketoimintaan tai ammattiin. (GDPR, 2016.)

2.3.1 Yksilö ja data

GDPR tavoitteet ovat hyvin yksilökeskeiset, ennen kaikkea tarkoitus on parantaa yksilöiden tietoturvaa digitalisoituneessa maailmassa. Kuten todettu tietosuojalain muutosten avulla pyritään muun muassa yhtenäistämään ja päivittämään yleisesti kaikkien EU:n jäsenmaiden yksityishenkilöiden tietoturva vastaamaan nykyajan tarpeita. (GDPR, 2016.) Myöskin yksilöiden oikeudet laajenivat säädöksen mukana. Yksilöille on ilmoitettava dataa kerättäessä toimenpiteestä, ja yksilön on annettava suostumus henkilötietojen keräykselle ja käsittelylle. Datan käsittelyllä tarkoitetaan tilanteita, joissa henkilötietoja tarkastellaan ja prosessoidaan oikeutetuin perustein. (Tankard, 2016.) Tietosuojavaltuutetun toimisto (2019) listasi yksilöiden tietosuojaoikeudet, jotka GDPR on määrittänyt:

- Oikeus tietää mitä henkilötietoja organisaatiolla on yksilöstä.
- Oikeus tietää miten ja mihin henkilötietoja käytetään.
- Oikeus virheellisten, epätarkkojen ja puutteellisten henkilötietojen korjaamiseen.
- Oikeus henkilötietojen poistamisen pyytämiseen.
- Oikeus vastustaa henkilötietojen käsittelyä.
- Oikeus pyytää henkilötietojen käsittelyn rajoittamista.
- Oikeus siirtää henkilötiedot toiseen organisaatioon.
- Oikeus olla joutumatta kohteeksi perusteetta automaattisen päätöksenteossa.

Organisaation käsitellessä yksilön henkilötietoja on yksilöllä oikeus tietoihin pääsyyn. Lisäksi yksilöllä on oikeus tietää tietojenkäsittelystä, kuten mihin tietoja hyödynnetään. (Tietosuojavaltuutetun toimisto, 2019.) GDPR toi mukanaan myös yksilöille oikeuden tulla unohdetuksi. Tämä tarkoittaa sitä, että mikäli yksilö ilmaisee tahtonsa olevan hänen tietojensa poistaminen, tulee datan kerääjien ja prosessoijien pystyä poistamaan kaikki tiedot, jotka eivät enää ole tarpeellisia säilyttää. Tämä aiheuttaa myös organisaatioille omat haasteensa, sillä yksilön unohtaminen edellyttää sitä, että organisaatiolla on täsmällisesti tiedossa missä kaikkialla dataa oikeastaan on yksilöstä säilössä. (Tankard, 2016.)

2.3.2 GDPR ja organisaation sisäiset uhat

GDPR asettaa myös vaatimuksia organisaatioiden teknologisille kontrollitoimille. Hyvänä periaatteena pidetään myös mahdollisimman vähäisen datamäärän keräämistä yksilöistä. Periaatteen tarkoituksena on varmistaa, että henkilöistä kerätään vain välttämättömät tiedot. Tällä pyritään myös helpottamaan datan hallinnointia ja varmistamaan, että dataa käytetään vain varsinaiseen tarkoitukseen, jota varten se on kerätty. Lisäksi on nostettu esille, että dataa tulisi säilyttää ja siirtää vain salattuna sen turvaamiseksi. On kuitenkin tunnistettu, että datan salaaminen yksinään ei riitä, sillä se auttaa turvaamaan dataa vain ulkoisilta uhkatekijöiltä. Organisaatioilta edellytetään, että heidän säilyttämäänsä dataan asetetut pääsykontrollit ja säätelyt suojaavat dataa silloin, kun se ei ole salatussa tilassa. Työntekijöillä, joilla ei ole roolinsa puolensa tarvetta tiettyyn tietoon, ei tulisi päästä siihen käsiksi ja järjestelmät, joiden ei tarkoitusperänsä perusteella kuuluisi sisältää tietynlaista dataa, ei sitä sisältäisi. Käyttäjien ja järjestelmien auditoinnilla voidaan pyrkiä monitoroimaan ja seuraamaan niin sanottuja organisaation sisäisiä uhkatekijöitä. (Tankard, 2016.)

Tietoturvaloukkauksen sattuessa organisaatiolla on velvollisuus ilmoittaa tietoturvaloukkauksesta. GDPR:n mukaan loukkauksesta on ilmoitettava tietosuojaviranomaisille 72 tunnin kuluessa loukkauksen tietoon tulosta, mikäli loukkaus luo riskin yksilön oikeuksille ja vapauksille. Mikäli tietoturvaloukkauksen voidaan katsoa aiheuttavan suurta riskiä yksilölle, voi organisaatio olla velvollinen myös ilmoittamaan loukkauksesta yksilölle. (GDPR, 2016.) On tärkeää tunnistaa, että GDPR on tuonut mukanaan myös uusia tarpeita tietosuojan varmistamiseen. Tietosuoja-asetus edellyttää muun muassa tietosuojaselosteen laatimista, aina kun henkilötietoja tallennetaan järjestelmiin. (Kyberturvallisuuskeskus, 2020.)

GDPR:n näkökulmasta on tärkeää tallentaa tieto kirjautumisyrityksistä, niin onnistuneista, kuin myös epäonnistuneista, jotta voidaan seurata ketkä ovat pyrkineet käsiksi dataan (Weir, Aßmuth, Whittington & Duncan, 2017). Toisaalta on syntynyt entistä keskeisempi tarve saada talteen käyttäjien CRUD-toiminnot, jotka kohdistuvat henkilötietoon (Stan & Miclea, 2019). CRUD-termi muodostuu englannin kielen sanoista create, read, update, delete, eli luoda, lukea, päivittää ja poistaa. Käyttäjäkontrolli koskien henkilötietojen käsittelyä ei ole saavuttanut toivottua tasoa GDPR:stä huolimatta. Organisaatioiden haasteet valvoa sitä, kuka tekee mitä millä henkilödatalla ovat edelleen keskeinen kysymys. (Silva, ym., 2020.)

3 KATSAUS LOKITIETOIHIN

Teknologian hyödyntäminen organisaatioissa on arkipäivää, käytännössä teknologia ja tietojärjestelmät ovat välttämätön osa jokapäiväistä toimintaa. Tietojärjestelmissä voidaan myös säilyttää organisaatiolle ja asiakkaille arvokasta tietoa. Järjestelmiin ja tietoihin saatetaan tehdä muutoksia, tietoja saatetaan hakea ja poistaa tai virhetilanteita voi muodostua. Näiden tapahtumien monitorointi olisi käytännössä mahdotonta ilman, että tietojärjestelmissä tapahtuvat toimet kirjattaisiin lokiin. Lokin voidaan määritellä olevan tallenne tai rekisteri tapahtumista ja toimista, jotka ovat tapahtuneet. Tässä luvussa tarkastellaan yleisesti tietojärjestelmien lokitietoja, niiden käyttötarkoituksia, sekä tietojen analysointia.

3.1 Lokien käyttötarkoitukset

Tietojärjestelmissä, sovelluksissa, tietoverkoissa ja tietosisällöissä tapahtuu muutoksia ja niissä toteutetaan erilaisia toimenpiteitä. Jotta näistä tapahtumista jäisi jälki, tulisi toteutetut toimenpiteet ja tapahtumat, sekä niiden aiheuttajat kirjata lokiin aikajärjestyksessä, eli lokittaa. Lokin voidaan siis kuvata olevan tallennetta, ja lokituksen olevan itse tapahtuma, jossa tallenteen kerääminen tapahtuu. On huomioitava, että lokitus ei siis ole kerran toteutettava toimenpide, vaan se on jatkuva prosessi. (Kyberturvallisuuskeskus, 2020.) Henin, Zhunin, Henin ja Lyun (2020) mukaan voitaisiin todeta, että yleisesti lokit ovat strukturoimatonta tekstiä, jota tuotetaan lokituskomennoilla lähdekoodissa. He esittivät esimerkki lokiviestiksi seuraavanlaista muotoilua:

> 2008-11-09 20:46:55,556 INFO dfs.DataNode$PacketResponder:
> Received block blk_3587508140051953248 of size 67108864 fr
> om /10.251.42.84

Edellä esitetyssä lokiviestissä esiintyvät kentät aikaleimalle, eli tapahtuman toteutumisajalle, vakavuusasteelle, eli tapahtuman tyypille, tässä tapauksessa

INFO, sekä tapahtuman kuvaus vapaana tekstinä. (He, ym., 2020.) On kuitenkin huomioitava, että lokiviestit voivat vaihdella merkittävästi järjestelmän mukaan, sekä määriteltyjen lokiin kerättävien tietojen mukaan.

Lokin kerääminen on suhteellisen vakiintunut toimenpide. Voidaan jopa odottaa, että järjestelmä tallentaa tiedot esimerkiksi virhetilanteista ja järjestelmään kirjautumisista. (Basis, Schaller & Schläpfer, 2011.) Li (2016) totesi, että lokit ovat tosielämän hetkellisiä tapahtumia, jotka on kerätty talteen. Linin mukaan tapahtumat, jotka kirjataan lokiin kertovat muun muassa komponenttien statuksen ja toiminnot, kuten operationaaliset muutokset, turvallisuustapahtumat, sekä esimerkiksi järjestelmävirheet. Lokien välillä on eroja, eivätkä kaikki lokit ole tarkoitettu samaan tarkoitukseen. Kyberturvallisuuskeskus (2020) esitteli kuusi erilaista lokia (Taulukko 2), jotka on luokiteltu lokien muotojen, käyttötarkoituksen ja käyttötavan perusteella.

TAULUKKO 2 Lokityypit tarkoitusperän mukaan

Lokityyppi	**Selite lokitettavasta tiedosta**
Tapahtumaloki (käyttöloki)	Tarkoituksena tallentaa muun muassa käyttäjien kirjautumistietoja, sekä järjestelmän suorittamien prosessien tietoja. Järjestelmän moduulit kutsuvat toisia moduuleja, ja jättävät näin jäljen lokiin. Voidaan pitää yleisimpänä ja välttämättömimpänä lokimuotona.
Ylläpitoloki	Tarkoituksena ylläpitää tietoa esimerkiksi järjestelmän toiminnan ja käyttöoikeuksien muutoksista. Voidaan hyödyntää myös virhetilanteiden sattuessa. Keskeinen kokonaisarkkitehtuurin ja versionhallinnan osalta.
Muutosloki	Tarkoituksena tallentaa järjestelmässä tapahtuneet muutokset. Esimerkiksi tietojen lisäykset, muutokset ja poistot. Keskeinen loki muutosten oikeellisuuden varmentamisessa.
Virheloki	Tarkoituksena tallentaa virheiden syyt lokiin mahdollisimman yksityiskohtaisesti. Keskeinen virhetilanteiden ratkaisemisessa ja virheiden aiheuttajien korjaamisessa.
Viestintäloki	Tarkoituksena tallentaa tiedot järjestelmässä kulkeneesta viestinnästä. Lokiin tallennetaan esimerkiksi viestin alkuperä, kohde, sekä ajankohta. Esimerkiksi teletunnistetiedot ovat viestintälokitietoa ja useat sähköpostipalvelimet kirjaavat viestintälokia.
Haltijaloki	Tarkoituksena tallentaa tiedot, jotka kertovat kenelle esimerkiksi nettiosoite, puhelinnumero tai verkkodomain on kuulunut tietyllä ajanhetkellä.

On tärkeää huomioida, että erilaisia lokeja on todella paljon, ja kaikkia lokityyppejä ei ole siis listattu edeltävässä taulukossa. Taulukossa on kuitenkin esitetty tyypillisimpiä lokityyppejä, ja taulukon tarkoituksena voidaan pitää sitä, että sillä kyetään antamaan selkeä kuva lokien monipuolisuudesta ja siitä, että lokeista ja niiden keräämisestä puhuttaessa voidaan tarkoittaa hyvin montaa eri näkökulmaa ja tarkoitusperää. Lokit voivat toimia tärkeänä työkaluna niin tietojärjestelmien toimivuuden varmistamisessa, käytön tilastoinnissa, kuin tietoturvan takaamisessa. (Viestintävirasto, 2016.) On selvää, että tietoja lokitetaan erilaisten tarkoitusperien ja tavoitteiden perusteella, kuten Taulukko 1:ssä on esitetty. Esimerkiksi virhelokin on tarkoitus kerätä tietoa järjestelmässä esiintyneistä virheistä, ja sitä lokia hyödynnetäänkin eri tilanteessa, kun esimerkiksi käyttölokia. (Kyberturvallisuuskeskus, 2020.) Lokitapahtumien tyyppien jakoja voidaan tehdä eri perusteiden mukaan. Esimerkiksi Marty (2011) jakoi lokituksen tarpeet pilvipalveluiden osalta neljään eri tarkoitusperäiseen kategoriaan:

- Liiketoimintaan liittyvät lokit
- Operationaalis pohjaiset lokit
- Turvallisuuteen liittyvät lokit
- Regulaatioiset ja standardien mukaiset lokit

Liiketoimintalokien avulla voidaan seurata ominaisuuksien käyttöä ja erilaisia liiketoimintamittareita. Operationaalisten lokien avulla voidaan puolestaan seurata esimerkiksi järjestelmien virheitä ja muutoksia. Turvallisuuteen liittyvien lokien avulla voidaan seurata turvallisuuteen liittyviä tapahtumia, kuten kirjautumisyrityksiä ja salasanojen vaihdoksia. Regulaatioiset ja standardien mukaiset lokit voivat puolestaan tarkoittaa esimerkiksi joidenkin tiettyjen alojen osalta sähköistentapahtumien seuraamista, kuten maksuliikenteen tapahtumien. On mahdollista, että kategorioiden sisäiset lokit voivat myös mennä päällekkäin ja palvella useampaa tarkoitusperää. (Marty, 2011.) Lokien tarkoitusperät ja hyödyntämisen muodot vaihtelevat siis merkittävästi. Osaa lokeista voidaan käyttää ongelmien ratkaisuun ja turvallisuuden takaamiseen. Kun taas esimerkiksi verkkoserverin lokia voidaan käyttää tietoliikenne trafiikin tutkimiseen, joka edelleen voi tuottaa hyödyn esimerkiksi markkinoinnissa. (Oliner, Ganapathi & Xu, 2012.) Yhteistä erilaisille lokeille on se, että niiden avulla tapahtumista ja toiminnoista, sekä niiden käsittämisestä voidaan tehdä avoimia, sekä kokonaisvaltaisia (Basis, Schaller & Schläpfer, 2011). Viestintäviraston (2016) mukaan ilman lokitietoja, esimerkiksi virhetilanteiden syiden selvittäminen olisi käytännössä mahdotonta ja näin ollen myös virhetilanteiden korjaaminen olisi haastavaa.

On tärkeää huomioida, että lokien käyttötarkoitukset ja muodot voivat myös muuttua, muuttuvien vaatimusten ja tarpeiden myötä. Esimerkiksi on todettu, että kasvavien turvallisuusuhkien takia turvallisuuslokien (eng. Security logs) tarve on lisääntyneet merkittävästi (Söderström & Moradian, 2013). Turvallisuuslokeihin tallennetaan yleisesti ottaen turvallisuuteen liittyviä tapauksia. Turvallisuuslokit voidaan generoida useasta lähteestä, sillä useat järjestelmät

sisältävät turvallisuuden kannalta oleellista tietoa. Esimerkiksi turvallisuuteen liittyvien järjestelmien ja palomuurien lokien sisältämät tiedot voivat olla keskeisiä. (Kent & Souppaya, 2006.) Yksi yleisesti tunnettu esimerkki turvallisuuslokista on Microsoft Windowsin turvallisuusloki. Tähän lokiin kirjataan tapahtumat, joiden katsotaan liittyvät turvallisuuteen tapahtumienvalvonnan määritysten mukaan, kuten esimerkiksi kirjautumisyritykset. Lokitettavia tietoja voidaan muotoilla myös tässä tapauksessa. Esimerkiksi organisaatiot, jotka käyttävät Microsoft Windowsia voivat itse määrittää auditoitavat tapahtuvat luoden auditointi käytänteen. Määritettyjä käytänteitä sovelletaan näin kaikkiin organisaation käytänteen piirin laitteisiin. (Microsoft, 2017.)

Kuten tässä luvussa on esitetty, lokityyppejä on informaatioteknologiassa monenlaisia ja lokeja on moneen eri tarkoitusperään. Keskeistä kuitenkin lokituksessa on huomioida se, että lokitiedot ovat oikeastaan ainoa keino selventää jälkikäteen mitä eri järjestelmissä ja ohjelmistoissa on tapahtunut, oli kyseessä sitten virhetilanne tai käyttäjän toteuttama tapahtuma. Lisäksi keskeistä on tunnistaa, että lokitus on jatkuva prosessi, jota tulisi säännöllisesti auditoida.

3.1.1 Lokitietojen laadun merkitys

Lokitiedot tarjoavat arvokasta informaatiota organisaatioille. On kuitenkin keskeistä huomioida, että mikäli lokiin ei kerätä oikeita asioita, voivat lokitiedot olla jopa hyödyttömiä. Usein organisaatiot ja järjestelmän kehittäjät, sekä ylläpitäjät voivat säädellä ja kehittää lokitettavia tietoja ja tapahtumia. Yleisesti voitaisiin sanoa, että on tärkeää, että lokitetuilla tiedoilla pystytään suorittamaan lokien auditointia, eli arviointia tai analysointia. Jotta auditointia voitaisiin suorittaa, on tärkeää, että tapahtumista on olemassa luotettava tallenne. (Roratto & Dias, 2014.) On kuitenkin todettu, että mikäli lokitetaan kaikki tapahtumat, tai suuri osa niistä, tallentuu huomattavasti merkityksetöntä ja tarpeetonta dataa. Tämä voi puolestaan edelleen sotkea Audit trailin eli tapahtumaketjun ja hankaloittaa analysointia ja järjestelmänylläpidon mahdollisuutta havaita anomaalisia tapahtumia. (King, Pandita & Williams, 2015.) Lisäksi on vaarana myös toisenlainen tilanne, jossa tietoja lokitetaan liian vähän ja yksityiskohdat puuttuvat. Tietojen vähyys voi myös asettaa haasteita lokitietojen analysoinnille ja anomaalisten tapahtumien havaitsemiselle. (Jayathilake, 2012.) Ylipäätään monissa organisaatioissa lokeilla saattaa olla useita eri lähteitä, epäjohdonmukaisuutta saattaa esiintyä lokien sisällössä, aikaleimoissa, sekä lokiformaateissa (Kent & Souppaya, 2006).

Mikäli lokitusmekanismien konfigurointi ei ole ollut käytännössä optimaalista johtaa se usein suureen määrään tarpeettomia lokitapahtumia, tärkeiden lokitapahtumien seassa. Tällaisessa tilanteessa on suuri mahdollisuus siihen, että keskeiset ja tärkeät tapahtumat lokissa jäävät huomaamatta. (Basis, Schaller & Schläpfer, 2011.) Lokeihin ei ole esitetty tiettyä standardisoitua formaattia. Tämä aiheuttaa myös haasteita, sillä lokitietojen formaatit voivat erota huomattavasti toisistaan. Lokitiedostojen formaatit voivat myös ajan kuluessa muuttua ja

kehittyä, mikä saattaa aiheuttaa lisähaasteista. (Jayathilake, 2012.) Lokitietoja yhdistellään usein useasta eri lokilähteestä, ja lokitietoja parsitaan. Mikäli järjestelmissä, jotka lokeja tuottavat on suuria eroja tallennetun tiedon välillä, voi lokien parsiminen olla haasteellinen prosessi, joka saattaa tuottaa puutteellista tietoa. Lisäksi lokitietojen formaateissa voi olla eroavaisuuksia. Esimerkiksi päivämäärät voivat olla eri muotoisia, mikä tekee lokien yhdistelemisestä monimutkaisempaa. (Kent & Souppaya, 2006.) Lokituksen tulisi siis olla suunniteltu prosessi, jossa otetaan jo suunnitteluvaiheessa huomioon se, missä, miten ja kuinka kauan tietoja tulisi säilyttää. Lisäksi lokituksen suunnittelussa on tärkeää arvioida, tarvitaanko muita turvallisuustoimia, kuten lokitietojen salausta tai varakopioiden luomista. Hyvä laatuisella lokilla on paljon käyttötarkoituksia turvallisuuden ja organisaation toiminnan näkökulmasta, esimerkiksi operatiivisten ongelmien tai kyberhyökkäysten havaitseminen, hyökkääjien toimintatapojen tunnistaminen, hyökkäysten vaikutusten analysointi, sekä hyökkäyksen alkuperän tunnistaminen. (Basis, Schaller & Schläpfer, 2011.)

Lokituksen kohdistamisen ja rajaamisen haasteiden lisäksi myös muut asiat vaikuttavat lokitietojen laatuun ja käytettävyyteen. Lokitietojen luotettavuus on yksi keskeisimpiä kysymyksiä, jotka liittyvät lokien hallintaan. Lokitietojen säilytyksen ja siirron tulisi olla suojattu niin, että kukaan ei pääse muuntelemaan dataa valtuuttamattomasti. On mahdollista, että esimerkiksi kyberhyökkääjillä on motiiveja muokata lokidataa, esimerkiksi piilottaakseen hyökkäyksen jäljet. Lokitietoja voidaan hyödyntää myös rikostutkinnassa, jolloin luotettavuus on erityisen kriittistä. (Accorsi, 2009.) Lokien korruptoituminen ja epäjohdonmukaisuus ovat myös haasteita, jotka vaikuttavat lokitietojen luotettavuuteen (Jayathilake, 2012). Lokien turvallisuus on myös keskeinen kysymys niiden sisältämien arkaluontoisten tietojen takia. Mikäli lokille esimerkiksi tallentuu tahattomasti käyttäjän salasana, voi se aiheuttaa tietoturvariskin, niin valtuutetun, kuin valtuuttamattoman lokitietoihin pääsyn osalta. (Kent & Souppaya, 2006.)

Marty (2011) painotti lokien hallinnan (eng. Log Management) tärkeyttä keskeisiin lokien hyödyntämisen haasteisiin liittyen. Lokien asianmukaisen hallinnan avulla voidaan auttaa niin suojaamaan lokitiedostoja, kuin myös kohentamaan lokien laatua. Hän nosti esiin pilvipalveluiden lokien hallinnan toimenpiteinä esimerkiksi Audit Trailin, data-analytiikan, sekä lokien keskittämisen. Lokien hallinta pitää sisällään kuitenkin myös useita haasteita. Esimerkiksi on todettu, että useilla organisaatioilla on haasteita tasapainottaa jatkuvasti kasvavien lokimassojen ja lokien hallinnan resurssien välinen suhde. Toinen keskeinen asia lokien hallinnoimisessa, on taata lokien saatavuus. Lokien maksimikoot voivat tulla vastaan suurien lokimassojen osalta. Velvoitteet Data Retention-tietokantaan voivat myös pidentää tarvetta tietojen säilytykseen, mikä voi poiketa lokien alkuperäisestä säilytysajasta. Tämä voi aiheuttaa haasteita lokitukseen ja lokidatan käsittelyyn, mikäli tietojen asianmukaisesta säilytyksestä ja tallennuksesta ei ole huolehdittu. (Kent & Souppaya, 2006.)

Kyberturvallisuuskeskus (2020) selvensi, että lokiselosteesta tulisi selvitä, mihin tarkoitukseen tietoja kerätään ja kuka tietoja voi käyttää. Kyberturvallisuuskeskus korosti myös sitä, että lokitietojen keräämiseen on oltava

asianmukainen peruste. Lisäksi olisi huomioitava se, että lokituksen kertaalleen suunnittelu ei riitä, vaan lokien turvallisuutta ja asianmukaisuutta tulisi auditoida säännöllisesti. Eli voidaan todeta, että kerran hyväksi ja toimivaksi suunniteltu lokiprosessi ei välttämättä ole lopullinen ratkaisu, vaan säännöllisesti kehitettävä. Auditoinnilla voidaan varmistua siitä, että lokitus vastaa edelleen sille esitettyjä vaatimuksia ja tarpeita, jotka myös voivat ajan kuluessa muuttua. Kent ja Souppaya (2006) totesivat, että organisaation johdon tulisi osoittaa tukensa lokituksen toteutukselle, jotta lokien hallinta voitaisiin taata olevan tehokasta jokaisella organisaation osa-alueella. Tuen tulisi sisältää niin tiedon jakamista, kouluttamista, yhteyshenkilöiden asettamista, teknologista ohjeistusta, kuin työkalujen ja dokumentaation tarjoamista.

3.1.2 Lokitiedot ja lakivelvoitteet

Tarkastelemalla erilaisia useita lokityyppejä, voidaan tehdä toteamus, että lokeja kerätään ja käytetään monen eri tarkoitusperän ja tarpeen perusteella. Lokeja voidaan pitää jopa järjestelmän tai laitteen oletusarvoisena ominaisuutena. Lokien hyöty organisaation kannalta on kiistämätön ja luo luonnollisesti perustaa sille, että lokitietoja tulisi kerätä. Sen lisäksi, että lokit tarjoavat organisaatioille arvokasta tietoa, lokitusvelvollisuuden voidaan katsoa tulevan myös Suomessa sovellettavista laista. Myöskin lokien käsittelyvaatimukset perustuvat useaan eri säädäntöön. (Kyberturvallisuuskeskus, 2020.)

GDPR (2016) 5 artikla, joka säätää henkilötietojen käsittelyä koskevista periaatteista velvoittaa organisaatiot varmentumaan henkilötietojen käsittelyn luvallisuudesta, sekä tietojen eheydestä ja luottamuksellisuudesta. Myös GDPR:n 25 artikla, joka säätää rekisterinpitäjän vastuusta ja artikla 32, joka säätää käsittelyn turvallisuudesta voidaan nähdä olevan vankka peruste lokitietojen keräämiseen. Voidaan myös katsoa, että lokitietojen analysoinnin velvoite tulee laista. Asetus ei kuitenkaan suoraan velvoita lokitietojen analysoimiseen, mutta se velvoittaa valvomaan henkilötietojen käsittelyä ja varmistumaan siitä, että vain käsittelyn kannalta oleellisia henkilötietoja käsitellään.

GDPR:n lisäksi muun muassa Tietosuojalaki (5.12.2018/1050) asettaa käsittelyvaatimuksia lokitiedoille. Myös lakien Yksityisyyden suojasta työelämässä (13.8.2004/759) ja Sähköisen viestinnän palveluista (7.11.2014/917) voidaan katsoa asettavan lokien käsittelyyn erilaisia vaatimuksia. (Kyberutrvallisuuskeskus, 2020.) Asetukset tai laki ei suoraan määrää tai kerro kokonaisuutena ja yksityiskohtaisesti millaisella tasolla lokitietojen keräämisen tulee organisaatioissa olla, tai mitä kaikkia tietoja tulisi kerätä. Laki kuitenkin asettaa joitakin vaatimuksia lokien sisällölle, säilytysajalle, tiedon eheydelle, sekä käyttötarkoituksille. (Viestintävirasto, 2016.)

Lisäksi on huomioitava, että toimialakohtaiset säädökset voivat myös vaikuttaa lokitietojen keräykseen. Esimerkiksi terveydenhuollossa, teleoperaattoreiden toiminnassa ja pankkitoiminnassa on velvoitteita, jotka eroavat muista toimialoista lokitietojen suhteen. Lisäksi on huomioitava, että moni muukin lainsäädäntö koskettaa lokitietojen keräämistä ja lainsäädännöt voivat olla tässä kontekstissa hyvin paljolti maa kohtaisia.

3.2 Lokianalytiikka

Lokitiedostot pitävät sisällään paljon informaatioita. Usein tärkeät ja merkitykselliset lokikirjaukset saattavat jäädä suuren lokimassan seassa huomaamatta. Joskus myöskään yhden lokin tiedot eivät ole riittäviä, vaan tietoja on yhdisteltävä useasta lähteestä, jotta syntyisi selkeä kokonaiskuva tapahtumien kulusta. (Basis, Schaller & Schläpfer, 2011.) Lokien sisältämä tieto voi olla merkityksellistä tietoa järjestelmänhaltijoille, kuin myös järjestelmän ylläpitäjille. Lokien analysoinnin nähdäänkin olevan keskeinen osa organisaatioiden verkkojen ja järjestelmien hallintaa. Lokien avulla voidaan esimerkiksi hahmottaa järjestelmän nykytilannetta, sekä ymmärtää erilaisia turvallisuustapahtumia, jotka tapahtuvat järjestelmän sisällä, kuten kirjautumisyritykset ja käyttäjien suorittamat toimenpiteet. (Söderström & Moradian, 2013.) Lokitietojen tehokasta analysointia voidaan pitää lokitietojen hallinnan vaativimpana, mutta myös tärkeimpänä osa-alueena. Valtionhallinnon tietoturvallisuuden johtoryhmän koostama VAHTI (2009) lokiohjeen mukaan organisaation tulisi lokitietojen analysointia suunnitellessa määrittää ainakin seuraavat asia:

- Miten usein lokitietoja tulee analysoida.
- Mitä lokitietoja tulee analysoida.
- Kenellä tulee olla pääsy lokitietoihin, ja kuinka lokien käsittely tulee kirjata.
- Miten toimitaan, mikäli lokitiedoissa havaitaan poikkeama.
- Kuinka käsitellään luottamuksellisen tiedon tahaton paljastuminen.

Yllä esitetyt lokien analysoinnin raamit edistävät analysointiprosessia. Lokien analysoinnin tulisi olla säännöllistä, mikä myös mahdollistaa ajan kuluessa järjestelmän normaalien tapahtumien määrittelyn, mikä puolestaan edistää epätavallisten tapahtumien tunnistamista. Lokien analysoinnissa voidaankin pitää keskeisenä järjestelmien normaalin toiminnan tuntemusta. Lokitietojen täysin manuaalinen analysointi voi osoittautua organisaatioille erittäin työlääksi, ellei jopa mahdottomaksi. Lokianalyysissa onkin keskeistä pyrkiä mahdollisimman automatisoituun toimintatapaan, joka mahdollistaa haitallisten toimintojen tunnistamisen. (VAHTI, 2009.)

Duncan ja Whittington (2016) totesivat, että yksinään tehokkaan Audit Trailin omaaminen, ei ole riittävää. Organisaation on osattava hyödyntää tapahtumaketjua ja ymmärtää sen todelliset hyödyt analysoimalla sitä tehokkaasti ja säännöllisesti. Keskeistä on myös lokitietojen analysoinnin tarvesidonnaisuus (VAHTI, 2009). Lokianalytiikassa on keskeistä pystyä tapahtumia tutkiessa vastaamaan kysymyksiin: milloin, mitä, kuka ja miksi. Lokiin tulisi tallentaa muun muassa aikaleima, järjestelmä, sekä käyttäjätunnus. Tallentuvien tapahtumien

aikaleima kertoo milloin mikäkin tapahtuma on toteutunut. Järjestelmätieto puolestaan kertoo, mistä tietojärjestelmästä lokitapahtuma on peräisin. Käyttäjätietoon tulisi tallentua kuka toimen on toteuttanut. Tämän kentän osalta olisi merkityksellistä, että käyttäjät ovat yksilöitävissä käyttäjänimensä perusteella. (Marty, 2011.) Lokitietojen säilyttäminen riittävän mittaisen ajan on keskeistä, jotta niitä voidaan analysoida myös myöhemmin tarpeen tullen. Esimerkiksi tietoturvapoikkeamien havaitseminen lokianalytiikalla voidaan toteuttaa niin reaaliajassa, kuin myös jälkeenpäin. (Viestintävirasto, 2016.) Tapahtumien kulun tunnistamisen lisäksi lokianalytiikka voi auttaa myös tapahtumien vahinkojen korjaamisessa. Esimerkiksi Viestintävirasto (2016) esitti, että lokien analysointia voidaan hyödyntää vahinkojen korjaamisessa yhdistämällä usean eri lokin tietoja. Esimerkiksi pääsynvalvontalokista voidaan selvittää järjestelmään kohdistuneet murtautumiset ja muutoslokin avulla puolestaan voidaan korjata murron kohteena olleet tiedot.

Kuten todettu, lokitietojen analysointia voidaan pitää lokien hallinnan vaativimpana osa-alueena. Lokitietojen analysointimetodit ovatkin hyvin organisaatio kohtaisia. Lokitietojen manuaalisen analysoinnin tehostamiseksi on havaittu tarve työkaluille, joilla voidaan suorittaa suuren luokan monitorointia. Tällainen monitorointi puolestaan mahdollistaa edelleen tarkemman analysoinnin ja tapahtumien tutkimisen. (Legg, Buckley, Goldsmith & Creese, 2015.) Kehitetyt lokien analysointijärjestelmät voidaan jakaa karkeasti kahteen eri tyyppiin: Määritetyin väliajoin lokitetoja hakeviin Offline-järjestelmiin, sekä reaaliajassa lokitietoja hakeviin Online-järjestelmiin. Offline-analysointi työkaluja ovat esimerkiksi logwatch ja SLAPS. Online-analysointi työkaluja puolestaan ovat esimerkiksi LoGS ja Splunk. (Ambre & Shekokar, 2015.) Järjestelmät mahdollistavat ja helpottavat suurempien lokimassojen analysointia useasta eri lähteestä. Järjestelmät myös tekevät lokidatasta helppolukuisempaa. Lisäksi esimerkiksi Splunk mahdollistaa erilaiset visuaaliset esitystavat järjestelmässä olevalle datalle, kuten esimerkiksi kaavioiden ja kuvioiden luomisen datasta. Lisäksi Splunk mahdollistaa myös datan vertailun, esimerkiksi kuukausien välillä, sekä tietyn tyyppisten tapahtumien etsimisen. (Carasso, 2012.)

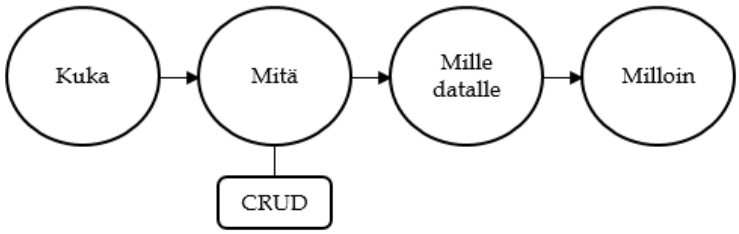
3.2.1 Audit Trail lokitiedoissa

Usein kuullaan puhuttavan lokitapahtumien yhteydessä Audit Trailistä tai Audit lokista. Audit Trailin juuret juontavat taloushallintoon. Audit Trailillä kuvattaan taloushallinnon yhteydessä kirjauksiin tehtyjä muutoksia. (Jiang & Cao, 2011.) Informaatioteknologian yhteydessä Audit Trailillä tarkoitetaan toimia, jotka muodostavat katkeamattoman tapahtumaketjun tai kirjausketjun. Voidaan katsoa, että Audit loki esittää aikajärjestyksessä selkeästi tapahtumaketjun, joka on koostettu usean muun lokin tapahtumista. Tapahtumat esitetään siis kronologisessa järjestyksessä tapahtumaketjussa. Yksittäinen tapahtumaketju keskittyy usein yhteen käyttäjään, tai turvallisuustapahtumaan. Audit lokia pidetäänkin tärkeänä työkaluna esimerkiksi tietovuotojen sattuessa. (Chapple, 2020.)

Voidaankin pitää lähes välttämättömänä, että järjestelmät, jotka tallentavat tai joihin on tallennettu kriittistä dataa, tuottavat jonkin tasoista Audit lokia. Audit lokit auttavat tunnistamaan niin sisäisiä, kuin ulkoisia, sekä tarkoituksellisia, että tarkoituksettomia haitallisia toimintoja, jotka ovat tapahtuneet tietojärjestelmissä. (Roratto & Dias, 2014.)

Audit lokiin kirjattaviin tapahtumiin kirjataan usein tapahtumat toteuttajan jonkinlainen tunnus, esimerkiksi käyttäjänimi. Myöskin keskeistä tietoa on kirjata lokiin tapahtuman tyyppi ja tapahtumassa esiintyneet parametrit. (Lee, Zhang & Xu, 2013.) Yleisesti voitaisiin todeta, että Audit Trailistä tulisi selvitä tiedot siitä kuka toteutti toimen, minkä tyyppinen toimi oli, mille kohteelle toimi kohdistui ja minä ajankohtana tapahtuma toteutettiin (Olympia, 1994).

Lokituksen raameja punnittaessa kohdataan haaste siinä, millaisten tapahtumien tapahtumaketjuja tulisi lokittaa. Useat akateemiset, regulatiiviset, sekä ammatilliset ohjeistukset ovat suositelleen ohjenuoraksi CRUD-operaatioita. Voidaan pitää tietynlaisena oletuksena, että jos tapahtuma sisältää jonkin tiedon luomista, lukemista, päivittämistä tai poistamista, on siitä jäätävä lokitietoihin jälki. (King, Pandita & Williams, 2015.) Yhdistäen Leen, Zhangin ja Xun (2013, Olympian (1994) ja Kingin, Panditan ja Williamsin (2015) näkemykset on muodostettu yksinkertaistettu esimerkki siitä, mistä tekijöistä Audit Trail voi muodostua (Kuvio 1). Kuviossa esitetään, että Audit trailissä keskeistä tietoa on kuka, esimerkiksi käyttäjä, teki minkä CRUD:n mukaisen toimen, mille datakohteelle toimi toteutettiin ja mille ajankohdalle toimi sijoittuu.



KUVIO 2 Esimerkki Audit Trailistä (Lee, ym., 2013, Olympian, 1994, King, ym., 2015.)

Audit Trail voi esittää selkeästi tapahtumaketjun, jonka avulla voidaan selvittää, mitä tietyssä tapahtumassa on todellisuudessa tapahtunut. Kuten todettu, kyberhyökkäyksen tai tietoturvaloukkauksen sattuessa Audit lokeja voidaan pitää hyvin keskeisenä osana hyökkäysten juurisyyn tunnistamista lokianalytiikan keinoin. Audit lokeihin perustuvassa analyysissa on kuitenkin haasteensa. Yksi keskeisistä haasteista on lokitiedostojen suuri määrä, jolloin merkityksellisten tapahtumaketjujen tunnistaminen voi olla haasteellista. (Lee, ym., 2013.)

Duncan ja Whittington (2016) tunnistivat, että vaikka hyvin spesifioitu Audit Trail on voimakas ja simppeli työkalu kyberrikollisuutta vastaan, kuitenkin usein sen hyödyt sivuutetaan. Sivuutuksen syyksi he esittivät niin ymmärryksen, kuin pätevyyden puutetta. Myöskin väärin konfiguroidut lokit ja joissain tapauksissa jopa laiskuus asian suhteen voivat johtaa Audit Trailin hyötyjen sivuuttamiseen. He tunnistivat keskeiseksi ongelmaksi myös Audit lokien

asianmukaisen säilytyksen haasteet. Eli voimme todeta, että pelkästään lokien ja Audit Trailien olemassaolo ei ole riittävää, vaan tallennetun datan tulee olla sellaista, että sitä osataan ja voidaan hyödyntää turvallisuuskentän varmistamisessa.

3.2.2 Lokianalytiikan merkitys organisaatiolle

Lokitiedostot tarjoavat valtavasti informaatiota organisaatioille, mukaan lukien tietoa järjestelmien virhetilanteista, sekä tietomurroista. Niin tietojärjestelmävirheet, kuin tietomurtotilanteet voivat aiheuttaa organisaatioille ja kehittäjille merkittäviä kustannuksia, sekä kuluttaa arvokasta aikaa. (Das, ym., 2020.) Iso-Britannian kyberturvallisuuskeskus (2018) on korostanut lokitietojen merkitystä kyberturvallisuuteen liittyvien tapahtumien selvittämisessä. He korostivat lokituksen olevan perusta turvallisuuden monitoroinnille ja tilanteiden käsittämiselle. Heidän mukaansa lokitiedot voivat tarjota kyberturvallisuusloukkauksen sattuessa vastauksen ja tukea keskeisiin kysymyksiin. Tällaisia kysymyksiä ovat esimerkiksi mitä on tapahtunut, mitkä tapahtuman vaikutukset ovat, mitä tulisi tehdä seuraavaksi, toimivatko korjaustoimenpiteet ja toimivatko turvallisuustoimenpiteet ylipäätään.

King ja Williams (2014) mukaan analysoimalla tietomurtoja lokien kautta, voidaan saada selville: kuka teki mitä, milloin tapaus sattui ja missä, sekä miten tietomurto oikeastaan tapahtui. Heidän mukaansa lokitusmekanismeilla voidaan myös ennaltaehkäistä uhkia, jotka liittyvät esimerkiksi käyttäjiin, jotka kieltävät suorittaneensa joitakin toimia järjestelmän sisällä. Näitä toimia ei voitaisiin todistaa ilman lokidataa. Lisäksi he nostivat esille, että sopivilla lokitusmekanismeilla voidaan vahvistaa järjestelmän kykyä tunnistaa väärää kiistämistä, sekä käyttäjien vastuuta toiminnastaan. On myös tutkittu, että auditointi lisää yksilön halua toimia hyväksyttyjen tapojen mukaisesti, joka edelleen vähentää yksilön toimia, jotka ovat turvallisuusohjeiden vastaisia. (King, ym., 2015).

Sisäisten uhkien on havaittu olevan yksi merkittävimmistä uhista organisaatioiden tietoturvallisuudelle. Kuitenkin samaan aikaan sisäiset uhat ovat myös hankalimpia uhkia havaita. Tämä johtuu suurilta osin siitä, että organisaation sisäisillä toimijoilla on tietämys organisaation turvallisuustoimista. Sisäiset toimijat voivat helposti ohittaa tai muuten toimia niin, että heidän toimintaansa ei organisaation asettamien turvallisuustoimien rajoissa pystytä havaitsemaan. (Myers, Grimaila & Mills, 2009.) Sisäisellä uhalla voidaan viitata niin työntekijöihin, kuin myös muihin liiketoiminnan sidosryhmiin, joilla on ollut tai on edelleen oikeutettu pääsy tietoon. Jatkuva tietojärjestelmien tapausten monitorointi lokianalytiikan keinoin onkin välttämätöntä, jotta näihin uhkiin voidaan pureutua. (Ambre & Shekokar, 2015.)

Useat organisaatiot hyödyntävätkin sisäisten uhkien analysoinnissa juuri lokitietoja (Roy, Sengupta & Mazumdar, 2021). Esimerkiksi King ja Williams (2014) esittivät että terveydenhuollossa suurin osa tietoturvaloukkauksista syntyi työntekijöiden toimesta, kun he valtuuttamattomasti hakivat muiden terveystietoja. King ja Williams (2014) myös esittivät, että lokituksen ja lokitietojen analysoinnin avulla tällaisia tapauksia voidaan paremmin havaita ja osoittaa todeksi.

He totesivat, että lokien auditoinnilla voidaan proaktiivisesti tunnistaa valtuuttamattomia hakuja, kuin myös reaktiivisesti jäljittää käyttäjien toimet liittyen tietoturvaloukkauksiin.

Lokianalytiikka voi tarjota organisaatiolle tietoturvatapahtumien lisäksi myös muunlaista arvokasta informaatiota. Myös järjestelmänhaltijat saavat arvokasta tietoa lokianalytiikasta, esimerkiksi teknologioiden monitorointia ja hallintaa varten. (Hamooni, Debnath, Xu, Zhang, Jiang & Mueen, 2016.) On myös esitetty, että lokitietoja analysoimalla voidaan myös tutkia järjestelmän tulevia tapahtumia. Tämä mahdollistaa korjaavien toimien toteuttamisen, ennen haitallisen tapahtuman toteutumista. (Das, ym., 2020.)

Kuten aiemmin todettu, GDPR:n näkökulmasta tapahtumalokien kerääminen on tärkeää tietoturvaloukkausten tunnistamiseksi. Kuitenkin GDPR:n näkökulmasta tarkasteltuna, voidaan todeta lokien keräämisen olevan muutenkin organisaatioille hyödyllistä. Lokitietoja hyödyntäen voidaan suorittaa henkilötietoja koskettavien monimutkaisten prosessien tehokasta auditointia ja kehittämistä. Tällaisen auditoinnin ja kehittämisen ansiosta prosessit asettuvat paremmin tietosuoja-asetuksen vaatimuksiin. (Gonçalves-Ferreira, Leite, Santos-Pereira, Correia, Antunes & Cruz-Correia, 2018.) Kuitenkin kasvavien lokimassojen johdosta täysin manuaalinen lokianalytiikka on käymässä resursseja kuluttavaksi, sekä aikaa vieväksi. Lokianalytiikan toteutukseen ja automatisointiin on tutkittu tekoälyn hyödyntämisen mahdollisuuksia. (He, ym., 2020.)

Voidaan todeta, että lokit tarjoavat organisaatioille valtavasti arvokasta tietoa, jonka hyödyllisyys nousee tosissaan esiin vasta lokianalytiikan myötä. Alun perin lokeja on hyödynnetty lähinnä järjestelmävirheiden tunnistamisessa ja jäljittämisessä. Kuitenkin nykyisin lokeja hyödynnetään organisaatioissa laajemmalla skaalalla. Rutiininomainen lokianalytiikka tarjoaa mahdollisuuden havaita turvallisuuspoikkeamia, käytänteiden vastaista toimintaa, petollista toimintaa, sekä operationaalisia ongelmia. Lokitiedot auttavat myös auditoinnissa, sekä rikkomusten tutkimisessa. Lokitiedot tarjoavatkin tukea organisaation sisäisille tutkimuksille, jonka lisäksi lokitiedot mahdollistavat lähtökohtien hahmottamisen, sekä operationaalisten ja pitkäkestoisten ongelmien tunnistamisen. (Singh, Tomar & Roy, 2010.)

4 LOKIANALYYTIIKKA VÄÄRINKÄYTÖSTEN TUNNISTAMISESSA

Organisaatioiden sisäiset uhkatekijät asettavat organisaatiot usein hankalaan tilanteeseen, sillä valtuutetun käyttäjän toteuttama tietoturvaloukkaus voi olla hankala tunnistaa. Tämä johtuu siitä, että usein yrityksen työntekijöillä on pääsy dataan oikeutetusti. Loukkaus tapahtuu siinä vaiheessa, kun työntekijä toteuttaa organisaation tietojärjestelmissä toimen, johon hänellä ei kuitenkaan ole ollut oikeutettua perustetta. Tietoja, joita yrityksen työntekijät saattavat hakea valtuuttamattomasti voivat olla esimerkiksi asiakkaiden henkilötiedot. GDPR on kuitenkin rajannut, että henkilötietoja saa käsitellä vain tietyissä tilanteissa ja käsittelyyn tarvitaan peruste.

Muutamia keskeisiä tapoja tunnistaa sisäisiä tietoturvaloukkauksia on esitetty. On kuitenkin havaittu, että lokitietojen kerääminen järjestelmistä, on ainoa tapa monitoroida sen tapahtumia. Analysoimalla näitä tietoja tietoturvapoikkeamien havaitseminen on mahdollista. Järjestelmien, ja niiden sisältävän datan määrä on kokenut merkittävän kasvun viime vuosikymmenten aikana ja luonnollisesti tämän mukana myös lokitettavan tiedon määrä on kasvanut merkittävästi. Näin ollen myös lokien sisältämät datamäärät ovat kasvaneet. Tämä aiheuttaa haasteita lokianalytiikan tehokkuuteen ja edelleen hankaloittaa tietoturvaloukkausten erottamista suuresta tapahtumien massasta. Tässä osassa tutkielmaa perehdytään tieteellisen kirjallisuuden tarjoamiin näkökulmiin sisäisten uhkien tunnistamisesta lokianalytiikan keinoin, sekä reaalimaailman näkemyksiin GDPR:n luomasta tarpeesta lokianalytiikkaan.

4.1 Poikkeavuuksien tunnistaminen lokitiedoista

Sisäisten toimijoiden toteuttamat valtuuttamattomat toimet voivat helposti hukkua valtuutettujen toimien suureen joukkoon (Liu, Del Vel, Chen, Zhang & Xiang, 2018). Kuitenkin järjestelmien lokeista saatavilla olevaa tietoa analysoimalla on

mahdollista tunnistaa tietystä ympäristöstä poikkeavia tapahtumia, jotka voivat puolestaan paljastaa valtuuttamattomia tapahtumia. Poikkeavia tapahtumien tunnistusta käytetään niin ulkoisten, kuin sisäisten uhkien tunnistamisessa. Liu, Qin, Guan, Jiang ja Wang (2018) esittelivät uuden ryhmittely-lajittelu-jalostus viitekehyksen, jota noudattamalla on helpompi tunnistaa suurista lokimassoista poikkeavuuksia. Viitekehys koostuu viidestä vaiheesta: lokien keräämisestä, ominaisuuksien evaluointi, ryhmittelystä, lajittelusta, sekä jalostuksesta. Keskeistä viitekehyksessä on käyttäjien ja istuntojen piirteiden ymmärtäminen lokitietojen ominaisuuksia evaluoimalla (Taulukko 3).

TAULUKKO 3 Ominaisuuksien erittely ryhmittely-lajittelu-jalostus viitekehyksessä

Ominaisuus	**Selite**	**Loki**
Käyttäjänimi	Pyrkii tunnistamaan käyttäjän, joka on kirjautunut järjestelmään.	Turvallisuusloki
Kirjautumisen isäntä	Pyrkii tunnistamaan, miltä isännältä yhteys on muodostettu, IP-osoite.	Turvallisuusloki
Kirjautumisen fyysinen tieto	Pyrkii tunnistamaan kirjautumisen fyysisen sijainnin, esimerkiksi kaupungin.	Turvallisuusloki
Yhteyden alkamisajankohta	Pyrkii kertomaan tarkan ajan yhteyden alulle.	Turvallisuusloki
Yhteyden kesto	Pyrkii kertomaan yhteyden keston, alkuajasta terminointi hetkeen.	Turvallisuusloki
Komentosekvenssien tiheys	Pyrkii tunnistamaan komentosekvenssien toteutukset.	Viestiloki
Virheelliset kirjautumisyritykset	Pyrkii kertomaan, miten usein väärä salasana on syötetty kirjautumisen yhteydessä tietyn aikamääreen sisällä.	Turvallisuusloki
Operationaalisten toimien tiheys	Pyrkii kuvaamaan operationaaliset tapahtumat, kuten tietojen lukemisen, kirjoittamisen, luomisen, tai poistamisen.	Auditloki
Arkaluontoisten tai kryptattujen tiedostojen avaamisen tiheys	Pyrkii kuvaamaan miten usein käyttäjä avaa arkaluontoisia tai salattuja tiedostoja.	Auditloki
Oikeuksien korottamisen frekvenssi istunnoissa	Pyrkii kertomaan, miten usein käyttäjä korottaa oikeuksiaan operaatioihin servereillä.	Viestiloki

Ensimmäiset viisi ominaisuutta pyrkivät kuvaamaan enemmän perinteisiä istunnon tietoja, kun taas viisi jälkimmäistä ominaisuutta ovat enemmän keskittyneet

kuvaamaan istuntokohtaista käyttäytymistä. Istuntokohtainen käyttäytyminen voi auttaa paljastamaan poikkeavia toiminteita. Esimerkiksi tiheä ja suuri virheellisten kirjautumisyritysten määrä voi kertoa siitä, että käyttäjän tiliä on yritetty murtaa. Myös operaationaalisia toimia analysoimalla voidaan havaita poikkeavaa käytöstä, mikäli käyttäjä esimerkiksi pyrkii poistamaan paljon tiedostoja, tai mikäli käyttäjä hakee paljon sensitiivistä dataa. Viimeisellä kolmella ominaisuudella, eli operationaalisten toimintojen tiheydellä, arkaluontoisten tai kryptattujen tietojen avaamisen tiheydellä, sekä oikeuksien korottamisen frekvenssillä istuntojen aikana voidaan mitata käyttäjän toiminnan mahdollista uhan tasoa. (Liu, Qin, Guan, Jiang & Wang, 2018.)

Ryhmittely-lajittelu-jalostus viitekehyksen näkökulmasta poikkeavuuksia pyritään tunnistamaan edellä esiteltyjen ominaisuuksien avulla. Ryhmittelyalgoritmeilla pyritään louhimaan suuresta datamassasta mahdollisesti poikkeavat tapahtumat. Viitekehys hyödyntää K-prototyyppi algoritmia, jonka on havaittu olevan erityisen tehokas keskenään erilaisen datan, perustuen datan samankaltaisuuteen. Lajittelu-vaiheessa valitaan alustavasti poikkeavaksi tunnistetut tapahtumat. Tämä vaihe pyrkii jättämään selvästi normaalit toiminnot pois jatkotutkimuksesta. Viitekehyksen mukaan poikkeavia tapahtuma kandidaatteja voidaan katsoa olevan tapahtumat, joita ei ole suuria määriä. Jalostus vaiheessa pyritään määrittelemään jokaisesta poikkeama kandidaatista, onko tapahtuma todellisuudessa poikkeava. (Liu, Qin, Guan, Jiang & Wang, 2018.)

Poikkeavuuksien tunnistamisessa on keskeistä määrittää ensin mikä toiminta on normaalia toimintaa kyseisessä ympäristössä. Normaalia toimintaa voidaan pitää lähtökohtana yksittäisen toimijan tai ryhmän toimijoita arvioinnissa. Kaikkea toimintaa, joka eroaa lähtökohtana pidettävästä käytöksestä, voidaan pitää poikkeavana. Poikkeavien tapahtumien määrittely voi kuitenkin olla haastavaa. Esimerkiksi auditointi data voi olla monimutkaista ja epälineaarista, jolloin se ei välttämättä paljasta täysin käyttäjän toimintoja. Esimerkiksi tietokoneelle kirjautumisajat eivät välttämättä kerro tietokoneen todellisesta käyttöaikaa istunnon aikana. Normaalin toiminnan määrittelemiseksi on ehdotettu hyödynnettävän esimerkiksi neuroverkkojen autoenkoodaajaa. (Liu,Del Vel, Chen, Zhang & Xiang 2018.) Voidaan siis nähdä, että käyttäjien toiminnan analysointi on keskeistä poikkeavuuksien tunnistamisessa ja analytiikan kehittämisen tulee kiinnittää huomiota käyttäjien toiminnan tarkasteluun ja analysoimiseen.

4.2 Käyttäjä- ja roolipohjainen analytiikka

Sisäisten uhkien huomioiminen merkittävänä riskinä on yleistynyt niin organisaatiollisella, kuin valtiollisella tasolla. Sisäiset uhat voivat aiheuttaa merkittäviä riskejä tietoturvallisuudelle, kuin myös taloudelliselle vakaudelle. (Zhang, ym., 2018.) Web-ympäristössä toimiessaan käyttäjän toimista generoituu merkittävät määrät dataa. (Singh, Tomar & Roy, 2010.) Käyttäjien toimien lokittamisen on havaittu olevan keskeisessä asemassa mahdollisten uhkien tunnistamisessa. Lokitiedoista selviää esimerkiksi mitä tietoja käyttäjät ovat hakeneet ja mihin aikaan.

Lokitettavan tiedon määrä on kuitenkin merkittävän suuri, sillä järjestelmien sisällä tapahtuu tyypillisesti paljon aktiviteettejä päivittäin. (Legg, Buckley, Goldsmith & Creese, 2015.) Yhtenä vaihtoehtona sisäisten uhkien tehokkaampaan tunnistamiseen on ehdotettu roolipohjaista lokianalytiikkaa (Zhang, ym., 2018; Legg, ym., 2015).

Legg (2015) esitti, että sisäinen uhka voidaan tunnistaa yleisesti lokeilla esiintyvien erilaisten poikkeavien toimien kautta, jotka eroavat normaalista toiminnasta (Taulukko 4). Legg jakoi poikkeavat toimet kolmeen hänen mukaansa yleisimpään kategoriaan: uusiin observointeihin, observoinnin aikaan ja observointien tiheyteen.

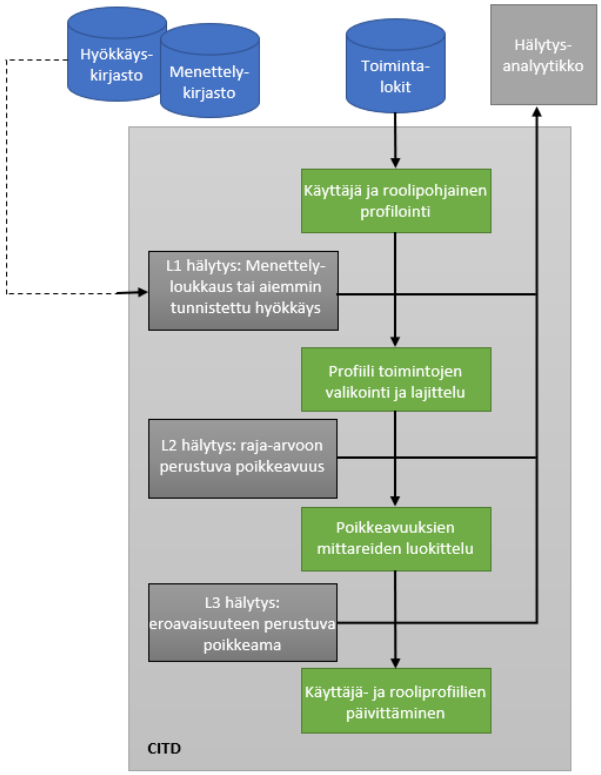
TAULUKKO 4 Observointiin perustuva lajittelu

Observointi	**Selite**
Uusi observointi	Käyttäjä toteuttaa uuden poikkeavan aktiviteetin, tai käyttäjä toteuttaa aktiviteetin poikkeavilla attribuuteilla. Mikäli toiminto/attribuutit ovat käyttäjän normaaleista toimista poikkeavat, tulee tarkastella ovatko toimet poikkeavia roolista (esimerkiksi onko kukaan muu saman roolin toimija toteuttanut vastaavaa toimea aiemmin).
Observoinnin aika	Käyttäjä toteuttaa toimen/attribuutin poikkeavaan aikaan päivästä, verrattuna normaaliin toimintaan. Mikäli toiminto/attribuutit ovat toteutettu käyttäjän normaaleista ajoista poikkeavalla ajalla, tulee tarkastella ovatko ajat poikkeavia roolista.
Observoinnin tiheys	Käyttäjä toteuttaa toimen / attribuutin tiheämmin, verrattuna normaaliin toimintaan. Mikäli toiminto/attribuutit ovat toteutettu käyttäjän normaaleista toimista poikkeavalla tiheydellä, tulee tarkastella, onko samassa roolissa toimivilla käyttäjillä vastaavaa toimintaa.

Peruste siihen, miksi edellä olevassa taulukossa esitetyillä kategorioilla pystyttäisiin tunnistamaan sisäisiä uhkia tehokkaammin, perustuu siihen, että tutkimusten mukaan pahantahtoista toimintaa toteuttaessa työntekijöiden käytös muuttuu ja väärinkäytösten toimet eroavat käyttäjän normaaleista toimista. Lisäksi poikkeavuuksien tunnistamisessa on keskeistä vertailla, miten muut samassa roolissa työskentelevät henkilöt toimivat ja poikkeaako tapahtumat heidän toimistaan. (Legg, 2015.) Roolipohjaista arviointia voidaankin pitää lokianalytiikassa tärkeänä. Esimerkiksi myös Malin, Nyemba ja Paulett (2011) esittivät, että sairaanhoidon parissa työskentelevien henkilöiden hakujen

oikeellisuutta voitaisiin arvioida esimerkiksi osaston perusteella, jolla henkilö työskentelee.

Leggin ym. (2015) mukaan erilaisten hälytyspohjaisten lokianalytiikan arviointien perusteella voidaan mahdollistaa tarkempi lokianalytiikka suuristakin lokimassoista. He esittivät poikkeavuuksien tunnistuksessa käytettäväksi mittareita, jotka johtavat oleellisten piirteiden ryhmittelyistä. Heidän esittelemän mittareita hyödyntävän järjestelmän avulla voitaisiin esimerkiksi merkitä käyttäjiä, jotka ylittävät määritellyn raja-arvon poikkeamille määritellyissä mittareissa. Lisäksi järjestelmä voitaisiin esimerkiksi konfiguroida tunnistamaan tapaukset, joissa käyttäjä poikkeaa määritellyistä mittareista verrattuna vastaavassa roolissa toimiviin henkilöihin. He esittelivät poikkeavuuksien tunnistamisen, joka koostuu poikkeavuudet tunnistavista mittareista, jotka hyödyntävät rinnakkain kulkevaa koordinaatiokäyrää. Lopputuloksena poikkeavuudet johtavat edelleen hälytyslistaan. Yrityksen sisäisten uhkien tunnistamisen työkalu CITD (eng. Corporate Insider Threat Detection) koostuu kolmesta erillisestä kerroksesta, jotka edellä kuvatun mukaisesti luovat hälytyksiä (Kuvio 3).



KUVIO 3 CITD työkalun kuvaus (Legg, ym., 2015)

CITD-järjestelmän tarkoitus on mahdollistaa sisäisten uhkien toiminnan tunnistaminen organisaation tietovarannoissa. Hälytyksen luovia tasoja on kuvattuna kuviossa kolme. Ensimmäinen kerros perustuu vakiintuneiden toimintatapojen rikkomuksiin, sekä aiemmin tunnistettuihin hyökkäyksiin. Toinen hälytyskerros puolestaan perustuu raja-arvot ylittäviin poikkeuksiin. Kolmannessa

hälytyskerroksessa hälytys perustuu eroavaisuuksista johtuviin poikkeamiin. Järjestelmän perimmäinen toimintaperiaate on luoda käyttäjä- ja roolipohjaisia profiileja seuratuista toimintalokeista. Profiilien toimintojen valikointi perustuu profiilien sisältöön, jotta voidaan seurata jokaista asetettua poikkeamien mittaria. Luokittelu toteutetaan poikkeamien mittareille, jotta voidaan määrittää, poikkeaako tapahtuma merkittävästi normaalin toiminnan raameista. Mikäli yksikään hälytys ei aktivoidu, päivittäin päivitetään käyttäjä- ja rooliprofiilit sen mukaisesti. Edellä esitettyjen toimintaperiaatteiden lisäksi järjestelmän olisi keskeistä pystyä kehittymään analyytikon toiminnasta. Järjestelmän tulisi oppia hälytystoimien kehitystä perustuen luotujen hälytysten jälkikäsittelyyn. (Legg, ym., 2015.)

Käyttäjien haitallinen toiminta perustuu usein toimiin, jotka ovat valtuuttamattomia. On kuitenkin huomioitava, että normaalista poikkeavia tapahtumia voidaan määritellä olevan merkittävä määrä. Lokiin tallentuvia poikkeavia tapahtumia voivat olla esimerkiksi epäonnistuneet kirjautumiset tai kirjautumiset normaalien työskentely kellonaikojen ulkopuolella. (Singh, Tomar & Roy, 2010.) Valtuuttamattomien toimien tunnistamiseksi on ehdotettu, että rooli- ja käyttäjäpohjaista lokianalytiikka voisi tehostaa koneoppimisen avulla. Koneoppimisessa voitaisiin valjastaa esimerkiksi neuroverkkoarkkitehtuurin pitkäkestoinen lyhytkestomuistiarkkitehtuuri LSTM (eng. Long Short-Term Memory) muotoilemaan käyttäjälokia luonnollisen kielen sekvenssillä. Jolloin se lopulta saavuttaisi roolipohjaisen luokittelun. Käytännössä koneoppiminen analysoisi käyttäjien toimintasekvenssejä ja louhisi näin käyttäjien käyttäytymiskaavioita. Työntekijöiden haitallinen toiminta voitaisiin edelleen tunnistaa louhimalla edellä mainittuja käyttäytymiskaavioita, milloin lokeista parsitut poikkeavuudet eroavat aiemmista malleista. Toisin sanoen, koneoppimisen LSTM mallin avulla opituista käyttäytymismalleista ja -kaavoista poikkeavat toiminnot havaittaisiin lokitiedoista, paljastaen mahdollisen sisäisen uhan toiminnan organisaatiossa. (Zhang, ym., 2018.) Zargar, Nowroozi ja Halili (2016) puolestaan esittelivät XABA-metodin, joka pyrkii analysoimaan raakalokeja, sekä tietoverkkojen liikennettä reaaliajassa, päämääränään tunnistaa sisäisiä uhkia. Metodi oppii käyttäjien roolien toimintaperiaatteita ja vertaa lokitiedoista poimittua toimintaa epänormaalin toiminnan kaavoihin, jonka jälkeen mahdollinen hälytys muodostuu.

Erilaisia malleja ja metodeja hyödyntämällä voidaan siis nähdä olevan mahdollista tunnistaa käyttäjien normaalien toimien ulkopuolelle asettuvaa toimintaa. Edellä esitettyjen lisäksi on tärkeää tunnistaa, että lokitietojen analysoimalla voidaan tehdä myös muita tietoturvallisuutta parantavia käyttäjäpohjaisia huomioita. Lokitietoja analysoinnin avulla voidaan myös helpottaa loppukäyttäjän käytöksen ymmärtämistä, lokitiedot esimerkiksi helpottavat käyttäjien kategorisointia. Lokitiedostojen avulla voidaan pyrkiä myös tunnistamaan käyttäjätunnuksille tyypillisiä tapahtumia, jotka voivat paljastaa puutteita toiminnassa. Nämä tapahtumat voivat kertoa tai tarkentaa missä rooleissa lisäkoulutus tai kurinpidolliset toimet ovat tarvittuja, tietoturvallisuuden takaamiseksi. Lokianalytiikka voi siis parantaa tietoturvallisuutta osoittamalla organisaation sisäisiä tietoturvallisuuden vaarantajia ja epäkohtia. (Singh, Tomar & Roy, 2010.)

4.3 GDPR ja valtuuttamattomat haut

Yksilöiden tietoturva on organisaatioissa tärkeässä roolissa, jo lakivelvoitteidenkin, kuten GDPR:n takia. GDPR:n näkökulmasta henkilötiedon voidaan nähdä olevan kaikista sensitiivisintä dataa, joka turvallisuutta organisaatioiden on kyettävä suojaamaan. Henkilötietoja tulisi säilyttää ja käsitellä niin, että tietojen eheys ja luottamuksellisuus on taattu. Organisaation tulisi kyetä turvaamaan henkilötietojen käsittelyn oikeellisuus ja laillisuus teknisten ja organisaationaalisten toimien kautta. (Ávila, Khoury, Khoury, & Petrillo, 2021.) Organisaatiolla on lisäksi osoitusvelvollisuus, tarkoittaen, että organisaation on pyydättäessä kyettävä osoittamaan, että henkilötietoja on käsitelty asianmukaisella tavalla. Lisäksi henkilötietojen käsittelylle tulee aina olla peruste. Henkilötietoja ei saa edes katsella, ilman asianmukaista perustetta. (JUDO, 2019.)

Sisäiset toimijat voivat usein toimia hyvin huomaamattomasti, koska tuntevat organisaation toimintaperiaatteet ja turvallisuustoimet (Maloof & Stephens, 2007). Esimerkiksi sairaanhoidon voidaan nähdä kamppailevan potilaiden tietoturvan varmistamisen ha mahdollisimman tehokkaan sairaanhoidon turvaamisen välissä (Malin, ym., 2011). Tarkoittaen, että tasokkaan asiakaspalvelun varmistamiseksi toiminnan on oltava myös tehokasta. Asiakaspalvelijoiden, sekä asiakkaiden tietoja työkseen käsittelevien on organisaatiossa päästävä käsiksi asiakkaiden henkilötietoihin, mutta organisaatioiden tulisi varmistua siitä, että tietoja käsitellään vain valtuutetusti.

Ajatellen GDPR:n näkökulmasta, oikeuksien hallinta on keskeistä. Työntekijöillä tulee olla pääsy ainoastaan asiakkaiden dataan ainoastaan heidän työtehtävien välttämättömyydestä johtuen. Oikeuksien hallinnalla voidaan rajata se, että työntekijöillä ei ole pääsyä muuhun kuin heidän työtehtäviensä kannalta merkitykselliseen dataan. Yleisen tietosuoja-asetuksen myötä on tullut yhä keskeisemmäksi painottaa työntekijöiden toiminnan olevan linjassa tietosuojan kanssa. Onkin tärkeää, että GDPR on työntekijöille esitelty ja tutuksi tehty, jotta työntekijät ymmärtävät toimensa GDPR:n näkökulmasta. (Hossain, 2019.)

Lokitietojen keräämisellä voidaan nähdä olevan keskeinen merkitys henkilötietojen turvaamisessa. On todettu, että puutteellinen lokitietojen kerääminen ja puutteelliset monitorointi mekanismit voivat lisätä tahallista, sekä tahatonta prosessien ja resurssien väärinkäyttöä, joka johtaa edelleen henkilötietojen vaarantumiseen. Henkilötietoja käsittelevien toimien, kuten katselun, muokkauksena ja poistamisen lokittamisen voidaan siis nähdä olevan keskeistä valtuuttamattomien hakujen estämiseksi, sekä tunnistamiseksi. GDPR:n näkökulmasta onkin tärkeää, että lokitietoja kerätään kaikista järjestelmistä, joissa henkilötietoja voidaan käsitellä. (ENISA, 2017.) Tietysti on keskeistä huomioida GDPR:n näkökulmasta, että lokeille ei tallenneta turhaa henkilötietoa ja vain oikeutetuilla lokien käsittelijöillä on valtuutettu pääsy lokidataan. Henkilötietoja sisältävän

lokin käsittelyssä onkin noudatettava GDPR:n mukaisia toiminteita. (Kyberturvallisuuskeskus, 2020.)

4.3.1 GDPR ja lokitiedot

GDPR:n artiklan 32 mukaan jokaisen henkilötietojen rekisterinpitäjän, sekä tietojenkäsittelijän tulee voida taata heidän alaisuudessaan toimivien luonnollisten henkilöiden, joilla on pääsy henkilötietoihin asianmukainen käsittely. Tämä tarkoittaen sitä, että henkilötietoihin pääsyn omaavien henkilöiden tulee käsitellä henkilötietoja vain rekisterinpitäjän ohjeistamisen mukaisesti, tai lain määrittelemällä tavalla. (GDPR, 2016.) Kuten todettu, on tärkeää, että organisaatiot huolehtivat pääsykontrolleista dataan. Tarkoittaen, että työntekijöillä, joilla on tarve työkuvansa puolesta päästä käsiksi henkilötietoa sisältäviin järjestelmiin, on ainoastaan käyttöoikeudet näihin järjestelmiin. Kuitenkin lokitietojen kerääminen ja monitorointi on keskeinen tekijä tietosuoja-asetuksen noudattamisen valvomisessa. Tämä tapahtuu seuraamalla ja varmistumalla työntekijöiden toimista, jotka koskevat henkilötietojen käsittelyä tietojärjestelmissä. (ENISA, 2021.)

Euroopan unionin Kyberturvallisuusvirasto ENISA (2021) listasi viisi keskeistä lokituksen toimea, jotka edistävät henkilötietohakujen valvomisen tehokkuutta:

- Lokitietoja tulee kerätä kaikista järjestelmistä, joissa käsitellään henkilötietoa. Tietoja tulee kerätä kaikista datan käsittelyn tapahtumista (kuten katsominen, muokkaus, poistaminen).
- Lokitietojen tulisi sisältää johdonmukainen aikaleima, sekä tietojen tulisi olla suojattu asianmukaisesti peukalointia tai valtuuttamatonta pääsyä vastaan.
- Lokeihin tulisi tallentua myös järjestelmän omistajien ja hallinnoijan toimet, kuten lisäykset, poistot ja oikeuksien muutokset.
- Lokitietoja ei pitäisi pystyä poistamaan tai muokkaamaan. Lokitietojen käsittelystä tulisi myös tallentua lokijälki monitorointia ja poikkeuksellisten tapahtumien tunnistamista varten.
- Järjestelmän tulisi monitoroida lokitiedostoja ja tuottaa raportteja järjestelmän tilasta, kuin myös luoda potentiaalisia hälytyksiä poikkeuksista.

Edellä esitetyt toimet ovat keskeisiä lokitoimia GDPR:n noudattamisen seuraamisen kannalta. Audit Trailin muodostumisen voidaan nähdä olevan tärkeää turvallisuuden ylläpitämisessä, sekä monitoroinnissa. (ENISA, 2021.) Perustuen edellä esitettyyn, voidaan lokianalytiikan roolia GDPR:n noudattamisen varmentaja pitää keskeisenä, ellei jopa ehdottomana.

4.3.2 PoSeID-on projekti

GDPR oli tarvittu ja merkittävää lisäturva henkilötiedoille. GDPR myös muutti merkittävästi organisaatioiden tarvetta rajoittaa ja monitoroida heidän asiakkaidensa henkilötietojen käsittelyä yrityksen sisällä. Valtuuttamattomien hakujen tunnistamisen ja ennaltaehkäisemisen on havaittu olevan keskeinen osa GDPR:n todellista toteutumista organisaatioissa. Tämä on havaittu myös Euroopan Unionin tasolla. Euroopan Komission tilaama ja Euroopan Union rahoittaman H2020 projektin PoSeID-onin tarkoitus on pureutua henkilötietojen turvan ja organisaatioiden kokemien haasteiden aihepiiriin juuri GDPR:n näkökulmasta. Projektin tarkoitus on luoda skaalautuva, integroitava ja kokonaisvaltainen alusta, jonka voidaan suojella yksityishenkilöiden henkilötietoja, sekä tukea organisaatiota datan hallinnassa samalla varmistaen GDPR:n mukainen toiminta. (Poseidon, 2020.)

Projektin kokonaisuus perustuu teknologioihin, kuten lohkoketjuihin, älysopimuksiin, sekä pilvipalveluihin. Näitä teknologioita hyödyntämällä on tarkoitus mahdollistaa käyttäjien omien henkilötietojen hallinnointi ja henkilötietojen käsittely. (Silva, ym., 2020.) Tarkemmin kuvattuna, alustan kautta yksilöllä on vahvempi, läpinäkyvämpi, ymmärrettävämpi, ja helpompi pääsy tietoihin, sekä parempi mahdollisuus seurata, kontrolloida ja hallita tietoja, joita julkiset ja yksityiset organisaatiot heistä käsittelevät. Yksilöt pystyvät alustan kautta lisäksi määrittelemään sen, kuka tietojenkäsittelijä heidän tietojaan oikeastaan voi käsitellä, tämän luotettavuuden perusteella. Lisäksi yksilöillä olisi mahdollisuus perua annettuja datan käsittelyoikeuksia, sekä pyytää poistamaan pysyvästi heitä koskevat tiedot, tai rajata datan määrä minimointiperiaatteen mukaisesti. (Cordis, 2021.)

PoSeID-on:in avulla pyritään lisäämään teknologisten palveluiden toiminnan GDPR:n mukaisuutta, tarjoamalla kokonaisvaltainen ICT-pohjainen työkalu, joka skaalautuu organisaatioiden toimintaympäristöön (Cordis, 2021). Toteutus sisältää riskien hallinta moduulin, henkilötietojen analysoijan, sekä yksityisyyttä edistävän alustan. Alustan on siis tarkoitus hallinnoida yksilön henkilötietojen ja rekisterinpitäjien, sekä tiedonkäsittelijöiden välisiä toimia. Lisäksi järjestelmä luo varoituksia tietojen kohteelle, mikäli havaitaan mahdollisia tietoturvaloukkauksia. Tässä ratkaisussa poikkeavat toiminnat, jotka kohdistuvat henkilötietoihin havaitaan koneoppimisen ja lokianalytiikan kautta. Poikkeamien havainnointi tapahtuu riskien hallinta moduulissa. Poikkeavuuksia havaitaan järjestelmistä syötetyistä lokeista. Yksityisyydenloukkaukset puolestaan pyritään havaitsemaan henkilötietojen analysoijaosiossa. Henkilötietojen analysoija monitoroi henkilötietoja sisältäviä toiminteita ja lohkoketjualustan luomia varoituksia. Varoitukset perustuvat ennalta määritettyihin ehtoihin, kuten onko datan käsittelijälle annettu lupa datan prosessointiin. (Silva, ym., 2020.)

He, Zhu, He ja Lyu (2016) esittelivät viitekehyksen, jonka avulla lokiperusteinen poikkeamien tunnistaminen olisi tehokkaampaa. Viitekehys koostuu neljästä keskeisestä vaiheesta: lokien keräämisestä, lokien parsimisesta, piirteiden erottelemisesta, sekä poikkeuksien tunnistamisesta. Casaleiron ym. (2019)

mukaan poikkeamien tunnistaminen PosSeID-On alustalla seuraa tätä kyseistä viitekehystä (Taulukko 5) ja perustuu järjestelmälokien analytiikkaan.

TAULUKKO 5 Viitekehyksen soveltaminen PoSeID-On alustaan

Viitekehyksen vaihe	**Vaiheen selite**
Lokien kerääminen	Lokit toimitetaan viestibussin kautta suoraan riskienhallinta moduuliin, hyödyntäen kustomoitua viestintä protokollaa. Viestit sisältävät Graylogin laajennetun lokiformaatin (GELF) standardoinnin mukaisen lokitiedoston, joka sisältää myös ylimääräisen kentän tapahtumille, jotka käsittelevät henkilötietoja. Kentät sisältävät tietoa tiedon kohteesta, sekä käsittelijästä. Henkilötiedot lisätään, vain kun tälle on suostumus.
Lokien parsiminen	Lokit koostuvat vapaamuotoisesta tekstistä, niiden parsiminen on välttämätöntä eheiden tapahtumamallien, jotka sisältävät yleisimmät lokien segmentit saavuttamiseksi. Kun tapahtumamallit on asetettu, tätä seuraavat lokit voidaan joko nähdä johdettuina tapahtumina, tai ne voivat luoda uuden tapahtumamallin. Lokien parsimisvaiheessa käytetään lähestymistapana Drain-metodia, joka on todistettu ylivoimaiseksi verrattuna muihin avoimen lähdekoodin vaihtoehtoihin. Drainin integroimiseksi riskienhallintamoduuliin vaadittiin Java-implementaatio.
Ominaisuuksien erottelu	Lokien parsinnan jälkeen on tärkeää luoda numeraaliset ominaisuusvektorit, jotka välitetään koneoppimisen malleille, jotka suorittavat poikkeavuuksien tunnistamisen. Drainin lokiparsinta-algoritmissa lokit ryhmitellään, jonka jälkeen lasketaan tapahtuman esiintymisten määrä. Tuloksena muodostetaan ryhmä tapahtumia ja näitä vastaava määrä esiintymisiä. Kehityksen mukana ryhmittelyssä vakavuustasoa testataan parhaan lähestymistavan löytämiseksi.
Poikkeuksien tunnistaminen	Järjestelmän aikaisessa vaiheessa valvomattomat oppimismallit ovat suotuisampia, operationaalisen datan vähäisyyden johdosta. Sparkin koneoppimisen kirjastoa hyödynnetään tässä vaiheessa. Kehitys vaatii myös lisätestausta, jotta voidaan saavuttaa paras valvomaton oppimisalgoritmi. Kun poikkeavuuksia tunnistetaan, tästä lähetetään ilmoitus datan kohteelle, datan käsittelijöille, sekä PoSeID-onin hallinnolle.

Esitetyn viitekehyksen ja PoSeID-onin yhteensovittamisessa on huomioitava, että taulukossa esitetyt vaiheet perustuvat projektivaiheen suunnitelmaan. PoSeID-onin riskienhallintamoduulia pohdittaessa on keskeistä ymmärtää, että alustan on oltava skaalautuva ja suorituskykyinen, sillä käsiteltäviä datan kohteita voi olla miljoonia. (Casaleiro, 2019.)

PoSeID-on projektin mukaisella alustalla, lokitietojen analysointi perustuu vahvasti GDPR:n mukaisen toiminnan varmistamiseen. Lokitietoja käsitellään riskien hallinta moduulissa, joka mahdollistaa normaalien ja poikkeavien käyttäytymiskaavojen tunnistamisen. Myös henkilötietojen analysoijalla on keskeinen rooli GDPR:n toteutumisessa, sillä se varoittaa käyttäjiä, mikäli jotain heidän dataansa käsitellään, vaikka siihen ei ole valtuutusta. Keskeistä ratkaisussa on monitasoinen lokianalytiikka, jossa arvioidaan useista lähteistä tulevaa informaatiota, niin käyttäjätasoisesti, kuin järjestelmätasoisesti. Riskien hallinta moduuli hallinnoi myös tietojenkäsittelijöiden riskitilannetta, joka puolestaan voi jälleen tarjota tietoa siitä, missä palveluissa henkilötiedot ovat riskissä. (Casaleiro, ym., 2019.) Vaikka PoseID-on ei tarjoa suoraa vastausta siihen, miten lokitietoja tulisi analysoida valtuuttamattomien hakujen tunnistamiseksi, se antaa arvokasta näkökulmaa siihen, miten tärkeää lokitietojen analysointi on. Lisäksi projekti tarjoaa erilaista näkökulmaa siihen, miten lokianalytiikka voidaan toteuttaa.

5 TUTKIMUSMENETELMÄT

Tutkimuksen tutkimusongelma, keskeinen tutkimuskysymys ja aihetta tukevat tutkimuskysymykset johtavat tutkimuksen toteuttamista. Toimeksiantajaorganisaation työntekijät käsittelevät päivittäin työtehtävissään suuria määriä asiakkaiden henkilötietoja, ja näiden tietojen turvaaminen on organisaatiolle keskeistä. Organisaation valtuutettujen toimijoiden valtuuttamattomien toimien tunnistaminen olisi jo pelkästään GDPR:n näkökulmasta merkityksellistä. Yhdeksi keskeiseksi tavaksi tunnistaa työntekijöiden valtuuttamattomia toimia on tunnistettu lokianalytiikka. Kuitenkin miljoonia asiakkaita omaavassa toimeksiantajayrityksessä henkilötietohakuja suoritetaan jatkuvaksi ja näin ollen analysoitavat lokimassat ovat suuria ja niiden joukosta voi olla haastavaa tunnistaa mahdollisia valtuuttamattomia hakuja.

Tutkielman tavoitteena on luoda artefaktina malli, jonka avulla valtuutetun käyttäjän poikkeavia ja valtuuttamattomia henkilötietojahakuja on mahdollista löytää tehokkaammin lokianalytiikan keinoin. Tutkielman keskeisimpään tutkimuskysymykseen vastaaminen on siis toteutettavan tutkimuksen keskiössä, ja siksi tutkimus toteutetaan konstruktiivisena tutkimuksena. Tutkielman keskeisenä motivaationa toimii sisäisten uhkien tunnistamisen haastavuus, GDPR, sekä havaittu tarve lokianalytiikan tehostamiseen. Lisäksi tarve lokianalytiikan tärkeyden ymmärtämiseen ja lokien laadun vaikutusten ymmärtämiseen ovat keskeisiä. Tämä tutkielman osa esittelee tutkielman tavoitteiden saavuttamiseksi käytetyt tutkimusmenetelmät.

5.1 Toimeksiantajayrityksen esittely

Tämän tutkimuksen toimeksiantajayrityksenä toimii yksi Suomen kolmesta suurimmasta teleoperaattorista. Tutkielma on toteutettu toimeksiantajayrityksen kontekstissa, ja näin ollen tarjoaa tukea organisaation käytännön ongelmaan.

Kuten kaikkia organisaatioita, jotka käsittelevät eurooppalaisten henkilöiden henkilötietoja, myös toimeksiantajayritystä koskee yleisessä tietosuoja-

asetuksessa GDPR:ssä esitetyt velvoitteet. Organisaation on taattava henkilötietojen asianmukainen käsittely myös työntekijöidensä toiminnan osalta (GDPR, 2016). Tutkimuksen alkuperäinen motivaatio tuli toimeksiantajayrityksen halusta ja tarpeesta edelleen kehittää organisaation sisäisten henkilötietohakujen lokianalytiikkaa. Suomen suurimmilla teleoperaattoreilla on kaikilla miljoonia asiakkaita, joten henkilötietojen määrä organisaation tietojärjestelmissä on massiivinen ja lisäksi suurimmat teleoperaattorit työllistävät lisäksi tuhansia henkilöitä. Näiden kahden tekijän takia asiakkaiden henkilötietojen hakujen määrä on massiivinen pelkästään jo asiakaspalvelutehtävissä. Toimeksiantajayrityksessä oli tutkielman toteutuksen aikana käynnissä lokianalytiikkaa kehittävä projekti ja tämän tutkielman tarkoitus oli tukea kyseistä projektia ja sen lopputuotoksia.

5.2 Suunnittelutieteellinen tutkimusmenetelmä

Tutkimus on toteutettu konstruktiivisella tutkimusotteella, tarkemmin sanottuna suunnittelutieteellisen tutkimusmenetelmän (eng. Design Science) keinoin. Tutkimus seuraa Peffersin, Tuunasen, Genlerin, Rossin, Huin, Virtasen ja Braggen (2007) esittelemää suunnittelutieteellistä mallia informaatiosysteemien tutkimuksessa. Malli, jotka tutkielma noudattaa, koostuu kuudesta eri vaiheesta:

1) Ongelman tunnistaminen ja motivaatio
2) Ratkaisun objektiivit
3) Suunnittelu ja toteutus
4) Demonstrointi
5) Arviointi
6) Kommunikointi

Ensimmäisessä vaiheessa määritelmät tutkimusongelma ja perustellaan tutkimuksen toteuttaman ratkaisun arvo, eli mikä motivoi tutkimuksen toteuttamista. Toisessa vaiheessa ratkaisun objektiivit johdetaan tutkimuskysymysten pohjalta, eli kuvataan mitä ratkaisulla tavoitellaan. Kolmannessa vaiheessa luodaan artefakti. Neljännessä vaiheessa demonstroidaan luodun artefaktin kelpoisuus tutkimusongelman ratkaisuun. Viidennessä vaiheessa arvioidaan, kuinka hyvin luotu artefakti soveltuu kyseisen tutkimusongelman ratkaisuun. Tässä vaiheessa on myös mahdollista toteuttaa iterointi ja edelleen kehittää artefaktia. Kuudennessa vaiheessa ongelma, sen tärkeys, sekä ongelmaan ratkaisu kommunikoidaan. On tärkeää huomioitava, että mallia ei tarvitse toteuttaa ensimmäisestä vaiheesta alkaen, vaan mallia seuraava suunnittelutieteellinen tutkimus voidaan aloittaa neljästä eri vaiheesta. Mallia seuraava suunnittelutieteellinen tutkimus voidaan aloittaa joko ensimmäisestä kohdasta, eli ongelman tunnistamisesta ja motivaatiosta, kohdasta kaksi, eli ratkaisun objektiiveista, kohdasta kolme, eli suunnittelusta ja toteutuksesta, tai kohdasta neljä, eli demonstroinnista. (Peffers, ym., 2007.)

Tässä tutkimuksessa lähestymistapa on hyvin ongelmakeskeinen, sillä selkeään reaalimaailman ongelmaan pyritään löytämään tapoja ratkaista se. Siksi

tämä tutkimus seuraa Peffersin ym. (2007) esittämää mallia kohdasta yksi alkaen. Ensimmäisen kohdan jälkeen esitellään kohdan kaksi mukaisesti toteutuksen ratkaisun objektiivit. Kohdassa kolme toteutetaan puolistrukturoiduthaastattelut, analysoidaan data ja peilataan sitä aiemmin tutkittuun tieteelliseen kirjallisuuteen, joiden pohjalta luodaan mallin ensimmäinen versio. Neljäs ja viides kohta, eli demonstrointi ja arviointi toteutetaan iteroinnin yhteydessä, jonka jälkeen palataan takaisin suunnittelu ja toteutus kohtaan. Lopuksi ratkaisu kommunikoidaan tämän tutkielman kautta. Nämä vaiheet on esitelty luvussa 6.

Suunnittelutieteellisen tutkimusmenetelmä soveltuu yleisesti hyvin tietojärjestelmätieteen alueen tutkimuksiin, joissa pyritään luomaan jotain uutta. Konstruktiivisen tutkimusotteella pyritään ratkaisemaan reaalimaailman ongelmia. Kaikkia ihmisen luomia artefakteja voidaankin pitää konstruktioina. (Lukka, 2014.) Konstruktiivista tutkimusotetta voidaankin pitää tutkielmaan soveltuvana, sillä tutkielmassa pyritään ratkaisemaan reaalimaailman ongelma luomalla artefakti. Tutkimuksen tavoite on luoda artefaktina suodatinmalli, jonka avulla lokianalytiikan rajaukset tehostavat valtuutetun käyttäjän tekemien poikkeavien hakujen tunnistamista, ja näin ollen edelleen valtuuttamattomien hakujen tunnistamista.

5.3 Datan keräystavat ja analysointi

Data kerättiin toimeksiantajayrityksen lokianalytiikan parissa työskenteleviltä tai aiemmin työskennelleiltä työntekijöiltä, joilla on näkemys jostakin tutkimuksen kannalta oleellisesta lokianalytiikan aspektista. Näin ollen haastateltavilla on perusymmärrys lokianalytiikan tavoitteista kyseisessä kontekstissa. Haastateltavat valittiin niin, että haastattelujoukolla oli mahdollisimman monipuoliset lähestymistavat ja näkökulmat lokianalytiikkaan, mikä edelleen mahdollistaa mahdollisimman monipuolisen otannan. Tällä pyrittiin saavuttamaan mahdollisimman laajoja näkökulmia, jotka mahdollistavat parhaan mahdollisen lopputuloksen.

Haastattelut toteutettiin puolistrukturoituina yksilöhaastatteluina, tarkoittaen, että haastatteluun on luotu runko kysymyksiä teemoittain, mutta myös vapaa keskustelu ja kysymysten sivusta tulevat näkökulmat ja keskustelu olivat mahdollisia. Kysymyspatteristo esitetään liitteessä yksi (LIITE 1). Haastattelut rakentuvat seuraavan kolmen teeman ympärille:

- Teema 1: Lokianalytiikan haasteet
- Teema 2: Suodattimien määrittäminen
- Teema 3: Lokianalytiikan merkitys

Haastattelut suoritettiin etäyhteyksien kautta. Haastattelut toteutettiin Microsoft Teams-alustalla, ja haastattelut nauhoitettiin Microsoft Teamsin sisäisellä nauhoitustyökalulla myöhempää litterointia varten. Haastatteluja toteutettiin yhteensä kahdeksan kappaletta, kahdeksalle eri henkilölle. Haastattelujen ja

tieteellisen kirjallisuuden perusteella luotavan artefaktin iterointi tapahtui sähköpostitse. Haastatellut henkilöt ja heidän organisaatiolliset taustatietonsa on esitetty taulukossa 6.

TAULUKKO 6 Haastateltavien organisaatiolliset taustatiedot

Haastateltava	**Positio**	**Aika positiossa**	**Aika organisaatiossa**	**Yhteys lokianalytiikkaan organisaatiossa**
Haastateltava 1	Security Manager	9 vuotta	23 vuotta	Tiimin johtaminen, jonka vastuulla on lokianalyysijärjestelmän tekemien havaintojen manuaalisen analysointi. Analysointityökalujen kehittämiseen ja arviointiin osallistuminen
Haastateltava 2	Senior Consultant	20 vuotta	33 vuotta	GDPR määräystenmukaisuuden varmistaminen lokien toiminnassa, sekä lokianalytiikan kehitys
Haastateltava 3	Security Specialist	3 vuotta	15 vuotta	Lokianalytiikan järjestelmän kehitys ja ylläpitopuoli
Haastateltava 4	Security Specialist	5 vuotta	26 vuotta	Lokianalytiikasta nostettujen tapausten manuaalinen analysointi
Haastateltava 5	Software Developer	1 vuosi	1 vuosi	Lokianalytiikkatyökalujen rakentaminen, integraatiot ja lokilaadun parantaminen
Haastateltava 6	Software Developer	13 vuotta	13 vuotta	Lokien pohjadatan analysoiminen, työkalujen rakentaminen, integraatiot, lokilaadun parantaminen
Haastateltava 7	Corporate Security Manager	5 vuotta	5 vuotta	Tietoturvaprojektin johtaminen, joka käsittelee myös lokianalytiikan toteuttamista
Haastateltava 8	Head of Privacy Operations	7 vuotta	20 vuotta	Tiimin tehtävinä varmistaa riittävät toimet henkilön tietosuojan toteutumiseen ja käsittelyn oikeellisuuteen, jossa lokianalytiikka on keskeistä.

Haastatteluista ja iteroinnissa kerätty data analysoitiin laadullisen tutkimuksen perinteisen sisältöanalysoinnin kategorioihin jakamisen keinoin. Kerätystä datasta tunnistetaan siellä esiintyviä teemoja ja näiden perusteella vastauksia ryhmitellään. Kategorioiden tunnistamisen avulla voidaan datan määrä vähentää oleelliseen dataan ja sen analysoiminen on näin helpompaa. Perinteisessä sisältöanalyysissa kategoriat tunnistetaan datasta sen analysoinnin aikana. Tämä menetelmä mahdollistaa käsiteltävän ilmiön syvällisemmän ymmärtämisen

tutkijalle. (Hsieh & Shannon, 2005.) Teemat, jotka tunnistettiin haastatteludatasta analysoinnin vaiheessa ovat:

- Lokianalytiikan tarve: GDPR, työntekijöiden turva ja prosessipuutteet, sekä muunlaiset lokianalytiikan tarpeet, jotka nousivat esille
- Lokien laadun ongelmat
- Suodattimet poikkeavien hakujen esille nostamiseen

Ensimmäisen haastattelukierroksen datan analysoinnin perusteella syntyneelle suodatinmallin ensimmäiselle versiolle suoritettiin iterointi. Iterointiin ja arviointiin osallistuivat samat henkilöt, joita ensimmäisessä vaiheessa haastateltiin. Peffersin ym. (2007) esittämän suunnittelutieteellisen mallin mukaisesti iterointi suoritetaan arviointivaiheessa. Iterointi vaiheen tärkeys on tunnistettu erityisesti suunnittelu- ja tuotekehitykseen keskittyvässä tutkimuksessa. Iterointi mahdollistaa muun muassa tarvittavien muutosten integraation. (Wynn & Eckert, 2017.)

Haastateltaville tarjottiin mahdollisuus tutustua malliin rauhassa ja antaa kommentteja koskien mallin edelleen kehitystä. Tässä vaiheessa toimeksiantaja-organisaation haastateltavilta pyydettiin palautetta mallista, sekä arviointia siitä, kuinka esitetty artefakti sopisi lokianalytiikan kontekstiin. Lisäksi haastateltavilta kysyttiin muutama tarkentava kysymys, aiheista jotka havaittiin ensimmäisellä haastattelukierroksella merkityksellisiksi, mutta joille ei saatu riittävää selkeyttä vielä ensimmäisissä haastatteluissa. Iterointivaihe toteutettiin sähköpostin välityksellä, jossa haastateltaville toimitettiin saatekirje (LIITE 2), sekä kuvaus suodatinmallin ensimmäisestä versiosta, sekä tarkentavat jatkokysymykset (LIITE 3).

Iteroinnin tulosten analysoinnin vaiheessa havaittiin kolme keskeistä yläkategoriaa, jotka ilmenivät suodattimista. Nämä kategoriat muodostivat selkeät teemat:

- Suodatinmallin hyvä soveltuvuus
- Esiteltyjen suodattimien kehittäminen
- Uudet suodattimet

Iterointivaiheessa kerätyn datan perusteella suodatinmallia ja sen soveltuvuutta valtuuttamattomien henkilötietohakujen tunnistamiseen arvioitiin ja mallia, sekä suodattimia itsessään edelleen kehitettiin. Tämän perusteella luotiin tämän tutkielman puitteissa lopullinen versio suodatinmallista.

6 SUODATINMALLIN LUOMINEN

Tämä luku sisältää kuvauksen tutkimusmenetelmän soveltamisesta tässä tutkimuksessa, sekä itse tutkimuksen. Ensin kuvataan tutkimuksen ongelma, sekä ratkaisun objektiivit. Tämän jälkeen toteutetaan ensimmäinen haastattelukierros, jonka tuloksia ja aikaisempaa kirjallisuutta hyödyntäen luodaan ensimmäinen versio artefaktista. Mallin luomisen jälkeen sen soveltuvuutta arvioidaan ja suoritetaan iterointi mallin edelleen kehittämiseksi. Haastateltaville esitellään ensimmäinen luotu versio mallista ja arvioidaan sen soveltuvuutta ongelman ratkaisuksi. Iteroinnin pohjalta mallia edelleen kehitetään. Edelleen kehityksen tuotoksena syntyy tämän tutkimuksen rajoissa toteutettavan suodatinmallin lopullinen versio.

6.1 Ongelma, motivaatio ja ratkaisun objektiivit

Suodatinmallin kehittäminen seuraa Peffersin ym. (2007) esittämää konstruktiivista suunnittelutieteen tutkimusmallia. Lähtökohtana tutkimuksessa toimii reaalimaailman ongelman keskeisyys, joten ensin tässä alaluvussa esitellään tutkimuksen pohjana oleva ongelma, sekä sen motivointi. Lisäksi tässä luvussa esitellään tutkimuksen artefaktin objektiivit.

GDPR loi vuonna 2018 entistä vankamman tarpeen varmistua henkilötietojen käsittelyn oikeellisuudesta. Organisaatioiden on kyettävä varmentamaan, että henkilötietoja käsitellään organisaation sisällä vain todellisesta ja oikeudellisesta syystä. Toimeksiantajayritys toteuttaa jo nykyhetkellä lokianalytiikkaa, jonka avulla pyritään tunnistamaan valtuutetun käyttäjän valtuuttamattomia hakuja yrityksen tietojärjestelmistä. Kuitenkin, kun otetaan huomioon, että kyseessä on miljoonia asiakkaita omaava teleoperaattori, on selvää, että lokitietoihin tallentuu paljon valtuutettuja henkilötietohakuja, joiden seasta on haastavaa tunnistaa mahdolliset valtuuttamattomat haut. Tämän tutkimuksen tavoite on luoda suodatinmalli, jonka avulla olisi mahdollista rajata lokimassasta

nostettavia tapauksia niin, että poikkeavat ja näin ollen mahdollisesti valtuuttamattomat haut nousisivat tarkasteltavaksi. Organisaation sisäisten uhkien toiminta on yleisesti hankalammin havaittavaa, kuin ulkoisten. Tämä johtuu valtuutetulle toimijalle myönnetyistä oikeuksista toimia ympäristössä, sekä sisäisen toimijan omaamasta tiedosta koskien yrityksen turvallisuustoimenpiteitä. (Colwill, 2009.) Lokianalytiikan on havaittu olevan yksi tehokkaimmista keinoista tunnistaa sisäisten toimijoiden valtuuttamaton toiminta (Myers, Grimaila & Mills, 2009).

Suodatinmallin luomisen lisäksi tämän tutkielman tukevat tutkimuskysymykset koskien lokianalytiikan toteuttamisen perusteita, sekä lokien laadun merkitystä lokianalytiikan näkökulmasta otetaan huomioon haastatteluissa. Näiden kysymysten käsittely on keskeistä, sillä ne tukevat onnistuneen lokianalytiikan toteuttamista, sekä luovat perusteen sille, miksi tällaista lokianalytiikkaa tulisi ylipäätään toteuttaa. Tutkimuskysymyksien käsittely avaa perustelut sille miksi lokianalytiikkaa on tärkeää, ja miksi se pitäisi ottaa huomioon jo tietojärjestelmien lokien suunnitteluvaiheessa.

Tämän suunnittelutieteellisen tutkimuksen tuotoksena syntyvää artefaktia voidaan pitää suodatinmallina, joka tarjoaa näkökulman suodatuksiin, jotka voivat auttaa nostamaan esiin poikkeavia ja valtuuttamattomia hakuja suuresta toteutettujen hakujen lokitietomassasta. Tätä artefaktia, joka antaa keinot rajata lokimassaa valtuuttamattomien hakujen tunnistamiseksi voidaan pitää tutkimuksen selkeänä tavoitteena. Suodatinmalli tulee vastaamaan tutkimuskysymykseen ”Millaisilla suodatuksilla valtuuttamattomien henkilötietohakujen tunnistamista lokianalytiikassa voidaan tehostaa?”. Ratkaisun keskeisin objektiivi on mahdollistaa valtuuttamattomien hakujen tunnistaminen entistä tehokkaammin, nostamalla rajauksilla poikkeavia hakuja todennäköisemmin esiin, kuin ilman suodatuksia toteutetussa lokianalytiikassa.

6.2 Lokianalytiikkaa ja sen merkitys

Tässä alaluvussa käsitellään haastattelujen analysoinnissa esille noussutta lokianalytiikan merkitystä ja syitä siihen, miksi lokianalytiikkaa tulisi tehdä organisaatioissa. Näkökulma keskittyy haastattelun luonteen mukaisesti erityisesti valtuuttamattomien henkilötietohakujen ja lokianalytiikan suhteeseen.

Haastatteluissa kysyttiin, että onko haastateltavien tietoon tullut, että työntekijöiden toimesta olisi toteutettu valtuuttamattomia hakuja. Kysymys koski niin organisaatioita, jossa he työskentelevät parhaillaan, kuin myös organisaatioita, joissa he ovat aiemmin työskennelleet. Haastateltavista puolet olivat tietoisia, että valtuuttamattomia hakuja on työntekijöiden toimesta toteutettu jossakin heidän työpaikoistaan. Suurin osa lopuista haastateltavista epäili tällaisia hakuja olleen, mutta heillä ei ollut varmaa tietoa asiasta. Voidaan siis päätellä, että valtuuttamattomat haut ovat reaalimaailman ilmiö, johon organisaatioiden on puututtava ja joita organisaatioiden on pyrittävä ennaltaehkäisemään tietoturvallisuuden säilymisen takia.

Kun haastateltavilta kysyttiin, onko valtuuttamien henkilötietohakujen tunnistamiseen tehokkaampaa keinoa kuin lokianalytiikka, muutamassa haastattelussa nousi esiin, että mikäli olisi mahdollista, että tietojärjestelmistä ei pääsisi hakemaan tietoa ilman, että se vaatisi valtuutuksen olisi se mahdollisesti tehokkaampi valtuuttamattomia hakuja estävä keino. Kuitenkin haastateltavat kokivat, että lokianalytiikka tarvittaisiin joka tapauksessa varmistamaan, että väärinkäytöksiä ei voi syntyä. Haastateltavista suurin osa koki, että valtuuttamattomien henkilötietohakujen tunnistamiseen ei ole olemassa tehokkaampaa keinoa, kuin lokianalytiikka.

> ”No varmaan semmoinen, että se itse työkalu olisi jo siinä, jotta pääsee katsomaan henkilön tietoja, niin siinä pitää olla tietyt kontrollit jo paikallaan. Että, esimerkiksi se ei salli sinun lähteä surffailemaan noin vain, vaan sinulla pitää olla siihen esimerkiksi joku asiakastiketti tai asiakasyhteydenotto, ennen kuin se järjestelmä sallii sinut sinne sisään. Niin se nyt voisi olla yksi sellainen kontrolli. Mutta, aina pahan tekijä keinot keksivät, että kyllä tällä lokianalytiikalla pystytään niin kuin varmistamaan se, ja erityisesti sillä pystytään todistamaan sitten dataan perustuen se, että jos on jotakin tapahtunut, niin mitä on tapahtunut.”

Haastateltavista kaikki kokivat, että tieto organisaatiossa toteutettavasta valtuuttamattomiin hakuihin kohdistuvasta lokianalytiikasta ennaltaehkäisee valtuuttamattomia hakuja työntekijöiden toimesta. Osa haastateltavista piti tietoa lokianalytiikasta jopa tehokkaimpana ennaltaehkäisevänä tapana. Tämä omalta osaltaan luo perustaa miksi lokianalytiikkaa tulisi tehdä jo ennaltaehkäisevänä toimenpiteenä.

> ”Senhän pitäisi kaikille olla itsestäänselvyys, että ei mene katsomaan sellaista tietoa mikä ei ole sallittua, mutta aina se varmasti vahvistaa sitä, että on tieto siitä, että näitä myös sitten tutkitaan tarvittaessa.”

Osa haastateltavista kuitenkin pohti, onko tieto valtuutettujen hakujen tunnistamisen lokianalytiikasta riittävän yleisessä tiedossa organisaatiossa. Haastatteluissa ilmeni, että tietoa lokianalytiikasta pidetään tärkeänä ennaltaehkäisevänä tekijänä, ja näin ollen tietoa lokianalytiikan toteuttamisesta tulisi organisaatiossa jakaa työntekijöille aktiivisesti.

6.2.1 GDPR lokianalytiikan tarpeen vahvistajana

Haastattelujen analysoinnin kautta esille nousi näkökulmia, jotka keskittyivät lokianalytiikan tärkeyteen tietoturvallisuuden varmentajana, erityisesti GDPR:n näkökulmasta. Kaikki haastateltavat kokivat, että GDPR on tuonut mukanaan entistä merkittävämmän tarpeen varmentua työntekijöiden toteuttamien henkilötietohakujen oikeellisuudesta, jotka toteutetaan organisaation tietojärjestelmistä. Teleoperaattoreita on jo aiemmin koskenut välitystietoihin kohdistuva lainsäädäntö, joka on vaatinut omalta osaltaan vastaavanlaista valtuuttamattomien hakujen monitorointia ja tunnistamista. Tämä on toiminut pohjana GDPR:n

tuomalle tarpeelle. Haastatteluissa nousi ilmi, että toimeksiantajaorganisaatio on GDPR:n voimaan astumisen myötä varmistanut tietojärjestelmiensä, sekä lokiensa asetuksen mukaisuuden. Havaittiinkin, että sen lisäksi, että GDPR on tuonut vankemman tarpeen lokianalytiikan suorittamiseen, se on tuonut myös tarpeen lokien, niiden toiminteiden ja tallennuksen oikeellisuuden kartoittamiseen.

> ”GDPR:n kautta korostus tapahtumien tallentumiseen tietojärjestelmistä asianmukaisella tavalla. Panostettu, kun lainsäädäntö astunut voimaan ja silloin on viimeistään varmistettu, että tämä lokien tallennus on kunnossa.”

> ”On tuonut tarpeen lokianalytiikkaan, täysin oikeastaan se GDPR. Sehän tulee niin kun lisänä siihen mitä meillä nyt on ollut Suomessa jo. se Suomen oma tämä tietoyhteiskuntakaari, missä määritellään niin kuin tämän välitystiedon osalta operaattorille tietyt vaatimukset, mutta sitten GDPR myötä se on laajentunut, että se käsittää sen koko henkilötiedon käsittelyn. Ja varmistetaan sillä lokianalytiikalla, se GDPR lain vaade, että asiakkaiden henkilötiedon käsittely on asianmukaista, ja sitä käsitellään vain tarpeen vuoksi.”

> ” […] ja ennen kaikkea johtuu siitä, että tämmöisessä niin kuin ICT-maailmassa, digimaailmassa, missä meillä on niin kuin valtava määrä paitsi henkilötietoa, niin valtava määrä käsittelyä, niin se ainoa omasta mielestä uskottava tapa on se että meillä on voimallisia tietoteknisiävälineitä ja analyysivälineitä näiden mahdollisten vaikkapa väärinkäytösten tai virheiden havaitsemiseen.”

Haastatteluissa nousi esiin, että yksilön oikeuksien lisäys GDPR:n myötä koskien tietojen käyttöä ja poistamista on tuonut myös organisaatiolle enemmän ylläpidollista vastuuta, sekä tarvetta keskittyä lokien laatuun, jotta kaikki tarpeellinen todella saadaan lokeilta. Lisäksi havaittiin, että GDPR on tuonut myös uudenlaista tarvetta todella pohtia ja ymmärtää prosessit tietojen käsittelyn ympärillä.

> ”Ja kyllä me ollaan todettu, että maailmassa GDPR tyyppinen asia, missä meidän pitää ymmärtää miten me käsitellään henkilötietoa ja laajemmin oikeastaan mitä vain tietoa, eli sopii laajemminkin.”

6.2.2 Lokianalytiikka työntekijöiden turvana

Haastatteluissa nousi esille myös lokianalytiikan merkitys työntekijän näkökulmasta. Tärkeä havainto oli, että lokianalytiikka voi toimia myös työntekijän turvana. Lokianalytiikan keinoin voidaan tarvittaessa mahdollisesti todentaa työntekijän toimineen oikeellisesti tapauksissa, joissa joku esittää epäilyksen työntekijän toimintaa kohtaan. Lokianalytiikka voi olla keskeistä myös, jos syntyy tilanne, missä pitää pystyä varmistumaan johonkin tietoon kohdistuneista tapahtumista tietojärjestelmän sisällä.

> ”[…] tai poissulkemaan sitten tilannetta, jossa syytetään suotta, että lokilla ei kyllä tällaista tapahtumaa näy, tai sen on tehnyt joku muu. Eli se on myös turva sille työntekijälle, kuin myös sille asiakkaan datalle.”

> ”Ihan yhtä lailla ollaan selitetty ihmisille, että se on heidän oman oikeussuojan kannalta hyvä. Että, jos tulee joku tilanne, missä epäillään joitain asiaa ja me lokeilta nähdään, että se joku tietty henkilö ei ole hakenut niitä tietoja tai on muuten käyttäytynyt ihan normaalisti, niin se on hänellekin hyvä asia.”

Lokianalytiikan hyödyissä keskeistä on, että sen avulla pystytään todeksi näyttämään, mitä tietojärjestelmissä on todellisuudessa tapahtunut. Ilman lokianalytiikkaa tietojärjestelmissä tapahtuneita toimia ei voitaisiin jälkikäteen todentaa. Lokianalytiikka suojaa siis niin asiakkaiden dataa, kuin myös työntekijöiden toimintaa ja turvaa.

6.2.3 Prosessipuutteiden havaitseminen

Haastatteluissa nousi esiin lokianalytiikan toinen keskeinen hyöty, prosessipuutteiden havaitseminen. Osa haastateltavista koki, että valtuuttamattomien hakujen tunnistamisen lisäksi lokianalytiikka voi ja on paljastanut prosessipuutteita. Yhdeltä haastateltavalta kysyttäessä, että onko hän tietoinen, että valtuuttamattomia hakuja olisi työntekijät suorittaneet hänen nykyisessä, tai aiemmissa työpaikoissaan hän totesi, että näiden tunnistaminen ei ole hänen työkuvaansa, mutta prosessipuutteita hän on lokianalytiikan kautta havainnut.

Lokianalytiikan kautta voidaan myös havaita puutteita esimerkiksi käytössä olevissa työkaluissa. Muutamassa haastattelussa nousi esiin käyttötapauksena niin sanottujen ”villikorttien” tai massahakujen käyttö. Tarkoittaen, että tehdään haku, joissa haetaan esimerkiksi tunnisteen alkuosalla, jolloin tuloksia nousee tarkasteltavaksi enemmän, kuin mitä yksittäinen tunniste olisi tuonut. Esimerkiksi tällaisten toimintatapojen paljastuminen voi edelleen paljastaa prosessivirheen kahdelta kulmalta. Prosessivirheeksi voi paljastua liian laaja haku, jolloin työtavoissa on menettelyvirhe. Tai virheeksi voi paljastua myös puute työkaluissa, mikäli tällaisia massahakuja pitää työtehtävissä tehdä, mutta niihin ei ole luotu omaa työkalua, jonka puitteissa toiminta ei olisi poikkeavaa.

> ”Mutta just siltä kannalta, että oleellistahan olisi, että jos joku ongelma vaikka ratkaistaan ja on oikeus hakea sillä kuka sitä ratkaisee ongelmaa sen yhden käyttäjän, niin silloin ei yleensä ole tarpeellista niitä massahakuja käyttää, että näkisi vahingossakaan niitä muita käyttäjiä”

> ”Että, onko ongelma, että meillä ei ole toista työkalua, jolla voitaisiin tehdä se haku. Jos se esimerkiksi liittyy vaan johonkin työtehtävään ja tarvitaan semmoinen tieto ja ei ole muuta työkalua kuin tämä. Niin sitten se voisi paljastaa jopa semmoisenkin poikkeavuuden tai puutteen ehkä enemmän.”

Lisäksi esille nousi, että käyttöoikeuksien seuranta on oleellista valtuuttamattomien hakujen tunnistamisessa. Mikäli lokeilta havaittaisiin, että käyttäjä on toteuttanut toimen järjestelmään, johon hänelle ei ole oikeuksia annettu virallisen

prosessin mukaisesti käyttöoikeuksien hallinnan kautta, voisi se paljastaa myös prosessin menettelyvirheen.

> ”Eli käyttöoikeuksien se, että me tarkistetaan se, että se henkilö, joka siellä sitä henkilötietoa käsittelee, että onko sillä oikeus päästä siihen järjestelmään, että onko se saanut jotakin muuta kautta sen pääsyn, kuin virallista prosessia myöten.”

Edellä esitettyjen lisäksi on keskeistä huomioida, että myös tietoturvakoulutuksessa voidaan havaita puutteita, mikäli työntekijä tekee valtuuttamattomia hakuja. Organisaatiossa tulisi laajalti olla tieto siitä, että hakuja saa tehdä vain oikeellisella perusteella ja, että valtuuttamattomien hakujen lokianalytiikkaa toteutetaan.

6.2.4 Muunlaisten valtuuttamattomien hakujen lokianalytiikka

Haastatteluissa nousi esille myös muutamia aiheita, joiden huomioimiselle lokianalytiikassa voisi olla tarve, mutta jotka eivät kuitenkaan istuneet täysin tässä tutkielmassa esiteltävän suodatinmallin näkökulmaan. Esimerkiksi yksi haastateltavista esitti, että tällaiselle lokianalytiikalle, jota tutkielma käsittelee, voisi olla tarve myös kehittäjien työtä kohtaan. Tähän kohdistuen selvisi, että esimerkiksi asiakaspalvelutyötä tekevillä on selkeä ohjeistus siihen, milloin dataa voidaan ja saadaan käsitellä, mutta sitten kovemman IT-puolen toimijoiden hakuihin, jotka menevät esimerkiksi suoraan tietojärjestelmän kantaan ei välttämättä ole yleisesti olemassa yhtä selkeää ohjeistusta.

> ” […] mutta minä mietin vain ylipäätään sitä IT-puolen käsittelyä, missä et välttämättä mene edes käyttöliittymästä sisään, niin se on ehkä sellainen mistä toivoisi niin kuin kovastikin tarkempaa jälkeä siitä.”

Lisäksi yhdessä haastattelussa nousi myös esille mahdollinen tarve tarkastella organisaatioiden sisäisiin toimijoihin kohdistuvia toimia. Esimerkiksi nousi esille työntekijöistä mahdollisesti muodostuvat suoriutumisraportit. Haastattelussa havaittiin, että osalla esimiehistä on pääsy myös muiden kuin oman tiiminsä suoritusraportteihin, johtuen muun muassa siitä, että esimiehet sijaistavat usein toisiaan. Haastattelussa nousi esille, että olisi mielenkiintoista tutkia sitä, että ketkä raportteja tarkastelevat ja miksi. Tämä edelleen kehittäisi työntekijöiden tietoturvaa ja nostaisi oikeellista käsittelyä ja sen merkitystä suuremmin esille, mikäli perusteettomien raporttien tarkastelukulttuuri havaittaisiin.

> ”[…] pitäisi privacy-näkökulmasta nimenomaan pitää huoli, että myös työntekijöiden tieto käsitellään oikein ja muuta ja toki siellä aika tiukat ohjeet onkin ja mitä kaikkea muuta.”

6.3 Lokien laadun vaikutus analytiikkaan

Tässä alaluvussa käsitellään lokien laadun vaikutusta lokianalytiikkaan. Näkemys on erityisen keskittynyt lokianalytiikkaan, jonka avulla pyritään tunnistamaan valtuuttamattomia hakuja ja lokien laadun vaikutuksiin näiden tunnistamisessa. Haastatteluissa selvisi, että suurin osa koki lokien laadun ongelmalliseksi lokianalytiikan näkökulmasta. Osa haastateltavista nosti esille, että lokien tulisi täyttää GDPR:n asettamat vaatimukset, myös lokianalytiikan toteuttamisen kannalta.

> ”Pitäisi saada ne tavallaan henkilötietojen väärinkäytökset ilmi ja aikaisemminhan se on ollut vähän vaikeata, tietysti kun lokituksen laatu on ollut mitä on ollut, niin sitä täytyy parantaa, että saadaan ne väärinkäytökset sieltä esiin.”

Haastateltavilta kysyttäessä, ovatko suuret määrät valtuutettuja hakuja hankaloittaneet mahdollisten valtuuttamattomien hakujen tunnistamista, osalla haastateltavista nousi esiin lokien laatu. Osa haastateltavista koki lokien laadun jopa isoimmaksi haasteeksi, kuin suuret määrät valtuutettuja hakuja.

> ”No toki se vaikuttaa, se määrä vaikuttaa aina, tai antaa lisähaasteen. Ja oikeastaan ei välttämättä edes se määrä, vaan kun me käsitellään oikeastaan kaikki työtehtävissämme henkilötietoa, mutta siinä määrin, että se suurin haaste on sen lokin laatu, ei ehkä se määrä.”

> ”No kyllä, lähinnä just se että meillä ehkä just sen lokilaadun mukaan on vähän ollut vaikea yhdistää sitten niitä valtuutuksia, mitkä ollaan tuotu sitten muualta tuonne. Niin pitäisi olla ne tietyt tiedot, millä me pystytään yhdistää ne valtuutukset niihin lokitapahtumiin.”

> ”Tietysti suuret määrät hakuja voi johtua kahdesta syystä, siitä että hakuja tehdään oikeasti tai tänä päivänä siitä, että lokilaatu ei ole kauhean hyvä.”

Haastatteluissa nousi esille myös se, että lokianalytiikan tehostamisessa oleellista olisi lokin laadun parantaminen. Haastatteluissa selvisi, että nykyhetkellä lokin laatua pidetään joiltakin osin haasteena lokianalytiikan tehokkuudelle, sekä näin ollen valtuuttamattomien hakujen tunnistamiselle. Haasteita kuvattiin olevan myös lokien muotoilun kanssa, sillä järjestelmiä on toteutettu eri aikaan ja kaikissa ei ole käytetty lokeihin samoja standardointiperusteita. Lisäksi kaikki eivät käsitä lokitietoa samalla tavalla. Esimerkiksi järjestelmänhaltijan näkökulmasta jokin lokitieto voi olla paljon tärkeämpää tallentaa, kuin lokianalytiikkaa toteuttavan henkilön mielestä keskeisten asioiden tallentaminen lokiin. Ylipäätään lokitus saatetaan nähdä prosessina hyvin erilaisista näkökulmista, osittain johtuen myös standardoinnin vähyydestä.

> ”[…] ihmiset ei ymmärrä lokitusta samalla tavalla, monelle loki on loki ja ne puhuu silloin yleensä siitä operatiivisesta lokista, joka syntyy tyypillisesti sinne järjestelmän levylle ja josta etsitään sitten niitä ongelmatilanteiden mahdollisia syitä.”

> ”Se oikeastaan olisi tärkein asia, että se pohjadata, se lokidata olisi yksiselitteistä. Melkein voisi sanoa, että mahdollisimman lyhyttä, että siinä olisi kaikki se oleellinen, mutta ei mitään ylimääräistä. Ja hyvässä järjestyksessä esimerkiksi asiat ja niin edelleen. Mitä sekavampaa se loki on, mitä se joku järjestelmä tuottaa, niin sitä haasteellisempaa sitä on saada sellaiseen järjestykseen, että sieltä saa ne oleelliset asiat irti.”

> ”Se lokinlaatu itsessään on niin kuin tosi tärkeä asia, että se analytiikka olisi helpompaa. Se on helpoin korjata se laatu siellä alkupäässä.”

> ”Siellä on monenlaisia haasteita ja ne lähtee ihan niin yksinkertaisesta asiasta, kuin siitä aikaleimasta. Se voi olla ihan rikki, väärin muotoiltu ja se on epäselvä, ei tiedetä mikä on kuukausi, mikä on vuosi. Se ei ole yksiselitteinen, hyvin monessa järjestelmässä on defaulttina se tilanne.”

Sen lisäksi, että lokien laatua tulisi korjata sen merkittävän vaikutuksen johdosta lokianalytiikkaan, haastatteluissa havaittiin myös tarve lokien auditointiin, ja siihen, että lokien laatu saataisiin ylläpidettyä tarvittavalla tasolla. On selkeää, että lokien muotoilua ja toimivuutta tulee ylläpitää lokilaadun ja lokien hyödyn takia. Kuten myös GDPR:n voimaan astumisen myötä havaittiin, tulee lokeja päivittää myös ympäristön muuttuvien tarpeiden mukaan.

> ”Aina ei ole niin yksiselitteistä mihin se raja vedetään, milloin lokitapahtuma täytyy pystyä siirtämään järjestelmään, jossa tällaista jatkoanalysointia pystytään tekemään. Siinä on ensimmäinen haaste, mutta niihin on kyllä löydetty hyviä ratkaisuja. Tällä hetkellä ainakin meillä pystytään hyvin määrittelemään ne tilanteet joissa sitä lokia pitää syntyä. Mutta silti sitten, jos järjestelmässä tehdään jotakin muutoksia ja se menee rikki osittain tai kokonaan. Jos menee kokonaan rikki se huomataan helposti, mutta jos jokin tietty käyttötapaus ei enää ei tuotakaan lokia siinä mihin se on alun perin suunniteltu niin se on hankala havaita. Se on suurin haaste, järjestelmässä tehdään jotakin päivityksiä, muutoksia ominaisuuksiin ja sen jälkeen jokin tietty käyttötapaus onkin jäänyt huomioimatta lokien tallennuksen, siirtämisen muuhun järjestelmään kannalta. Se on se haaste mikä on nyt ja tulevaisuudessa ja pitämään laatu korkealla, että kaikki tarvittavat lokieventit syntyisivät ja viedään talteen toiseen järjestelmään.”

6.4 Suodatinmallin ensimmäisen version luominen

Tässä alaluvussa esitetään ensimmäinen haastattelukierroksen analysointi suodatinmallin luomisen näkökulmasta. Haastattelun tulosten perusteella ja käsitellyn tieteellisen kirjallisuuden perusteella luodaan ensimmäinen versio suodatinmallista. Teoriaosiossa havaittiin, että työntekijöiden valtuuttamaton toiminta poikkeaa usein jollain tapaa valtuutetusta toiminnasta. Haastattelun avulla pyrittiin siis myös lisäksi tunnistamaan muita suodattimia, joiden avulla poikkeuksellista ja näin ollen mahdollisesti valtuuttamattomia hakuja voitaisiin tunnistaa lokitiedoista niiden analysoinnin kautta.

Haastateltavia pyydettiin arvioimaan, sopisivatko tieteellisestä kirjallisuudesta johdetut suodattimet organisaation valtuuttamattomien lokianalytiikkaan. Suodattimet, jotka haastattelussa esitettiin, olivat johdettu tässä tutkielmassa esitellystä kirjallisuudesta. Esiteltyihin suodattimiin lukeutui Leggin (2015) esittelemät observointiin ja rooliin perustuvat poikkeavuudet: uusi observointi, observoinnin aika, sekä observoinnin tiheys. Uudella observoinnilla tarkoitetaan käyttäjän roolista poikkeavaa hakua tai haun attribuuttia. Observoinnin ajalla tarkoitetaan käyttäjän roolista poikkeavaan aikaan toteutettua toimea. Observoinnin tiheydellä puolestaan tarkoitetaan havaintoa, jossa käyttäjä roolistaan poiketen hakee attribuuttia tai toteuttaa jonkin toimin tiheämmin, verrattuna normaaliin toimintaan. Lisäksi haastattelussa esitettyihin suodattimiin lukeutui Leggin ym. (2015) CITD-työkalussa esittämässä hyökkäyskirjastossa käytetty suodatin, joka arvio aiemmin valtuuttamattoman toimen tehneen työntekijän toimia. Jokainen haastateltava piti kaikkia suodattimia mahdollisesti sopivina organisaation valtuuttamattomien hakujen tunnistamisen lokianalytiikkaan. Kuitenkin suodattimia pitäisi muotoilla sopimaan organisaation ja roolin kontekstiin. Nousikin esille, että suodattimia tulisi yhdistellä, esimerkiksi poikkeava observoinnin aika tulisi suhteuttaa työntekijän rooliin ja siihen rooliin tyypilliseen työaikaan. Kuitenkin todettiin myös, että suodattimet tulee suhteuttaa tilanteeseen, esimerkiksi aiemmin kerran valtuuttamattoman haun suorittanut työntekijä ei välttämättä toteuta valtuuttamattomia hakuja uudelleen.

> ”Joo kyllä ne siis ovat kaikki ihan päteviä, tällaisia mahdollistajia sille, että löydetään massasta jotain poikkeavia. Ja niin kuin sanottu, niin ollaan aika uudella alueella, ja tuon lisäksihän niitä voi olla vaikka kuinka paljon lisää mitä me ei olla vielä keksitty. Mutta kyllä nuo kaikki on semmoisia, että niillä varmasti saadaan hyötyä. Ja se olennaisen varma asia on, että meillä on tuollaisia suodattimia, joilla vähennetään sitä turhaa kohinaa ja nostetaan ihmissilmille vain sellaista, jonka ihminen osaa sitten helpommin ratkaista ja oppia mikä niistä oli, jolloin voidaan kehittää niitä suodattimia, kun todetaan että suodatusperiaatteista huolimatta massasta on edelleen puolet jotakin sellaista aivan selvää, joka voitaisiin poistaa. Silleen on aika iteratiivinen tuo asia vielä tänä päivänä.”

> ”Tilanteesta riippuen, yhdistelmiä: haku outoon aikaan yhdistettynä hakijan rooliin. Meillä on firmassa henkilöitä, jotka tekevät hakuja ihan tota 24/7 tavallaan, heidät pitäisi saada eroteltua, mutta jos rooli on sellainen, että tehdään niin sanottua toimistotyöaikaa ja sitten rupeaa tulla hakuja yölliseen aikaan, niin se on sitten sellainen suodatin, että kannattaa lähteä tutkimaan tapausta tarkemmin.”

> ”Toi viimeinen oli sellainen mitä jäin eniten ehkä arpomaan, että jos joku on jossakin kohtaan tehnyt jostakin syystä vaikkapa jonkun virheen ja asia on käyty henkilön kanssa läpi, niin tämähän ei nyt kerro siitä, että henkilö jatkossa toimisi samalla tavalla, niin kuin tämän tyyppinen seuranta ei mutta sitten taas toisaalta miettii sitä, että voi olla tilanteita, joissa joku systemaattisesti toimisi jostakin syystä sillä tavalla, niin ehkä se olisi kuitenkin sekin joissakin tilanteissa validi.”

Haastattelussa nousi esille myös se, että tutkimushetkellä vallitsevan pandemiatilanteen takia suuri osa organisaatioiden työntekijöistä työskentelee kotoa käsin, mikä puolestaan on tuonut myös organisaation tietojärjestelmät mukaan kotiin.

> ”Myös esimerkiksi kellonaikaan sidottu, varsinkin nyt korona-aikana, kun tehdään etätöitä ja vielä helpommin voi tuo vapaa-aika ja työ sekoittua ja on kaikki nämä joka päivä käytettävissä kotona työkalut ja tietojärjestelmät, jos siellä epänormaaliin aikaan tehdään hakuja, niin herää kysymys ollaanko silloin työtehtävien vai henkilökohtaisten intressien takia tekemässä hakua.”

Yksi haastateltava nosti esiin myös sen, että poikkeava observointitiheys voi jakautua pitkälle aikavälille, eli tulisi huomioida lyhyellä aikavälillä tiheästi haettavien tapausten lisäksi myös pidemmälle aikavälille asettuva haun iterointi.

> ”Kyllä se korostaa sitä, että katsotaan riittävä aika taaksepäin. Ja siltä kannalta tutkitaan niitä, että onko jotain tiettyä tavallaan jotain käyttäjää, tai sitten hakijaa tai haettavaa. Että, niin kun se jotenkin toistuu siellä, että minkä takia sitä haetaan niin.”

Haastatteluissa nousi esille, että valtuutettu toiminta on sellaista, mihin löytyy lähtökohtaisesti työ- tai asiakaslähtöinen peruste. Tällaisella tarkoitetaan esimerkiksi asiakasyhteydenotosta löytyvää kirjausta. Tällaisella suodattimella voitaisiin erottaa varmasti valtuutetut haut lokiin tallentuneiden hakujen joukosta. Tämä puolestaan pienentäisi analysoitavien tapahtumien lukumäärää ja helpottaisi mahdollisten valtuuttamattomien hakujen esiin nostamista.

> ”Panostaisin siihen, että otetaan selville tietojärjestelmien tietojen avulla, onko asiakas ollut yhteydessä tai onko ollut tarvetta olla yhteydessä asiakkaaseen. Jos ei kumpaakaan löydy, niin se on silloin mielenkiintoinen asia, miksi on tietoja haettu.”

Edellä esitetyn havainnon kautta pystyttiin toteamaan, että suodattimet, joilla pyritään nostamaan lokitiedoista mahdollisesti valtuuttamattomia toimia esiin voivat olla tarkemman tarkastelun arvoa nostavia tai laskevia suodattimia. Edellä esitetty asiakaslähtöinen selite haulle laskisi näin ollen tarkempaan tarkasteluun nostamisen arvoa tapahtumalle.

Keskeiseksi suodattimeksi haastatteluissa pystyttiin tunnistamaan jo kirjallisuudesta johdettu roolipohjainen poikkeavuus. Haastatteluissa nousi esiin, että olisi tärkeää, että pystyttäisiin pysymään työntekijöiden roolien muutoksissa mukana, ja päivittämään roolin mukaan siihen tarvittavia oikeuksia. Näin varmistaen, että työntekijöille ei jää oikeuksia tietojärjestelmiin, jossa olevia tietoja he eivät enää tarvitse uudessa roolissaan. Haastatteluissa nousi esille myös se, että haku on tehty jotenkin poikkeavalla hakuparametrilla.

Yhdessä haastattelussa nousi esiin suodattaminen perustuen poikkeavaan toimintajärjestykseen. Esimerkiksi siihen, jos tietyssä työtehtävässä tyypillisesti käytetään esimerkiksi tiettyjä järjestelmiä tietyssä järjestyksessä, niin poikkeavuus syntyisi, kun työntekijä hakisikin suoraan tietoja esimerkiksi tyypillisesti viimeiseksi käytettävästä järjestelmästä.

> ”Ja sitten pystyttäisiin vertamaan tavallaan niitä, et jos tuleekin silleen, että ensin mennään tai se normaali tapahtumajärjestys oisikin vaikka, että ensin järjestelmään A, sitten mennään järjestelmään B, sitten mennään järjestelmään C, jos se olisikin sitten, että mennään suoraan järjestelmään C, ja ei tehdä muuta. Tämmöinen tavallaan nostaisi sitten hälytyksen, että tämä oli poikkeava tapaus. Niin tällaista pitäisi miettiä, että pystyisi suodattamaan niitä ja löytämään niitä poikkeuksellisia tapahtumia.”

Jo aiemmin esille nostettu käyttöoikeuksien seuraaminen havaittiin haastatteluissa myös tärkeäksi. Suodattimena voitaisiin hyödyntää työntekijän toimien vertaamista siihen, onko työntekijällä oikeellista ja virallista prosessia myöten myönnetty käyttöoikeus järjestelmään. Tämä voi nostaa esille mahdollisesti valtuuttamattomia toimia, tai kuten aiemmin esitetty myös prosessipuutteita ja menettelyvirheitä.

Suodatinvaihtoehdoista ja poikkeavista hauista puhuttaessa muutamassa haastattelussa nousi esiin, että olisi hyödyllistä, jos pystyttäisiin päättelemään, kuuluuko valtuutetun käyttäjän hakema henkilö esimerkiksi hakijan lähipiiriin, tai onko kyseessä esimerkiksi julkisuudenhenkilö. Tarkoittaen, että pystyttäisiin tunnistamaan, onko hakijalla haettavaan muita sidosryhmäyhteyksiä, kuin ammatilliset sidokset. Tällainen näkökulma johtaa siitä, että osasta haastateltavia tuntui, että väärinkäytökset ovat sellaisia, joissa haetaan jonkun läheisen tietoja. Lisäksi haastateltavat pohtivat olisiko tämä edes mahdollinen toteuttaa teknisesti, sekä työntekijän oikeuksien turvaamisen näkökulmasta.

> ”Jos pystyttäisiin jotenkin päättelemään, kuuluuko liittymä käyttäjän lähipiiriin, siis sen, joka on tietoja hakenut. Onko se sukulainen, tuttava, ystävä tai onko se julkisuuden henkilö. Jos meillä olisi siihen jotakin työkaluja, että pystyttäisiin siihen. Onko sidosryhmä yhteyksiä, se olisi hyvin hyödyllinen.”

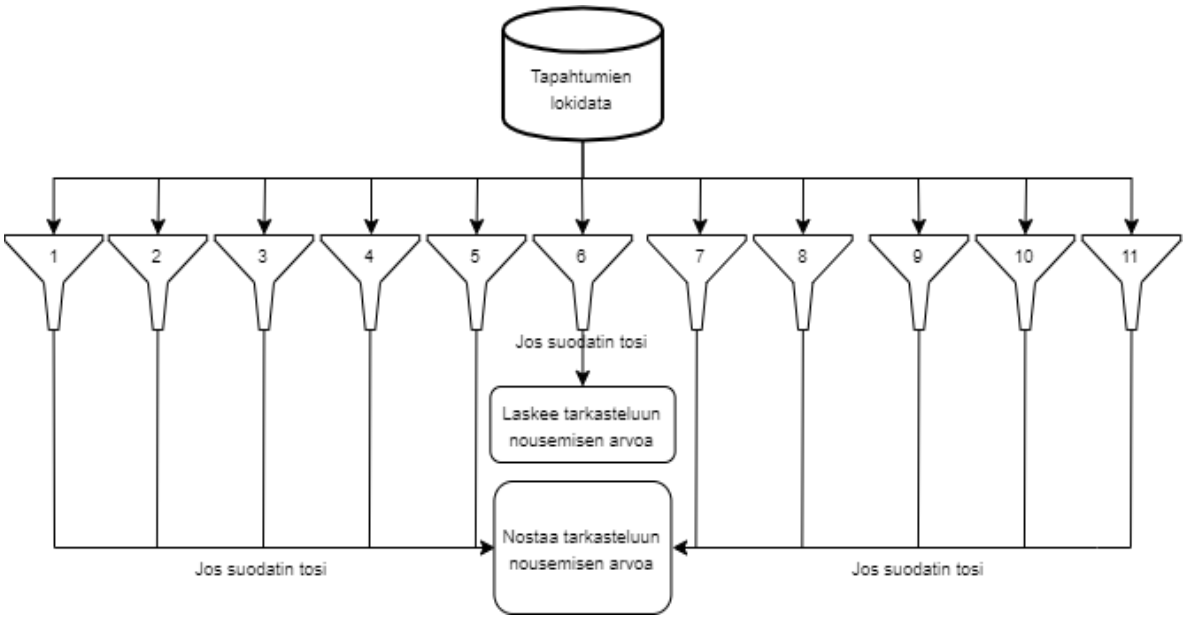
> ”Mutta yks sellainen oleellisimpia juttuja on tavallaan hakijan sukulaissuhteet ja ystävyyssuhteet, jos sellaisia tietoja haetaan, niin ne olisivat minun mielestäni sellaisia, jotka pitäisi pystyä suodattamaan, niin että niitä tutkittaisiin vähän tarkemmin.”

Kuitenkaan haastatteluissa ei noussut esiin suoraan rajauksia, joilla tällaiset henkilökohtaisiin sidosryhmiin kohdistuvat valtuutetun käyttäjän suorittamat haut voitaisiin tunnistaa. Koska muutamassa haastattelussa nousi esille tällaisen suodattimen merkityksellisyys, otettiin suodatin mukaan suodatinmallin ensimmäiseen versioon ja iteroinnissa pyydetään erikseen kehitysehdotuksia tähän suodattimeen.

Haastatteluissa havaittiin myös koneoppimisen hyöty valtuuttamattomien toimien tunnistamisen lokianalytiikassa. Haastatteluissa ilmeni, että tähän lokianalytiikkaan on tuotu mukaan jo jonkin verran koneoppimista ja että, koneoppimista on tarkoitus edelleen kehittää niin, että poikkeavuudet olisi entistä paremmin ymmärrettävissä, sekä tunnistettavissa. Koneoppimisen havaittiin olevan keskeistä suodattimien taustatietojen yhdistelyn kannalta. Koneoppimisen avulla on helpompi esimerkiksi määrittää mitkä ovat normaalit toimintatavat, jolloin edelleen pystytään helpommin tunnistamaan poikkeavat tapaukset. Havaittiin myös, että koneoppiminen edesauttaa lokianalytiikan tehokkuutta.

> ”Kyllä se on haasteellista, kun niitä on paljon kuitenkin niitä hakuja. Toisaalta tuo mitä ollaan vähän koneoppimista tuotu tuohon sovellukseen [lokianalytiikkasovellus], mikä on vähän helpottanut sitä, miten pystytään löytämään niitä, mitkä on valtuutettuja ja mitkä valtuuttamattomia todennäköisesti.”

Suodatinmallin ensimmäiseen versioon on kerätty haastatteluissa esiteltyjä suodattimia, sekä haastatteluista esiin nousseita suodattimia. Suodatinmallin periaatteista on koostettu visuaalinen mallikuvio (Kuvio 4). Kuviossa nähdään lokidataan tallentuvien tapahtumien menevän suodattimien läpi. Mikäli tapahtuma vastaa suodattimien 1-5 ja 7-11 suodatusta, nostaa se tapahtuman manuaaliseen tarkasteluun nousemisen arvoa. Nämä suodattimet indikoivat toteutuessaan, että tapahtuma on poikkeava normaaleista toimintatavoista. Puolestaan suodatin 6 laskee tapahtuman tarkasteluun nousemisen arvoa suodattimen ollessa tosi. Suodatin 6 indikoi, että tapahtumalle löytyy oikeutettu peruste, eli työtai asiakaslähtöinen työtilaus.



KUVIO 4 Suodatinmallin ensimmäinen versio

Suodattimia on mallin ensimmäisessä versiossa yhteensä 11 kappaletta. Suuri osa suodattimista on roolipohjaisia, ja niitä on siis sovellettava työntekijän rooliin nähden. Suodatinmallin esittämät suodatukset pyrkivät rajaamaan tapahtumien lokimassaa niin, että sieltä nousisi manuaaliseen tarkasteluun poikkeuksellisia tapahtumia, jotka edelleen voivat paljastaa valtuuttamattomia tapahtumia. Suodatinmallin suodattimet on esitelty tarkemmin selitteineen alla taulukossa (Taulukko 7).

TAULUKKO 7 Suodatinmallin ensimmäisen version selitteet

Suodatin	**Selite**	**Nostaa/laskee tarkemman tarkastelun arvoa**
1. Poikkeava observoinnin aika	Observoinnin aika poikkeaa työntekijän rooliin suhteutetusta työajasta. Esimerkiksi toimistoaikana työskentelevä työntekijä toteuttaa haun yöaikaan	Nostaa
2. Poikkeava observoinnin tiheys yhteen kohteeseen lyhyellä aikavälillä	Työntekijä observoi poikkeavan tiheästi kohdetta lyhyen aikavälin sisällä	Nostaa
3. Poikkeava observoinnin tiheys yhteen kohteeseen pitkällä aikavälillä	Työntekijä observoi poikkeavan tiheästi kohdetta pitkällä aikavälillä. Esimerkiksi työntekijä hakee puolen vuoden aikana joka kuukausi saman henkilön tietoja	Nostaa
4. Poikkeava observoinnin tiheys useaan kohteeseen	Hakukentän iterointi. Työntekijä hakee tiheään tahtiin eri henkilötiedoilla tietojärjestelmistä	Nostaa
5. Poikkeava observoinnin kohde	Työntekijä observoi hänen roolistaan poikkeavaa kohdetta. Esimerkiksi yritysasiakaspalvelija hakee kuluttaja-asiakkaan tietoja	Nostaa
6. Observoinnille löytyy asiakaslähtöinen peruste	Työntekijän toteuttamalle haulle löytyy esimerkiksi asiakkaan yhteydenotto perusteeksi	Laskee
7. Aiemmin valtuuttamattoman haun tehnyt työntekijä toteuttaa haun	Aiemmin tunnistetusti valtuuttamattomia hakuja tehnyt työntekijä tekee haun	Nostaa
8. Poikkeava hakuparametri	Työntekijä toteuttaa haun poikkeavalla parametrilla, jolla ei tyypillisesti haeta tietojärjestelmistä	Nostaa
9. Poikkeava tapahtumajärjestys	Työntekijä toteuttaa toimia poikkeavassa tapahtumajärjestelmässä. Esimerkiksi työntekijä hakee normaalisti ensin järjestelmästä A ja sitten B, sitten C. Poikkeavassa haussa työntekijä menee suoraan järjestelmään C	Nostaa
10. Käyttöoikeudet järjestelmään puutteelliset	Työntekijälle ei ole annettu virallisen prosessin mukaan käyttöoikeutta järjestelmään, vaan pääsy järjestelmään on saatu muuta kautta. Myös tilanne, jossa työntekijälle on jäänyt vanhasta roolista käyttöoikeudet järjestelmään, johon ei niitä uudessa roolissaan tarvitse	Nostaa
11. Observointi on kohdistunut valtuutetun käyttäjän henkilökohtaiseen sidosryhmään	Työntekijä on tunnistetusti hakenut esimerkiksi hänen läheisensä tietoja organisaation tietojärjestelmästä	Nostaa

Edellä olevassa taulukossa (Taulukko 7) esitetyt suodattimet muodostavat tutkielman ensimmäisen version suodatinmallista, jota iteroidaan tutkielman seuraavassa alaluvussa.

6.5 Suodatinmallin arviointi ja edelleen kehitys

Tässä alaluvussa esitellään kohdassa 6.4 esitellyn suodatinmallin ensimmäisen version iterointi ja arviointi. Nämä toteutettiin keräten palautetta, arviointia ja ajatuksia jo ensimmäisellä kierroksella haastatelluilta henkilöiltä. Iteroinnin tarkoituksena oli edelleen kehittää mallia, jotta lopputulos olisi paras mahdollinen. Iterointivaiheessa pyrittiin myös löytämään merkittäväksi havaitulle sidosryhmiin kohdistuvalle suodattimelle toimintaraameja.

Suodatinmallin iteroinnin datan analysointivaiheessa tunnistettiin kolme keskeistä teemaa mallin edelleen kehityksen kannalta. Teemat keskittyivät esitellyssä mallissa olevien suodattimien kehittämiseen, uusien suodattimien luomiseen, sekä mallin soveltuvuuden arviointiin.

Iterointikierroksella haastateltavia pyydettiin arvioimaan esitetyn mallin soveltuvuutta toimeksiantajaorganisaation valtuuttamattomien hakujen tunnistamisen lokianalytiikan kontekstiin. Haastateltavat kokivat, että suodatinmalli soveltuu edellä esitettyyn tarkoitukseen ja ympäristöön. Osa haastateltavista toi ilmi, että tutkimus on lukijoilleen aiheesta ajatuksia herättävä.

> ”Suodatinmallista on varmasti hyötyä hakujen tunnistamisessa. Kun erilaiset tilanteet on eri suodattimiksi eritelty, se helpottaa valtuuttamattomien henkilötietohakujen tunnistamista.”

> ”Kyllä varmasti tämä suodatinmalli auttaa tunnistamaan valtuuttamattomia tapahtumia, tai ainakin niitä tapahtumia jotka suuremmalla todennäköisyydellä ovat valtuuttamattomia.”

> ”Suodatinmalli näyttää hyvältä ja selkeyttää huomattavasti keinoja, millä perusteella valtuuttamattomia henkilötietohakuja voidaan tunnistaa.”

Suodatinmallin todettiin olevan soveltuva poikkeavien ja edelleen valtuuttamattomien hakujen tunnistamiseen. Kuitenkin todettiin, että organisaatioiden, jotka mallia hyödyntäisivät tulisi tehdä tarkemmat määritellyt jokaisen suodattimen kohdalle kontekstin mukaisesti. Lisäksi on selkeää, että suodattimia tulee tarkentaa vielä sen mukaan, miten niiden havaitaan käytännössä toimivan.

6.5.1 Esiteltyjen suodattimien kehitys

Iterointivaiheessa nousi esille pohdintoja artefaktin ensimmäisessä versiossa esitellyistä suodattimista. Selvisi, että osaa suodattimista voisi edelleen kehittää, laajentaa, tai tarkentaa, sekä osan kohdalla voitaisiin punnita niiden oikeellisuutta.

Ensimmäisessä versiossa suodattimeen 1. Poikkeava observoinnin aika -kohdistettiin kehitysehdotus sen laajentamisesta. Laajennuksen mukana suodattimessa huomioitaisiin myös työntekijän loma- ja poissaoloajat, sekä mahdolliset vuorolistat. Eli lähtökohtaisesti observointeja ei pitäisi tapahtua loma- ja poissaoloaikana. Vuorolistat olisivat lisätarkennus roolista poikkeavaan työaikaan.

Ensimmäisen version suodattimen 6. observoinnille löytyy asiakaslähtöinen peruste – liittyen nousi useampikin huomio esille, että peruste voi olla myös työnkuva lähtöinen, ja myös näille voi löytyä kirjaus. Tämä tulisi tuoda selkeämmin ilmi suodattimessa. Lisäksi esille nousi huomio siitä, että asiakaslähtöisen perusteen tulee sijoittua lähelle observoinnin ajankohtaa. Toisaalta tätä pidettiin myös jossain määrin itsestäänselvyytenä, jotta asiakaslähtöinen peruste olisi pätevä.

> ”Ehkä kohtaan 6 voisi myös lisätä työperusteiset tapahtumahaut. Tällöin ei välttämättä ole asiakaslähtöistä perustetta.”

> ”Tiketillä tarkoitan mitä vaan virallista tietojärjestelmään kirjattua työtoimeksiantoa. […] Pointti siinä, että joku valtuuttaa työntekijän hoitamaan jonkun asiakkaaseen liittyvän tehtävän.”

Ensimmäisen mallin version suodattimesta 11. Observointi on kohdistunut valtuutetun käyttäjän henkilökohtaiseen sidosryhmään – pyydettiin erikseen tarkennuksia haastateltavilta henkilöiltä. Iteroinnista oli keskeistä saada kyseisen suodattimen kannalta palautetta siitä, millaisilla märittelyillä suodatinta voitaisiin hyödyntää, eli miten sidosryhmiä olisi mahdollista tunnistaa. Tämä suodatin synnytti pohdintoja siitä, miten se voitaisiin todellisuudessa toteuttaa, rikkomatta kenenkään yksityisyyttä. Yksi haastateltava nosti esille, että todennäköisesti tällaiset haut, joissa haetaan sidosryhmään kuuluvia henkilöitä, tulisivat esille jo suodattimien 1-4 kautta, mikäli haku olisi systemaattista. Suodattimen määrittelyiksi nousi esille asiakkaan sukunimen vertaaminen hakijan sukunimeen, jolloin sukulaissuhteita olisi mahdollista tunnistaa. Kuitenkin esille nousi myös se, että myös sukulaisten tietoja voidaan käsitellä valtuutetusti. Lisäksi nousi esiin myös työpaikan ulkopuolisiin sidosryhmiin yhdistävä katuosoite, sekä tiimi ja käyttäjätunnus tilanteessa, jossa haetaan kollegan tietoja käyttäjätunnuksella. Suodattimen tarkastelun arvoja ja niiden pätevyyttä myös lain näkökulmasta tulisi ehdottomasti tarkastella syvemmin organisaation sisällä. Haastateltavien pohdinnoissa nousi esille, että suodatin voisi olla vasta niin sanottu toisen asteen suodatin, jonka kautta suodatettaisiin vain jos tapahtuman painoarvo on kasvanut toisessa suodattimessa.

> ”Lähipiiri konseptina on tietysti laaja ja sitäkin asiaa voisi pohtia pitkään ja hartaasti. Sukulaiset, kaveri, tai jopa julkkikset. Näissä kaikissa on se ongelma, että miten generalisoida lähipiirin ilman tarkkoja ”nimilistoja”.”

> ”Tätä ei ehkä voi käyttää suoraan vaan vasta kun jokin muu suodatin nostaa oudon käytöksen esille. Tämä vaatii selvästi pohdintaa, mutta kannattaa pitää mukana. Mahdollisia rajoja voi/pitää hakea myös yrityksen privacy/legal osastojen kanssa.”

> ”Tämä olisi erittäin tarpeellinen, mutta siinä tulee vastaan lainsäädännöllisiä esteitä. Mikäli Suomen lainsäädäntö mahdollistaa, tämän kehittäminen edes jossain muodossa olisi hyvin tarpeellista.”

Edellä esitetyt sidosryhmän tunnistamisen esimerkit tulisi käydä läpi esimerkiksi organisaatioiden lakiosastojen kanssa ja oleellisten toimijoiden tulisi yhdessä määrittää, mitkä sidosryhmän tunnistamisen metodit ovat mahdollisesti hyödynnettävissä suodatinmallin puitteissa.

Ensimmäisen version suodatin 10 osalta nousi myös keskustelua sen laajentamisesta niin, että se ottaisi huomioon myös organisatorisen tekijän. Tarkoittaen, että työntekijän roolia verrattaisiin käytettävän järjestelmän käyttöoikeuteen poikkeavuuksien havaitsemiseksi. Havaintoja voitaisiin tehdä esimerkiksi vertaamalla yksikön työntekijöiden rooleja ja käyttöoikeuksia toisiinsa. Myös haastateltava tunnisti suodattimen taustoittamisen monimutkaisuuden, ja on selkeää, että koneoppiminen olisi keskeisessä roolissa myös tämän suodattimen tehostamisessa.

> ”Yksinkertaisesti ajatellen voi outoutta olla esim. se, että laskutussysteemeihin olisi oikeus myyntiorganisaatiosta tai johtajalla olisi jokin tietokanta-admin rooli.”

Esitellyt suodattimien edelleen kehitys ideat on huomioitu suodatinmallin lopullisessa versiossa. Suodatinmallin lopullinen versio esitellään osiossa 6.5.3.

6.5.2 Uudet suodattimet ja mallin jatkokehitys

Suodatinmallin lopullisen versioon löydettiin myös lisättäviä suodattimia. On kuitenkin selvää, että suodatinmallia tulisi edelleen kehittää myös jatkossa. Keskeiseksi havaittiin ymmärrys siitä, mitä kaikkea dataa oikeastaan voisikaan yhdistellä lokianalytiikkaan. Lisäksi tärkeä havainto oli, että suodattimien toiminnan monitoroinnin kautta niiden hyödyllisyyttä voidaan edelleen kehittää ja tehostaa.

> ”Uusia ideoita voisi syntyä, jos pystyisi hahmottamaan mitä kaikkea muuta dataa yritys tuottaa, ja jonka voisi korreloida tuottamaan tätä lisäarvoa.”

> ”Tarkennuksia ja algoritmin ehtoja on syytä tarkastella jatkuvasti, ja niihin voidaan tehdä muutoksia joustavasti. Tarkennuksia tulee yleensä esimerkiksi suodattimen herkkyyteen, eli miten helposti ja suuria määriä se nostaa tiettyä tapausta manuaalisen analysoinnin piiriin. Todellisten käyttötapausten ja analysoitavien esimerkkien avulla tarkastelua tehdään säännöllisesti.”

Analysointivaiheessa havaittiin, että myös suodatin, joka nostaa esille tapahtumia, joita on observoinnit poikkeuksellisen moni työntekijä, olisi tärkeä. Suodatin voisi nostaa esille tapahtumia, joissa on esimerkiksi haettu jonkin julkisuudessa esillä olleen henkilön tietoja.

> ”Jos samaa kohdetta katsoo lyhyellä aikavälillä poikkeavan suuri määrä työntekijöitä. Liittyen esim. mediassa pinnalla olevien tapausten ja/tai kuuluisien henkilöiden tietojen aiheettomiin hakuihin.”

Esille nousi tarve tarkistella myös järjestelmän ominaisuuksien mukaan suodattimista, sen mukaan mitkä ominaisuudet järjestelmässä koettaisiin keskeisiksi valtuuttamattomien hakujen tunnistamisen kannalta. Eli järjestelmän ominaisuudet toimisivat luokittelijoina, jotka saattaisivat nostaa tai laskea tapahtumien manuaalisen tarkastelun arvoa, riippuen järjestelmän ominaisuuksien painoarvosta. Esimerkiksi järjestelmät, joissa on sensitiivisempää dataa tallessa, olisivat korkeamman prioriteetin järjestelmiä tarkasteluun. Järjestelmän ominaisuuksien mukaan voitaisiin mahdollisesti tehdä useampiakin suodattimia, tai yksi suodatin johon määritellään eri järjestelmien ominaisuuksien mukaisia tapauksia, jotka joko nostavat tai laskevat tapahtuman manuaalisen tarkastelun painoarvoa. Kuitenkaan tämä suodatin ei suoraan indikoi poikkeavuuksia, joten se ei täysin asetu tämän suodatinmallin skaalaan. Kuitenkin on noteerattava, että tällaista järjestelmäominaisuuksien mukaista luokittelua tulisi tarpeen mukaan tehdä lokianalytiikassa.

> ”Ehkä voisi mainita mahdolliset järjestelmäkohtaiset suodattimet liittyen järjestelmän ominaisuuksiin, henkilötiedon määrään järjestelmässä, käyttäjien määrään etc.”

Analysointivaiheessa havaittiin, että lokianalytiikkaa olisi tärkeää kohdentaa myös ensimmäisissä haastatteluissa esiin nousseeseen ”villikortti”-tarkastelutapaan. Tällä tarkoitetaan järjestelmiin tehtäviä laajoja hakuja, kuten esimerkiksi pelkällä puhelinnumeron alkuosalla tehtävää hakua. Lisäksi esille nousi, että lokianalytiikassa tulisi tarkastella myös mahdollisia suuria määriä järjestelmästä ulos vietäviä tapahtumia. Nämä eivät kuitenkaan asetu tämän suodatinmallin skaalaan. Kuitenkin näidenkin huomioon ottaminen lokianalytiikassa tulee noteerata.

Analysoinnissa selvisi myös, että suodattimia tulisi ehdottomasti tarkentaa suodatinmallia hyödyntävän organisaation sisällä. Suodattimiin tulee tehdä tarkemmat määrittelyt niin, että ne ovat oikeat organisaation kontekstiin. Lisäksi esille nousi myös se, että organisaation sisällä suodattimien painoarvoa voitaisiin säädellä niin, että joistakin suodattimista manuaalisen tarkastelun painoarvo olisi korkeampi, kuin jostakin toisesta.

> ”Suodattimilla voisi olla erilaisia painoarvoja. Esim. 5. Poikkeava observoinnin kohde (yritysasiakaspalvelija hakee kuluttaja-asiakkaan tietoja) ja 9. Poikkeava tapahtumajärjestys voisivat nostaa tarkemman tarkastelun arvoa vähemmän kuin esim 1. Poikkeava observoinnin aika.”

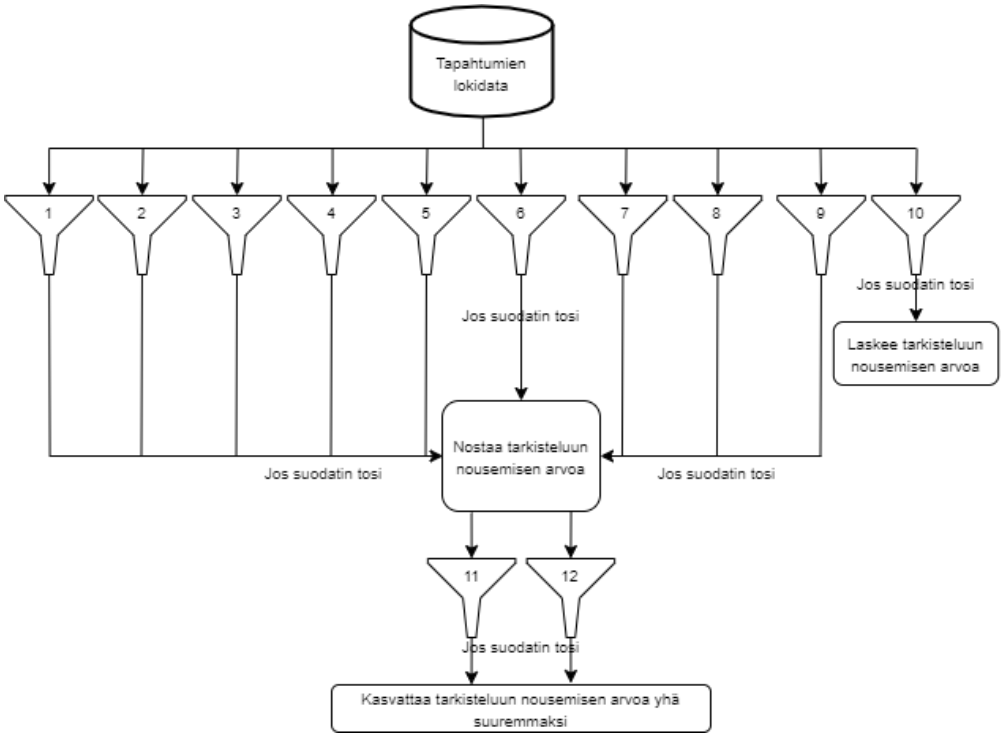
6.5.3 Suodatinmallin lopullinen versio

Suodatinmalli on luotu kokoamalla yhteen teoriasta johdetut suodattimet, ensimmäisistä haastatteluista tunnistetut suodattimet ja iteroinnin kautta

tunnistetut suodattimet. Osaa suodatinmallin ensimmäisessä versiossa esitellyistä suodattimista on edelleen jalostettu iterointivaiheen perusteella parhaan lopputuloksen aikaan saamiseksi. On tärkeää tunnistaa, että suodattimet eivät itsessään määritä minkään tapahtuman olevan valtuuttamaton. Suodattimet mahdollistavat poikkeavien tapahtumien tunnistamisen ja nousemisen manuaaliseen tarkasteluun, jolloin tarkistetaan, löytyykö henkilötietojen haulle valtuutettua perustetta.

Suodatinmallin iteroinnin datan analysointivaiheessa selvisi, että osa suodattimista tulisi mahdollisesti olla niin sanottuja toisen vaiheen suodattimia, jolloin ensimmäisestä suodattimesta manuaaliseen tarkasteluun nousemiseen painoarvoa kerännyt tapahtuma menisi toisen suodattimen läpi, mikä voisi edelleen nostaa painoarvoa. Tällaisessa tapauksessa tapahtuma nousisi suurella todennäköisyydellä tarkasteltavaksi. Tällaisia suodattimia olivat aiemmin tunnistetusti valtuuttamattoman haun tehneen työntekijän suodatin, sekä työntekijän henkilökohtaisen sidosryhmän suodatin.

Suodatinmallin lopullinen versio koostuu 12 suodattimesta (kuvio 5). Suodatinmallin lopullisessa versiossa suodattimia on tarkennettu, lisätty, sekä uudelleen järjestelty ensimmäiseen versioon verraten. Suodatinmallin suodattimet 1-9 nostavat manuaaliseen tarkasteluun nousemisen arvoa, mikäli tapahtumat indikoivat poikkeaman havaitsemista. Suodatin 10 puolestaan laskee manuaaliseen tarkasteluun nousemisen arvoa, mikäli suodatin löytää tapahtumalle kirjatun perusteen ja indikoi näin ollen tapahtuman olleen valtuutettu. Mikäli suodattimista 1-9 on noussut esille poikkeama, menevät tapahtumat vielä suodattimien 11 ja 12 läpi, jotka voivat edelleen nostaa manuaalisen tarkastelun painoarvoa.



KUVIO 5 Suodatinmallin lopullinen versio

Edellä esitetyssä kuviossa on esitetty suodatinmallin lopullinen versio. Kuvio antaa visuaalisen taustoituksen suodatinmallin toiminnasta ja helpottaa näin ollen suodatinmallin toiminnan ymmärretystä. Suodattimet on kuvattu tarkemmin alla esitetyssä taulukossa (Taulukko 8), jossa kerrotaan kunkin suodattimen nimi, selite, sekä manuaalisen tarkastelun painoarvon vaikutus. Suodattimien selitteet kuvaavat mihin pohjautuvia poikkeavuuksia suodatin pyrkii tapahtumien lokimassasta havaitsemaan.

TAULUKKO 8 Lopullisen suodatinmallin selitteet

Suodatin	**Selite**	**Nostaa/laskee tarkemman tarkastelun arvoa**
1. Poikkeava observoinnin aika	Observoinnin aika poikkeaa työntekijän rooliin suhteutetusta työajasta. Esimerkiksi toimistoaikana työskentelevä työntekijä toteuttaa haun yöaikaan. Myöskin vuorolista, sekä loma- ja poissaoloaikojen kanssa ristiriidassa olevat haut. Nämä mahdollistavat sellaisen observoinnin tunnistamisen, joka on toteutettu aikana, jolloin työntekijä ei ole mahdollisesti ollut työtehtävissä	Nostaa
2. Poikkeava observoinnin tiheys yhteen kohteeseen lyhyellä aikavälillä	Työntekijä observoi poikkeavan tiheästi kohdetta lyhyen aikavälin sisällä	Nostaa
3. Poikkeava observoinnin tiheys yhteen kohteeseen pitkällä aikavälillä	Työntekijä observoi poikkeavan tiheästi kohdetta pitkällä aikavälillä. Esimerkiksi työntekijä hakee puolen vuoden aikana joka kuukausi saman henkilön tietoja	Nostaa
4. Poikkeava observoinnin tiheys useaan kohteeseen	Hakukentän iterointi. Työntekijä hakee tiheään tahtiin eri henkilötiedoilla tietojärjestelmistä	Nostaa
5. Poikkeavan suuri määrä työntekijöitä observoi samaa kohdetta	Samaa kohdetta observoi poikkeavan suuri määrä työntekijöitä lyhyellä aikavälillä	Nostaa
6. Poikkeava observoinnin kohde	Työntekijä observoi hänen roolistaan poikkeavaa kohdetta. Esimerkiksi yritysasiakaspalvelija hakee kuluttaja-asiakkaan tietoja	Nostaa
7. Poikkeava hakuparametri	Työntekijä toteuttaa haun poikkeavalla parametrilla, jolla ei tyypillisesti haeta tietojärjestelmistä	Nostaa
8. Poikkeava tapahtumajärjestys	Työntekijä toteuttaa toimia poikkeavassa tapahtumajärjestelmässä. Esimerkiksi työntekijä hakee normaalisti ensin järjestelmästä A ja sitten B, sitten C. Poikkeavassa haussa työntekijä menee suoraan järjestelmään C	Nostaa

(jatkuu)

Taulukko 8 (jatkuu)

9. Käyttöoikeudet järjestelmään puutteelliset tai poikkeavat	Työntekijälle ei ole annettu virallisen prosessin mukaan käyttöoikeutta järjestelmään, vaan pääsy järjestelmään on saatu muuta kautta. Myös tilanne, jossa työntekijälle on jäänyt vanhasta roolista käyttöoikeudet järjestelmään, johon ei niitä uudessa roolissaan tarvitse, tai käyttöoikeus poikkeaa roolille tyypillisestä oikeudesta	Nostaa
10. Observoinnille löytyy työ- tai asiakaslähtöinen peruste	Työntekijän toteuttamalle haulle löytyy asiakkaan yhteydenotto perusteeksi, tai muu kirjattu työtoimeksianto	Laskee
11. Aiemmin valtuuttamattoman haun tehnyt työntekijä toteuttaa haun	Aiemmin tunnistetusti valtuuttamattomia hakuja tehnyt työntekijä tekee haun. Toisen asteen suodatin.	Nostaa
12. Observointi on kohdistunut valtuutetun käyttäjän henkilökohtaiseen sidosryhmään	Työntekijä on mahdollisesti hakenut esimerkiksi hänen läheisensä tietoja organisaation tietojärjestelmästä. Toisen asteen suodatin.	Nostaa

Edellä olevassa taulukossa (Taulukko 8) esitetyt suodattimet muodostavat tämän tutkielman puitteissa lopullisen suodatinmallin eli tutkielman artefaktin. Suodattimet ovat pyrkineet huomioimaan kokonaisvaltaisesti mahdollisia käyttötapauksia, joissa valtuutettu käyttäjä voisi suorittaa valtuuttamattomia hakuja. Mikäli suodattimia sovelletaan käytännössä lokianalytiikkaan, tulee niitä tarkentaa organisaation kontekstiin sopivaksi, tarjolla olevien mahdollisuuksien ja datan perusteella. Lisäksi organisaatioiden tulee itse tehdä määritellyt siitä, miten suuren osan tapahtumia he ottavat manuaaliseen tarkasteluun ja tarkistellaanko manuaalisesti myös tapahtumia, joita suodatinmalli ei nosta esille. Lisäksi on huomioitava, että suodattimien painoarvot voidaan määritellä eri tasoisiksi, kuten kerätyn datan analysointivaiheessa on esitetty.

7 JOHTOPÄÄTÖKSET JA POHDINTA

Tässä luvussa kootaan yhteen tämän Pro gradu-tutkielman oleelliset tulokset, sekä niiden myötä myös tutkielman kontribuutio. Ensimmäisessä alaluvussa käsitellään ja kootaan yhteen tutkielman tulokset teoreettisen ja empiirisen osion osalta. Alaluvussa 7.2 käsitellään tulosten merkityksellisyyttä. Alaluvussa 7.3 puolestaan pohditaan tutkielman rajoitteita, sekä esitetään mahdollisia relevantteja jatkotutkimusaiheita.

7.1 Tutkielman tulokset

Maailman digitalisoituminen on tuonut paljon mahdollisuuksia, mutta näiden mahdollisuuksien myötä myös paljon haasteita. Suuri osa organisaatioista toimii nykyisin päivittäin digitaalisten työkalujen ympärillä ja käsiteltävän datan määrä on moninkertaistunut. Tämä murros on vaikuttanut myös yksilöihin, sillä heistä kerätään enemmän dataa, kuin koskaan aiemmin. Euroopan Unionin Yleinen tietosuoja-asetus GDPR on tuonut paljon tarvittua turvaa yksilön henkilötiedoille, mutta se on myös tuonut organisaatioille tarpeen varmistua asiakkaidensa tietojen asianmukaisesta käsittelystä, ja siitä etteivät työntekijät toteuta valtuuttamattomia henkilötietohakuja. Keskeiseksi keinoksi valtuuttamattomien hakujen monitoroinnissa on havaittu lokianalytiikka. Lokianalytiikan yhdeksi haasteeksi on kuitenkin tunnistettu suuret lokimäärät, joiden seasta valtuuttamattomien hakujen tunnistaminen voi olla haastavaa. Tämä Pro gradu-tutkielma toteutettiin toimeksi antona yhdelle Suomen suurimmista teleoperaattoreista.

Tutkielmassa esitettiin kolme tutkimuskysymystä. Varsinaisen tutkimuskysymyksen motivaatiota tukevat ja taustoittavat tutkimuskysymykset:

- Miksi sisäiseen toimintaan kohdistuvaa lokianalytiikka tulisi tehdä?
- Miksi lokien laadulla on vaikutus lokianalytiikan toteuttamiseen?

Sekä tutkielmaa johdattanut, edellä kuvattuun ongelmaan pureutuva tutkimuskysymys, jolla pyrittiin löytämään suodattimia, jotka voisivat auttaa nostamaan valtuuttamattomia hakuja esille suuresta lokitietojen massasta:

- Millaisilla suodattimilla valtuuttamattomien henkilötietohakujen tunnistamista lokianalytiikassa voidaan tehostaa?

Seuraavissa alaluvuissa tullaan kokoamaan yhteen tutkimuskysymyksiin vastaukset, jotka on johdettu tutkielmassa empiirisessä osiossa kerätystä datasta ja sen analyysistä, peilaten tuloksia teoriaosion havaintoihin.

7.1.1 Lokianalytiikan keskeisyys

Ensiksi tutkielmassa tarkasteltiin organisaation sisäisiä toimijoita ja heidän asettamiaan uhkia tietoturvallisuudelle. Pystyttiin havaitsemaan, että sisäiset toimijat ovat yksi oleellisimmista uhista, joita organisaatioiden tietoturvallisuus kohtaa. Tämän jälkeen käsiteltiin, miten Yleinen tietosuoja-asetus GDPR on tuonut entistä vankemman tarpeen organisaatioille pystyä varmistumaan työntekijöidensä asianmukaisesta toiminnasta koskien asiakkaidensa tietoja. Seuraavaksi tutkielma käsitteli lokitietoja, lokianalytiikkaa ja sen merkitystä organisaatioille.

Sekä tutkielmassa esitetty teoria, että empiirisen osion analyysi osoittivat lokianalytiikan keskeisen merkityksen organisaation tietoturvallisuudelle. Tunnusten- ja pääsynhallinta on keskeistä tietoturvallisuuden kannalta, mutta yksinään ne eivät ole riittävä toimenpide. Organisaatioiden tietojärjestelmissä työntekijöiden toteuttamien valtuuttamattomien toimien toteaminen olisi käytännössä mahdotonta ilman, että lokeille tallentuisi järjestelmissä toteutetut toimet ja että näitä toimia analysoitaisiin. Ilman lokianalytiikkaa olisi periaatteessa mahdotonta varmistua esimerkiksi osasta GDPR:n asettamia vaatimuksia.

Teoriassa ja empiirisessä osiossa havaittiin myös lokianalytiikan toteuttamisen merkitys ennaltaehkäisevänä tekijänä. Molemmissa osioissa voitiin todeta, että tieto siitä, että valtuuttamattomien hakujen tunnistamiseen kohdistuvaa lokianalytiikkaa tehdään, estää omalta osaltaan työntekijöitä toteuttamasta valtuuttamattomia hakuja.

Valtuuttamattomien toimien todentamisen ja ennaltaehkäisemisen lisäksi, lokianalytiikalla on myös muita kiistämättömiä hyötyjä organisaation näkökulmasta. Tällaiseksi havaittiin teoriaosiossa esimerkiksi lisäkoulutuksen tarpeen toteaminen, esimerkiksi tilanteissa, joissa lokianalytiikassa huomataan puutteita tietyn työntekijän tai osaston toimintatavoissa. Empiirisessä osiossa puolestaan nousi esiin prosessivirheiden tunnistaminen, jotka voivat kertoa esimerkiksi työkalujen tai tapojen päivittämisen tarpeesta. Erityisesti empiirinen tutkimus korosti myös työntekijöiden turvaa, jonka lokianalytiikka tuo. Tämä on keskeistä tilanteissa, joissa lokeja analysoimalla voidaan todistaa työntekijän toimineen oikein.

7.1.2 Lokien laadun merkitys lokianalytiikassa

Tutkielman kirjallinen osuus, sekä empiirinen osuus osoittivat molemmat, että lokien laatu todella sisältää paljon haasteita, jotka voivat olla merkittävä tekijä lokianalytiikan tehokkuuden kannalta. Lokien laadun ongelmien havaittiin johtuvan ainakin osittain lokimuodon standardisoinnin puutteesta. Lisäksi tietojärjestelmissä tallentuvien tapahtumien määrä on kasvanut merkittävästi, lisäten myös lokitetun tiedon määrää. Empiirinen tutkimus nosti esiin myös sen, että useat tietojärjestelmät ovat vanhoja ja niiden lokitusmekanismit on rakennettu kauan ennen GDPR:n julkaisua. Tämä on omalta osaltaan aiheuttanut haasteita, sillä lokitusta ei ole järjestelmissä rakennettu näistä näkökulmista ja lokitusmekanismeja on pitänyt tarkastella ja edelleen kehittää jälkikäteen, jotta ne vastaisivat nykypäivän tarpeita.

Tutkielman kirjallisen osuuden ja empiirisen osuuden tuloksena voidaan todeta, että lokitietojen laadun merkitys on todella keskeistä lokianalytiikan tehokkuuden kannalta. Lokituksen tulee olla ennalta suunniteltu prosessi, jossa on todella pohdittu mitä kaikkea tietoa tarvitaan. Suuret lokimassat aiheuttavat todella paljon haasteita lokien käsittelemiseen, mikäli lokitettu tieto ei ole relevanttia. Kaikkien tietojen lokittaminen johtaa helposti keskeisen ja tärkeän tiedon hukkumiseen suureen lokimassaan. Toisaalta lokitietojen hyödyntäminen vaarantuu myös, jos jokin relevantti tieto ei ole tallentunut lokiin. Lisäksi epäformaali tai epäjohdonmukainen lokidata voi aiheuttaa haasteita sen käsittelemiseen ja johtaa edelleen lisäresurssien tarpeeseen. Useat organisaatiot vievät järjestelmistä tulevat lokit vielä erilliseen järjestelmään, jossa lokeja analysoidaan ja yhdistellään. Tällainen jatkokäsittely on haasteellista, mikäli järjestelmistä tulevat lokit eroavat merkittävästi standardoinniltaan toisistaan. Myöskin Audit trailin seuraaminen voi vaarantua lokien laadun takia.

On tärkeää tunnistaa, että lokitietojen eheys ja oikeellisuus on taattava. Lokitiedot on havaittu keskeiseksi tekijäksi tietoturvallisuuden varmentamisessa, joten niiden analysoinnin on oltava mahdollisimman tehokasta ja varmaa. Pahimmassa tapauksessa lokidatan potentiaali ja hyödyntäminen on asetettu vaaraan, mikäli lokitietojen laatu ei ole vaaditulla tasolla. Empiirinen osuus korosti sitä, miten lokien laatu saattaa jopa aiheuttaa lokianalytiikassa suurempia haasteita, kuin lokimassojen suuruus. Lokianalytiikan ja tietoturvallisuuden kannalta lokien laatuun tulisi panostaa ja niiden keskeisyys pitäisi tunnistaa ja huomioida.

7.1.3 Suodatinmalli lokianalytiikan tukena

Tutkielman keskeisenä tuloksena voidaan esittää artefaktia, eli suodatinmallia, jonka keskeinen tavoite on mahdollistaa valtuuttamattomien henkilötietohakujen havaitseminen lokianalytiikan keinoin. Suodatinmallin tarkoituksena on esittää suodattimia, joita lokianalytiikassa tarkempaan tarkasteluun nostettavia tapauksia voitaisiin suodattaa niin, että valtuuttamattomat tapaukset nousisivat herkemmin esiin. Näin ollen suodatinmalli ja siinä esitetyt suodattimet vastaavat osaltaan tutkielman keskeisimpään ja johtavaan tutkimuskysymykseen.

Suodatinmallin muodostaminen oli kolmetasoinen prosessi. Suodatinmalliin päätyi tieteellisen kirjallisuuden tarjoamia suodattimia, joiden sopivuutta organisaation valtuuttamattomien hakujen lokianalytiikkaan arvioitiin haastatteluissa. Tämän lisäksi haastatteluiden pohjalta pyrittiin tunnistamaan suodattimia, jotka sopisivat kyseiseen kontekstiin. Näiden pohjalta luotiin suodatinmallin ensimmäinen versio, joka toimitettiin haastatelluille henkilöille iterointia varten. Haastateltavilla oli mahdollisuus antaa mallista palautetta, sekä esittää edelleen kehitysehdotuksia. Jo esiteltyjä suodattimia edelleen kehitettiin ja uusia suodattimia luotiin. Näiden perusteella suodatinmallista muodostettiin tämän tutkielman rajoissa valmis versio.

Suodatinmalli tarjoaa suodattimia, jotka voivat laskea tai nostaa lokianalytiikassa manuaaliseen tarkasteluun nousevien tapahtumien painoarvoa. Suodattimet auttavat tunnistamaan poikkeavuuksia tapahtumista, mitkä voivat edelleen indikoida valtuuttamatonta tapahtumaa. Suodatinmalli ei siis suoraan merkitse valtuuttamattomia tapahtumia, vaan se antaa indikaatiota siitä, onko jokin tapahtuma mahdollisesti ollut poikkeava, jolloin sen manuaalisen tarkastelun painoarvo kasvaa. Osa suodattimista toimii niin sanottuina toisen asteen suodattimina, jolloin ne edelleen nostavat jo suodatuksessa esille nousseen tapahtuman tarkastelun painoarvoa edelleen. Yhteensä suodattimia, jotka nostettiin suodatinmalliin, tunnistettiin 12 kappaletta. Suodattimet olivat: poikkeava observoinnin aika, poikkeava observoinnin tiheys yhteen kohteeseen lyhyellä aikavälillä, poikkeava observoinnin tiheys yhteen kohteeseen pitkällä aikavälillä, poikkeava observoinnin tiheys useaan kohteeseen, poikkeavan suuri määrä työntekijöitä observoi samaa kohdetta, poikkeava observoinnin kohde, poikkeava hakuparametri, poikkeava tapahtumajärjestys, käyttöoikeudet järjestelmään puutteelliset tai poikkeavat, observoinnille löytyy työ- tai asiakaslähtöinen peruste, aiemmin valtuuttamattoman haun tehnyt työntekijä toteuttaa haun, sekä observointi on kohdistunut valtuutetun käyttäjän henkilökohtaiseen sidosryhmään.

7.2 Tutkielman tulosten merkitys

Tämän Pro gradu-tutkielman keskeinen lopputuotos on esitelty poikkeavaa käyttäytymistä esiin nostava suodatinmalli, jonka avulla on mahdollista nostaa esiin valtuutetun käyttäjän suorittamia poikkeavia ja edelleen mahdollisesti valtuuttamattomia hakuja. Suodatinmalli vastaa tutkimusongelman tarpeeseen, mahdollistaen suuren lokimassa suodattamisen niin, että mahdollisesti valtuuttamattomat tapahtumat voivat nousta selkeämmin esiin. Iterointi vaiheessa haastateltavat arvioivat suodatinmallin soveltuvuutta valtuuttamattomien hakujen tunnistamiseen kohdistuvaan lokianalytiikkaan, ja olivat kaikki sitä mieltä, että suodatinmalli tuo hyötyä lokianalytiikkaan.

Tutkielman alussa esitetty tutkimuskysymys koskien lokianalytiikan toteuttamisen perusteluja havaittiin tutkimuksessa tärkeäksi. Tutkielma kattavasti miksi lokianalytiikan toteuttaminen on tärkeää. Tehdyt havainnot tukevat merkittävästi lokianalytiikan toteuttamista. Tutkielman voidaan nähdä tarjoavan

perusteen sille, miksi valtuuttamattomia hakuja tunnistavaan lokianalytiikkaan tulisi organisaatioiden keskittyä. Lokianalytiikan havaittiin olevan oikeastaan ainoa keino, jolla GDPR:n mukaisten henkilötiedon käsittelyn oikeellisuudesta voidaan varmentua. Myöskin Euroopan Unionin tilaama PoSeID-on projekti tukee lokianalytiikan merkityksellisyyttä GDPR:n näkökulmasta. Tämän todentaminen, sekä lokianalytiikan muiden hyötyjen, kuten prosessipuutteiden havaitseminen ja työntekijöiden turvana toimiminen ovat merkittäviä perusteita organisaatioissa toteuttaa lokianalytiikkaa.

Lisäksi tutkielma tarjosi näkökulman siihen, miksi lokien laatuun tulisi keskittyä, mikä edelleen tarjoaa perustelun tämän aihealueen kehittämiseen organisaatioiden kontekstissa. Tutkielma käsitteli lokien laatua GDPR:n kannalta erilaisesta näkökulmasta, kuin mihin suurin osa aiemmasta tieteellisestä kirjallisuudesta on keskittynyt. Aiempi kirjallisuus on fokusoitunut paljolti lokien GDPR:n mukaisuuteen, mutta aiempi kirjallisuus ei ole tarkastellut laajalti lokien laatua valtuutetun käyttäjän tekemien henkilötietohakujen lokianalytiikan näkökulmasta. Tutkielma toi tieteellistä kontribuutiota osoittaen, että vaikka lokien toiminta olisi GDPR:n myötäistä, voi lokien huono laatu edelleen asettaa organisaation tietoturvallisuuden myös tämän osalta riskinalaiseksi. Ylipäätään lokien laadun keskeisyyden osoittaminen lokianalytiikassa oli tutkielmassa lopulta hyvin tärkeässä asemassa.

Tutkielman teoria ja empiirisessä osiossa havaittiin, että koneoppiminen on tehokas keino ymmärtää organisaation työntekijöiden normaalia toimintaa ja näin edelleen määrittää epänormaalin toiminnan raamit. Koneoppiminen on iteratiivinen prosessi, jota esitelty suodatinmalli voi omalta osaltaan tukea. Suodatinmalli tarjoaa poikkeavuuksien tunnistamiseen hyödyllisiä näkökulmia, joita tarkastelemalla poikkeukselliset asiakastietohaut voidaan tunnistaa. Näin ollen tutkielman tulokset tarjoavat myös tukea myös koneoppimisen kehittämistä lokianalytiikassa.

Kokonaisuutena tutkielma toi uusia näkökulmia lokianalytiikan aihepiiriin, sekä tarjosi myös tarvittua perustelua lokianalytiikan toteuttamisen merkityksellisyydestä. Lisäksi tutkielma osoitti lokien laadun tärkeyden tietoturvallisuuden osana. Tutkielmassa esitetty artefakti toi kaivattuja suodattimia, joiden avulla lokianalytiikkaa voidaan tehostaa valtuuttamattomien hakujen tunnistamiseksi. Esitelty suodatinmalli koettiin toimeksiantajaorganisaation haastateltavien mielestä toimivaksi ja hyödylliseksi avuksi lokianalytiikkaan. Ylipäätään tutkielma voidaan nähdä ajatuksia ja uusia näkökulmia herättävänä kokonaisuutena sisäisten uhkakuvien ja lokianalytiikan saralla.

7.3 Tutkielman rajoitteet ja ehdotukset jatkotutkimusaiheiksi

Tutkielmaan kohdistuu joitakin rajoitteita. Samankaltaista suodatinmallia ei tutkielman kirjoittaja löytänyt, joten lopullisen suodatinmallin vertaaminen ei ole mahdollista vastaavanlaiseen tieteellisen kirjallisuuden tarjoamaan malliin.

Tieteellinen kirjallisuus ei ole käsitellyt laajasti tämän tyyppisiä lokianalytiikan työkaluja ja suoraan GDPR:n piirissä olevien toiminteiden käsittelyyn ei löytynyt lokianalytiikasta kovinkaan laajasti näkökulmia.

Tutkielma toteutettiin toimeksiantajaorganisaation organisaatiokontekstissa, ja artefaktin hyödyllisyyttä arvioitiin vain tämän organisaation kontekstissa, joten tutkielman artefaktina syntynyt suodatinmalli ei ole välttämättä yleistettävissä suoraan muihin organisaatiokonteksteihin. Lisäksi on huomioitava, että tutkielman aikana suodatinmallissa esiteltyjä suodattimia ei sovellettu käytännössä lokianalytiikkaan, vaan mallista suoritettiin pelkästään arviointia.

Jatkotutkimusaiheiksi ehdotetaan esitellyn suodatinmallin käytännön soveltamista valtuuttamattomien ja poikkeavien toimien tunnistamiseen kohdistuvaan lokianalytiikkaan, sekä sen toiminnan edelleen arviointia. Lisäksi ehdotetaan, että suodatinmallia edelleen jalostettaisiin ja kehitettäisiin mahdollisuuksien mukaan. Lisäksi jatkotutkimusaihe, jota ehdotetaan, keskittyy lokien laadun parantamiseen juuri tällaisen lokianalytiikan näkökulmasta. Lokien laadulla havaittiin tutkielmassa olevan merkittävä vaikutus lokianalytiikan tehokkuuteen, joten lokien laadun parantaminen on keskeinen tekijä onnistuneen lokianalytiikan toteuttamisessa, joten sen tutkiminen on merkityksellistä ja tärkeää. Lisäksi tutkimusta tulisi kohdentaa myös esille nousseisiin muihin lokianalytiikan osaalueisiin, esimerkiksi tulisi tutkia miten voidaan varmistua työntekijöiden tietojen asianmukaisesta käsittelystä organisaatioissa lokianalytiikan keinoin.

8 YHTEENVETO

Tämä Pro gradu-tutkielma koostuu teoreettisesta ja empiirisestä osiosta. Teoreettinen osuus pyrki tarjoamaan ymmärryksen aihealueeseen, sen tärkeyteen, sekä luomaan teoreettisen pohjan tutkittavalle aiheelle. Tutkielman empiirinen osio on toteutettu suunnittelutieteellisin metodein, pureutuen tutkimuskysymysten käsittelyyn ja artefaktin luomiseen konstruktiivisella tutkimusotteella. Empiirinen tutkimus toteutettiin puolistrukturoituina teemahaastatteluina, joiden perusteella luotiin ensimmäinen versio tutkielman artefaktista. Tätä seurasi vielä iterointi, jossa artefaktia arvioitiin ja edelleen kehitettiin haastateltavien avustuksella.

Tutkielma pureutui haasteeseen tunnistaa valtuutetun käyttäjän valtuuttamattomia hakuja organisaation tietojärjestelmissä. Valtuutetulla käyttäjällä on oikeellisesti pääsy dataan, mutta mikäli hän käsittelee dataa ilman oikeellista perustetta, on toimi valtuuttamaton. Erityisen tarpeen valtuutetun käyttäjän valtuuttamattomien henkilötietohakujen tunnistamiseen on tuonut Euroopan Unionin Yleinen tietosuoja-asetus GDPR. Johtuen kuitenkin valtuutetun käyttäjän oikeellisesti omaavasta pääsystä dataan, on valtuuttamattomia toimia hankala tunnistaa. Keskeiseksi keinoksi tällaisten toimien tunnistamisessa on havaittu lokianalytiikka.

Tutkielma osoitti lokianalytiikan keskeisen merkityksen sisäisten toimijoiden valtuuttamattomien hakujen tunnistamisessa. Ilman lokianalytiikkaa tietojärjestelmissä tapahtuvia toiminteita olisi käytännössä mahdoton monitoroida ja varmentaa, jolloin organisaatioiden, että niiden asiakkaiden tietoturva olisi uhattuna. Lisäksi tutkielma osoitti, että lokianalytiikka tarjoaa organisaatioille myös muista näkökulmista arvokasta tietoa, esimerkiksi paljastaen prosessivirheitä ja tarjoten työntekijöille turvaa tietojärjestelmissä oikeellisesti toimimiseen.

Tutkielma tarjosi myös vankat perustelut sille, miksi lokien laatu tulisi huomioida ja miksi lokien laatua voidaan nähdä jopa kriittisenä tekijänä lokianalytiikan tehokkuuden ja onnistumisen kannalta. Todettiin, että lokien heikko laatu uhkaa lokianalytiikan tehokasta toteuttamista, sekä näin ollen asettaa organisaation tietoturvallisuuden vaaraan.

Tutkielman keskeinen tarkoitus oli löytää käytännön suodattimia, joiden avulla valtuutetun käyttäjän toteuttamiin valtuuttamattomiin hakuihin voitaisiin päästä tehokkaammin kiinni lokianalytiikan keinoin. Teoriaosiossa havaittiin, että usein työntekijöiden valtuuttamattomat toimet poikkeavat normaalista käyttäytymisestä, joka edelleen indikoi, että poikkeavuuksien tunnistaminen tapahtumista on tärkeää. Keskeiseksi havaittiin suodattimien vertaaminen ja suhteuttaminen työntekijän rooliin. Malli esittää 12 suodatinta, joiden avulla voidaan havaita työntekijöiden toteuttamia poikkeuksellisia hakuja, jotka edelleen voivat paljastaa valtuuttamattomia toimia.

LÄHTEET

Accorsi, R. (2009, September). Safe-keeping digital evidence with secure logging protocols: State of the art and challenges. In *2009 Fifth International Conference on IT Security Incident Management and IT Forensics* (pp. 94-110). IEEE.

Ambre, A., & Shekokar, N. (2015). Insider threat detection using log analysis and event correlation. Procedia Computer Science, 45, 436-445.

Basin, D. k., Schaller, P. k. & Schläpfer, M. k. (2011). *Applied Information Security: A Hands-on Approach*. Springer Berlin Heidelberg.

Bishop, M., & Gates, C. (2008). Defining the insider threat. In *Proceedings of the 4th annual workshop on Cyber security and information intelligence research: developing strategies to meet the cyber security and information intelligence challenges ahead* (pp. 1-3).

Bulpett, B. (2020). Safeguarding against the insider threat. *Network security, 2020*(6), 14-17. https://doi.org/10.1016/S1353-4858(20)30068-4

Casaleiro, R. (2020). Protection and control of personal identifiable information: The PoSeID-on approach. *Journal of Data Protection & Privacy*, *3*(2), 199-228.

Carasso, D. (2012). *Exploring splunk*. New York, USA: CITO Research.

Chabinsky, S. (2018). GDPR: Will Your Company Be Fine or Fined? *Security, 55*(5), 30.

Chapple, M. (2020). *Access control, authentication, and public key infrastructure*. Jones & Bartlett Publishers.

Colwill, C. (2009). Human factors in information security: The insider threat–Who can you trust these days?. *Information security technical report*, *14*(4), 186-196.

CORDIS. (2021). Protection and control of Security Information by means of a privacy enhanced Dasboard. EU research results. Haettu 4.4.2021 osoitteesta: https://cordis.europa.eu/project/id/786713

Das, D., Schiewe, M., Brighton, E., Fuller, M., Cerny, T., Bures, M., ... & Tisnovsky, P. (2020, October). Failure Prediction by Utilizing Log Analysis: A Systematic Mapping Study. In *Proceedings of the International Conference on Research in Adaptive and Convergent Systems* (pp. 188-195).

Dodge Jr, R. C., Ferguson, A. J. & Cappelli, D. M. (2013). Introduction to Insider Threat Modeling, Detection, and Mitigation Minitrack. https://doi.org/10.1109/HICSS.2013.308

Doherty, N. F., & Fulford, H. (2005). Do information security policies reduce the incidence of security breaches: an exploratory analysis. *Information Resources Management Journal (IRMJ)*, *18*(4), 21-39.

Egress. (2020). Insider Data Breach Survey 2020. Opinion matters.

Elifoglu, I., Abel, I. & Taşseven, Ö. (2018). Minimizing Insider Threat Risk with Behavioral Monitoring. *Review of Business, 38*(2), 61-73.

ENISA. (18.4.2021). Risk level assessment – Security measures. Haettu osoitteesta https://www.enisa.europa.eu/risk-level-tool/help

ENISA, (2017). Handbook on Security of Personal Data Processing. European Union Agency For Network and Information Security.

Faraj, S. & Sambamurthy, V. (2006). Leadership of information systems development projects. *IEEE Transactions on Engineering Management*, 53(2), 238–249.

Ferreira, M. B., & Alonso, K. C. (2013). Identity management for the requirements of the information security. In *2013 IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 53-57). IEEE.

Foroughi, F. (2008, July). The application of system dynamics for managing information security insider-threats of IT organization. In *Proceedings of the World Congress on Engineering* (Vol. 1, pp. 2-4).

General Data Protection Regulation (GDPR). (2016). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Gonçalves-Ferreira, D., Leite, M., Santos-Pereira, C., Correia, M. E., Antunes, L. & Cruz-Correia, R. (2018). HS.Register - An Audit-Trail Tool to Respond to the General Data Protection Regulation (GDPR). *Studies in health technology and informatics, 247*, 81.

Hamooni, H., Debnath, B., Xu, J., Zhang, H., Jiang, G., & Mueen, A. (2016, October). Logmine: Fast pattern recognition for log analytics. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 1573-1582).

Harkins, M. k. (2013). *Managing Risk and Information Security: Protect to Enable*. Apress.

He, S., Zhu, J., He, P., & Lyu, M. R. (2020). Loghub: a large collection of system log datasets towards automated log analytics. *arXiv preprint arXiv:2008.06448*.

He, S., Zhu, J., He, P., & Lyu, M. R. (2016, October). Experience report: System log analysis for anomaly detection. In *2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE)* (pp. 207-218). IEEE.

Hossain, S. (2019). A case study on managing customer data to comply with GDPR.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. Qualitative health research, 15(9), 1277-1288.

Iso-Britannian kyberturvallisuuskeskus. (8.7.2018). Introduction to logging for security purposes. Haettu 25.2.2021 osoitteesta https://www.ncsc.gov.uk/guidance/introduction-logging-security-purposes

Jayathilake, D. (2012, May). Towards structured log analysis. In *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)* (pp. 259-264). IEEE.

Jiang, K., & Cao, X. (2011). Design and implementation of an audit trail in compliance with US regulations. *Clinical Trials*, *8*(5), 624-633.

Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: An empirical study. *MIS quarterly*, 549-566.

JUDO, (2019). Henkilötietojen käsittelyn yleisohje. Julkisen Hallinnon Digitaalisen Turvallisuuden Kehittämisohjelma.

Kandias, M., Mylonas, A., Virvilis, N., Theoharidou, M., & Gritzalis, D. (2010, August). An insider threat prediction model. In *International conference on trust, privacy and security in digital business* (pp. 26-37). Springer, Berlin, Heidelberg.

Kent, K., & Souppaya, M. (2006). Guide to computer security log management. *NIST special publication*, *92*, 1-72.

King, J., Pandita, R., & Williams, L. (2015). Enabling forensics by proposing heuristics to identify mandatory log events. In *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security* (pp. 1-11).

King, J., & Williams, L. (2014). Log your CRUD: design principles for software logging mechanisms. In *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security* (pp. 1-10).

Kyberturvallisuuskeskus. (21.07.2020). Näin keräät ja käytät lokitietoja. Haetty 25.2.2021 osoitteesta https://www.kyberturvallisuuskeskus.fi/fi/ajankohtaista/ohjeet-ja-oppaat/nain-keraat-ja-kaytat-lokitietoja?toggle=Lokeja%20koskeva%20lains%C3%A4%C3%A4d%C3%A4nt%C3%B6&toggle=Lokeja%20eri%20tarkoituksiin

Legg, P. A. (2015, October). Visualizing the insider threat: challenges and tools for identifying malicious user activity. In *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)* (pp. 1-7). IEEE.

Legg, P. A., Buckley, O., Goldsmith, M., & Creese, S. (2015, April). Caught in the act of an insider attack: detection and assessment of insider threat. In *2015 IEEE International Symposium on Technologies for Homeland Security (HST)* (pp. 1-6). IEEE.

Lee, K. H., Zhang, X., & Xu, D. (2013, November). LogGC: garbage collecting audit log. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (pp. 1005-1016).

Li, T. (2016). *Event mining: Algorithms and applications*. CRC Press, Taylor & Francis Group.

Liu, L., De Vel, O., Chen, C., Zhang, J. & Xiang, Y. (2018). *Anomaly-Based Insider Threat Detection Using Deep Autoencoders*. https://doi.org/10.1109/ICDMW.2018.00014

Liu, Z., Qin, T., Guan, X., Jiang, H., & Wang, C. (2018). An integrated method for anomaly detection from massive system logs. *IEEE Access*, *6*, 30602-30611.

Lukka, K. (2014). Kari Lukka: konstruktiivinen tutkimusote. Saatavilla https://metodix. fi/2014/05/19/lukka-konstruktiivinen-tutkimusote/.

Lundgren, B., & Möller, N. (2019). Defining information security. *Science and engineering ethics*, *25*(2), 419-441.

Malin, B., Nyemba, S., & Paulett, J. (2011). Learning relational policies from electronic health record access logs. *Journal of Biomedical Informatics*, *44*(2), 333-342.

Maloof, M. A., & Stephens, G. D. (2007, September). Elicit: A system for detecting insiders who violate need-to-know. In International Workshop on Recent Advances in Intrusion Detection (pp. 146-166). Springer, Berlin, Heidelberg.

Mamonov, S., & Benbunan-Fich, R. (2018). The impact of information security threat awareness on privacy-protective behaviors. *Computers in Human Behavior*, *83*, 32-44.

Marty, R. (2011, March). Cloud application logging for forensics. In *proceedings of the 2011 ACM Symposium on Applied Computing* (pp. 178-184).

Microsoft. (19.04.2017). Basic security audit policies. Haettu 07.03.2021 osoitteesta https://docs.microsoft.com/en-us/windows/security/threat-protection/auditing/basic-security-audit-policies

Mulligan, D. K., & Schneider, F. B. (2011). Doctrine for cybersecurity. Daedalus, 140(4), 70-92.

Myers, J., Grimaila, M. R., & Mills, R. F. (2009, April). Towards insider threat detection using web server logs. In *Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies* (pp. 1-4).

Nieles, M., Dempsey, K., & Pillitteri, V. Y. (2017). An introduction to information security. NIST special publication, 800, 12.

Oliner, A., Ganapathi, A. & Xu, W. (2012). Advances and challenges in log analysis. *Communications of the ACM, 55*(2), 55-61. https://doi.org/10.1145/2076450.2076466

Olympia, P. L. (1994). Hot on the audit trail: Audit files can reliable tell you the who, what, and when of a modified database record. (Set Expert On)(Column) (Tutorial). *DBMS, 7*(2), 91.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of management information systems, 24(3), 45-77.

PoSeID-on project. Protection and control of secured information by means of a privacy enhanced dashboard. https://www.posei don-h2020.eu Accessed Jan 2020.

Rickett, L. K. (2015). The data security breach and risk assessment: Imminent threat, high cost, and preparation is key. *Internal Auditing, 30*(6), 6.

Roratto, R. & Dias, E. D. (2014). Security information in production and operations: A study on audit trails in database systems. *Journal of Information Systems & Technology Management, 11*(3), 717. https://doi.org/10.4301/S1807-17752014000300010

Roy, P., Sengupta, A., & Mazumdar, C. (2021). A Structured Control Selection Methodology for Insider Threat Mitigation. *Procedia Computer Science*, *181*, 1187-1195.

Safa, N. S., Maple, C., Watson, T., & Von Solms, R. (2018). Motivation and opportunity based model to reduce information security insider threats in organisations. Journal of information security and applications, 40, 247-257.

Silva, P., Casaleiro, R., Simões, P., Antunes, N., Curado, M., & Monteiro, E. (2020). Risk management and privacy violation detection in the PoSeID-on data privacy platform. *SN Computer Science*, *1*, 1-10.

Singh, N. K., Tomar, D. S., & Roy, B. N. (2010). An approach to understand the end user behavior through log analysis. *International Journal of Computer Applications*, *5*(11), 27-34.

Stan, O. P., & Miclea, L. (2019). New era for technology in healthcare powered by GDPR and blockchain. In *6th International Conference on Advancements of Medicine and Health Care through Technology; 17–20 October 2018, Cluj-Napoca, Romania* (pp. 311-317). Springer, Singapore.

Söderström, O., & Moradian, E. (2013). Secure audit log management. *Procedia Computer Science*, *22*, 1249-1258.

Tankard, C. (2016). What the GDPR means for businesses. *Network Security*, *2016*(6), 5-8.

Tietosuojavaltuutetun toimisto. (2019). Usein kysyttyä EU:n tietosuoja-asetuksesta. Haettu 21.3.2021 osoitteesta: https://tietosuoja.fi/gdpr

VAHTI (2009). Lokiohje. Valtionvarainministeriö. Edita Prima Oy Helsinki.

Viestintävirasto. (2016). Lokien keräys ja käyttö. Haettu 7.3.2021 osoitteesta https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/file/Lokitusohje.pdf

Von Solms, R. (1998). Information security management (1): Why information security is so important. *Information Management & Computer Security, 6*(4), 174-177. https://doi.org/10.1108/EUM0000000004533

Zargar, A., Nowroozi, A. & Jalili, R. (2016). *XABA: A zero-knowledge anomaly-based behavioral analysis method to detect insider threats*. https://doi.org/10.1109/ISCISC.2016.7736447

Zhang, D., Zheng, Y., Wen, Y., Xu, Y., Wang, J., Yu, Y., & Meng, D. (2018, January). Role-based log analysis applying deep learning for insider threat detection. In *Proceedings of the 1st Workshop on Security-Oriented Designs of Computer Architectures and Processors* (pp. 18-20).

Wall, D. S. (2013). Enemies within: Redefining the insider threat in organizational security policy. Security journal, 26(2), 107-124.

Walton, R. (2006). Balancing the insider and outsider threat. *Computer fraud & security*, *2006*(11), 8-11.

Warkentin, M., & Willison, R. (2009). Behavioral and policy issues in information systems security: the insider threat. *European Journal of Information Systems*, *18*(2), 101-105.

Weir, G., Aßmuth, A., Whittington, M., & Duncan, B. (2017, August). Cloud accounting systems, the audit trail, forensics and the EU GDPR: how hard can it be?. In British Accounting & Finance Association (BAFA) Annual Conference 2017.

Whitman, M. E., & Mattord, H. J. (2011). Principles of information security. Cengage Learning.

Wynn, D. C., & Eckert, C. M. (2017). Perspectives on iteration in design and development. Research in Engineering Design, 28(2), 153-184.

LIITE 1 HAASTATTELUN KYSYMYSPATTERISTO

Puolistrukturoidun haastattelun runko

HAASTATELTAVAN TAUSTATIEDOT

1) Mikä on työnimikkeenne?
2) Kuinka kauan olet työskennellyt nykyisessä työtehtävässäsi?
3) Kuinka kauan olet työskennellyt nykyisen työnantajasi palveluksessa?
4) Kerro lyhyesti, miten työtehtäväsi liittyvät lokianalytiikkaan?

TEEMA 1: LOKIANALYTIIKAN HAASTEET

5) Koetko, että GDPR on tuonut tarpeen tehokkaampaan lokianalytiikkaan, jos kyllä niin mistä tämä johtuu?
6) Koetko, että suuret määrät valtuutettuja henkilötietohakuja asettavat haasteen tunnistaa valtuuttamattomia henkilötietohakuja, joita työntekijät saattavat suorittaa, jos kyllä niin miten tämä ilmenee?
7) Oletko havainnut haasteita eri järjestelmistä tulevien lokien laadun kanssa, jotka voivat vaikuttaa lokianalytiikan tehokkuuteen, jos kyllä, kuvaile mitä haasteita laatu on aiheuttanut?
8) Onko tietoosi tullut, että organisaatiossa, jossa työskentelet tällä hetkellä tai organisaatioissa, joissa olet työskennellyt aiemmin, olisi työntekijöiden toimesta toteutettu valtuuttamattomia hakuja?

TEEMA 2: SUODATTIMIEN MÄÄRITTÄMINEN

9) Mitkä seuraavista tieteellisestä kirjallisuudesta johdetuista suodattimista kokisit soveltuvan organisaation valtuuttamattomien henkilötietohakujen lokianalytiikkaan: poikkeukselliseen aikaan tehty haku, työntekijän roolista poikkeava haun kohde, roolista poikkeava hakutiheys, haku, jonka on toteuttanut aiemmin tunnistetusti valtuuttamattoman haun tehnyt työntekijä?
10) Minkälaiset suodatukset lokimassasta tarkempaan tarkasteluun poimittavista valtuutetun käyttäjän suorittamista tapahtumista voisi mielestäsi auttaa nostamaan mahdollisesti valtuuttamattomia hakuja esille?
11) Minkälaiset henkilötietohaut organisaation tietojärjestelmistä koet poikkeaviksi työntekijöiden toteuttamina?

TEEMA 3: LOKIANALYTIIKAN MERKITYS

12) Koetko, että valtuuttamattomien henkilötietohakujen tunnistamiseen olisi olemassa tehokkaampi keino, kuin lokianalytiikka? Jos kyllä, niin mikä?
13) Koetko, että tieto valtuuttamattomiin hakuihin kohdistuvan lokianalytiikan suorittamisesta organisaation sisällä ennaltaehkäisee työntekijöiden toteuttamia valtuuttamattomia henkilötietohakuja?

LIITE 2 SAATEKIRJE HAASTATELTAVILLE ITEROINTI

Hei,

Kiitos vielä Pro gradu-tutkielmani haastatteluun osallistumisesta. Haastattelut on nyt analysoitu ja ensimmäinen versio suodatinmallista on luotu.

Tässä vaiheessa mallia on tarkoitus iteroida ja edelleen kehittää. Sähköpostin liitteenä löytyy suodatinmallin ensimmäinen versio, johon toivoisin sinun tutustuvan ja antavan palautetta. Liitteestä löytyy myös muutama tarkentava jatkokysymys.

Koostamani suodatinmallin on tarkoitus tarjota suodattimia, joiden avulla suuresta ja osittain monimutkaisesta tapahtumien lokimassasta saataisiin suodatettua esille valtuutetun käyttäjän mahdollisia poikkeavia tapahtumia, ja edelleen mahdollisesti valtuuttamattomia tapahtumia. Ensimmäinen versio suodatinmallista on koostettu suoritettujen haastattelujen, sekä tieteellisen kirjallisuuden perusteella. Suodatinmalli itsessään ei luonnollisesti ole riittävä työkalu, vaan se vaatisi tehokkaasti toimiakseen roolien erottelua, tietorajapintojen rakentamista, sekä mahdollisesti koneoppimisen edelleen kehittämistä. Suodatinmalli tarjoaa näkökulmia siihen, millaiset suodatukset voisivat auttaa tunnistamaan poikkeavia toimia. Toivoisin palautetta viimeistään keskiviikkoon, 26.05.2021 mennessä.

Ystävällisin terveisin
Jenny Hornborg

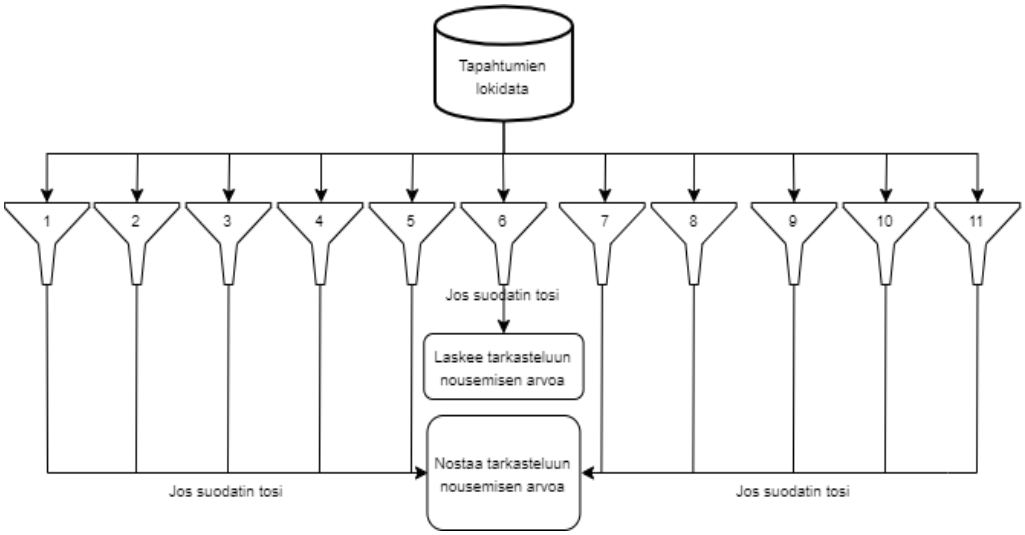
LIITE 3 SUODATINMALLIN ESITTELY ITEROINTIA VARTEN

Suodatinmallin ensimmäisen version esittely ja edelleen kehitys

Suodatinmallin perusperiaatteet: Lokianalytiikan järjestelmässä olevasta henkilötietohakujen tapahtumien lokidatasta pyritään suodattamaan esiin poikkeavia tapahtumia. Mikäli suodattimista 1-5 tai 7-11 jokin on tosi, nostaa se tapahtuman manuaalisen tarkastelun painoarvoa, indikoiden tapahtuman olevan poikkeava. Mikäli suodatin 6 on tosi, vähentää se manuaalisen tarkastelun painoarvoa tapahtumalle, indikoiden, että tapahtuma on ollut valtuutettu. Alta löydät suodatinmallin toimintaa kuvaavan kuvion, ja seuraavalta sivulta löydät suodattimet ja näiden selitteet.

Kysymykset:

1) Onko suodatinmallissa jokin/joitakin suodattimia, jotka ei ole valtuuttamattomien hakujen kontekstiin sopivia, jos kyllä mikä/mitkä ja miksi?
2) Puuttuuko suodatinmallista jokin tai joitakin suodattimia, mitkä voisivat auttaa tunnistamaan poikkeavia tapahtumia, jos kyllä, kuvaile mitä suodattimia malliin tulisi lisätä?
3) Arvioisitko suodatinmallin tarjoavan hyötyä valtuuttamattomien henkilötietohakujen tunnistamisessa?
4) Millaisilla määrittelyillä kokisit suodattimen 11 olevan mahdollinen, eli miten voitaisiin tunnistaa mahdolliset työntekijän henkilökohtaiset sidosryhmät?
5) Tulisiko mielestäsi jotakin muuta suodatinta tarkentaa, jos kyllä niin mitä ja miten?
6) Onko sinulla muita kehitysehdotuksia tai palautetta suodatinmallista?



Suodatin	**Selite**	**Nostaa/laskee tarkemman tarkastelun arvoa**
1. Poikkeava observoinnin aika	Observoinnin aika poikkeaa työntekijän rooliin suhteutetusta työajasta. Esimerkiksi toimistoaikana työskentelevä työntekijä toteuttaa haun yöaikaan	Nostaa
2. Poikkeava observoinnin tiheys yhteen kohteeseen lyhyellä aikavälillä	Työntekijä observoi poikkeavan tiheästi kohdetta lyhyen aikavälin sisällä	Nostaa
3. Poikkeava observoinnin tiheys yhteen kohteeseen pitkällä aikavälillä	Työntekijä observoi poikkeavan tiheästi kohdetta pitkällä aikavälillä. Esimerkiksi työntekijä hakee puolen vuoden aikana joka kuukausi saman henkilön tietoja	Nostaa
4. Poikkeava observoinnin tiheys useaan kohteeseen	Hakukentän iterointi. Työntekijä hakee tiheään tahtiin eri henkilötiedoilla tietojärjestelmistä	Nostaa
5. Poikkeava observoinnin kohde	Työntekijä observoi hänen roolistaan poikkeavaa kohdetta. Esimerkiksi yritysasiakaspalvelija hakee kuluttaja-asiakkaan tietoja	Nostaa
6. Observoinnille löytyy asiakaslähtöinen peruste	Työntekijän toteuttamalle haulle löytyy esimerkiksi asiakkaan yhteydenotto perusteeksi	Laskee
7. Aiemmin valtuuttamattoman haun tehnyt työntekijä toteuttaa haun	Aiemmin tunnistetusti valtuuttamattomia hakuja tehnyt työntekijä tekee haun	Nostaa
8. Poikkeava hakuparametri	Työntekijä toteuttaa haun poikkeavalla parametrilla, jolla ei tyypillisesti haeta tietojärjestelmistä	Nostaa
9. Poikkeava tapahtumajärjestys	Työntekijä toteuttaa toimia poikkeavassa tapahtumajärjestelmässä. Esimerkiksi työntekijä hakee normaalisti ensin järjestelmästä A ja sitten B, sitten C. Poikkeavassa haussa työntekijä menee suoraan järjestelmään C	Nostaa
10. Käyttöoikeudet järjestelmään puutteelliset	Työntekijälle ei ole annettu virallisen prosessin mukaan käyttöoikeutta järjestelmään, vaan pääsy järjestelmään on saatu muuta kautta. Myös tilanne, jossa työntekijälle on jäänyt vanhasta roolista käyttöoikeudet järjestelmään, johon ei niitä uudessa roolissaan tarvitse	Nostaa
11. Observointi on kohdistunut valtuutetun käyttäjän henkilökohtaiseen sidosryhmään	Työntekijä on tunnistetusti hakenut esimerkiksi hänen läheisensä tietoja organisaation tietojärjestelmästä	Nostaa